Optimal Test for Markov Switching*

Marine Carrasco†
University of Montreal

Liang Hu‡
University of Leeds

Werner Ploberger§
Washington University at St. Louis

January, 2007
Comments welcome

Abstract

We propose a new class of tests for the stability of parameters. We cover the class of Hamilton models, where regime changes are driven by an unobservable Markov chain. We derive a class of information matrix-type tests and show that they are equivalent to the likelihood ratio test. Hence, our tests are asymptotically optimal. Moreover these tests are easy to implement as they do not require the estimation of the model under the alternative. They are also very general. Indeed, the underlying process driving the regime changes may have a finite or continuous state space, as long as it is exogenous. The model itself need not be linear. It may be a GARCH model, for instance.

We use this test to investigate the asymmetric pattern in stock price dynamics. Using US data, we find evidence in favor of nonlinearities, which is consistent with the asymmetric behavior in stock prices.

---

* Carrasco gratefully acknowledges financial support from the National Science Foundation under grant SES-0211418.
† Email: marine.carrasco@umontreal.ca
‡ Email: lh@lubs.leeds.ac.uk
§ Email: werner@ploberger.com
1. Introduction

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

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

$$\frac{1}{\sqrt{T}} \sum_t h_t \left[ \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta} \right) + \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_t}{\partial \theta} \right) ^\prime + 2 \sum_{s<t} \rho^{(t-s)} \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_s}{\partial \theta} \right) ^\prime \right] h_t$$

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

There are few papers proposing tests for Markov-switching. Garcia (1998) studies the asymptotic distribution of a sup-type Likelihood ratio test. Hansen (1992) treats the likelihood as a empirical process indexed by all the parameters (those identified and those unidentified under the null). His test relies on taking the supremum of LR over the nuisance parameters. Both papers require estimating the model under the alternatives, which may be cumbersome. None investigates local powers. Gong and Mariano (1997) reparametrize their linear model in the frequency domain and construct a test based on the differences in the spectrum between null and alternative. Although they do not discuss the asymptotic power of their tests, a closer reading of the paper shows that their test shares certain features with our test. Some work has been done on testing for independent mixtures, Chesher (1984), Lee and Chesher (1986), Davidson and MacKinnon (1991), and recently Cho and White (2003).

It should be emphasized that testing parameter stability against a Markov switching alternative is much more challenging than testing for Structural change or Threshold. They have in common that they involve nuisance parameters that are not identified under the null hypothesis. The latter have been investigated in many papers: Davies (1977, 1987), Andrews (1993), Andrews and Ploberger (1994), Hansen (1996) among others. There is, however, some difference to the classical situation: the “right” local alternatives are of order $T^{-1/4}$. Hence, to study the properties of this test, we need to do expansions of the likelihood at the fourth order.
To illustrate the applicability of our test, we use it to detect the asymmetry in stock markets. We regress the stock price on dividends and we expect the residual to exhibit periods of slow increases and sharp declines. Using US data, we find that the residuals are stationary, which could be hastily interpreted as evidence against the asymmetric behaviors. However, our Markov switching test strongly rejects the linearity, suggesting that at least two regimes should be used to fit the data. Estimating a three-state Markov switching model reveals that one regime is near unit root, the other has an explosive root with negative drift, while in the third regime a large increase or decrease in the previous period is reinforced, which illustrates that a Markov Switching model can capture very well the asymmetric pattern of slow increases and sharp decreases. It is worth mentioning another application of our test. In a recent paper, Hamilton (2004) argues that a linear statistical model cannot capture the recurring cyclical pattern observed in economic aggregates. He applies our test to show that there are nonlinearities in the unemployment rate over the business cycle and that a Markov switching model is particularly well designed to capture these nonlinearities.

The outline of the paper is as follows. Section 2 describes the test statistic. Section 3 establishes the admissibility. In Section 4, we describe simulation results. Finally in Section 5, we use this test to investigate the asymmetric pattern in stock markets. In Appendix A, we define the tensor notations used to derive the fourth order expansion of the likelihood. These notations are interesting in their own as they could be used in other econometric problems involving higher-order expansions. The proofs are collected in Appendix B.

2. Assumptions and test statistic

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

\[
H_0 : \theta_t = \theta_0, \text{ for some unspecified } \theta_0
\]

against

\[
H_1 : \theta_t = \theta_0 + \eta_t, \quad (2.1)
\]

where the switching variable \( \eta_t \) is not observable.

**Assumption 1.** (i) \( \eta_t \) is stationary and \( \beta \)-mixing with geometric decay. It implies in particular that there exist \( 0 < \lambda < 1 \) and a measurable non-negative function \( g \) such that

\[
\sup_{|h| \leq 1} \left| E \left[ h(\eta_{t+m}) | \eta_t, ... \right] - E \left[ h(\eta_t) \right] \right| \leq \lambda^m g(\eta_t, ...). \quad (2.2)
\]

and

\[
Eg(\eta_t, ..) < \infty. \quad (2.3)
\]
Furthermore we assume that

\[ E \eta_t = 0, \max_t \| \eta_t \| \leq M < \infty, \]

\[ f (y_t | \eta_t, y_{t-1}, \eta_{t-1}, \ldots, y_1, \eta_1) = f (y_t | \eta_t, y_{t-1}, \ldots, y_1). \]

**Remark 1.** The assumption \( E \eta_t = 0 \) is not restrictive as the model can always be reparametrized to ensure this condition. \( \eta_t \) \( \beta \)-mixing is satisfied by e.g. irreducible and aperiodic Markov chain with finite state space. \( \max_t \| \eta_t \| \leq M < \infty \) will also be satisfied by any finite state space Markov chain, however it will not be satisfied by an AR(1) process with normal error. This condition could be relaxed to allow for distributions of \( \eta_t \) with thin tails but this extension is beyond the scope of the present paper. Although some form of mixing is necessary for the \( \eta_t \), one should be able to relax condition (2.3).

**Assumption 2.** The distribution of \( \eta_t \) may depend on some unknown parameters \( \beta \). They are nuisance parameters that are not identified under \( H_0 \). We assume that \( \beta \) belongs to a compact set \( B \), and that \( \lambda \), the constant \( M \), and the function \( g \) defined in Assumption 1 are independent of \( \beta \).

**Assumption 3.** \( y_t \) is stationary under \( H_0 \) and the following conditions on the conditional log-density of \( y_t \) given \( y_{t-1}, \ldots, y_1 \) (under \( H_0 \)), \( l_t \), are satisfied. \( l_t = l_t (\theta) \), as a function of the parameter \( \theta \), is at least 5 times differentiable. Moreover, let us denote by \( l_t^{(k)} \) the \( k \)-th derivative of the likelihood with respect to the parameter \( \theta \).

\[
\sup_{t, \theta \in \mathcal{N}} E \left( \left\| l_t^{(1)} (\theta) \right\|^{24} \right) < \infty,
\]

\[
\sup_{t, \theta \in \mathcal{N}} E \left( \left\| l_t^{(2)} (\theta) \right\|^{12} \right) < \infty,
\]

\[
\sup_{t, \theta \in \mathcal{N}} E \left( \left\| l_t^{(3)} (\theta) \right\|^{8} \right) < \infty,
\]

\[
\sup_{t, \theta \in \mathcal{N}} E \left( \left\| l_t^{(4)} (\theta) \right\| \right) < \infty,
\]

\[
\sup_{t, \theta \in \mathcal{N}} E \left( \left\| l_t^{(5)} (\theta) \right\| \right) < \infty.
\]

where \( \mathcal{N} \) is a neighborhood around \( \theta_0 \).

**Remark 2.** The expectations in the above formulae are to be understood as expectations with respect to the probability measure corresponding to the parameter \( \theta_0 \).

**Remark 3.** For the “norm” of the derivatives we can e.g. take the usual \( L^2 \) norm

\[
\left\| l_t^{(\ell)} (\theta) \right\| = \sqrt{\sum_{0 \leq i_1, i_2, \ldots, \leq \ell, \sum_{j=1}^{k} i_j = \ell} \left( \frac{\partial^{i_1} l_t}{\partial \theta_1^{i_1} \partial \theta_2^{i_2} \ldots \partial \theta_k^{i_k}} \right)^2}. \tag{2.5}
\]
Usually the first derivatives of the likelihood is associated with the vector of scores and the second one with the Hessian. This interpretation is sufficient for a statement of the results. For the proofs of our theorems, however, we need derivatives of higher order. Their precise nature will be discussed in Appendix A.

**Remark 4.** We do not impose restrictions on the moments of \( y_t \). For instance \( y_t \) could be a stationary IGARCH process. However, we rule out the case where \( y_t \) is a random walk. To deal with unit root, we would have to alter the test statistic by proper rescaling and its asymptotic distribution would be different. We leave this extension for future research. As in Andrews and Ploberger (1994, Section 4.1.), the vector of observable variables \( y_t \) may include exogenous variables.

The test statistic, for a given \( \beta \), is of the form:

\[
TS_T (\beta) = TS_T (\beta, \hat{\theta}) = \Gamma_T - \frac{1}{2T} \tilde{\varepsilon} (\beta)' \tilde{\varepsilon} (\beta)
\]

where

\[
\Gamma_T = \frac{1}{2} \left( \frac{1}{\sqrt{T}} \sum_t tr \left( \left( \ell_t^{(2)} + \ell_t^{(1)} l_t^{(1)\prime} \right) E (\eta_t \eta_t') \right) + \frac{2}{\sqrt{T}} \sum_{t>s} tr (l_t^{(1)} l_s^{(1)\prime} E (\eta_t \eta_s')) \right)
\]

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

As \( \beta \) is unknown and can not be estimated consistently under \( H_0 \), we use sup-type tests like in Davies (1987)

\[
\supTS = \sup_{\beta \in B} TS_T (\beta)
\]

or exponential-type tests as in Andrews and Ploberger (1994)

\[
\expTS = \int_B \exp (TS_T (\beta)) dJ (\beta)
\]

where \( J \) is some prior distribution for \( \beta \) with support on \( B \) a compact subset of \( B \). We will establish admissibility for a class of \( \expTS \) statistics.

**Remark 5.** The asymptotic distribution of the tests will not be nuisance parameter free in general. Therefore we have to rely on parametric bootstrap to compute the critical values.
Remark 6. The test statistic $TS$ depends only on the score and derivative of the score under the null and on the estimator of $\theta$ under $H_0$. Therefore it does not require estimating the model under the alternative. This is a great advantage over competing tests like those of Garcia (1998), Hansen (1992) because estimating a Markov switching model is particularly burdensome (Hamilton, 1989) or even intractable if the model is highly nonlinear as in the GARCH model.

Remark 7. The test relies on the second Bartlett identity (Bartlett, 1953a,b). It is related to the Information Matrix test introduced by White (1982). Cheshier (1984) shows that the Information Matrix test has power against models with random coefficients. He shows that a score test of the hypothesis that parameters have zero variance is close to the Information Matrix test. Davidson and McKinnon (1991) derive information-matrix-type tests for testing random parameters. The main difference with our setting is that they assume that the parameters are independent, whereas we assume that the parameters are serially correlated and we fully exploit this correlation. Recently, Cho and White (2003) have proposed a test for independent mixture.

Remark 8. The form of our test is insensitive to the dynamic of the latent process $\eta_t$. It depends only on the form of the autocorrelation of $\eta_t$.

Remark 9. We assume throughout the paper that the model under the null is correctly specified. The issue of misspecification is not addressed here.

Remark 10. The main difference with Structural change and threshold testing is that here the local alternatives are of order $T^{-1/4}$. This is due to the fact that the regimes $\eta_t$ are unknown and one needs to estimate them at each period. It is also linked to the singularity of the information matrix under the null hypothesis.

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

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

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

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

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

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

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

and \( \bar{B} = \{ c^2, h, \rho : c^2 > 0, \|h\| = 1, \rho < \rho < \bar{\rho} \} \) and \( -1 < \rho < \bar{\rho} < 1 \).

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

\[
\sum \sup_{T} TS = \sup_{\left\{ h, c^2, \rho \mid \|h\| = 1, \rho < \rho < \bar{\rho} \} } \sqrt{\rho} \frac{\rho^t}{\sqrt{\bar{\rho}} \bar{\rho}^\bar{s}} \left( \max \left( 0, \frac{\Gamma_T^*}{\sqrt{\bar{\rho}} \bar{\rho}^\bar{s}} \right) \right)^2
\]

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

3. Local alternatives and asymptotic optimality

First of all let us discuss some general principles for the construction of admissible tests. A test is admissible if there is no other test that has uniformly higher (or equal) power. Consider a general testing problem of testing a null \( H_0 \) against an alternative \( H_1 \) and let \( \mu_0 \) and \( \mu_1 \) be two measures concentrated on \( H_0 \) and \( H_1 \), respectively. Furthermore assume that the probability measures for our models are given by densities \( f_{\xi} \), (with respect to a common dominating measure), where the parameter \( \xi \in H_0 \cup H_1 \). Then a test rejecting when

\[
\int \frac{f_{\xi} d\mu_1}{f_{\xi} d\mu_0} > K
\]

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

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

The null hypothesis for a given \( \theta \) is denoted as

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

and the sequence of local alternatives is given by

\( H_{1T}(\theta) : \theta_t = \theta + \frac{\eta_t}{\sqrt{T}}. \)

(3.2)
Let $Q^\beta_T$ denote the joint distribution of $(\eta_1, ..., \eta_T)$, indexed by the unknown parameter $\beta$. Let $P_{\theta, \beta}$ be the probability measure on $y_1, y_2, ..., y_T$ corresponding to $H_{1T}(\theta)$, and $P_\theta$ be the probability measure on $y_1, y_2, ..., y_T$ corresponding to $H_0(\theta)$. The ratio of the densities under $H_0(\theta)$ and $H_{1T}(\theta)$ is given by

$$
\ell^\beta_T(\theta) \equiv \frac{dP_{\theta, \beta}}{dP_\theta} = \int \frac{T}{\prod_{t=1}^T f_t(\theta + \eta_t/T^{1/4})} \frac{dQ^\beta_T}{\prod_{t=1}^T f_t(\theta)}.
$$

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

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

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

The proof of the theorem is rather complicated, so we give it in Appendix B.

**Remark 11.** Although Gong and Mariano(1997) never evaluate the asymptotic power of their test, a closer look at their results shows that it is compatible with our theory. In their paper, the process representing "our" $\eta_t$ is of the form $\alpha_1 S_t$, where $S_t$ is a process taking only values 0 and 1. They test for $\alpha_1 = 0$ by constructing an LM-test for another parameter (in their notation) $\delta = \alpha_1^2$. Hence their test should have power against alternatives for which $\delta = O(1/\sqrt{T})$, which implies that $\alpha_1 = O(1/\sqrt{T})$.

We can easily see from (2.6) that $\mu_{2,t}(\beta, \theta_0)$ is a stationary and ergodic martingale difference sequence, hence the central limit theorem applies. Moreover, for each sequence

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

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

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

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

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

Using a Taylor expansion around $\theta_0 + \frac{1}{\sqrt{T}} d$, we obtain the following lemma.
Lemma 3.3. For all $\theta_0 \in \mathcal{N}$, and for all vectors $d$

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

(3.5)

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

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

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

we can conclude that with

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

(3.6)

$$
\frac{dP_{\theta_T, \beta}}{dP_{\theta_0}} \to 1
$$

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

$$
\to 1
$$

where the convergence is again uniform in probability with respect to $P_{\theta_0}$.

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

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

(3.7)

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

Definition 3.4. A sequence $\varphi_T$ of tests is asymptotically better than $\psi_T$ at $\theta_0$ if it is “better” on the null

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

(3.8)

and “better” on the alternatives, that is, for all $\theta_T$ and $\beta$

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

(3.9)
This definition is essentially the same as used by Andrews and Ploberger (1994) and a bit different from the one in Strasser (1995). Although the latter can be very useful when analyzing the asymptotic behavior of possible power functions for testing problems, our definition here proved more practical in econometric analysis because it directly deals with the asymptotic behavior of tests. Our definition here is, however, close enough to the one in Strasser (1995) so that we can use the standard proofs of optimality.

**Definition 3.5.** A test \( \psi_T \) is said to be admissible if there exists no asymptotically better test.

Let \( \varphi_T \) be some test statistics that has asymptotic level \( \alpha \) (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 \( \nu \) be an arbitrary, but finite measure concentrated on a compact subset of \( B \times R^k \). Without limitation of generality we can assume that \( \nu (B \times R^k) = 1 \). Then let us define the loss function

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

By Fubini’s theorem, we have

\[
L (\varphi_T) = \int (K - \frac{dP_{\theta_0 + d / \sqrt{T}, \beta}}{dP_{\theta_0}}) \varphi_T dP_{\theta_0} d\nu (\beta, d) = \int \left(K - \left\{ \int \frac{dP_{\theta_0 + d / \sqrt{T}, \beta}}{dP_{\theta_0}} d\nu (\beta, d) \right\} \right) \varphi_T dP_{\theta_0} \tag{3.11}
\]

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

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

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

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

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

Let us now assume that we have a competing sequence of tests \( \varphi_T \). Note that (3.13) does not uniquely determine a test: We do not care about the behavior of the test on the event \( \left\{ \int \frac{dP_{\theta_0 + d / \sqrt{T}, \beta}}{dP_{\theta_0}} d\nu (\beta, d) \right\} = K \}. Hence the following definition will be useful:
Definition 3.6. The tests \( \varphi_T \) and \( \varphi'_T \) are asymptotically equivalent (with respect to our loss function \( L \)) if and only if for all \( \varepsilon > 0 \)
\[
\lim E_{\theta_0} |\varphi_T - \varphi'_T| I \left[ \left\{ \int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d) \right\} - K \right] > \varepsilon = 0. \tag{3.15}
\]

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

**Theorem 3.7.** Suppose \( \varphi_T \) and \( \psi_T \) are asymptotically equivalent, where \( \psi_T \) is defined by (3.13). Then
\[
\lim (L(\psi_T) - L(\varphi_T)) = 0. \tag{3.16}
\]
If \( \varphi_T \) and \( \psi_T \) are not asymptotically equivalent (in the above sense), then
\[
\lim \inf (L(\psi_T) - L(\varphi_T)) < 0. \tag{3.17}
\]
Hence (3.16) implies that \( \psi_T \) and \( \varphi_T \) are asymptotically equivalent.

**Proof.** We can easily see that \( L(\psi_T) - L(\varphi_T) = \int (K - \left\{ \int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d) \right\}) (\psi_T - \varphi_T) dP_{\theta_0} \).

The construction of \( \psi_T \) and the fact that \( 0 \leq \varphi_T \leq 1 \) imply that the integrand is nonpositive. Let \( \varepsilon > 0 \) be arbitrary. Let us define
\[
r = (K - \left\{ \int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d) \right\}). \tag{3.18}
\]

Then
\[
L(\psi_T) - L(\varphi_T) = \int r I ([r] > \varepsilon) (\psi_T - \varphi_T) dP_{\theta_0} + \int r I ([r] \leq \varepsilon) (\psi_T - \varphi_T) dP_{\theta_0}. \tag{3.19}
\]

Since \( |\psi_T - \varphi_T| \leq 1 \), we have
\[
\left| \int r I ([r] \leq \varepsilon) (\psi_T - \varphi_T) dP_{\theta_0} \right| \leq \varepsilon. \tag{3.20}
\]

For asymptotically equivalent tests, \( \int r I ([r] > \varepsilon) (\psi_T - \varphi_T) dP_{\theta_0} \to 0 \), which proves (3.16).

For (3.17), observe that if \( \varphi_T \) and \( \psi_T \) are not asymptotically equivalent, then there exists an \( \eta > 0 \) so that
\[
\limsup E_{\theta_0} |\varphi_T - \psi_T| I \left[ \left\{ \int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d) \right\} - K \right] > \eta > \varepsilon. \tag{3.21}
\]

The construction of \( \psi_T \) guarantees that \( r (\psi_T - \varphi_T) \leq 0 \). Hence - \( |\varphi_T - \psi_T| r I ([r] > \varepsilon) = r I ([r] > \varepsilon) (\psi_T - \varphi_T) \leq r I ([r] > \eta) (\psi_T - \varphi_T) \) if \( \eta \geq \varepsilon \), hence for all \( \varepsilon \) small enough \( \lim inf \)
\[ \int r I \left[ r > \varepsilon \right] \left( \psi_T - \varphi_T \right) dP_{\theta_0} < - \lim \sup E_{\theta_0} \left| \varphi_T - \psi_T \right| I \left[ \left\{ \int \frac{dP_{\theta_0+d}/\sqrt{T}}{dP_{\theta_0}} d\nu(\beta, d) \right\} - K \right] > \eta, \]

and together with (3.20) this proves our theorem. \[ \Box \]

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

For this purpose, let us first observe that the processes

\[ Z_T(\beta, \theta) = \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta) - \frac{1}{8}E \left( \mu_{2,t}(\beta, \theta)^2 \right) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d'd_{t}^{(1)}(\theta) + \frac{1}{2}E \left( \left( d'd_{t}^{(1)}(\theta) \right)^2 \right), \]

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

We now want to show that the tests \( \psi_T \) and the tests based on

\[ \int \exp(Z_T(\beta, \theta_T))d\nu(\beta, d) \] (3.23)

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

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

We know that for all finite sets \( \beta_i, di \)

\[ \exp(Z_T(\beta_i, \theta_0 + di/\sqrt{T}))/ \frac{dP_{\theta_0+di}/\sqrt{T}}{dP_{\theta_0}} \to 1. \] (3.25)

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

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

Let the tests \( \phi_T \) reject when \( \int \exp(Z_T(\beta, \theta_T))d\nu(\beta, d) > K \) and accept when \( \int \exp(Z_T(\beta, \theta_T))d\nu(\beta, d) < K \). Then these tests are asymptotically equivalent to the tests \( \psi_T \). Consequently, we have the following result:
Theorem 3.8. Let \( \varphi_T \) be a sequence of tests that is asymptotically better (in the sense of definition 3.4) than \( \phi_T \). Then \( \varphi_T \) is asymptotically equivalent to \( \phi_T \).

Proof. We just have shown that the \( \phi_T \) are equivalent to the \( \psi_T \), hence

\[
\lim (L(\phi_T) - L(\psi_T)) = 0. \tag{3.26}
\]

Since \( \psi_T \) are the tests with minimal loss function, we also have

\[
\lim \inf (L(\varphi_T) - L(\phi_T)) \geq 0. \tag{3.27}
\]

If \( \delta \) 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 \( \lim \inf h_n \delta \leq \lim \inf \int h_n \delta \). The definition 3.4 guarantees that \( \lim \inf (\int \varphi_T dP_{\theta_T,\beta} - \int \phi_T dP_{\theta_T,\beta}) \geq 0 \) and

\[
\lim \sup (\int \varphi_T dP_{\theta_0} - \int \phi_T dP_{\theta_0}) \leq 0. \text{ Since } L(\varphi_T) - L(\phi_T) =
\]

\[
K \left( \int \varphi_T dP_{\theta_0} - \int \phi_T dP_{\theta_0} \right) - \left( \int \varphi_T dP_{\theta_0 + d/\sqrt{T},\beta} \right) d\nu(\beta, d), \text{ we can conclude that}
\]

\[
\lim \sup (L(\varphi_T) - L(\phi_T)) \leq 0. \tag{3.28}
\]

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

We now are able to construct asymptotically optimal tests for each parameter \( \theta_0 \). The problem, however, is that we do not know \( \theta_0 \). Hence we will try to find for each \( \theta_0 \) a measure \( \nu_{\theta_0} \) so that the corresponding test statistic

\[
\int \exp(Z_T(\beta, d))\nu_{\theta_0}(\beta, d) \tag{3.29}
\]
does not depend on \( \theta_0 \). For this purpose, define

\[
d(\beta) = d(\beta, \theta_0) = (I(\theta_0))^{-1} \text{cov} \left( \frac{1}{2} I_{2,1} (\beta, \theta_0), I_{1} (\theta_0) \right) \tag{3.30}
\]

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

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

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

Let \( \hat{\theta} \) be the maximum likelihood estimator for \( \theta \) under \( H_0 \), i.e.

\[
\hat{\theta} = \arg \max \sum l_i(\theta). \tag{3.32}
\]
Then

\[ \exp (TS - ST(\theta_0)) \to 0 \]  \hspace{1cm} (3.33)

in probability under \( P_{\theta_0} \), where

\[ \exp (TS) = \int (\exp (TS_T(\beta, \hat{\theta}))) \, dJ(\beta), \]  \hspace{1cm} (3.34)

and

\[ TS_T(\beta, \hat{\theta}) = \frac{1}{2T} \sum \mu_{2,t} (\beta, \hat{\theta}) - \frac{1}{2T} \tilde{\varepsilon} (\beta)' \tilde{\varepsilon} (\beta), \]  \hspace{1cm} (3.35)

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

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

\[ P_\theta \left( \left\{ \exp (TS) < K(\hat{\theta}) \right\} \right) \leq 1 - \alpha \]  \hspace{1cm} (3.36)

\[ P_\theta \left( \left\{ \exp (TS) > K(\hat{\theta}) \right\} \right) \leq \alpha \]  \hspace{1cm} (3.37)

and assume \( K(\hat{\theta}) \to K \). Then the tests \( \varphi_T \), which reject if \( \exp (TS) > K(\hat{\theta}) \), and accept if \( \exp (TS) < K(\hat{\theta}) \), are for all \( \theta_0 \) asymptotically equivalent under \( P_{\theta_0} \) to tests rejecting if \( ST(\theta_0) > K \), and accepting if \( ST(\theta_0) < K \). Moreover, we have

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

and

\[ P_{\theta_0} \left( \left\{ ST(\theta_0) > K \right\} \right) \leq \alpha \]  \hspace{1cm} (3.39)

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

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

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

\[ TS_T(\beta, \hat{\theta}) - \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{\mu_{2,t} (\beta, \theta_0)}{2} - d(\beta)' l_t^{(1)} (\theta_0) \right) - \frac{1}{2} E_{\theta_0} \left( \left( \frac{\mu_{2,t} (\beta, \theta_0)}{2} - d(\beta)' l_t^{(1)} (\theta_0) \right)^2 \right) \right) \to 0 \]  \hspace{1cm} (3.40)
uniformly on all compact sets. Moreover under $H_0$, we have

$$TS_T (\beta, \hat{\theta}) \overset{D}{\to} G(\beta),$$

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

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

Under $H_1$, $TS_T (\beta, \hat{\theta})$ converges to a Gaussian process with mean $k(\beta, \beta_0) - \frac{1}{2} k(\beta, \beta)$ and variance $k(\beta_1, \beta_2)$, where $\beta_0$ is the true value of the parameter $\beta$ under the alternative.

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

It is, however, also possible that

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

This case is not that implausible. Indeed we have

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

$$= E_{\theta_0} \left( \left( \frac{\mu_{2,t}}{2} \right)^2 \right) - E_{\theta_0} \left( l_t^{(1)} \frac{\mu_{2,t}}{2} \right) \left( \frac{\partial l_t^{(1)}}{\partial \theta} \right) \left( \frac{\partial l_t^{(1)}}{\partial \theta} \right)' E_{\theta_0} \left( \left( \frac{\partial l_t^{(1)}}{\partial \theta} \right) \right)^{-1} E_{\theta_0} \left( l_t^{(1)} \frac{\mu_{2,t}}{2} \right)$$

using (3.30). Hence (3.41) is satisfied if and only if $\mu_{2,t}$ belongs to the linear span of the components of $l_t^{(1)}$. Assume for a moment that $\rho = 0$ and all the other prerequisites of Assumptions 3 and 4 are fulfilled. Then $\mu_{2,t}$ is a linear functional of the second-order derivatives of the log-likelihood, namely $h' \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta'} + \left( \frac{\partial l_t}{\partial \theta} \right)' \left( \frac{\partial l_t}{\partial \theta} \right) \right) h$. Then (3.41) means that the second order derivatives can be written as a linear combination of the scores. This is a geometric condition, which has profound statistical implications: E.g. in Murray and Rice (1993), p. 16 it is used to characterize linear exponential families. A typical
example would be the normal distribution. We have two parameters, mean and variance, and we can easily see that if we take $h = (1, 0)'$ (our first parameter should be the mean) (3.41) is fulfilled. This corresponds to testing for independent mixture of two normals with different unknown means and same unknown variance. This same effect was noticed in Gong and Mariano(1997): They remark that their test does not work in this situation.

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

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

**Proposition 3.11.** Suppose Assumptions 1 to 4 hold. Assume furthermore that for all $t, s$, $h' \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta'} + \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta} + 2 \sum_{s < t} \rho^{(t-s)} \frac{\partial l_t}{\partial \theta} \frac{\partial l_s}{\partial \theta} \right)' h \text{ can not be represented as a linear combination of components of } \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta} \right).$ Then for each $h$, there exist at most finitely many $\rho$ so that (3.41) is fulfilled.

**Proof.** First of all let us observe that

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

Let us assume that for one $h$ there exist infinitely many values of $\rho$ so that (3.41) is fulfilled. We can easily see that $\mu_{2,t} (\beta, \theta)$, and hence $d$, are analytic functions of $\rho$. Therefore $E_{\theta_0} \left( \frac{\mu_{2,t} (\beta, \theta_0)}{2} - d'_{l_t} (\theta_0) \right)^2$ 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

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

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

$$
2c^2 h' \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_s}{\partial \theta} \right)' h = d'_{t-s} (\theta_0) \frac{\partial l_t}{\partial \theta},
$$

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

First of all let us observe that
\[
\mu_{2,t}(\beta, \theta) = c^2 h' \left[ \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta'} + \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_t}{\partial \theta} \right)' \right) + 2 \sum_{s < t} \rho^{(t-s)} \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_s}{\partial \theta} \right)' \right] h \quad (3.42)
\]
Let us assume that for one \(h\) there exist infinitely many values of \(\rho\) so that (3.41) is fulfilled. We can easily see that \(\mu_{2,t}(\beta, \theta)\), and hence \(d\), too are analytic functions of \(\rho\). Therefore \(E_{\theta_0} \left( (\mu_{2,T}(\beta, \theta_0) - l_t^{(1)}(d))^2 \right)\) must be an analytic function two. We did assume that this function has infinitely many zeros in a finite interval, hence it must be identically zero. Hence
\[
c^2 h' \left[ \left( \frac{\partial^2 l_t}{\partial \theta \partial \theta'} + \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_t}{\partial \theta} \right)' \right) + 2 \sum_{s < t} \rho^{(t-s)} \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_s}{\partial \theta} \right)' \right] h = d(c, h, \rho)' \left( \frac{\partial l_t}{\partial \theta} \right) \quad (3.43)
\]
for all \(\rho\). Since both sides of the equation are analytic functions, their derivatives must be equal, too. Hence
\[
2c^2 h' \left( \frac{\partial l_t}{\partial \theta} \right) \left( \frac{\partial l_s}{\partial \theta} \right)' h = d'_{t-s} \left( \frac{\partial l_t}{\partial \theta} \right), \quad (3.44)
\]
where \(d'_{t-s}\) is the partial derivative with respect to \(\rho\) of order \(t-s\) of \(d(\ldots, \ldots)\). In case \(c^2 \neq 0\), this contradicts our assumption.

The restriction to prior measures with compact support might be a bit restrictive. In most cases, we should be able to approximate prior measures with noncompact support by ones with compact support. In cases where (3.41) is fulfilled, we will, however, encounter a difficulty. For our test statistic, we have to compute \(expTS = \int \left( \exp(TS_T(\beta, \hat{\theta})) \right) dJ(\beta)\).

For the values of \(\beta\) where (3.41) holds the corresponding \(TS_T(\beta, \hat{\theta})\) will converge to zero. It is, however, difficult to get uniform convergence. Hence we will not derive theorems for these measures here.

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

4. Monte Carlo study

In some simple cases, if the asymptotic distribution of the test is nuisance parameter free, we are then able to tabulate the critical values for our test statistic. For instance, we look at a very simple model used in Garcia (1998) with switching intercept and an uncorrelated and homoscedastic noise component,
\[
y_t = \alpha_0 + \alpha_1 S_t + \omega_0 \varepsilon_t
\]
where
\[
P(S_t = 1 \mid S_t = 1) = p \quad P(S_t = -1 \mid S_t = -1) = q
\]
and \( \varepsilon_t \sim iid \mathcal{N}(0, 1) \).

For simplicity, we assume \( \omega_0 \) is known to be 1.

Garcia (1998) gives the limiting distribution of \( LR_n \), which is \( \sup C = \sup_{\gamma=(p,q) \in \Gamma} C(\gamma) \),

\[
C(\gamma) = g(\gamma)^2 \pi (1 - \pi)
\]

where

\[
g(\gamma) = \frac{\min(\pi_1, \pi_2) - \pi_1 \pi_2}{\pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2)}
\]

and

\[
\pi = E(S_t) = \frac{1 - q}{2 - p - q}
\]

Critical values of the asymptotic distribution is tabulated in his paper.

To look at the null asymptotic distribution of our test, with

\[
m_{2,t}(\beta, \hat{\theta}) = \left[ \left( \frac{\partial^2 l_{t}}{\partial \theta \partial \theta'} + \left( \frac{\partial l_{t}}{\partial \theta} \right) \left( \frac{\partial l_{t}}{\partial \theta} \right)' \right) + 2 \sum_{s<t} \rho^{t-s} \left( \frac{\partial l_{s}}{\partial \theta} \right) \left( \frac{\partial l_{s}}{\partial \theta} \right)' \right]
\]

and

\[
\Gamma^*_T = \frac{1}{2 \sqrt{T}} \sum m_{2,t}(\beta, \hat{\theta})
\]

with \( \varepsilon^* \) denoting the OLS residual of \( \frac{1}{2} m_{2,t}(\beta, \hat{\theta}) \) on \( l_{t}^{(1)}(\hat{\theta}) \), our sup test is defined as

\[
\text{supTS} = \sup_{\{\rho, \rho<q<\rho\}} \left( \max \left( 0, \frac{\Gamma^*_T}{\sqrt{\varepsilon^* \tilde{\varepsilon}^*}} \right) \right)^2
\]

To satisfy the centering assumption, we need to reparametrize (??). Define \( S_t' \) such that \( E(S_t') = 0 \). We then have

\[
y_t = \alpha_0 + \alpha_1 S_t + \omega_0 \varepsilon_t
\]

\[
= \alpha_0 + \alpha_1 \cdot E(S_t) + \alpha_1 \left( S_t - E(S_t) \right) + \varepsilon_t
\]

\[
= \alpha_0 + \frac{1 - q}{2 - p - q} + \alpha_1 S_t' + \varepsilon_t
\]

\[
= a + \alpha_1 S_t' + \varepsilon_t
\]

Then we have

\[
l_{t}^{(1)} = y_t - a
\]

and

\[
l_{t}^{(2)} = -1
\]

\( l_{t}^{(1)} \) at ML estimator is

\[
l_{t}^{(1)}(\hat{\theta}) = y_t - \bar{Y}
\]
Then
\[
\Gamma_T^* = \frac{1}{2\sqrt{T}} \sum_t \left[ (y_t - \bar{Y})^2 - 1 + 2 \sum_{s<t} \rho^{(t-s)} (y_t - \bar{Y}) (y_s - \bar{Y}) \right]
\]

To find the asymptotic distribution of \( \Gamma_T^* \), it’s easy to see that
\[
\frac{1}{2\sqrt{T}} \sum_t \left[ (y_t - \bar{Y})^2 - 1 \right] \Rightarrow N(0, \frac{1}{2})
\]

The second part converges weakly to a linear combination of Gaussian processes following from Andrews and Ploberger (1996). That is,
\[
\frac{1}{\sqrt{T}} \sum_t \sum_{s<t} \rho^{(t-s)} (y_t - \bar{Y}) (y_s - \bar{Y}) \Rightarrow \sum_{i=1}^{\infty} \rho^i Z_i
\]

where \( Z_i \) are iid standard Gaussian processes.

The correlation between the two parts is zero. So \( \Gamma_T^* \) converges weakly to a stochastic process \( G(\rho) \), where
\[
G(\rho) = \frac{1}{\sqrt{2}} Z_0 + \sum_{i=1}^{\infty} \rho^i Z_i
\]

where \( Z_0 \) is also standard normal and independent to all the \( Z_i \).

Moreover,
\[
\text{var} (G(\rho)) = \frac{1}{2} + \frac{\rho^2}{1-\rho^2}
\]

From Continuous Mapping Theorem,
\[
\sup \text{TS} \Rightarrow \sup_{\rho, \rho < \rho^*} \frac{1}{2} \left( \max \left( 0, \frac{1}{\sqrt{2}} Z_0 + \sum_{i=1}^{\infty} \rho^i Z_i \right) \right)^2
\]

(4.1)

So we could simulate the asymptotic critical values based on (4.1). As in Andrews and Ploberger (1996), we truncate the series \( \sum_{i=1}^{\infty} \rho^i Z_i \) at a large value \( TR \). We pick \( TR = 50 \) and we draw 100 \( \rho^i \)’s from an equispaced grid on \((-0.98, 0.98)\) and \((-0.7, 0.7)\) respectively. We run 10,000 replications and asymptotic critical values are tabulated.
<table>
<thead>
<tr>
<th>% of Dist. interval</th>
<th>Emp c.v. $[-0.98,0.98]$</th>
<th>Asym. c.v. $[-0.98,0.98]$</th>
<th>Emp. c.v. $[-0.7,0.7]$</th>
<th>Asym. c.v. $[-0.7,0.7]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>3.95</td>
<td>4.34</td>
<td>3.17</td>
<td>3.87</td>
</tr>
<tr>
<td>95%</td>
<td>2.40</td>
<td>2.81</td>
<td>2.06</td>
<td>2.27</td>
</tr>
<tr>
<td>90%</td>
<td>1.82</td>
<td>2.10</td>
<td>1.54</td>
<td>1.61</td>
</tr>
<tr>
<td>80%</td>
<td>1.26</td>
<td>1.42</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>70%</td>
<td>0.91</td>
<td>1.00</td>
<td>0.56</td>
<td>0.62</td>
</tr>
<tr>
<td>50%</td>
<td>0.37</td>
<td>0.50</td>
<td>0.18</td>
<td>0.21</td>
</tr>
<tr>
<td>10%</td>
<td>0</td>
<td>1.24e-3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Now we compare the power performance of our test with García’s. The power of our test is based on the asymptotic critical values tabulated above. Table 1A in García (1998) is used for his test. We generate the data with $\alpha_0 = 0$ and $\alpha_1 = c / \sqrt{T}$. Sample size is 100 and we iterate 1000 times.

The problem of local maxima may arise in García’s test (see Hamilton (1989) and García and Perron (1996)). We follow the standard method in the literature and estimate the model using EM algorithm with 7 sets of starting values and take the maximum over the values obtained.

In our test, we draw $\rho$ 100 times from the corresponding equispaced grid.

We plot the power as a function of $c$. To be fair, we put the power based on smaller interval for $\pi$ and $\rho$ in Figure 7.1, larger interval for both $\pi$ and $\rho$ in Figure 7.2.

Both tests tend to under reject. Yet our test perform better than Garcia’s in both cases.

In more general models, the asymptotically distribution of our test may not be nuisance parameter free. We rely on parametric bootstrap to compute the critical values. To find the maximum over $h$ and $\rho$ in (2.8), we generate $h$ uniformly over the unit sphere and $\rho$ is selected from an uniformly spaced grid of $(-0.7,0.7)$. The number of values for $h$ is 30 and that of $\rho$ is 60. We obtain the empirical critical values with 1000 iterations and sample size is taken to be 100. Then we plot the size-corrected power with the same number of iterations and same sample sizes for the following models.

Linear model with an intercept term:

$$y_t = x_t' \left( \beta + \frac{C\eta_t}{\sqrt{T}} \right) + \varepsilon_t$$

$$\varepsilon_t \sim iid \mathcal{N}(0,1)$$

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

20
In the simulations, we set $c_1 = c_2 = c$ and vary them. The size-corrected power as a function of $c$ is plotted in Figure 7.3.

ARCH(1) model:

$$y_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \left( \frac{1}{4} + \frac{c_1 \eta_t}{\sqrt{T}} \right) + \left( \frac{1}{4} + \frac{c_2 \eta_t}{\sqrt{T}} \right) y_{t-1}^2$$
$$\varepsilon_t \sim iid \mathcal{N}(0,1)$$

$\eta_t$ is a two-State Markov chain that takes the values 1 and $-1$ with transition probabilities $P(\eta_t = 1|\eta_{t-1} = 1) = 0.75$ and $P(\eta_t = -1|\eta_{t-1} = -1) = 0.75$. The size-corrected power is shown in Figure 7.4 as a function of $c = c_1 = c_2$.

IGARCH(1,1):

The model is as follows:

$$y_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \left( \frac{1}{2} + \frac{c_1 \eta_t}{\sqrt{T}} \right) + \left( \frac{1}{2} + \frac{c_2 \eta_t}{\sqrt{T}} \right) \sigma_{t-1}^2 + \left( \frac{1}{2} + \frac{c_3 \eta_t}{\sqrt{T}} \right) y_{t-1}^2$$
$$\varepsilon_t \sim iid \mathcal{N}(0,1)$$

Note that $\alpha_1 + \beta_1 = 1$. Here, we let $\eta_t$ take the values 0 and $-1$ with transition probabilities $P(\eta_t = 0|\eta_{t-1} = 0) = 0.75$ and $P(\eta_t = -1|\eta_{t-1} = -1) = 0.75$. $c_1$, $c_2$, and $c_3$ are taken to be equal. See size-corrected power in Figure 7.5.

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

5. Asymmetry in stock price dynamics

As noted by Neftci (1984) and Hamilton (2004), the major cyclical variables display an asymmetric behavior over various phases of the business cycle. In this section, we are interested in the stock price dynamics. Let $P_t$ and $D_t$ be the stock price index and dividend at time $t$. We use a two-step approach. First we estimate the following cointegration relationship between $\ln(P_t)$ and $\ln(D_t)$

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

(5.1)

by ordinary least-squares. As $D_t$ plays the role of fundamentals (in the spirit of Lucas, 1978, Blanchard and Watson, 1982, and Froot and Obstfeld, 1991), we expect the residual $y_t$ to exhibit periods of slow increases (expansions) and sharp declines (recessions). Then we fit on $y_t$ the Markov-switching model:

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

(5.2)
where \( \varepsilon_t \sim iid \mathcal{N}(0, \sigma^2) \). \( S_t \) is an exogenous three-state Markov chain that takes the values 1, 2, and 3 and has for transition probabilities \( 0 < p_{ij} < 1 \). Because the labels of the regimes are interchangeable, we set \( \beta_1 \geq \beta_2 \geq \beta_3 \). The parameter of interest is \( \theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \phi_{1i}, \phi_{2i}, \phi_{3i} : i = 1, \ldots, l, p_{ij} : i, j = 1, 2, 3)' \). The following proposition justifies our two-step approach.

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

**Data**

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

**Results**

Applying a BIC criterion on an autoregressive model reveals that 2 lags are best, hence we set \( l = 1 \) in Model (5.2). The augmented Dickey Fuller test rejects the null of a unit root on \( y_t \) at a 1% level. This permits to conclude that \( y_t \) is stationary. However, we know from Yao and Attali (2000) that markov-switching process may be stationary even if there is an explosive root in one of the regimes. Therefore testing the stationarity of \( \{y_t\} \) alone does not preclude the possibility of lapses of explosive behaviors or booms.

Now we wish to test whether \( y_t \) is better described by an AR(2) process with fixed coefficients or with markov-switching coefficients. We apply the supTS test (described in (2.7) and (2.8)) where the maximum over \( h \) and \( \rho \) is obtained by drawing \( h \) uniformly over the unit sphere (30 values used) and by taking the values of \( \rho \) in an equally spaced grid over \((-0.7, 0.7)\) (60 values used). Empirical critical values are computed from 1000 iterations for a sample size of 1576. The values of the parameters used to simulate the series are those obtained when estimating the model under \( H_0 \). The critical values are 5.6577635, 4.2483499, 3.7680360 at 1%, 5% and 10% respectively. The test statistic for our data is 22.938546. Hence our linearity test rejects strongly the null of a linear model versus a Markov-switching alternative, suggesting that at least two regimes should be used to fit the data. We estimate model (5.2) by maximum likelihood using the EM algorithm described in Hamilton (1989). We use 12 sets of starting values and select the one corresponding to the largest value of the likelihood.
<table>
<thead>
<tr>
<th>estimate</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>-0.100</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.038</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.195</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.002</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.010</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.321</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.057</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.216</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>1.431</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.001</td>
</tr>
</tbody>
</table>

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

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

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

Regime 1 ($S_t = 1$) has an explosive root with negative drift. In this regime, the trend (-0.1) dominates corresponding to declines of 10%. The second regime ($S_t = 2$) corresponds to a near unit-root with a slight positive drift. In this regime the process is stationary because the null hypothesis $H_0 : \beta_2 = 0$ is rejected. 94% of the data lies in this regime, which is very persistent. In Regime 2, $y_t$ increases slowly. Finally, in Regime 3 ($S_t = 3$), there are two complex roots. The coefficient of $\phi_3$ (1.43) is very large, as a result a large increase or decrease in the previous period is reinforced. By filtering, we compute the probabilities to be in Regime 1 conditional on the data: $P(S_t = 1 | y_1, ..., y_T)$. When $P(S_t = 1 | y_1, ..., y_T) > 0.5$, it is considered that the process at date $t$ is in Regime 1. The following months lie in Regime 1:

1873 (9-11), 1880 (4), 1893 (5-7), 1907 (3.8,10,11), 1917 (11), 1929 (10, 12), 1930 (5,6,10,12), 1931 (9,10,12), 1932 (4-6,10), 1933 (2), 1934 (5), 1937 (4,6,9,10), 1939 (4), 1940 (5), 1946 (9), 1950 (7), 1962 (5), 1970 (5), 1973 (11), 1974 (7,9), 1980 (3), 1981 (9), 1987 (10).

Similarly, we isolate the dates where Regime 3 dominates:

1873 (12), 1877 (7), 1880 (5), 1893 (9), 1907 (9, 12), 1929 (11), 1930 (11), 1931 (1,11), 1932 (1,7,8,11), 1933 (3,4,5), 1937 (5,7,11), 1938 (7,10), 1940 (6), 1950 (8), 1962 (6), 1973 (12), 1974 (8, 10), 1975 (1), 1987 (11).

In Regime 1, we recognize the big crashes such as October 1929 and October 1987. We can compare our results with those of Pagan and Sossounov (2003) on bull and bear markets. We see that Regime 1 identifies the month just preceding a trough of the US stock market cycles as reported in Pagan and Sossounov (1962/6, 1970/6, 1974/9, and 1987/11). Regime 1 corresponds mainly to big declines in the stock price. The
interpretation of Regime 3 is not as clear. While some dates correspond to big drops in the stock price (1929/11, 1962/6, 1987/11), other dates correspond to large increases (1932/8, 1933/5).

In Figure (7.6), we plot the graphs of the probabilities of the three regimes for the period from 1950 to 2002. In the background of each plot, there is the graph of the log stock price centered by its mean over the full dataset and rescaled so that the maximum value equal 1. The process \( y_t \) spends most of the time in the near unit-root regime 2. It follows an asymmetric pattern exhibiting slow increases and fast decreases that is well captured by a Markov-switching model.

6. Appendix A: Notations

6.1. 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 will get really complicated. Hence we will adopt some elements from multilinear algebra, which will facilitate our computations.

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

\[
\begin{align*}
\lambda M(x^{(1)}, x^{(2)}, \ldots, x^{(i)}_1, \ldots, x^{(p)}) + \mu M(x^{(1)}, x^{(2)}, \ldots, x^{(i)}_2, \ldots, x^{(p)}) \\
= M(x^{(1)}, x^{(2)}, \ldots, \lambda x^{(i)}_1 + \mu x^{(i)}_2, \ldots, x^{(p)}).
\end{align*}
\]

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

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

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

\[
x \to Df(x),
\]

24
which maps $O$ into $L(V,F)$, so we may use the concept of Frechet-differentiability again and differentiate $Df$. We then get the second derivative $D^2f(x)$. This second derivative at a point is a linear mapping from $V$ to $L(V,F)$ (an element from $L(V,L(V,F))$). That means that, for each $h \in V$, $D^2f(x)(h)$ is an element of $L(V,F)$, so for $k \in V$ $D^2f(x)(h)(k)$ is an element of $F$. Moreover, we can easily see that - by construction - the expression $D^2f(x)(h)(k)$ is linear in $h$ and $k$. Hence $D^2f(x)$ maps each pair $(h,k)$ into $F$ and is linear in each of the arguments, so we can think of $D^2f(x)$ as a bilinear form from $V \times V$ into $F$.

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

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

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

Most importantly, we have again a Taylor formula

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

with

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

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

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

$$D^nf \text{ is symmetric}$$

i.e.

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

for every permutation $\pi$.

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

$$g^{(n)}(0) = D^nf(x)(h,\ldots,h).$$

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

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

(6.11)
Symmetry implies that the left hand side of the above equation equals $2B(h, k) = 2B(k, h)$.

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

$$D(f \circ g) = Df(Dg).$$

(6.12)

The algebra of multilinear forms is often treated as a special case of tensor algebra. Although this branch of mathematics is well developed, it is rarely used in econometrics. Furthermore, many of the advanced concepts are of no use to us. Hence we will 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 $b_1, \ldots b_k$ is a basis for this space. Although the basis is arbitrary, we will from now on assume this basis to be fixed. It is essential for our approach that we fix the underlying vector space and the basis, since all of our definitions relate in one way or another to our chosen basis. It should be noted that we follow this approach not out of necessity - coordinate independent definitions of tensors are commonplace in differential geometry and mathematical physics, but purely out of convenience. E.g. we do not need to distinguish between co- and contravariant tensors - so we do not have to distinguish between “upper” and “lower” indices.

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

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

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 a multilinear form to be scalar. Let now $M$ be such a multilinear form Then, using linearity, we have

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

(6.14)
where the sum symbol corresponds to \( p \) sums extending over all values of \( i_1, \ldots, i_p \) between 1 and \( k \). So we can easily see that there is a one-to-one correspondence between the \( k^p \) numbers \( M(b_{i_1}, \ldots, b_{i_p}) \) 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 \( M(b_{i_1}, \ldots, b_{i_p}) \). 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, \( \ldots, K \), but our index set consists of the \( p \)-tuples \((1, \ldots, 1), (2, 1, \ldots), (k, k, \ldots, k)\)

This way we can work with multilinear forms and related mathematical objects without having to discuss tensor algebra. We can easily see that bilinear forms (forms of order two) are \( k \times k \) matrices.

1. We can easily see that multilinear forms form a vector space, and the mapping attaching each multilinear form its coordinates is an isomorphism. Hence we do not need to distinguish between multilinear forms and \( k^p \) numbers indexed by a multiindex \((i_1, \ldots, i_p)\).

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

\[
c_{i_1, \ldots, i_p} = c_{\pi(i_1), \ldots, \pi(i_p)}.
\]  

(6.15)

We can easily see that this property is equivalent to our definition above, (6.9)). For a form \( C \) defined by coordinates \((c_{i_1, \ldots, i_p})\) define its symmetrization \( C^{(S)} \) by

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

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

\[
C(h, \ldots, h) = C^{(S)}(h, \ldots, h),
\]

(6.16)

and, for any form \( C \), \( C^{(S)} \) is the only symmetrical form with the property (6.16).

3. Another special case of multilinear forms are our derivatives of scalar functions defined on open subsets of our space \( V \). We can easily see that the coordinates \( D^p f \) can be calculated in the following way. Define the function \( g \) by

\[
g(x_1, \ldots, x_p) = f(\sum x_i b_i),
\]

(6.17)

where the \( b_i \) are our fixed basis vectors. Then the corresponding coordinates of the derivative are given by \( \left( \frac{\partial^p g}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} \right)_{(i_1, \ldots, i_n)} \).
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

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

(6.18)

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

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

(6.19)

Then - due to (6.16) and symmetry of the derivative - we can conclude that $D^n f = C^{(S)}$.

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

$$a_{i_1, \ldots, i_p} b_{i_{p+1}, \ldots, i_{p+q}}.$$  

(6.20)

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

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

(6.21)

The tensor product of symmetric forms, however, in general is not 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 $(t_{i_1, \ldots, i_p})$, $C$ is a form with coordinates $(c_{i_1, \ldots, i_p})$ we have

$$\langle T, C \rangle = \sum t_{i_1, \ldots, i_p} c_{i_1, \ldots, i_p}.$$  

(6.22)

7. This scalar product is useful in computing expectation of multilinear forms with random arguments. First of all let us observe that each vector $h \in V$ has exactly $k$ coordinates. Since (6.14) defines for each set of coordinates a form, we can identify $h$ with an 1-form (i.e. a linear form with one argument). We will use the same symbol $h$ for this form. Now let $h_1, \ldots, h_p \in V$. Then we can use (6.20) to define $h_1 \otimes \ldots \otimes h_p$.

Now suppose we want to compute the value of the multilinear form $T(h_1, \ldots, h_p)$. Then we can see from (6.14),(6.22) that $T(h_1, \ldots, h_p)$ equals $\langle T, h_1 \otimes \ldots \otimes h_p \rangle$. Let $H_1, \ldots, H_p$ 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

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

(6.23)
Since \( T(H_1, ..., H_p) = \langle T, H_1 \otimes ... \otimes H_p \rangle \), and since \( T \) is independent of the \( H_i \), we can easily see that

\[
ET(H_1 \otimes ... \otimes H_p) = \langle T, E(H_1 \otimes ... \otimes H_p) \rangle,
\]

provided the expectations exist (a sufficient condition is e.g. \( E \|H_1\|...\|H_p\| < \infty \): \( H_1 \otimes ... \otimes H_p \) 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, we can easily see that (6.24) is valid for conditional expectations, too. Moreover, we can easily see that we have an analogous result if \( T \) and the \( H_1, ..., H_p \) are independent. In the sequel, we will use this type of identities rather freely.

8. Most of the proof of our theorem will be an evaluation of some kind of expectations multilinear forms representing derivatives. The notation using the bracket \( \langle .., .. \rangle \) would be rather clumsy. So we propose to use a more suggestive notation: Instead of

\[
\langle T, C \rangle
\]

we will use

\[
T(C),
\]

i.e. we use the form \( C \) as an argument. With this notation, we can write (6.24) as

\[
E(T(H_1 \otimes ... \otimes H_p)) = T(E(H_1 \otimes ... \otimes H_p)).
\]

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 we can easily see the for every \( T \),

\[
T(A) = T^{(S)}(A).
\]

In particular, if we have an arbitrary random vector \( H \) (with a sufficient number of moments) then \( E(H \otimes ... \otimes H) \) is symmetrical, hence (6.28) implies that, for all forms \( T \),

\[
T(E(H \otimes ... \otimes H)) = T^{(S)}(E(H \otimes ... \otimes H))
\]

\( ).

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 which norm we use. Of particular interest, however, is the norm

\[
\|T\| = \sqrt{\sum t_{i_1,...,i_p}^2}
\]

(6.30)
where the $t_{i_1, \ldots, i_p}$ are the coordinates of $T$. Cauchy-Schwartz inequality and (6.22) imply that for all $T, C$:

$$|T(C)| \leq \|T\| \|C\|. \quad (6.31)$$

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

### 6.2. Other notations

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

The sample is split in the following way:

$$t = 1, 2, \ldots, T_1, T_1 + 1, \ldots, T_2, \ldots, T_{i-1} + 1, \ldots, T_i, \ldots, T_{B_N-1} + 1, \ldots, T_{B_N}$$

$1$st block $2$nd block $i$th block $B_N$th block

There are $B_N$ blocks and each block has $B_L$ or $B_L - 1$ elements. $i$ is the index for the block $i = 1, \ldots, B_N$. 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.

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

$$D^k l_t = l_t^{(k)} \quad (6.32)$$

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

$$\ell_t(u) = \log f_t(\theta_T + uh) \quad (6.33)$$

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

- **1st derivative:** $\ell_t^{(1)} = \frac{f'}{f}$.
- **2nd derivative:** $\ell_t^{(2)} = \frac{f'^2}{f} - \frac{f'}{f^2} f'$. 
- **3rd derivative:** $\ell_t^{(3)} = \frac{f'^3}{f} - \frac{f'^2}{f^2} f' - \frac{2 f'^2}{f^2} f^{(2)} + \frac{2 f'^2}{f^3} f'$. 
- **4th derivative:** $\ell_t^{(4)} = \frac{f'^4}{f} - \frac{f'^3}{f^2} f' - \frac{3 f'^3}{f^2} f^{(2)} + \frac{3 f'^3}{f^3} f^{(2)} f' + \frac{6 f'^3}{f^3} f^{(2)} f' + \frac{6 f'^2}{f^3} f^{(2)} f' - \frac{6 f'^2}{f^4} f'$. 

30
According to the formalism outlined previously, we can conclude that \( \ell_t^{(k)} = \ell_t^{(k)}(h, \ldots, h) \) and that \( f^{(k)} = D^k f(h, \ldots, h) \). Taking into account our characterization of the tensor product (6.21), and the techniques described above, we can conclude that

\[
\begin{align*}
    \ell_t^{(1)} &= (1/f_t) D f_t, \\
    \frac{1}{f_t} D^2 f_t &= \ell_t^{(2)} + \ell_t^{(1)} \otimes \ell_t^{(1)}, \\
    \frac{1}{f_t} D^3 f_t &= \left( \ell_t^{(3)} + 3 \ell_t^{(2)} \otimes \ell_t^{(1)} + \ell_t^{(1)} \otimes \ell_t^{(1)} \right)^{(S)}, \\
    \frac{1}{f_t} D^4 f_t &= \left( \ell_t^{(4)} + 6 \ell_t^{(2)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} + 4 \ell_t^{(3)} \otimes \ell_t^{(1)} + 3 \ell_t^{(2)} \otimes \ell_t^{(2)} + \ell_t^{(1)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} \right)^{(S)}.
\end{align*}
\]

We can easily see that we do not need to symmetrize (6.34), since the form on the right hand side is symmetrical. Let us now denote by \( \mathcal{F}_t \) the \( \sigma \)-algebra generated by the data \( y_t, y_{t-1}, \ldots \). Note that \( \mathcal{F}_t = \mathcal{H}_{0,t} \). Then one can easily see that for \( k \leq 4 \), we have for arbitrary \( h \) \( E(\frac{1}{f_t} D^k f_t(h, \ldots, h) | \mathcal{F}_{t-1}) = \int \frac{1}{f_t} D^k f_t(h, \ldots, h) f_t d\mu(y_t) = \int D^k f_t(h, \ldots, h) d\mu(y_t) \), where \( \mu \) is the dominating measure defined in Section 2. Since we assumed \( f_t \) to be at least 5 times differentiable (and the 5th derivative to be uniformly integrable), we can easily see (Bartle, 1966, Corollary 5.9) that we can interchange integral and differentiation, and conclude that \( \int D^k f_t(h, \ldots, h) d\mu(y_t) = D^k(\int f_t d\mu(y_t))(h, \ldots, h) = 0 \), since all the \( f_t \) as conditional densities integrate to one.

Let us define

\[
\begin{align*}
    m_{2,t} &= \left( \ell_t^{(2)} + \ell_t^{(1)} \otimes \ell_t^{(1)} + 2 \ell_t^{(1)} \otimes L_{t-1} \right)^{(S)}, \\
    m_{3,t} &= \left( \ell_t^{(3)} + 3 \ell_t^{(2)} \otimes \ell_t^{(1)} + \ell_t^{(1)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} \right)^{(S)}, \\
    m_{4,t} &= \left( \ell_t^{(4)} + 6 \ell_t^{(2)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} + 4 \ell_t^{(3)} \otimes \ell_t^{(1)} + 3 \ell_t^{(2)} \otimes \ell_t^{(2)} + \ell_t^{(1)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} \otimes \ell_t^{(1)} \right)^{(S)}.
\end{align*}
\]

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

\[
M_j = \sum_{t=T_{t-1}+1}^{T_t} m_{j,t}, j = 2, 3, 4.
\]

It follows that \( \ell_t^{(1)} \), \( m_{2,t} \), \( m_{3,t} \), \( m_{4,t} \) are martingale difference sequences with respect to the \( \mathcal{F}_t \). Furthermore, the \( m_{j,t}, j = 2, 3, 4 \) are defined as symmetrizations of the multilinear form on the rhs of the above equations. Finally, we denote \( L_1 = \sum_{t=T_{t-1}+1}^{T_t} \ell_t^{(1)} \).

In the sequel, we will heavily rely on (6.28), both for the evaluation of the \( m_{j,t}, j = 2, 3, 4 \) and the derivatives as well. Note that when we evaluate forms with symmetrical arguments, it is irrelevant whether we use the forms themselves or the nonsymmetrical expressions used in the above definitions.
7. Appendix B: Proofs

The first theorem we want to prove is 3.1. The statement of the theorem involves some uniform convergence in probability of a parametrized family of random variables. First assume the theorem would not be true. There would exist a compact subset $K \subset \Theta \times B$ 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

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

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

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

or

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

In the sequel, we will prove this relationship. To simplify our notation, however, we will suppress the parameters $(\theta_T, \beta_T)$ and $(\theta_0, \beta_0)$. When analyzing expressions related to a sample of length $T$, we simply write $E$ and $P$ instead of $E_{\theta_T}$ and $P_{\theta_T}$. Moreover, we also will drop the argument from expression like $l_t(\theta_T), ...$ and simply use $l_t, ...$ The proper argument should be evident from the context. This simplification of 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 $\theta_T$, which is a bilinear form and $l_t^{(2)}(h, k)$ is the evaluation of this bilinear form with the arguments $h$ and $k$. In the sequel, $\sum_i = \sum_{t=T_i-1+1}$ and $\sum_i = \sum_{i=1}^{B_N}$ where $B_N$ is the number of blocks as defined in Appendix A.

The following lemmas are used in the proof of Theorem 3.1.

**Lemma 7.1.** Assume that for any $\varepsilon > 0$, we can find $1 - \varepsilon \leq \frac{\ell_T}{f_T} \leq 1 + \varepsilon$ on some set $A_T^\varepsilon$ so that $\lim_{T \to \infty} \sup P\left(A_T^\varepsilon \right) = 1$ where $A_T^\varepsilon$ is $\mathcal{H}_{0,T}$-measurable and independent of $\beta$. Then

$$
\sup_{\beta} E\left(f_T | \mathcal{H}_{0,T}\right) \to 1.
$$

Note that a sufficient condition for Lemma 7.1 is

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

32
where $C_T$ is $\mathcal{H}_{0,T}$-measurable and independent of $\beta$ and $C_T \xrightarrow{P} 0$.

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

$$\sup_{\beta} \left| \sum_{i=1}^{B_N} \Delta_{i,T} \right| \leq C_T \quad (7.4)$$

and

$$\sup_{\beta} \sum_{i=1}^{B_N} \Delta_{i,T}^2 \leq D_T. \quad (7.5)$$

Then

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

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

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

and $B_N \lambda^{B_L} \to 0$. Then (7.6) is satisfied.

**Lemma 7.4.** Let $a_1, a_2, ..., a_N$ be a sequence of numbers for some integer $N \geq 1$. Then

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

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

$$\sum_i E |\Delta_i| \to 0. \quad (7.7)$$

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

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

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

(7.8)
for some \( m \geq 1 \) and all \( j = 1, 2, 3, 4 \). Let \( k \leq 4 \) and \( \mathcal{D} = \{d_1, \ldots, d_k\} \) be any \( k \)-partition of the integers 1, 2, 3, 4. Assume that

\[
\sum_{j \in \mathcal{D}} \alpha_j > 1,
\]

and let \( B_L \) be such that

\[
\frac{B_L^{m-1}}{T \sum_{j \in \mathcal{D}} \alpha_j} = o(1).
\]

Then, Conditions (7.4) and (7.5) are satisfied for \( \Delta_{i,T} = E \left( \prod_{j \in \mathcal{D}} x_{ij} \mathcal{H}_{T_i,T} \right) \).

**Lemma 7.7.** Assume Assumption 4 holds. Let \( \text{const} \) denotes a constant independent of \( \beta \), we have

\[
E \left( \|\bar{x}_{i1}\|^4 \right) \equiv E \left( \|L_1\|^4 \right) \leq \text{const} B_L^4,
\]

\[
E \left( \|\bar{x}_{i2}\|^4 \right) \equiv E \left( \|M_2\|^4 \right) \leq \text{const} B_L^8,
\]

\[
E \left( \|\bar{x}_{i3}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} m_{2,t} \right\|^4 \right) \leq \text{const} B_L^8,
\]

\[
E \left( \|\bar{x}_{i4}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} m_{2,t} \right\|^4 \right) \leq \text{const} B_L^8,
\]

\[
E \left( \|\bar{x}_{i5}\|^4 \right) \equiv E \left( \|M_3\|^4 \right) \leq \text{const} B_L^4,
\]

\[
E \left( \|\bar{x}_{i6}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} m_{3,t} \right\|^4 \right) \leq \text{const} B_L^4,
\]

\[
E \left( \|\bar{x}_{i7}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} L_{t-1}^2 \right\|^4 \right) \leq \text{const} B_L^{12},
\]

\[
E \left( \|\bar{x}_{i8}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} L_{t-1}^2 \right\|^4 \right) \leq \text{const} B_L^{12},
\]

\[
E \left( \|\bar{x}_{i9}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} L_{t-1}^3 \right\|^4 \right) \leq \text{const} B_L^{16},
\]

\[
E \left( \|\bar{x}_{i10}\|^4 \right) \equiv E \left( \left\| \sum_t l_t^{(1)} L_{t-1} m_{2,t} \right\|^4 \right) \leq \text{const} B_L^{16}.
\]
Proof of Lemma 7.1. Let $\eta$ be an arbitrary positive number and $0 < \varepsilon < \eta$.

\[
\sup_{\beta} E(f_T|\mathcal{H}_{0,T}) = \sup_{\beta} E\left( \frac{f_T}{f_T^*} f_T^* |\mathcal{H}_{0,T} \right) = I_{A_T^\varepsilon}(1 - \varepsilon) \sup_{\beta} E(f_T|\mathcal{H}_{0,T}) + I_{(A_T^\varepsilon)^c} \sup_{\beta} E(f_T|\mathcal{H}_{0,T}) \cdot
\]

Under the assumptions of the lemma:

\[
I_{A_T^\varepsilon}(1 - \varepsilon) \sup_{\beta} E(f_T^*|\mathcal{H}_{0,T}) + I_{(A_T^\varepsilon)^c} \sup_{\beta} E(f_T|\mathcal{H}_{0,T}) \
\leq \sup_{\beta} E(f_T|\mathcal{H}_{0,T}) \
\leq I_{A_T^\varepsilon}(1 + \varepsilon) \sup_{\beta} E(f_T^*|\mathcal{H}_{0,T}) + I_{(A_T^\varepsilon)^c} \sup_{\beta} E(f_T|\mathcal{H}_{0,T}).
\]

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

\[
P[|X_T - 1| < \eta] = P[1 - \eta < X_T < 1 + \eta] \
\geq P \left( \left\{ I_{A_T^\varepsilon}(1 + \varepsilon) + I_{(A_T^\varepsilon)^c} X_T < 1 + \eta \right\} \cap \left\{ I_{A_T^\varepsilon}(1 - \varepsilon) + I_{(A_T^\varepsilon)^c} X_T > 1 - \eta \right\} \right] \
= P(A_T^\varepsilon) + P((A_T^\varepsilon)^c) P[1 - \eta < X_T < 1 + \eta] \
\geq P(A_T^\varepsilon) \rightarrow 1
\]

where the last equality follows from the law of total probability. Hence $X_T \overset{P}{\rightarrow} 1$.

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

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

uniformly in $\beta$, or more precisely

\[
1 - \varepsilon \leq \prod_{i=1}^{B_N} (1 + \Delta_{i,T}) \leq 1 + \varepsilon
\]

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

35
Using iterated expectations and the definition of \( \Delta_{i,T} \), we obtain

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

Hence on \( A^c_T \), we have

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

or equivalently

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

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

**Proof of Lemma 7.3.**

\[
\ln E \left[ \prod_{i=1}^{B_N} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right] = \sum_{l=1}^{B_N} \left\{ \ln E \left[ \prod_{i=1}^{l} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right] - \ln E \left[ \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right] \right\}
\]

where

\[
u_l = E \left[ \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right],
\]

\[
h_l = E \left[ \prod_{i=1}^{l} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right] - E \left[ \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right]
\]

\[
\quad = E \left[ x_l \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right]
\]

\[
\quad = E \left[ \left( x_l | \mathcal{H}_{T_{l-1},T} \right) \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right]
\]

\[
\quad = E \left[ \Delta_{l,T} \prod_{i=1}^{l-1} (1 + x_i) \left| \mathcal{H}_{0,T} \right. \right].
\]

36
Using a Taylor expansion, we have

\[ \left| \sum_{l=1}^{B_N} \left\{ \ln (u_l + h_l) - \ln (u_l) - \frac{h_l}{u_l} \right\} \right| \leq \sum_{l=1}^{B_N} \frac{h_l^2}{2 \left| u_l \right|^2 - \left| h_l \right|^2} \]

\[ = \frac{1}{2} \sum_{l=1}^{B_N} \left( \frac{h_l}{u_l} \right)^2 \frac{1}{1 - \left( \frac{h_l}{u_l} \right)^2} \]

\[ \leq \frac{1}{2} \sum_{l=1}^{B_N} \left( \frac{h_l}{u_l} \right)^2 \frac{1}{1 - \sum_{l=1}^{B_N} \left( \frac{h_l}{u_l} \right)^2}. \quad (7.10) \]

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

\[ \sum_{l=1}^{B_N} \delta_l^P \to 0, \quad (7.11) \]

\[ \sum_{l=1}^{B_N} \delta_l^2 \to 0, \quad (7.12) \]

then we have

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

which itself implies

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

(7.11) will follow from (7.12) and the fact that \( \delta_l \) is a martingale as \( \Delta_{l,T} \) is itself a martingale. Now we want to show that

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

We have

\[ \delta_l = \frac{E \left[ \Delta_l \prod_{i=1}^{l-1} (1 + x_i) \mid \mathcal{H}_{0,T} \right]}{E \left[ \prod_{i=1}^{l-1} (1 + x_i) \mid \mathcal{H}_{0,T} \right]} \]

\[ = \frac{E \left[ \Delta_l \prod_{i=1}^{l-2} (1 + x_i) (1 + x_{l-1}) \mid \mathcal{H}_{0,T} \right]}{E \left[ \prod_{i=1}^{l-2} (1 + x_i) (1 + x_{l-1}) \mid \mathcal{H}_{0,T} \right]} \]

\[ \leq \frac{3E \left[ \Delta_{l,T} \prod_{i=1}^{l-2} (1 + x_i) \mid \mathcal{H}_{0,T} \right]}{E \left[ \prod_{i=1}^{l-2} (1 + x_i) \mid \mathcal{H}_{0,T} \right]} \]
because $\|x_{l-1}\| \leq 1/2$ by assumption. Note that $E \left( \Delta_{i,T} | \mathcal{H}_{T_{i-2},T} \right) \leq E \left( \Delta_{i,T} | \mathcal{H}_{T_{i-2},T} \right)$ and it follows from the geometric ergodicity of $\eta_t$ that

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

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

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

Let

$$\delta_i = \frac{3E \left[ \Delta_{i,T} \right] \prod_{l=1}^{l-2} \left( 1 + x_i \right) | \mathcal{H}_{0,T} \right]}{E \left[ \prod_{l=1}^{l-2} \left( 1 + x_i \right) | \mathcal{H}_{0,T} \right]} \leq 3 \lambda^{B_L} \frac{E \left[ g \left( \mathcal{H}_{T_{i-2},T} \right) \prod_{l=1}^{l-2} \left( 1 + x_i \right) | \mathcal{H}_{0,T} \right]}{E \left[ \prod_{l=1}^{l-2} \left( 1 + x_i \right) | \mathcal{H}_{0,T} \right]} + 3E \left( \Delta_{i,T} | \mathcal{H}_{0,T} \right).$$

Therefore,

$$\delta_i^2 \leq O \left( \lambda^{B_L} \right) + 9E \left( |\Delta_{i,T}|^2 | \mathcal{H}_{0,T} \right).$$

We get

$$\sum_{l=1}^{B_N} E \left( \delta_i^2 \right) \leq O \left( B_N \lambda^{B_L} \right) + 9 \sum_{l=1}^{B_N} E \left( |\Delta_{i,T}|^2 \right).$$

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

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

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

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

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

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

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

Hence

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

38
The last statement of the lemma follows from

\[ E \Delta \leq E \left[ E \left( \prod_{j=1}^{k} x_{ij} \mid \mathcal{H}_{T_{i-1}, T} \right) \right] \leq E \left( \prod_{j=1}^{k} x_{ij} \right) = E \left( \prod_{j=1}^{k} x_{ij} \right) . \]

**Proof of Lemma 7.7.** Term \( \bar{x}_{i1} : \)

\[ E |L_t|^4 \leq B_L^3 \sum_{t=T_{i-1}+1}^{T_i} E \|l_t^{(1)}\|^4 \leq B_L^4 \sup_t E \|l_t^{(1)}\|^4 \]

by Lemma 7.4.

Term \( \bar{x}_{i2} : \)

\[ E (|M_2|)^4 \leq B_L^3 \sum_t E (|m_{2,t}|)^4 \]
\[ \leq B_L^3 \sum_t E \left( \|l_t^{(2)}\|^4 + \|l_t^{(1)}\|^8 + 2^4 \|l_t^{(1)}\|^4 \|L_{t-1}\|^4 \right) \]
\[ \leq B_L^3 \sum_t \left( E \|l_t^{(2)}\|^4 + E \|l_t^{(1)}\|^8 + 2^4 \left( E \|l_t^{(1)}\|^8 E (\|L_{t-1}\|^8) \right)^{1/2} \right) \]
\[ \leq \text{const} B_L^8 \]

provided \( \sup E \|l_t^{(1)}\|^8 < \infty \) and \( \sup E \|l_t^{(2)}\|^4 < \infty \), which is true by Assumption 3.

Term \( \bar{x}_{i3} : \)

\[ E \left( \sum_t l_t^{(1)} m_{2,t} \right)^4 \]
\[ = E \left( \sum_t (l_t^{(1)} l_t^{(2)} + l_t^{(1)} l_t^{(1)} + 2 l_t^{(1)} L_{t-1}) \right)^4 \]
\[ \leq B_L^3 \sum_t E \left( \|l_t^{(1)} l_t^{(2)}\|^4 + \|l_t^{(1)}\|^8 + 2^4 \|l_t^{(1)} L_{t-1}\|^4 \right) \]
\[ \leq B_L^3 \sum_t \left( E \|l_t^{(1)}\|^8 E \|l_t^{(2)}\|^8 \right)^{1/2} + E \|l_t^{(1)}\|^8 + 2^4 \left( E \|l_t^{(1)}\|^8 E (\|L_{t-1}\|^8) \right)^{1/2} \]
\[ \leq \text{const} B_L^8 \]

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

Term \( \bar{x}_{i4} : \)

39
\[ E \left( \sum_t l_t^{(1)} m_{2,t} \right)^4 \]
\[ = E \left( \sum_t l_t^{(1)} l_t^{(2)} + 2 l_t^{(1)} L_{t-1} \right)^4 \]
\[ \leq B_L^3 \sum_t E \left( \| l_t^{(1)} \|^4 + \| l_t^{(1)} \|^{16} + 24 \| l_t^{(1)} L_{t-1} \|^4 \right) \]
\[ \leq B_L^3 \sum_t \left( E \| l_t^{(1)} \|^{16} E \| l_t^{(2)} \|^{16} \right)^{1/2} + E \| l_t^{(1)} \|^{16} + 24 \left( E \| l_t^{(1)} \|^{24} E \left( \| L_{t-1} \|^8 \right) \right)^{1/2} \]
\[ \leq \text{const} B_L^3 \]

provided that \( \sup E \| l_t^{(1)} \|^{24} < \infty \) and \( \sup E \| l_t^{(2)} \|^{16} < \infty \).

Term \( \bar{x}_{i6} \):

\[ E \| M_3 \|^4 \leq B_L^3 \sum_t E \left( \| l_t^{(3)} \|^4 + \| l_t^{(2)} l_t^{(1)} \|^4 + \| l_t^{(1,3)} \|^4 \right) \]
\[ \leq B_L^3 \sum_t E \| l_t^{(3)} \|^4 + \left( E \| l_t^{(2)} \|^8 E \| l_t^{(1)} \|^8 \right)^{1/2} + E \| l_t^{(1)} \|^{12} \]
\[ \leq \text{const} B_L^4 \]

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

Term \( \bar{x}_{i6} \):

\[ E \left( \sum_{t=T_{i+1}}^{T_{i+1}} m_{3,t} l_t^{(1)} \right)^4 \leq B_L^2 \sum_t E \left( m_{3,t} l_t^{(1)} \right)^4 \]
\[ \leq B_L^2 \sum_t E \left( \| l_t^{(3)} l_t^{(1)} \|^4 + \| l_t^{(2)} l_t^{(1)} \|^4 + \| l_t^{(1,4)} \|^4 \right) \]
\[ \leq \text{const} B_L^4 \]

provided that \( \sup E \| l_t^{(1)} \|^{16} < \infty \), \( \sup E \| l_t^{(2)} \|^{8} < \infty \) and \( \sup E \| l_t^{(3)} \|^8 < \infty \).
Term $\tilde{x}_{17}$:

$$E \left( \sum_t l_t^{(1)} L_{t-1}^2 \right)^4 \leq B_L^3 \sum_t E \left| l_t^{(1)} L_{t-1}^2 \right|^4$$

$$\leq B_L^3 \sum_t \left( E \left\| l_t^{(1)} \right\|^8 \frac{E \left\| L_{t-1} \right\|^{16}}{2} \right)^{1/2}$$

$$\leq \text{const} B_L^{12}$$

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

Term $\tilde{x}_{18}$:

$$E \left( \sum_t l_t^{(1)2} L_{t-1}^2 \right)^4 \leq B_L^3 \sum_t E \left| l_t^{(1)2} L_{t-1}^2 \right|^4$$

$$\leq B_L^3 \sum_t \left( E \left\| l_t^{(1)} \right\|^{16} \frac{E \left\| L_{t-1} \right\|^{16}}{2} \right)^{1/2}$$

$$\leq \text{const} B_L^{12}$$

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

Term $\tilde{x}_{19}$:

$$E \left( \sum_t l_t^{(1)} L_{t-1}^3 \right)^4 \leq B_L^3 \sum_t E \left| l_t^{(1)} L_{t-1}^3 \right|^4$$

$$\leq B_L^3 \sum_t \left( E \left\| l_t^{(1)} \right\|^8 \frac{E \left\| L_{t-1} \right\|^{24}}{2} \right)^{1/2}$$

$$\leq \text{const} B_L^{16}$$

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

Term $\tilde{x}_{10}$:
\[
E \left( \sum_t l_t^{(1)} L_{t-1} m_{2,t} \right)^4 = E \left| \sum_t l_t^{(1)} l_t^{(2)} L_{t-1} + l_t^{(1)3} L_{t-1} + 2 l_t^{(1)2} L_{t-1}^2 \right|^4 \\
\leq B_L^3 \sum_t E \left| l_t^{(1)} l_t^{(2)} L_{t-1} \right|^4 + E \left| l_t^{(1)3} L_{t-1} \right|^4 + 24 E \left| l_t^{(1)2} L_{t-1}^2 \right|^4 \\
\leq B_L^3 \sum_t \left\{ E \left| l_t^{(1)} \right|^{12} + E \left| l_t^{(2)} \right|^{12} + E \left| L_{t-1} \right|^{12} \\
+ \left( E \left| l_t^{(1)} \right|^{24} E \left| L_{t-1} \right|^{8} \right)^{1/2} + 24 \left( E \left| l_t^{(1)} \right|^{16} E \left| L_{t-1} \right|^{16} \right)^{1/2} \right\} \\
\leq \text{const} B_L^{16}
\]
provided \( \sup E \left| l_t^{(1)} \right|^{24} < \infty \) and \( \sup E \left| l_t^{(2)} \right|^{12} < \infty \).

**Proof of Theorem 3.1**

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

\[
T E_T = \sum_{t=1}^T \left[ \frac{1}{\sqrt{T}} l_t^{(1)} (\eta_t) + \frac{1}{2 \sqrt{T}} l_t^{(2)} (\eta_t, \eta_t) + \frac{1}{6 \sqrt{T}} l_t^{(3)} (\eta_t, \eta_t, \eta_t) + \frac{1}{24 \sqrt{T}} l_t^{(4)} (\eta_t, \eta_t, \eta_t, \eta_t) \right] \\
\equiv \sum_{t=1}^T T E_t, \quad (7.15)
\]

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

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

**Step 1.** Using Lemma 7.1, we show that

\[
\frac{\ell_T^\beta (\theta)}{E \left[ \exp (T E_T) \left| \mathcal{H}_{0,T} \right. \right]} \stackrel{P}{\rightarrow} 1
\]

uniformly in \( \beta \).

**Step 2.** Using Lemma 7.1, we show that

\[
\frac{E \left[ \exp (T E_T) \left| \mathcal{H}_{0,T} \right. \right]}{E \left[ \exp \left( \widehat{S}_T + \sum_{i=1}^{B_N} \sum_{j=1}^{d} \ln (1 + x_{ij}) \right) \left| \mathcal{H}_{0,T} \right. \right]} \stackrel{P}{\rightarrow} 1
\]

uniformly in \( \beta \) for some \( x_{ij} \).
Step 3. Using Lemma 7.2, we prove that

\[
E \left[ \frac{\exp \left( \tilde{T}S_T + \sum_{i=1}^{B_N} \sum_{j=1}^{i} \ln (1 + x_{ij}) \right) | \mathcal{H}_{0,T}}{\exp \left( \tilde{T}S_T \right)} \right] \xrightarrow{p} 1
\]

uniformly in \( \beta \).

Then, result (3.3) follows from

\[
\frac{\ell^3_T (\theta)}{\exp \left( \tilde{T}S_T \right)} = \frac{E \left[ \exp (T \eta) | \mathcal{H}_{0,T} \right]}{E \left[ \exp (T \tilde{E} | \mathcal{H}_{0,T}) \right] E \left[ \exp \left( \tilde{T}S_T + \sum_{i=1}^{B_N} \sum_{j=1}^{i} \ln (1 + x_{ij}) \right) | \mathcal{H}_{0,T} \right]} \times \frac{E \left[ \exp \left( \tilde{T}S_T + \sum_{i=1}^{B_N} \sum_{j=1}^{i} \ln (1 + x_{ij}) \right) | \mathcal{H}_{0,T} \right]}{\exp \left( \tilde{T}S_T \right)}.
\]

Step 1. Using a Taylor expansion, we obtain

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

by Assumption 4. The result follows from Lemma 7.1.

In the sequel, we will use sup instead of \( \sup_{t, \theta \in \mathcal{N}} \) to simplify notation.

Step 2.
Let \( TE_T = \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} TE_{it} \), \( \tilde{T}S_T = \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} \tilde{T}S_{it} \).

\[
TE_T - \tilde{T}S_T = \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (TE_{it} - \tilde{T}S_{it}) \\
= \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (TE_{it} - \tilde{T}S_{it}) + \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (\tilde{T}S_{it} - \tilde{T}S_{it}) \quad (7.16)
\]

43
where
\[ \widehat{T}_S_{it} = \frac{1}{2\sqrt{T}} E \left( m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) - \frac{1}{8T} \left[ E \left( m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) \right]^2. \]

In the sequel \( \eta_t \) is split in the following manner
\[ \begin{align*}
\eta_t &= \xi_t + \alpha_t, \\
\xi_t &= \eta_t - E(\eta_t | \mathcal{H}_{T_{i-1},T}), \\
\alpha_t &= E(\eta_t | \mathcal{H}_{T_{i-1},T}).
\end{align*} \]

\[ \sum_t \widehat{T}_S_{it} \] can be decomposed as follows:
\[ \sum_{t=T_{i-1}+1}^{T_i} \widehat{T}_S_{it} = \sum_{t=T_{i-1}+1}^{T_i} \widehat{T}_S_{it}(\xi) + \sum_{t=T_{i-1}+1}^{T_i} \widehat{T}_S_{it}(\alpha), \]
\[ \sum_{t=T_{i-1}+1}^{T_i} \widehat{T}_S_{it}(\xi) = \frac{1}{2\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E \left( m_{2,t} (\xi) | \mathcal{H}_{T_{i-1},T} \right) - \frac{1}{8T} \left[ E \left( m_{2,t} (\xi) | \mathcal{H}_{T_{i-1},T} \right) \right]^2, \]
\[ \sum_{t=T_{i-1}+1}^{T_i} \widehat{T}_S_{it}(\alpha) = \frac{1}{2\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E \left( m_{2,t} (\alpha) | \mathcal{H}_{T_{i-1},T} \right) - \frac{1}{8T} \left[ E \left( m_{2,t} (\alpha) | \mathcal{H}_{T_{i-1},T} \right) \right]^2. \]

The mixed terms vanish because
\[ E \left( \alpha_t \otimes \xi_t | \mathcal{H}_{T_{i-1},T} \right) = 0. \]

Similarly, the Taylor Expansion in (7.14) can be rewritten as the sum of three parts, namely, the pure part w.r.t. \( \xi_t \), the pure part w.r.t. \( \alpha_t \), and the mixed part. That is,
\[ \begin{align*}
TE_{it}(\xi_t) &= \frac{1}{\sqrt{T}} j^{(1)}(\xi_t) + \frac{1}{\sqrt{T}} j^{(2)}(\xi_t, \xi_t) + \frac{1}{\sqrt{T^3}} j^{(3)}(\xi_t, \xi_t, \xi_t) + \frac{1}{2\sqrt{T^4}} j^{(4)}(\xi_t, \xi_t, \xi_t, \xi_t), \\
TE_{it}(\alpha_t) &= \frac{1}{\sqrt{T}} j^{(1)}(\alpha_t) + \frac{1}{\sqrt{T}} j^{(2)}(\alpha_t, \alpha_t) + \frac{1}{\sqrt{T^3}} j^{(3)}(\alpha_t, \alpha_t, \alpha_t) + \frac{1}{2\sqrt{T^4}} j^{(4)}(\alpha_t, \alpha_t, \alpha_t, \alpha_t)
\end{align*} \]
and
\[ \begin{align*}
TE_{it}(\xi_t, \alpha_t) &= \frac{1}{2\sqrt{T}} j^{(2)}(\xi_t, \alpha_t) + \frac{1}{6\sqrt{T^3}} j^{(3)}(\xi_t, \xi_t, \alpha_t) + \frac{1}{6\sqrt{T^3}} j^{(3)}(\xi_t, \alpha_t, \alpha_t) + \frac{1}{24T} j^{(4)}(\xi_t, \xi_t, \xi_t, \alpha_t) + \frac{1}{24T} j^{(4)}(\xi_t, \xi_t, \alpha_t, \alpha_t) + \frac{1}{24T} j^{(4)}(\xi_t, \alpha_t, \alpha_t, \alpha_t)
\end{align*} \]

Remark that for any linear function \( g \) and using the convention \( \alpha_t = 0 \) for \( t > T \), we can write
\[ \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} \left( g(\eta_t) \right) = \sum_{i=1}^{B_N} \left( \sum_{t=T_{i-1}+1}^{T_i} \left( g(\xi_t) \right) + \sum_{t=T_{i+1}}^{T_{i+1}} \left( g(\alpha_t) \right) \right) \]
using \( E(\eta_t|\mathcal{H}_{0,T}) = E(\eta_t) = 0 \) with \( \xi_t = \eta_t - E(\eta_t|\mathcal{H}_{T_{i-1},T}) \) and \( \alpha_t = E(\eta_t|\mathcal{H}_{T_{i},T}) \).

To summarize, we have the following decomposition

\[
\sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (TE_{it} - \tilde{T}_S_{it}) = \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (TE_{it}(\xi_t) - \tilde{T}_S_{it}(\xi_t)) \\
+ \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_{i+1}} (TE_{it}(\alpha_t) - \tilde{T}_S_{it}(\alpha_t)) \\
+ \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} TE_{it}(\xi_t, \alpha_t).
\]

Now we examine successively the two terms of (7.16).

1) Term \( \sum_{t=T_{i-1}+1}^{T_i} (\tilde{T}_S_{it} - \tilde{T}_{S_{it}}) \) :

Noting that \( \mu_{2,t} = E(m_{2,t}|\mathcal{H}_{0,T}) \), we have

\[
\sum_{t=T_{i-1}+1}^{T_i} (\tilde{T}_S_{it} - \tilde{T}_{S_{it}}) = \frac{1}{2\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} \left[ E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) - E(m_{2,t}|\mathcal{H}_{0,T}) \right] \tag{T1}
\]

\[
- \frac{1}{8T} \sum_{t=T_{i-1}+1}^{T_i} \left\{ \left[ E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) \right]^2 - \left[ E(m_{2,t}|\mathcal{H}_{0,T}) \right]^2 \right\} \tag{7.17}
\]

We establish that (a) the term (7.17) converges to 0 uniformly in probability. (b) \( \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_{i+1}} (E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) - E(m_{2,t}|\mathcal{H}_{0,T}) - \ln(1 + x_{11}(\xi))) \xrightarrow{P} 0 \) uniformly in \( \beta \) with \( x_{11}(\xi) = \frac{1}{2\sqrt{T}} [E(m_{2,t}(\xi)|\mathcal{H}_{T_{i},T}) - E(m_{2,t}(\xi)|\mathcal{H}_{0,T})] \). Then, from

\[
\sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} (\tilde{T}_S_{it} - \tilde{T}_{S_{it}}) = \frac{1}{2\sqrt{T}} \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} \left[ E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) - E(m_{2,t}|\mathcal{H}_{0,T}) \right] + o_p(1)
\]

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

it follows that \( \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_{i+1}} \left( \tilde{T}_S_{it}(\xi) - \tilde{T}_{S_{it}}(\xi) - \ln(1 + x_{11}(\xi)) \right) \xrightarrow{P} 0 \) uniformly in \( \beta \). The same is true for the term in \( \alpha \).

(a) First, we show that (7.17) converges to 0. Note that

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

and

\[
\|L_{t-1}\| \leq (t - T_{i-1}) \|l_t^{(1)}\|. \tag{7.18}
\]
Hence,

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

by the $\beta$-mixing property of $\eta_t$. Hence this term is negligible.

(b) $(T1)$ can be decomposed into a pure term in $\alpha_t$ and a pure term in $\xi_t$. Consider first the term in $\xi_t$. Using $|x - \log(1 + x)| \leq x^2$ and Assumption 1, we have

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

Moreover by

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

and $(7.18)$, we have

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

46
Hence, for the sum over all blocks, we have

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

Terms in \( \alpha_t \)?

2) Term \( T E_{it} - \hat{T} S_{it} \):

We analyze successively (a) the pure terms w.r.t. \( \xi_t \), (b) the pure terms w.r.t. \( \alpha_t \), and (c) the mixed terms.

(a) For simplicity, we drop \( \xi_t \) in the expressions. Using the notation described in Subsection 6, the pure terms \( \sum_t T E_{it} (\xi_t) \) can be rewritten as follows

\begin{align*}
\frac{1}{\sqrt{T}} L_1 - \frac{1}{2\sqrt{T}} L_2 + \frac{1}{3\sqrt{T^3}} L_3 - \frac{1}{4 T} L_4 \\ + \frac{1}{2\sqrt{T}} M_2 - \frac{1}{8 T} \sum_t m_{2,t}^2 \\ - \frac{1}{2\sqrt{T^3}} \sum_t l_{t}^{(1)} m_{2,t} \\ + \frac{1}{2 T} \sum_t l_{t}^{(1)} m_{2,t} \\ + \frac{1}{6 \sqrt{T^3}} M_3 \\ - \frac{1}{6 T} \sum_t m_{3,t} l_{t}^{(1)} \\ - \frac{1}{\sqrt{T^3}} \sum_t l_{t}^{(1)} L_{t-1}^2 \\ + \frac{1}{T} \sum_t l_{t}^{(1)} L_{t-1}^2 \\ + \frac{1}{T} \sum_t l_{t}^{(1)} L_{t-1}^3 \\ + \frac{1}{2 T} \sum_t l_{t}^{(1)} L_{t-1} m_{2,t} \\ + \frac{1}{24 T} M_4.
\end{align*}

(7.19)
And we add to (P2) the term \(-\sum_t \widehat{T}S_t(\xi_t)\):

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

(7.20)

Let \(x_{i1} = \bar{x}_{i1}/T^{1/4}\), \(x_{i2} = \bar{x}_{i2}/\sqrt{T}\), \(x_{i3} = \bar{x}_{i3}/T^{3/4}\), \(x_{i4} = \bar{x}_{i4}/T\), \(x_{i5} = \bar{x}_{i5}/T^{3/4}\), \(x_{i6} = \bar{x}_{i6}/T\), \(x_{i7} = \bar{x}_{i7}/T^{3/4}\), \(x_{ij} = \bar{x}_{ij}/T\), \(j = 8, 9, 10\), where \(\bar{x}_{ij}\) have been introduced in Lemma 7.7. Now we show that each term (P1) to (P10) can be approximated by a term of the form \(\ln(1 + x_{ij})\) (with \(j = 1, 2, \ldots, 10\)) provided

\[ B_L = o\left( T^{1/16} \right). \]  

(B1)

We also show that the sum over the blocks of (P11) goes to zero and therefore can be neglected.

(P1): We want to show that \(\left| \sum_i (P_{i1} - \ln(1 + x_{i1})) \right| \to 0\) in probability uniformly in \(\beta\) where \(P_{i1}\) is (P1) for the 1th block. By the triangular inequality, \(\left| \sum_i (P_{i1} - \ln(1 + x_{i1})) \right| \leq \sum_i |P_{i1} - \ln(1 + x_{i1})|\). Using a Taylor expansion we have

\[ \left| (P1) - \ln(1 + \frac{1}{\sqrt{T}}L_1) \right| \leq \text{const} \cdot \frac{1}{T^{3/4}} \|L_1\|^5 \]

Now we analyze the moment conditions needed.

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

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

\[ \frac{T}{B_L} \cdot \frac{1}{T^{3/4}} \cdot B_L^5 = o(1), \]

which is satisfied provided (B1) holds.

Consider term (P2)+(7.20):

\[
(P2)+(7.20) = \frac{1}{2\sqrt{T}} \left( M_2 - \sum_t E(m_{2,t} | \mathcal{H}_{T_{t-1},T}) \right) - \frac{1}{8T} \sum_t \left( m_{2,t}^2 - \sum_t \left[ E(m_{2,t} | \mathcal{H}_{T_{t-1},T}) \right]^2 \right)
\]

We want to show that (P2)+(7.20) can be approximated by

\[ \log(1 + \frac{M_2 - E(M_2 | \mathcal{H}_{T_{t-1},T})}{2\sqrt{T}} + \frac{K}{T}) \]

48
where

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

For arbitrary \( A \) and \( B \), a Taylor expansion gives:

\[
\left| \log \left( 1 + \frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{T} \right) - \left( \frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{T} \right) + \frac{1}{2} \left( \frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{T} \right)^2 \right| \\
\leq \frac{1}{3} \left| \frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{T} \right|^3
\]

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

\[
\left| \log(1 + \frac{A}{\sqrt{T}} + \frac{C}{T}) - \left( \frac{A}{\sqrt{T}} + \frac{B}{T} \right) \right| \\
\leq \frac{1}{2} \left| \frac{A^2}{2T} + \frac{1}{2} \left( \frac{A}{\sqrt{T}} + \frac{C}{T} \right)^2 + \frac{1}{3} \left| \frac{A}{\sqrt{T}} + \frac{C}{T} \right|^3 \right| (7.21)
\]

\[
= \frac{1}{2T^2} \left| \frac{C^2}{2} + \frac{AC}{T \sqrt{T}} \right| + \frac{1}{3} \left| \frac{A}{\sqrt{T}} + \frac{C}{T} \right|^3 (7.22)
\]

\[
\leq \text{const} \cdot \left( \frac{\|C\|^2}{T^2} + \left\| \frac{A}{\sqrt{T}} \right\| \cdot \frac{\|C\|}{T} + \frac{\|A\|^3}{T \sqrt{T}} + \frac{\|C\|^3}{T^3} \right) (7.23)
\]

We apply this result to \( A = (M_2 - E(M_2 | H_{T_{t-1}}, T)) / 2 \), \( B = \sum_t \left( m_{2,t}^2 - \sum_t \left[ E(m_{2,t} | H_{T_{t-1}}, T) \right]^2 \right) / 8 \) and \( C = K \). We want to establish that the expectation of the sum over the blocks of the r.h.s. of (7.23) goes to zero uniformly.

First we analyze \( \|A\|^3 \).

\[
\left\| \frac{A}{\sqrt{T}} \right\|^3 \leq \text{const} \cdot \left[ \frac{1}{T^{3/2}} \|M_2\|^3 + \frac{1}{T^{3/2}} E^3(\|M_2\| | H_{T_{t-1}}, T) \right] \\
\leq \text{const} \cdot \left[ \frac{1}{T^{3/2}} \|M_2\|^3 + \frac{1}{T^{3/2}} E(\|M_2\|^3 | H_{T_{t-1}, T}) \right] 
\]

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

$$E \left\| \frac{A}{\sqrt{T}} \right\|^3 \leq \text{const} \cdot \frac{1}{T^{3/2}} E \left\| M_2 \right\|^3 \leq \text{const} \cdot \frac{1}{T^{3/2}} E \left( \sum_t \left\| m_{2,t} \right\|^3 \right)$$

$$\leq \text{const} \cdot \frac{1}{T^{3/2}} B_L^2 \cdot E \left( \sum_t \left\| m_{2,t} \right\|^3 \right)$$

$$\leq \text{const} \cdot \frac{1}{T^{3/2}} B_L^2 \cdot E \sum_t \left( \left\| t^{(2)}_t \right\|^3 + \left\| t^{(1)}_t \right\|^6 + \left\| t^{(1)}_t \right\|^3 \cdot \left\| L_{t-1} \right\|^3 \right)$$

$$= \mathcal{O} \left( \frac{B_L^6}{T^{3/2}} \right)$$

where the third equality follows from Lemma 7.4 and the equality holds by Assumption 3. Hence

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

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

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

and

$$\left\| A \right\|^2 \leq \text{const} \cdot \left[ \left\| M_2 \right\|^2 + E^2 (\left\| M_2 \right\| \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right] \leq \text{const} \cdot \left[ \left\| M_2 \right\|^2 + E (\left\| M_2 \right\|^2 \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right]$$

Again, the second inequality comes from Jensen’s Inequality. Then we have

$$\left\| C \right\|^3 \leq \text{const} \cdot \left( \left\| M_2 \right\|^2 + E (\left\| M_2 \right\|^2 \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right)^3$$

$$\leq \text{const} \cdot \left( \sum_t \left\| m_{2,t} \right\|^2 + E (\sum_t \left\| m_{2,t} \right\|^2 \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right)^3$$

$$\leq \text{const} \cdot (B_L + 1)^3 \left( \sum_t \left[ \left\| m_{2,t} \right\|^2 + E (\left\| m_{2,t} \right\|^2 \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right] \right)^3$$

$$\leq \text{const} \cdot (B_L + 1)^3 \cdot B_L^2 \sum_t \left[ \left\| m_{2,t} \right\|^6 + E (\left\| m_{2,t} \right\|^6 \left\| \mathcal{H}_{T_{t-1},T} \right\|) \right]$$

Therefore,

$$E \left\| C \right\|^3 \leq \text{const} \cdot (B_L + 1)^3 B_L^2 \sum_t E \left\| m_{2,t} \right\|^6$$

$$\leq \text{const} \cdot \frac{(B_L + 1)^3}{T^3} B_L^2 \sum_t E \left( \left\| t^{(2)}_t \right\|^6 + \left\| t^{(1)}_t \right\|^6 + \left\| t^{(1)}_t \right\|^6 \cdot \left\| L_{t-1} \right\|^6 \right)$$

$$= \mathcal{O} \left( B_L^2 \right)$$

50
because sup $E \left\| t_t^{(2)} \right\|^6 < \infty$ and sup $E \left\| t_t^{(1)} \right\|^{12} < \infty$ by Assumption 3. Hence

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

by (B1). Moreover by Holder’s inequality

$$E \|C\|^2 \leq \left( E \|C\|^3 \right)^{2/3} = O \left( B_L^5 \right).$$

Hence

$$\sum_i \frac{E \|C\|^2}{T^2} = O \left( \frac{T}{B_L T^2} B_L^6 \right) = O \left( \frac{B_L^7}{T} \right) = o(1).$$

Now we analyze term $\left\| \frac{A}{\sqrt{T}} \right\| \cdot \frac{\|C\|}{T}$. Note that by Holder’s inequality,

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

By Cauchy-Schwartz inequality,

$$E \left( \left\| \frac{A}{\sqrt{T}} \right\| \cdot \frac{\|C\|}{T} \right) \leq \sqrt{E \left( \left\| \frac{A}{\sqrt{T}} \right\|^2 \right) \cdot E \left( \frac{\|C\|^2}{T^2} \right)} = O \left( \frac{B_L^6}{T^{3/2}} \right).$$

Hence

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

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

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

We need to show that the sum over the blocks of $E \|x^2\|$ converges to zero uniformly in $\beta$. To do so, we use the bounds given by Lemma 7.7. By Holder’s inequality, $E \left( \|x\|^2 \right) \leq E \left( \|x^4\| \right)^{1/2} E \left( \|x\| \right)^{1/2}.$

By Holder’s inequality, $E \left( \|x\|^2 \right) \leq E \left( \|x^4\| \right)^{1/2} E \left( \|x\| \right)^{1/2}.$

Term (P3):

$$E \left( \|P_3\|^2 \right) = \frac{1}{T^{3/2}} E \left( \|\tilde{x}_{i3}\|^2 \right) \leq \frac{1}{T^{3/2}} \left( B_L^8 \right)^{1/2} = B_L^4 \frac{B_L^3}{T^{3/2}},$$

$$\sum_i E \left( \|P_3\|^2 \right) \leq \frac{T}{B_L T^{3/2}} B_L^4 \frac{B_L^3}{T^{1/2}} = o(1).$$
Term (P4):

\[
E \left( \|P_4\|^2 \right) \leq \frac{1}{T^2} \left( B_L^8 \right)^{1/2} = \frac{B_L^4}{T^2},
\]

\[
\sum_i E \left( \|P_4\|^2 \right) \leq \frac{T \cdot B_L^4}{B_L \cdot T^2} = \frac{B_L^3}{T} = o(1).
\]

Term (P5):

\[
E \left( \|P_5\|^2 \right) \leq \frac{1}{T^{3/2}} \left( B_L^4 \right)^{1/2} = \frac{B_L^2}{T^{3/2}},
\]

\[
\sum_i E \left( \|P_5\|^2 \right) \leq \frac{B_L}{T^{1/2}} = o(1).
\]

Term (P6):

\[
E \left( \|P_6\|^2 \right) \leq \frac{1}{T^2} \left( B_L^8 \right)^{1/2} = \frac{B_L^4}{T^2},
\]

\[
\sum_i E \left( \|P_6\|^2 \right) \leq \frac{B_L^3}{T} = o(1).
\]

Term (P7):

\[
E \left( \|P_7\|^2 \right) \leq \frac{1}{T^{3/2}} \left( B_L^{12} \right)^{1/2} = \frac{B_L^6}{T^{3/2}},
\]

\[
\sum_i E \left( \|P_7\|^2 \right) \leq \frac{B_L^5}{T^{1/2}} = o(1).
\]

Term (P8):

\[
E \left( \|P_8\|^2 \right) \leq \frac{1}{T^2} \left( B_L^{12} \right)^{1/2} = \frac{B_L^6}{T^2},
\]

\[
\sum_i E \left( \|P_8\|^2 \right) \leq \frac{B_L^5}{T} = o(1).
\]

Term (P9):

\[
E \left( \|P_9\|^2 \right) \leq \frac{1}{T^2} \left( B_L^{16} \right)^{1/2} = \frac{B_L^8}{T^2},
\]

\[
\sum_i E \left( \|P_9\|^2 \right) \leq \frac{B_L^7}{T} = o(1).
\]
Term (P10):
\[ E \left( \| P_{10} \|^2 \right) \leq \frac{1}{T^2} \left( B_L^{10} \right)^{1/2} = \frac{B_L^8}{T^2}, \]
\[ \sum_i E \left( \| P_{10} \|^2 \right) \leq \frac{B_L^7}{T} = o(1). \]

Term (P11):
\[ \left\| \frac{1}{T} \sum_{t=1}^T m_{4,t} (\xi_t, \xi_t, \xi_t, \xi_t) \right\| \leq \left\| \frac{1}{T} \sum_{t=1}^T m_{4,t} \right\| M^4 \]
because \( \| \xi_t \| \leq M. \)
\[ E \| m_{4,t} \| \leq E \left( \| l_t^{(4)} + 6l_t^{(2)} \otimes l_t^{(1)2} + 4l_t^{(3)} \otimes l_t^{(1)} + 3l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)4} \| \right) \]
\[ \leq E \left( \| l_t^{(4)} \| + 6 \| l_t^{(2)} \| \| l_t^{(1)} \|^2 + 4 \| l_t^{(3)} \| \| l_t^{(1)} \| + 3 \| l_t^{(2)} \|^2 + \| l_t^{(1)4} \| \right) \]
\[ < \infty \]
provided \( \sup E \| l_t^{(4)} \| < \infty, \sup E \| l_t^{(3)} \| ^2 < \infty, \sup E \| l_t^{(2)} \| ^2 < \infty, \sup E \| l_t^{(1)} \| ^4 < \infty. \) As \( m_{4,t} \) is a martingale, \( \frac{1}{T} \sum m_{4,t} = o_p (1) \) and hence
\[ \frac{1}{T} \| M_4 \| = o_p (1) \]
uniformly in \( \beta. \) Therefore this term can be neglected.

Hence we have shown that \( \sum_{i=1}^{B_N} \sum_{t=T_i-1+1}^{T_i} \left( T E_{iT} (\xi_t) - T S_{iT} (\xi_t) - \sum_{j=1}^{10} \ln (1 + x_{ij}) \right) \xrightarrow{P} 0 \)
uniformly in \( \beta. \)

b) Second, we analyze the pure terms w.r.t. \( \alpha_t. \) We can do the same approximation as for the terms in \( \xi_t \) using the fact that for \( T_{i+1} \geq t > T_i \)
\[ \| \alpha_t \| = \| E (\eta_t | \mathcal{H}_{T_i,T}) \| \leq \chi^{T-T_i} \]
by the \( \beta \)-mixing property of \( \eta_t. \) Note that all the terms in \( 1/T \) can be neglected. For illustration, we treat the case of term (P9). Remark that
\[ \| L_{t-1} \|^3 = \left\| \sum_{t=T_{i+1}}^{t-1} l_s^{(1)} (\alpha_s) \right\|^3 \leq M^3 \sup \| l_t^{(1)} \|^3 (t - T_i)^3 \]

53
\[
\left\| \frac{1}{T} \sum_{t=T_{i+1}}^{T_i} l_t^{(1)} L_{t-1}^3 \right\| \leq M^3 \sup \left\| l_t^{(1)} \right\| \frac{1}{T} \sum_{t=T_{i+1}}^{T_i} (t - T_i)^3 \| \alpha_t \|
\]
\[
\leq \text{const} \frac{1}{T} \sum_{t=T_{i+1}}^{T_i} (t - T_i)^3 \lambda^{t-T_i}
\]
\[
\leq \text{const} \frac{1}{T}.
\]
Hence
\[
\left\| \frac{1}{T} \sum_{i} \sum_{t=T_{i+1}}^{T_i} l_t^{(1)} L_{t-1}^3 \right\| \leq \frac{T}{B_L T} \to 0.
\]
Therefore the term \((P9)\) is negligible. And so are the other terms in \(1/T\). The remaining terms in \(\alpha\) are \((P1), (P2), (P3), (P5)\) and \((P7)\). We have the same approximation in \(\ln(1+x)\) as for the terms in \(\xi_t\).

c) The mixed terms are as follows:
\[
\frac{1}{2\sqrt{T}} \sum_{\text{2 permutations}} l_t^{(2)}(\xi_t, \alpha_t) + \frac{1}{6\sqrt{T^3}} \sum_{\text{3 permutations}} l_t^{(3)}(\xi_t, \xi_t, \alpha_t) + \frac{1}{6\sqrt{T^3}} \sum_{\text{3 permutations}} l_t^{(3)}(\xi_t, \alpha_t, \alpha_t)
\]
\[
+ \frac{1}{24T} \sum_{\text{4 permutations}} l_t^{(4)}(\xi_t, \xi_t, \xi_t, \alpha_t) + \frac{1}{24T} \sum_{\text{6 permutations}} l_t^{(4)}(\xi_t, \xi_t, \alpha_t, \alpha_t) + \frac{1}{24T} \sum_{\text{4 permutations}} l_t^{(4)}(\xi_t, \alpha_t, \alpha_t, \alpha_t)
\]
For all the 4th-order terms,
\[
\sum_i E \frac{1}{T} \sum_{t=T_{i+1}}^{T_i} l_t^{(4)}(\cdot, \cdot, \cdot, \alpha_t) \leq \frac{T}{B_L T} \sup E \left\| l_t^{(4)} \right\| \cdot M^3 \cdot \frac{1}{1 - \lambda}
\]
which converges to zero uniformly provided
\[
\sup E \left\| l_t^{(4)} \right\| < \infty.
\]
For the 3rd-order terms, we apply the Bartlett Identity,
\[
M^{(3)}(a, b, c) = l^{(3)}(a, b, c) + l^{(2)}(a, b)l^{(1)}(c) + l^{(2)}(a, c)l^{(1)}(b) + l^{(2)}(b, c)l^{(1)}(a) + l^{(1)}(a)l^{(1)}(b)l^{(1)}(c)
\]
Hence the mixed-terms can be written as

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

$$+ \frac{1}{6\sqrt{T^3}} \sum \left( \begin{array}{c}
M_t^{(3)}(\alpha_t, \alpha_t, \xi_t) + M_t^{(3)}(\alpha_t, \xi_t, \xi_t) \\
\text{3 permutations} \\
\text{3 permutations}
\end{array} \right) \quad \text{(R2)}$$

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

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

$$- \frac{1}{6\sqrt{T^3}} \sum 3l_t^{(1,2)}(\alpha_t)l_t^{(1)}(\xi_t) \quad \text{(R5)}$$

$$- \frac{1}{6\sqrt{T^3}} \sum \left( 3l_t^{(1,2)}(\xi_t)l_t^{(1)}(\alpha_t) + 3l_t^{(2)}(\xi_t, \xi_t)l_t^{(1)}(\alpha_t) \right) \quad \text{(7.24)}$$

The term (7.24) is rewritten in the following way.

$$- \frac{1}{2\sqrt{T^3}} \sum l_t^{(1)}(\alpha_t)m_{2,t}(\xi_t, \xi_t) \quad \text{(7.25)}$$

$$+ \frac{1}{\sqrt{T^3}} \sum l_t^{(1)}(\alpha_t)l_t^{(1)}(\xi_t)L_{t-1}(\xi_t) \quad \text{(R6)}$$

Moreover, we have

$$- \frac{1}{2\sqrt{T^3}} \sum_{i=0}^{T_t} \sum_{t=T_{i-1}+1}^{T_t} l_t^{(1)}(\alpha_t)m_{2,t}(\xi_t, \xi_t)$$

$$= - \frac{1}{2\sqrt{T^3}} \sum_{i=0}^{T_t} \sum_{t=T_{i-1}+1}^{T_t} l_t^{(1)}(\alpha_t)m_{2,t} \left[ (\xi_t, \xi_t) - E \left[ (\xi_t \otimes \xi_t) | \mathcal{H}_{T_{i-1}, T} \right] \right] \quad \text{(R7)}$$

$$- \frac{1}{2\sqrt{T^3}} \sum_{i=0}^{T_t+1} \sum_{t=T_{i-1}+1}^{T_t} l_t^{(1)}(\alpha_t)m_{2,t} \left[ E \left[ (\xi_t \otimes \xi_t) | \mathcal{H}_{T_{i}, T} \right] - E \left[ (\xi_t \otimes \xi_t) \right] \right] \quad \text{(R8)}$$

$$- \frac{1}{2\sqrt{T^3}} \sum_{i=0}^{T_t+1} \sum_{t=T_{i-1}+1}^{T_t} l_t^{(1)}(\alpha_t)m_{2,t}E \left[ (\xi_t \otimes \xi_t) \right] \quad \text{(R9)}$$

Note that the sum in (R8) and (R9) is over $T_t + 1$ to $T_{t+1}$, this follows from a simple change of indice (replace $i$ by $i + 1$). Each term (R1) to (R9) (denoted $x$ for convenience) can be approximated by terms in $(1 + x)$. The terms $E(x^3)$ involve $\alpha_t$ and hence their sums converge to 0 uniformly in $\beta$.

**Step 3.**
As \( \exp(TS_T) \) is \( \mathcal{H}_{0,T} \)-measurable, we have

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

\( J \) is equal to 27, because there are 11 pure terms in \( \xi_t \) (corresponding to T1, and P1 to P10), 7 pure terms in \( \alpha_t \), and 9 mixed terms. The product can be rewritten as

\[
\prod_{j=1}^{J} (1 + x_{ij}) = 1 + \sum_{j=1}^{J} x_{ij} + \sum_{j \neq l} x_{ij} x_{il} + \sum_{j \neq l \neq p} x_{ij} x_{il} x_{ip} + \ldots + \prod_{j=1}^{J} x_{ij}
\]

where each \( x_{ij} \) is of the form \( \frac{1}{T} \sum_l \bar{x}_{ijl} \). Hence \( \prod_{j=1}^{J} (1 + x_{ij}) \) is 1 plus a sum of terms of the form \( \prod_{j \in d} x_{ij} \), where \( d \) is a partition of 1, 2, ..., \( J \). Each of these terms can be treated individually. We need to compute \( \Delta_{i,T} \) and check Conditions (7.4) and (7.5) in Lemma 7.2.

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

\[
E \left( \left\| \sum_t \bar{x}_{ijt} \right\|^4 \right) \leq \text{const} B^4_L T^4.
\]

Hence by Lemma 7.6, we have

\[
\frac{B^{m-1}_L}{T^{\sum \alpha_j - 1}} \leq \frac{B^{15}_L}{T^{1/2}} = o(1)
\]

for \( B_L = o \left( T^{1/30} \right) \). For this choice of \( B_L \), the conditions (7.4) and (7.5) are satisfied. If there are more than four terms, the conditions (7.4) and (7.5) are again satisfied. Indeed by Lemma 7.7 and Holder’s inequality, we have

\[
E (\|x_{ij}\|) \leq \text{const}\frac{B^4_L}{T^{\alpha_j}} = o(1).
\]

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

\[
\|\Delta_i\| = E \left[ \left\| \prod_{j \in d_1} x_{ij} \prod_{k \in d_2} x_{ik} \right\| \mathcal{H}_{T_{i-1},T} \right] \leq \frac{1}{2} E \left[ \left\| \prod_{j \in d_1} x_{ij} \right\| \mathcal{H}_{T_{i-1},T} \right]
\]

where cardinal of \( d_1 \leq 4 \). Hence the result follows from above.

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

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

$$x_{i1} = \frac{L_1}{T^{1/4}},$$

$$x_{i2} = x_{i20} + x_{i21},$$

$$x_{i20} = \frac{M_2 - E (M_2 | \mathcal{H}_{T_{i-1}, T})}{2\sqrt{T}},$$

$$x_{i2b} = \frac{1}{8T} [M_2 - E (M_2 | \mathcal{H}_{T_{i-1}, T})]^2 - \frac{1}{8T} \left[ \sum_{t} m_{2,t}^2 - \sum_{t} [E (m_{2,t} | \mathcal{H}_{T_{i-1}, T})]^2 \right].$$

$$x_{i3} = -\frac{1}{2\sqrt{T^3}} \sum_{t} l_{t}^{(1)} m_{2,t},$$

$$x_{i4} = \frac{1}{2T} \sum_{t} l_{t}^{(1)2} m_{2,t},$$

$$x_{i5} = \frac{1}{6\sqrt{T^3}} M_3,$$

$$x_{i6} = -\frac{1}{6T} \sum_{t} l_{t}^{(1)} m_{3,t},$$

$$x_{i7} = -\frac{1}{\sqrt{T^3}} \sum_{t} l_{t}^{(1)} L_{t-1}^2,$$

$$x_{i8} = \frac{1}{T} \sum_{t} l_{t}^{(1)2} L_{t-1}^2,$$

$$x_{i9} = \frac{1}{T} \sum_{t} l_{t}^{(1)} L_{t-1}^3,$$

$$x_{i10} = \frac{1}{2T} \sum_{t} l_{t}^{(1)} L_{t-1} m_{2,t},$$

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

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

(a) Terms for which $\sum_{j} \alpha_j = 1$.

Here is the list of such terms:

$x_{i2b}$,

$x_{i4} + x_{i10} + x_{i1} x_{i3},$

$x_{i6} + x_{i1} x_{i5},$

$x_{i8} + x_{i9} + x_{i4} x_{i7}.$
Terms $x_{i21}$:

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

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

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

\[
x_{i4} + x_{i10} + x_{i1} x_{i3}
= \frac{1}{2T} \sum_t l_t^{(1)} m_{2,t} + \frac{1}{2T} \sum_t l_t^{(1)} L_{t-1} m_{2,t} - \frac{1}{2T} \sum_{t,s} l_s^{(1)} l_t^{(1)} m_{2,t}
= -\frac{1}{2T} \sum_{t,s \geq t} l_s^{(1)} l_t^{(1)} m_{2,t}
\]

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

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

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

Hence

\[
\sum_i E \left( \Delta_i^2 \right) \sim \frac{B_t}{T} \to 0.
\]

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

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

58
We can treat separately the two terms on the r.h.s. They are both martingales. We get the same rate as for the previous case.

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

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

$\Delta_i$ is again a martingale, we obtain

$$
E(\Delta_i^2) \leq \text{const} \frac{1}{T^2} \sum_s E \left[ l_s^{(1)}^2 \left( \sum_{t<s} l_t^{(1)} L_{t-1}^2 \right)^2 \right] \chi^2(s-T_{i-1})
$$

$$
\leq \text{const} \frac{1}{T^2} \sum_s \left[ E \left( l_s^{(1)}^4 \right) E \left( \left( \sum_{t<s} l_t^{(1)} L_{t-1}^2 \right)^4 \right) \right]^{1/2} \chi^2(s-T_{i-1})
$$

$$
\leq \text{const} \frac{B^6_L}{T^2}
$$

Hence

$$
\sum_i E(\Delta_i^2) \sim \frac{B^5_L}{T} \to 0.
$$

(b) Terms for which $\sum_j \alpha_j < 1$.

The list of such terms is

- $x_{i1}$,
- $x_{i2}$,
- $x_{i3} + x_{i1}x_{i2}$,
- $x_{i5}$,
- $x_{i7}$,
- $x_{i11}$.

Term $x_{i1}$ :

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

$\Delta_i = 0$.  

59
Term $x_{i2a}$:

$$
\Delta_i = E \left[ x_{i2a} \big| \mathcal{H}_{T_{i-1}, T} \right] = 0.
$$

Term $x_{i3} + x_{i1}x_{i2a}$:

$$
x_{i3} + x_{i1}x_{i2a} = -\frac{1}{2\sqrt{T^3}} \sum_t l_{t}^{(1)} m_{2,t} + \frac{1}{2\sqrt{T^3}} \sum_{t,s} l_{t}^{(1)} (m_{2,s} - E (m_{2,s} \big| \mathcal{H}_{T_{i-1}, T}))
= -\frac{1}{2\sqrt{T^3}} \sum_{t \neq s} l_{t}^{(1)} (m_{2,s} - E (m_{2,s} \big| \mathcal{H}_{T_{i-1}, T})) - \frac{1}{2\sqrt{T^3}} \sum_t l_{t}^{(1)} E (m_{2,s} \big| \mathcal{H}_{T_{i-1}, T})
$$

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

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

$$
\Delta_i = E \left( \frac{1}{\sqrt{T}} \sum_{t=T_{i+1}}^{T_{i+1}} l_{t}^{(1)} (\alpha_t) \big| \mathcal{H}_{T_{i-1}, T} \right)
= \frac{1}{\sqrt{T}} \sum_{t=T_{i+1}}^{T_{i+1}} l_{t}^{(1)} E (\alpha_t \big| \mathcal{H}_{T_{i-1}, T})
= \frac{1}{\sqrt{T}} \sum_{t=T_{i+1}}^{T_{i+1}} l_{t}^{(1)} E (\eta_t \big| \mathcal{H}_{T_{i-1}, T})
$$

because $\alpha_t = E (\eta_t \big| \mathcal{H}_{T_{i}, T})$

$$
\|\Delta_i\| \leq \frac{1}{\sqrt{T}} \sum_{t=T_{i+1}}^{T_{i+1}} \|l_{t}^{(1)}\| \|E (\eta_t \big| \mathcal{H}_{T_{i-1}, T})\|
$$

$$
E \|\Delta_i\| \leq \frac{1}{\sqrt{T}} \sum_{t=T_{i+1}}^{T_{i+1}} (\sup E \|l_{t}^{(1)}\|) \lambda^{t-T_{i+1}}
$$

$$
\sum_{i=1}^{B_N} E \|\Delta_i\| \leq \frac{\text{const}}{\sqrt{T}} \sum_{i=1}^{B_N} \lambda^{B_L}
\leq \frac{\text{const} T}{\sqrt{T}} \frac{B_L}{B_L} \lambda^{B_L}
= \text{const} T^{3/4} \frac{B_{L-k}}{B_L}
$$

60
for any $k$. Hence $\sum_{i=1}^{B_N} E \| \Delta_i \| \to 0$.
The remaining terms can be treated similarly.

3) Mixed terms:
(R1) We have
\[
\Delta_i = \frac{1}{\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E \left[ \ell^{(2)}_t (\alpha_t, \xi_t) | \mathcal{H}_{T_{i-1},T} \right]
\]
\[
= \frac{1}{\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} \ell^{(2)}_t (\alpha_t \otimes E (\xi_t | \mathcal{H}_{T_{i-1},T}))
\]
\[
= 0
\]
because $E (\xi_t | \mathcal{H}_{T_{i-1},T}) = 0$. Hence, Lemma 7.2 applies.

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

For (R9), we can use the fact that $E (\xi_t \otimes \xi_t)$ is constant and $E (\alpha_t | \mathcal{H}_{T_{i-1},T})$ decays exponentially. Indeed we have
\[
\Delta_i = -\frac{1}{2\sqrt{T^3}} \sum_{t=T_{i-1}+1}^{T_{i+1}} \ell^{(1)}_t E (\alpha_t | \mathcal{H}_{T_{i-1},T}) m_{2,t} E [(\xi_t \otimes \xi_t)]
\]
\[
\| \Delta_i \| \leq \frac{1}{2\sqrt{T^3}} \sum_{t=T_{i-1}+1}^{T_{i+1}} \| \ell^{(1)}_t \| \| m_{2,t} \| \| E (\eta_t | \mathcal{H}_{T_{i-1},T}) \|
\]
\[
E \| \Delta_i \| \leq \text{const} \frac{1}{\sqrt{T^3}} \sum_{t=T_{i-1}+1}^{T_{i+1}} \lambda^{t-T_{i-1}} \leq \frac{\lambda^{B_L}}{\sqrt{T^3}}.
\]
Hence the conditions of Lemma 7.2 are satisfied.

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

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

4) Cross-products involving $\alpha_t$ and $\xi_t$:
Since the product has $\alpha_t$ involved, as far as $\sum \alpha_i \geq 1$, conditions (7.4) and (7.5) are satisfied. So we only need to concentrate on those terms with $\sum \alpha_i < 1$. They are, $x_{11}(\xi_t) \cdot x_{11}(\alpha_t), x_{11}(\xi_t) \cdot x_{20}(\alpha_t), x_{11}(\xi_t) \cdot x_{11}(\alpha_t), x_{120}(\xi_t) \cdot x_{11}(\alpha_t), x_{11}(\alpha_t) \cdot R1, x_{111}(\xi_t) \cdot x_{11}(\alpha_t)$ and $x_{11}(\xi_t) \cdot R1. \Delta_i$ is martingale in the first five cases. Hence we can apply Lemma 7.3. We treat the first case in details and omit the other cases.
Term $x_{i1}(\xi_t) \cdot x_{i1}(\alpha_t)$:

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

The associated $\Delta_i$ is a martingale. We can apply Lemma 7.3. Remark that

$$E(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1}, T}) = E E(\alpha_t | \mathcal{H}_{s, T}) \otimes \xi_s | \mathcal{H}_{T_{i-1}, T}$$

$$= E E(\eta_t | \mathcal{H}_{s, T}) \otimes \xi_s | \mathcal{H}_{T_{i-1}, T},$$

$$\|E(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1}, T})\| \leq E \|E(\eta_t | \mathcal{H}_{s, T})\| \|\xi_s\| | \mathcal{H}_{T_{i-1}, T}$$

$$\leq \text{const} \lambda^{t-s} g(\eta_{T_{i-1}}, ...)$$

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

$$\left| E \left[ l_t^{(1)}(\alpha_t) \sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}(\xi_s) | \mathcal{H}_{T_{i-1}, T} \right] \right|$$

$$\leq \sum_{s=T_{i-1}+1}^{T_i} \left| l_t^{(1)} l_s^{(1)} \right| \left| E(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1}, T}) \right|$$

$$\leq \text{const} \sum_{s=T_{i-1}+1}^{T_i} \left| l_t^{(1)} l_s^{(1)} \right| \lambda^{t-s} g(\eta_{T_{i-1}}, ...).$$

And

$$E(\Delta_i^2) \leq \frac{1}{T} \sum_{t=T_i+1}^{T_{i+1}} E \left\{ E \left[ l_t^{(1)}(\alpha_t) \sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}(\xi_s) | \mathcal{H}_{T_{i-1}, T} \right]^2 \right\}$$

$$\leq \frac{\text{const}}{T} \sum_{t=T_{i+1}}^{T_{i+1}} \sum_{s=T_{i-1}+1}^{T_i} \lambda^{2(t-s)} E \left[ \left( l_t^{(1)} l_s^{(1)} \right)^2 \right]$$

$$\leq \frac{\text{const}}{T} \left( \sup E \left\| l_t^{(1)} \right\| \right) \lambda^{2(t' - t)} \sum_{t=T_{i+1}}^{T_{i+1}} \sum_{s'=0}^{T_i-T_{i-1}} \lambda^{2(s' + t' - T_{i-1})}$$

$$\leq \frac{\text{const}}{T} \left( \frac{1}{1 - \lambda^2} \right)^2.$$

Therefore

$$\sum_i E(\Delta_i^2) \rightarrow 0.$$

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

62
\[ x_{i1} = \frac{1}{2\sqrt{T}} \sum_{t=T_{i}+1}^{T_{i+1}} [E(m_{2,t}|\mathcal{H}_{T_{i},T}) - E(m_{2,t}|\mathcal{H}_{0,T})] \]

\[ x_{i1}(\xi_t) \cdot x_{i1}(\alpha_t) = \frac{1}{2\sqrt{T^3}} \sum_{t=T_{i}+1}^{T_{i+1}} m_{2,t} [E(\xi_t \otimes \xi_t|\mathcal{H}_{T_{i},T}) - E(\xi_t \otimes \xi_t)] \sum_{s=T_{i}+1}^{T_{i+1}} l_{s}^{(1)}(\alpha_s) \]

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

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

\[ \frac{1}{\sqrt{T^3}} \sum_{t=T_{i}+1}^{T_{i+1}} I_t^{(2)}(\alpha_t, \xi_t) \sum_{s=T_{i}+1}^{T_{i+1}} l_{s}^{(1)}(\xi_s) \]

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

\[ \frac{1}{\sqrt{T^3}} \sum_{t} l_t^{(2)}(\alpha_t, \xi_t) \sum_{s \leq t} l_{s}^{(1)}(\xi_s) \]

\[ = \frac{1}{\sqrt{T^3}} \sum_{t} l_t^{(2)}(\alpha_t, \xi_t) l_{t}^{(1)}(\xi_t) \sum_{s \leq t} l_{s}^{(1)}(\xi_s) \]  

(7.26)

\[ + \frac{1}{\sqrt{T^3}} \sum_{t} l_t^{(2)}(\alpha_t, \xi_t) \sum_{s < t} l_{s}^{(1)}(\xi_s) \]  

(7.27)

(7.26) cancels out with (R4). (7.27) can be rewritten as

\[ \frac{1}{\sqrt{T^3}} \sum_{t} l_t^{(2)}(\alpha_t, \xi_t)L_{t-1}(\xi) \]

\[ = \frac{1}{\sqrt{T^3}} \sum_{t} \left( l_t^{(2)}(\alpha_t, \xi_t) + l_t^{(1)}(\alpha_t) \otimes l_{t}^{(1)}(\xi_t) \right) L_{t-1}(\xi) \]  

(7.28)

\[ - \frac{1}{\sqrt{T^3}} \sum_{t} l_t^{(1)}(\alpha_t) \otimes l_{t}^{(1)}(\xi_t)L_{t-1}(\xi) \]  

(7.29)

(7.28) is again a martingale which causes no problem. Finally, (7.29) cancels out with (R6).
**Proof of Corollary 3.2.**

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

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

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

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

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

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

**Proof of Theorem 3.9 and Lemma 3.10**

We have to analyze the difference between

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

(7.30)

where

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

(7.31)

and $d$ is chosen according to (3.30), and

$$TS_T(\beta, \hat{\beta}) = \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \hat{\beta}) - \frac{1}{2T} \widehat{\varepsilon}(\beta)' \widehat{\varepsilon}(\beta),$$

(7.32)

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

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 $\beta$ is restricted to a compact set.

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

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

(7.33)
Let
\[
\hat{d} = \hat{d}(\beta) = \left( \frac{1}{T} \sum_{t=1}^{T} l_t^{(1)}(\hat{\theta}) \otimes l_t^{(1)}(\hat{\theta}) \right)^{-1} \left( \frac{1}{2T} \sum_{t=1}^{T} \mu_{2,t}(\hat{\theta}, \beta) l_t^{(1)}(\hat{\theta}) \right).
\]
Denote \( y_t = \frac{1}{2} \mu_{2,t}(\hat{\theta}) \), \( x_t = l_t^{(1)}(\hat{\theta}) \), \( y = (y_1, \ldots, y_T)' \) and \( X = (x_1, \ldots, x_T)' \). Using these notations, \( \hat{d} = (X'X)^{-1} X'y \) and
\[
\frac{1}{4T} \sum_{t=1}^{T} \left[ \mu_{2,t}(\hat{\theta}, \beta) \right]^2 - \hat{d}' \hat{I}(\hat{\theta}) \hat{d} / T
= \left( y'y - y'X(X'X)^{-1} X'y \right) / T
= y' \left[ I - X(X'X)^{-1} X' \right] y / T
= y'M_X M_X y / T
= \varepsilon(\beta)' \varepsilon(\beta) / T
\]
where \( M_X = I - X(X'X)^{-1} X' \) is idempotent. Obviously our assumptions guarantee the consistency of the ML estimator. Then it is now an elementary exercise to show that
\[
\hat{d}(\beta) \to d(\beta)
\]
and consequently
\[
\frac{1}{2T} \varepsilon(\beta)' \varepsilon(\beta) \to \frac{1}{8} E\mu_{2,t}(\theta_0, \beta)^2 - \frac{1}{2} d'I(\theta_0) d
\]
(7.35)
\[
= \frac{1}{2} E \left[ \left( \frac{\mu_{2,t}(\theta_0, \beta)}{2} - d'l_t^{(1)}(\theta_0) \right)^2 \right]
\]
(7.36)
by (3.30). Hence it is sufficient for us to show that
\[
\frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta_T) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d'l_t^{(1)}(\theta_T) - \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \hat{\theta})
= \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta_T) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d'l_t^{(1)}(\theta_T) - \left( \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \hat{\theta}) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d'l_t^{(1)}(\hat{\theta}) \right)
\]
converges (uniformly in \( \beta \)) to 0. So define the function
\[
Y_T(\beta, \theta) = \frac{1}{2\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} d'l_t^{(1)}(\theta).
\]
(7.37)
Observe that our conditions guarantee that the ML estimator is \( \sqrt{T} \) consistent. Hence it is sufficient to show that for all \( M \)
\[
\sup_{\beta, \|\theta - \theta_0\| \leq M / \sqrt{T}} |Y_T(\beta, \theta) - Y_T(\beta, \theta_0)| \to 0
\]
(7.38)
65
Obviously \( Y_T \) is at least twice continuously differentiable as a function of \( \theta \), and we can easily see that its second derivative is \( O(\sqrt{T}) \). Hence to show (7.38) it is sufficient to show that the first derivative is \( o(\sqrt{T}) \) or equivalently

\[
\frac{\partial}{\partial \theta} \left( \frac{1}{2T} \sum_{t=1}^{T} \mu_{2,t}(\beta, \theta) - \frac{1}{T} \sum_{t=1}^{T} d''_{t}(\theta) \right) \to 0
\]  
(7.39)

**Remark 12.** Here we will use “conventional” calculus for partial derivatives, because the direct evaluation of the terms appearing in this proof is relatively easy.

**Remark 13.** Since the second derivative is \( O(\sqrt{T}) \), and the range of the arguments is \( O(1/\sqrt{T}) \), the changes in the first derivative are \( O(1) \). Hence it is sufficient to show the relationship (7.39) only for one value of \( \theta \).

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

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

**Lemma 7.8.** We have

\[
\frac{1}{T} \sum_{t} \frac{\partial \mu_{2,t}}{\partial \theta} = -\frac{1}{T} \sum_{t} \mu_{2,t} \frac{\partial l_{t}}{\partial \theta} + o_{p}(1)
\]

To establish (7.39), we have to show that

\[
\frac{1}{2T} \sum_{t} \frac{\partial \mu_{2,t}}{\partial \theta} - \frac{1}{T} \sum_{t=1}^{T} d''_{t}(\theta) \overset{P}{\to} 0.
\]  
(7.40)

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

\[
\frac{1}{T} \sum_{t=1}^{T} l''_{t}(\theta) \overset{P}{\to} -I(\theta)
\]  
(7.41)

and from Lemma 7.8, it follows that

\[
\frac{1}{T} \sum_{t} \frac{\partial \mu_{2,t}}{\partial \theta} \overset{P}{\to} -\text{cov} \left( \mu_{2,t}, \frac{\partial l_{t}}{\partial \theta} \right).
\]  
(7.42)
Then (7.39) is an easy consequence of the definition of \( d \) in (3.30).

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

\[
\log \left( \frac{dP_{\theta_0 + h_i/\sqrt{T}}}{dP_{\theta_0}} \right) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_i (l_t^{(1)}) (\theta_T) + \frac{1}{2} E \left( h_i (l_t^{(1)}) (\theta_T) \right)^2 \to 0. \tag{7.43}
\]

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

So it remains to show the lemma:

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

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

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

from the third Bartlett identity,

\[
m_{3,t} = \frac{\partial^3 l_t}{\partial \theta_i \partial \theta_j \partial \theta_k} + \frac{\partial l_t}{\partial \theta_i} \frac{\partial^2 l_t}{\partial \theta_j \partial \theta_k} + \frac{\partial l_t}{\partial \theta_j} \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_k} + \frac{\partial l_t}{\partial \theta_k} \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} \frac{\partial l_t}{\partial \theta_k}
\]
is a martingale difference sequence and therefore $\frac{1}{T} \sum_{t=1}^{T} m_{3,t} = o_p(1)$.

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

$$
= o_p(1) - \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} \right] \frac{\partial l_t}{\partial \theta_k}.
$$

$$
\frac{\partial}{\partial \theta_k} \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_{t-s}}{\partial \theta_j} = \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \left[ \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \right] \frac{\partial l_{t-s}}{\partial \theta_k}
$$

$$
= \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \left[ \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \right] \frac{\partial l_{t-s}}{\partial \theta_k} + \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \frac{\partial l_t}{\partial \theta_i} \frac{\partial^2 l_{t-s}}{\partial \theta_j \partial \theta_k}
$$

$$
- \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} \frac{\partial l_{t-s}}{\partial \theta_k}
$$

$$
= o_p(1) - \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} \frac{\partial l_{t-s}}{\partial \theta_k}
$$

because $\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} \frac{\partial l_{t-s}}{\partial \theta_k}$ are m.d.s. Therefore, we have

$$
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial \mu_{2,t}}{\partial \theta_k} = -\frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} + \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_t}{\partial \theta_j} + \frac{2}{T} \sum_{t=1}^{T} \sum_{s>0} \frac{\partial l_t}{\partial \theta_i} \frac{\partial l_{t-s}}{\partial \theta_k} \right] \frac{\partial l_t}{\partial \theta_k} + o_p(1)
$$

$$
= -\widehat{\text{cov}} \left( \mu_{2,t}, \frac{\partial l_t}{\partial \theta_k} \right) + o_p(1)
$$

where $\widehat{\text{cov}}$ denotes the empirical covariance. It is now an easy exercise to show that

$$
\widehat{\text{cov}} \left( \mu_{2,t}, \frac{\partial l_t}{\partial \theta_k} \right) \to \text{cov} \left( \mu_{2,t}, \frac{\partial l_t}{\partial \theta_k} \right).
$$

**Proof of Proposition 5.1.**

We do the proof for the case where $S_t$ takes two values only. The generalization to three regimes is immediate. We use the following notation $z_t = \ln (P_t)$ and $w_t = \ln (D_t)$ and we reparametrize slightly (5.2) so that

$$
z_t = a_0 + a_1 w_t + y_t
$$

$$
y_t = \alpha_{st} + \sum_{j=1}^{l} \gamma_{st,j} y_{t-j} + \varepsilon_t.
$$
In the two-step approach, the parameters are such that

\[ \sum \hat{y}_t = 0, \quad (7.44) \]
\[ \sum w_t \hat{y}_t = 0, \quad (7.45) \]
\[ \sum \hat{\varepsilon}_t^i P (S_t = i|\hat{y}_{t-1}, \ldots, \hat{y}_1) = 0, \quad (7.46) \]
\[ \sum \hat{y}_{t-j} \hat{\varepsilon}_t^i P (S_t = i|\hat{y}_{t-1}, \ldots, \hat{y}_1) = 0, \quad j = 1, \ldots, l, \ i = 0, 1. \quad (7.47) \]

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

\[ \hat{\varepsilon}_t^i = \hat{y}_t - \hat{\alpha}_i - \sum_{j=1}^l \hat{\gamma}_{i,j} \hat{y}_{t-j}, \]
\[ \hat{y}_t = z_t - \hat{\alpha}_0 - \hat{\alpha}_1 w_t \]
\[ = (z_t - \bar{z}) - \hat{\alpha}_1 (w_t - \bar{w}) \]

Note that there is a potential problem of identification as \( \sum \hat{y}_t = 0 \) by construction. Therefore, we do not estimate \( \hat{\alpha}_0 \) when we do global MLE, instead we demean the time series \( z_t \) and \( w_t \). To compute the global MLE, we use the equation

\[ \left( 1 - \sum_{j=1}^l \gamma_{s,t,j} L^j \right) (z_t - \bar{z}) = a_1 \left( 1 - \sum_{j=1}^l \gamma_{s,t,j} L^j \right) (w_t - \bar{w}) + \alpha_{s,t} + \varepsilon_t. \]

Hence the conditional log-likelihood equals

\[ \ln f (z_t|w_t, z_{t-1}, w_{t-1}, \ldots, z_1, w_1; s_t) \]
\[ = - \ln \left( \frac{1}{\sigma} \right) - \frac{1}{2 \sigma^2} \left\{ \left( 1 - \sum_{j=1}^l \gamma_{s,t,j} L^j \right) ((z_t - \bar{z}) - a_1 (w_t - \bar{w})) - \alpha_{s,t} \right\}^2 \]

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

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

Hence we have

\[ \frac{\partial L}{\partial \alpha_i} = \frac{1}{\sigma^2} \sum_t \hat{\varepsilon}_t^i P (S_t = i|z_{t-1}, w_{t-1}, \ldots, z_1, w_1) = 0, \quad i = 0, 1 \quad (7.48) \]
\[ \frac{\partial L}{\partial \gamma_{i,j}} = \frac{1}{\sigma^2} \sum_t \hat{y}_{t-j} \hat{\varepsilon}_t^i P (S_t = i|z_{t-1}, w_{t-1}, \ldots, z_1, w_1) = 0, \quad j = 1, \ldots, l, \ i = 0, 1 \quad (7.49) \]

As the relevant information (for \( S_t \)) contained in \( \sigma (z_{t-1}, w_{t-1}, \ldots, z_1, w_1) \) is the same as that contained in \( \sigma (\hat{y}_{t-1}, \ldots, \hat{y}_1) \), (7.48) and (7.49) coincide with (7.46) and (7.47).
\[
\frac{\partial L}{\partial a_1} = \frac{1}{\sigma^2} \sum_{t} \sum_{i=0,1} \left( (w_t - \bar{w}) - \sum_{j} \tilde{\gamma}_{i,j} (w_{t-j} - \bar{w}) \right) \tilde{\epsilon}_t^i P(S_t = i | z_{t-1}, w_{t-1}, \ldots, z_1, w_1) = 0. \tag{7.50}
\]

Note that \(\tilde{\gamma}_{i,j}\) is selected so that (7.48) and (7.49) hold. (7.50) will be guaranteed if

\[
\sum_{t} \sum_{i=0,1} (w_t - \bar{w}) \left( \hat{\alpha}_i - \sum_{j=1}^l \tilde{\gamma}_{i,j} \hat{y}_{t-j} \right) P(S_t = i | z_{t-1}, w_{t-1}, \ldots, z_1, w_1) = 0 \tag{7.51}
\]

\[
\sum_{t} \sum_{i=0,1} (w_{t-j} - \bar{w}) \tilde{\epsilon}_t^i P(S_t = i | z_{t-1}, w_{t-1}, \ldots, z_1, w_1) = 0 \tag{7.53}
\]

(7.51) holds if

\[
\sum_{t} (w_t - \bar{w}) \hat{y}_t = 0 \tag{7.52}
\]

\[
\sum_{t} w_t (\hat{y}_t - \bar{y}) = 0
\]

\[
\sum_{t} w_t ((z_t - \bar{z}) - \hat{\alpha}_1 (w_t - \bar{w})) = 0,
\]

corresponds to (7.45). The other equations overidentify the parameters but are satisfied in large sample as long as \(w_t\) is strictly exogenous. So far, we have shown that the two-step estimators coincide asymptotically with the global MLE. Now we turn our attention toward the independence.

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

\[
\frac{\partial^2 L}{\partial a_1 \partial \alpha_i} = -\frac{1}{\sigma^2} \sum_{t} \left( (w_t - \bar{w}) - \sum_{j} \gamma_{i,j} (w_{t-j} - \bar{w}) \right) P(S_t = i | z_{t-1}, w_{t-1}, \ldots, z_1, w_1)
\]

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

because

\[
E [(w_{t-j} - \bar{w}) \bar{y}_z | \hat{y}_t = i | w_{t-1}, \ldots, z_1, w_1]
\]

\[
= E [(w_{t-j} - \bar{w}) \bar{y}_z | \hat{y}_t = 1 | y_{t-1}, \ldots, y_1]
\]

\[
= E [(w_{t-j} - \bar{w}) \bar{y}_z | S_t]
\]

\[
= E (w_{t-j} - \bar{w}) E (S_t)
\]

\[
= 0, \quad j = 0, 1, \ldots, l,
\]
assuming that \( w_t \) is uncorrelated with \( y_t, \ldots, y_T \).

\[
\frac{\partial^2 L}{\partial a_1 \partial \gamma_{i,j}} = -\frac{1}{\sigma^2} \sum_t \left( (w_t - \bar{w}) - \sum_k \gamma_{i,k} (w_{t-k} - \bar{w}) \right) y_{t-j} P(S_t = i|z_{t-1}, w_{t-1}, \ldots, z_1, w_1) \\
- \frac{1}{\sigma^2} \sum_t (w_{t-j} - \bar{w}) \varepsilon_t^i P(S_t = i|z_{t-1}, w_{t-1}, \ldots, z_1, w_1)
\]

\[
E \left[ \frac{\partial^2 L}{\partial a_1 \partial \gamma_{i,j}} \right] = 0.
\]

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


[15] Davies, R.B. (1977) “Hypothesis testing when a nuisance parameter is present only under the alternative” Biometrika, 64, 247-254.
[16] Davies, R.B. (1987) “Hypothesis testing when a nuisance parameter is present only under the alternative” Biometrika, 74, 1, 33-43.


Figure 7.1: \( \pi \in [0.15, 0.85], \rho \in [-0.7, 0.7] \)

Figure 7.2: \( \pi \in [0.01, 0.99], \rho \in [-0.98, 0.98] \)
Figure 7.3: Linear model with intercept

Figure 7.4: ARCH(1)
Figure 7.5: IGARCH(1,1)
Figure 7.6: Regime Probabilities