Efficient Inference in Time Series Models with Conditional Heterogeneity

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Guido Markus Kuersteiner

Dissertation Director: Prof. Peter C.B. Phillips

May 1997

UMI Number: 9731016

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Abstract

Efficient Inference in Time Series Models with Conditional Heterogeneity

Guido Markus Kuersteiner

1997

This dissertation develops new techniques to improve the efficiency of estimators where regressors and errors are uncorrelated, but not independent. A natural case where such phenomena occur are linear time series models with martingale difference innovations. Prominent parametric examples mainly encountered in financial econometrics include models with heterogeneity in the conditional error distribution. Estimators not taking the dependence between the errors and regressors into account are generally inefficient. The form of this inefficiency however is not related to heterogeneity in the second moments as is the case in standard GLS type problems. Rather it is reflected by the appearance of fourth order cumulant terms in the asymptotic covariance matrix. The martingale property of the errors turns out to be the critical assumption allowing for a decomposition of the fourth order cumulant terms. This decomposition in turn is used to obtain a lower bound for the asymptotic covariance matrix.

Efficient estimators for linear time series models are developed by extending the scope of instrumental variable procedures to the case of conditional heterogeneity. The instrumental variables approach is used as a general technique to alter the statistical properties of the score function. Optimal instruments constructed from the reweighted innovation

sequence are used to invert the fourth order cumulant spectrum, heuristically speaking. This transformation is defined such that the variance of the score process matches the expectation of the first derivative of the score. It is shown that the optimal IV estimator has the lowest variance in the class of all instruments which are linear filters of the innovation process. Unobservability of the optimal instruments necessitates a semiparametric approach. The optimal filter is estimated from fourth order cumulants of consistent first stage regression residuals. The optimal instruments are then obtained by frequency domain techniques which convolute the filter with the residuals in a computationally efficient way.

While no parametric assumption for the generating mechanism leading to higher moment dependence of the errors is made, a number of well known parametric specifications fall into the class of processes considered. These include ARCH, GARCH and stochastic volatility models. Monte Carlo simulations are conducted to examine the finite sample properties of the instrumental variables estimator.

Contents

List of Tables

List of Figures

Acknowledgements

I wish to thank the members of my dissertation committee, Peter C.B. Phillips, Donald W.K. Andrews and Oliver Linton for their support and encouragement while I was working on this dissertation. I received many valuable comments and critical assessment of previous drafts. In particular I wish to thank Peter Phillips for spending many hours in conversations and for supporting me in all aspects of my work at Yale. I have also benefited from comments by Chris Sims, Robert Shiller and participants of the econometrics workshop at Yale. My colleagues Benoit Perron and Hyungsik Moon read an early draft of this dissertation and made helpful comments. Financial support from Yale University and from an Alfred P. Sloan Doctoral Dissertation Fellowship is gratefully acknowledged. Last but not least I wish to thank my parents for their continued encouragement during my time at Yale.

1. Introduction

This dissertation analyzes the asymptotic efficiency of linear time series models when the innovations have conditionally heterogenous distributions. New instrumental variables (IV) estimators for autoregressive time series models are constructed in a way to achieve an efficiency gain over traditional estimators based on Gaussian criterion functions.

The time series models considered include error processes which are conditionally heteroskedastic of unknown functional form. Efficiency gains are obtained without having to specify a model for the dependence in the errors. The setup is general enough to account for stylized facts in many economic time series displaying features such as thick tailed distributions and time dependent conditional variances.

Classical efficiency results for the quasi maximum likelihood estimator (QMLE) of the autoregressive parameters such as Hannan[42] depend on independence of the errors. With dependence in the errors the variance of the score process typically contains fourth order cumulant terms. This leads to a loss of efficiency of the QMLE. The asymptotic distribution of the QMLE in the case of martingale difference errors is shown to be a special case of the more general frequency domain estimators discussed in Hosoya and Taniguchi[54] and Keenan[59]. In the cases considered by these authors, no restrictions are imposed on the functional form of the fourth order cumulant spectrum of the process. Here it is shown, that under the martingale assumption the fourth order cumulant spectrum can be factorized. The factorization is used to obtain a lower bound for the covariance matrix. This result is the key to the construction of an optimal instrumental variables estimator.

An instrumental variables estimator using the suitably reweighted innovation process as instrument is shown to achieve the lower bound for the covariance matrix in the class of IV estimators with instruments that are linear filters of the errors. The distributional assumptions focus on the strict stationarity and ergodicity properties of the errors.

Since the optimal instruments are unobservable they need to be estimated nonparametrically. Assumptions about the generating mechanism of the volatility process or more generally the dependence in higher moments are replaced by smoothness assumptions for higher order cumulant spectra of the errors. This setup allows for the treatment of dependence in higher moments as a nuisance parameter. Nonparametric estimators of this nuisance parameter are used to construct the optimal instruments.

The term efficiency as used throughout this dissertation refers to the lowerbound within the class of linear instrumental variables estimators. Restricting the class of instruments to linear functions of the innovation process has considerable consequences for the efficiency properties of the estimator. The advantage of working with this small class of instruments lies in the fact, that the instrumental variables estimator can be expressed in the form of a linear filter applied to the data. This results in extreme computational stability of the procedure. Simulation results in Part III indicate that the IV estimator dominates the Gaussian estimator even in small samples.

It is worthwhile to review more general efficiency concepts. We proceed by first discussing estimators that are asymptotically minimax in a neighbourhood of the true parameter. An alternative, weaker concept of efficiency is in terms of lower bounds for estimators based on moment restrictions.

1.1. Efficiency Bounds for Parametric Models

Efficiency bounds for finite samples are characterized by Cramer Rao lower bounds which depend on the information matrix of the likelihood. This notion of efficiency can be extended to the information matrix of the asymptotic distribution when the class of estimators is sufficiently restricted.

The most commonly used formulation is the local asymptotic minimax criterion (LAM) of Hajek [38] and Fabian and Hannan [28]. This definition requires an estimator to have lowest possible expected risk for all bowl shaped risk functions in a neighbourhood of the true parameter. The definition rules out superefficient estimators which have good properties only on a dense subset of the parameter space. The LAM property is typically established by showing that the local asymptotic normality (LAN) condition of Le Cam [16] holds (see Fabian and Hannan [28]). The construction of the lower bound assumes knowledge of the true data generating model and is therefore fully parametric.

For expositional purposes we introduce notation following Le Cam [17]. Let $\{y_t\}_{t=1}^n$ be a discrete scalar valued time series defined on a measurable space $(\Omega_n, \mathcal{F}_n)$. For each n there is a filtration $\mathcal{F}_{n,k} \subseteq \mathcal{F}_{n,k+1}$. We consider a sequence of two probability measures $P_{\beta_0,n}$ and $P_{\beta_n,n}$ with restrictions to $\mathcal{F}_{n,k}$ denoted by $P_{\beta_0,n,k}$ and $P_{\beta_n,n,k}$. The likelihood ratio of $P_{\beta_n,n}$ with respect to $P_{\beta_0,n}$ is defined as the Radon-Nykodym derivative on $\mathcal{F}_{n,k}$ of the part of $P_{\beta_n,n,k}$ dominated by $P_{\beta_0,n,k}$. We write the likelihood ratio as $L_{n,k}(\beta_n, \beta_0) = dP_{\beta_n,n,k}/dP_{\beta_0,n,k}$ where reference to the two parameters emphasizes the fact that the two measures are indexed by points in the parameter space.

Fabian and Hannan [28] define the *LAN* property for the likelihood ratio by requiring the following conditions for (β_0, M_n, Δ_n): $\Delta_n \Rightarrow N(0, \Lambda_0)$ under $P_{\beta_0,n}$ probability; for any bounded sequence $h_n \in \mathbb{R}^p$ $\beta_n = \beta_0 + M_n^{-1/2} h_n$; and

$$\log dP_{\beta_n,n}/dP_{\beta_0,n} - h_n' \Delta_n + \frac{1}{2} h_n' \Lambda_0 h_n \rightarrow 0$$

in $P_{\beta_0,n}$ probability. A sequence of estimators T_n then is called regular if LAN (β_0, M_n, Δ_n) holds and $M_n^{1/2}(T_n - \beta_0) - \Delta_n \rightarrow 0$ in $P_{\beta_0,n}$ probability. Now define the class of loss functions $\mathfrak{L}$: $l \in \mathfrak{L}$ if for all $u, v \in \mathbb{R}^p$ $l(u) = l(-u)$, the set $\{u : l(u) < \zeta\}$ is convex for any $\zeta \in (0, \infty)$, $l(0) = 0$, $l(u)$ is continuous at $u = 0$ and $\int_{\mathbb{R}^p} l(u) e^{-1/2\lambda \|u\|^2} du < \infty$. Then by Fabian and Hannan [28, Theorem 6] the expected loss of any regular estimator T_n is bounded by

$$\lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{\left\| M_n^{1/2}(\beta - \beta_0) \right\| \leq K} E_{\beta,n} l(Q_n M_n^{1/2} \Lambda_0^{-1/2} (T_n - \beta_0)) \geq E_{N(0,I_p)} l(u)$$

where $E_{N(0,I_p)} l(u) = (2\pi)^{-p/2} \int_{-\infty}^{\infty} l(u) e^{-1/2\lambda \|u\|^2} du$ and Q_n is any sequence of orthogonal matrices.

If we assume a parametric structure for the time series model then we can apply arguments in Phillips [80]. Let y_t be generated by

$$y_t = m_t(\beta) + \varepsilon_t$$

where $m_t(\beta)$ is $\mathcal{F}_{n,t-1}$ measurable and ε_t is a martingale difference sequence. For a univariate *AR*(p) model for example, $m_t(\beta) = \beta_1 y_{t-1} + ... + \beta_p y_{t-p}$. The key assumption here is that ε_t is not *iid*. The conditional density of ε_t therefore depends on the whole innovation sequence in general. It is then convenient to express the log likelihood ratio as a telescoping sum using the fact that

$$L_n(\beta_1, \beta_0) = (L_n/L_{n-1}) \cdots (L_1/L_0)$$

i.e. $\log L_n(\beta_1, \beta_0) = \sum_{k=1}^{n} \log(L_k(\beta_1, \beta_0)/L_{k-1}(\beta_1, \beta_0))$. Using the definition of $L_n(\beta_1, \beta_0)$ the ratio of likelihoods is easily recognized as a ratio of conditional densities since

$$\frac{L_k(\beta_1, \beta_0)}{L_{k-1}(\beta_1, \beta_0)} = \frac{dP_{1,k}}{dP_{1,k-1}} \Big/ \frac{dP_{0,k}}{dP_{0,k-1}} = \frac{f_1(y_k \,|y_{k-1} \cdots y_1))}{f_0(y_k \,|y_{k-1} \cdots y_1))}$$

Correspondingly we can, as in Phillips [80], define the likelihood ratio for a model parametrized by β as $\log f_{n,\beta} = \sum \log f_\beta(y_k \,|y_{k-1} \ldots y_1)$. The score process can be written as

$$\partial/\partial\beta \log f_{n,\beta} = \sum \partial/\partial\beta \log f_\beta(y_k|y_{k-1} \cdots y_1) = n^{1/2} V_n(\beta).$$

By the chain rule $V_n(\beta)$ is $V_n(\beta) = n^{-1/2} \sum(\partial m_k/\partial\beta)(\partial/\partial m_k) \log f_\beta(y_k|y_{k-1} \cdots y_1)$ where $(\partial m_k/\partial\beta)$ is $\mathcal{F}_{n,t-1}$ measurable and $(\partial/\partial m_k) \log f_\beta(y_k|y_{k-1} \cdots y_1)$ is a martingale difference sequence under the $P_{\beta,n}$ measure. The matrix conditional quadratic variation process of $\log f_{n,\beta}$ can then be written as

$$B_n(\beta) = \frac{1}{n} \sum Z_k(\beta) Z_k(\beta)' h_k(\beta)$$

with $Z_k(\beta) = \partial m_k/\partial\beta$ and $h_k(\beta) = E(e_k^2|\mathcal{F}_{n,k-1})$ where $e_k = (\partial/\partial m_k) \log f_\beta(y_k|y_{k-1} \cdots y_1)$.

Now evaluate $\log f_{n,\beta_n}$ in a $\sqrt{n}$ neighbourhood of β_0 such that $\beta_n = \beta_0 + n^{-1/2}h_n$, where h_n is a bounded sequence. Then, under regularity conditions, it follows from the results in Le Cam and Yang [18] that the likelihood ratio can be approximated locally by a quadratic form (LAQ) as

$$\log L_n(\beta, \beta_0) \sim h_n' B_n(\beta_0)^{-1/2} V_n(\beta_0) - (1/2) h_n' B_n(\beta_0) h_n$$

in $P_{\beta_0,n}$ probability. Phillips [80] shows that this likelihood ratio can be embedded in a continuous time martingale under additional conditions. In particular if we define stopping times τ_k such that $E(\tau_k - \tau_{k-1}|\mathcal{F}_{n,k-1}) = E(e_k^2|\mathcal{F}_{n,k-1})$ a.s. in $P_{\beta_0,n}$ then $\Lambda(\tau_k) \sim h_n' \int_0^{\tau_k} SdW - (1/2)\, h_n'(\int_0^{\tau_k} SS')h_n$ with $S(r) = Z_k$ for $\tau_{k-1} < r < \tau_k$ and W a standard Brownian motion.

Under the stationarity and ergodicity assumptions it is also true that B_n converges to a constant positive definite matrix if $E(Z_k(\beta)Z_k(\beta)' h_k(\beta)) < \infty$. This follows from the ergodic theorem. Then, by the martingale difference CLT, $V_n(\beta_0) \Rightarrow N(0, B(\beta_0))$ under $P_{\beta_0,n}$ probability with $B(\beta_0) = E(Z_k(\beta_0)Z_k(\beta_0)' h_k(\beta_0))$. Under these conditions the LAQ property, as far as it exists, is strengthened to LAN (see Le Cam and Young [18]).

In our context it follows from Fabian and Hannan [28, Theorem 3] that, for a model with the LAN property, every regular estimator T_n such that $\sqrt{n}(T_n - \beta_0) - B(\beta_0)^{-1}V_n(\beta_0) = o_p(1)$ under $P_{\beta_0,n}$ probability is *LAM*, i.e. the following equality holds

$$\lim_{K\to\infty} \liminf_{n\to\infty} \sup_{\left\|M_n^{1/2}(\beta-\beta_0)\right\| \leq K} E_{\beta,n} l(Q_n M_n^{1/2} B(\beta_0)^{-1/2}(T_n - \beta_0)) = E_{N(0,I_p)} l(u).$$

Under additional conditions the LAN property can be exploited to construct asymptotically efficient estimators where the density is unknown and estimated nonparametrically. This technique was introduced by Bickel [7]. Examples of such estimators in the time series literature include Beran [6], Swensen's [91] adaptive estimator for autoregressive models and Kreiss [60] extension to the ARMA case. Linton [63] constructs an adaptive estimator for a regression model with ARCH error process. The common feature of these approaches is that they allow the unknown likelihood to be specified as a product of *iid* density functions.

In our example the conditional density $f_\beta(y_k|y_{k-1}\dots y_1)$ reduces to $f_\beta(\varepsilon_t)$ if the ε_t are independent. An adaptive estimator $\tilde{\beta}$ for β_0 can then be constructed from a consistent first step estimator $\hat{\beta}$ by $\tilde{\beta} = \hat{\beta} + \hat{B}(\hat{\beta})^{-1}\hat{V}_n(\hat{\beta})$ where $\hat{V}_n(\hat{\beta}) = n^{-1/2}\sum \hat{f}'_\beta(\hat{\varepsilon}_t)/\hat{f}_\beta(\hat{\varepsilon}_t)Z_t(\hat{\beta})$. Here $\hat{f}_\beta(.)$ and $\hat{f}'_\beta(.)$ are nonparametric estimates of the innovation density and the first derivative of the innovation density and we assume $\hat{B}(\hat{\beta}) \to B(\beta_0)$. Adaptiveness then obtains if $\hat{V}_n(\hat{\beta}) - V_n(\hat{\beta}) = o_p(1)$ in $P_{\beta_0,n}$ probability. In the context of conditional heterogeneity, i.e., when $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$ is not constant, adaptiveness can be achieved by transforming the model to $y_t/\sigma_t = m_t(\beta)/\sigma_t + \varepsilon_t/\sigma_t$. Such a transformation is feasible if σ_t can be estimated uniformly consistently for all t and the standardized innovation ε_t/σ_t is *iid*. An example of such an estimator is developed by Linton [63].

More generally however, ε_t is not independent of the conditioning variables even after standardization by σ_t. In this case, estimating the conditional density nonparametrically seems to be a less promising approach because of the potentially high dimensionality of the density. Under these conditions one could still perform the *GLS* transformation and achieve a GMM type lower bound to be discussed in the next section. This however is only possible if the σ_t can be estimated consistently.

1.2. Efficiency Bounds Implied by Moment Restrictions

We have seen in the previous section that information matrix lower bounds depend on knowledge of the likelihood function. This means that the statistical model determines the distribution of the data completely even though the true distribution may be unknown to the investigator. In certain special cases the density can be estimated nonparametrically leading to adaptive estimators. This concept was introduced by Stein [89]. It stands for achieving the same efficiency bound when the density is estimated nonparametrically rather than known up to a finite number of unknown parameters.

Here we look at statistical models which restrict only certain moments of the distribution. In particular we assume that there is a vector of observations $z_t = (y_t, x_t)$ and a known function $m(z_t, \beta) \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^p$ such that

$$E[m(z_t, \beta)|x_t] = 0 \tag{1.1}$$

where E denotes integration w.r.t the true underlying distribution F_0. In the time series context x_t contains lagged values of y_t and the expectation is conditional on past information. Levit [62] shows that estimators based on a finite number of unconditional moment restrictions have efficiency bounds characterized by the infimum of the information matrix over the class of all distributions satisfying the moment restrictions. Chamberlain [19] obtains a representation theorem for semiparametric estimators. Chamberlain [20] extends

Levits result to the case of conditional moment restrictions by approximating the space of distributions with multinomial distributions. This allows to reduce (1.1) to a finite set of unconditional moment restrictions. Efficiency bounds for estimators of model (1.1) are then characterized by the infimum of the information matrix over the class of all distributions satisfying the restriction (1.1). This efficiency bound is in general larger than the fully parametric bound if (1.1) is the only restriction imposed. Newey [73] shows that imposing additional restrictions on the error distribution such as independence or symmetry can lead to an efficiency bound corresponding to the full likelihood based lowerbound.

For the case where the statistical model is only restricted by (1.1) Chamberlain [20] shows under *iid* conditions that the GMM efficiency bound for an estimator T_n for the parameter β_0 is $\Lambda_0 = E(D_0(x)\Sigma_0^{-1}(x)D_0(x))^{-1}$ where $D_0(x) = E(\partial m(z_t, \beta_0)/\partial\beta|x)$ and $\Sigma_0(x) = E(m(z_t, \beta_0)m(z_t, \beta_0)'|x)$. As before, asymptotic efficiency is defined in the local minimax sense, i.e., for all bowl shaped loss functions l the following inequality holds

$$\liminf_{n\to\infty} \sup_{(F,\beta)\in\Gamma} E_F l(\sqrt{n}\Lambda_0^{-1/2}(T_n' - \beta_0)) \geq E_{N(0,I_p)} l(u) \tag{1.2}$$

where Γ is the set of all distributions satisfying the restrictions of the model. The arguments in Chamberlain [20] show that under the assumption of independence for the error distribution there is a one to one mapping between the moment restrictions and the true distribution locally at β_0. As a consequence, under these restrictions, Λ_0 corresponds to the semiparametric bound such that in this case (1.2) holds with equality for the efficient GMM estimator.

Newey [75] reviews the literature on semiparametric efficiency bounds. Semiparametric estimators achieving the bound Λ_0 were constructed for the linear regression case by Robinson [81] by accounting for conditional heteroskedasticity of unknown form in a nonparametric way. Newey [74] considers the instrumental variables estimator

$$\hat{\beta} = \arg\min_{\beta} \sum m(z_t, \beta_0)' a(x_t)' \left[\sum a(x_t) a(x_t)'\right]^{-1} \sum m(z_t, \beta_0)' a(x_t)' \qquad (1.3)$$

in an *iid* context. The optimal instrument is $a(x_t) = D_0(x_t)\Sigma_0^{-1}(x_t)$. Since the conditional expectations $D_0(x_t)$ and $\Sigma_0(x_t)$ are unknown functions of x_t they have to be estimated nonparametrically. Robinson [84] treats a similar problem in a time series context but assumes that $m(z_t, \beta_0)$ can be solved for y_t either analytically or numerically. This simplifies estimation of $D_0(x_t)$. The assumptions made do however not allow for heteroskedasticity. A frequency domain version of an optimal instrumental variables estimator in a linear time series framework is considered in Robinson [83].

An alternative approach to nonparametric estimation of the instrument $a(x_t)$ has been investigated by Cragg [22], Chamberlain [20], Newey [73] and Hansen [46]. The approach is based on the fact that the conditional moment restriction (1.1) implies an infinite number of unconditional moment restrictions of the form $E(m(z_t, \beta)g(x_t)) = 0$ where $g(x_t)$ is measurable and $E(g(x)^2) < \infty$. Chamberlain [20] shows that the optimal instrumental variables estimator can be approximated arbitrarily well by a GMM estimator. A complete sequence of functions $\{g_j(x)\}_{j=1}^{\infty}$ is defined such that for any $h \in L_2(F)$ and $\epsilon > 0$ there are real numbers $\alpha_1, ..., \alpha_k$ such that $\int [h(x) - \sum \alpha_j g_j(x)]^2 dF(x) < \epsilon$. A GMM estimator is constructed from $B_k(x) = I_p \otimes [g_1(x), ..., g_k(x)]'$ and $\psi_k(z, \beta) = B_k(x) m(z, \beta)$. The estimator $\hat{\beta}_k$ is defined as minimizing $\sum \psi_k(z_t, \beta) \Upsilon_k^{-1} \sum \psi_k(z_t, \beta)'$ with $\Upsilon_k = E[B_k(x)\Sigma_0(x)B_k(x)']$. Then by the arguments in Hansen [45], under regularity conditions, $\sqrt{n}(\hat{\beta}_k - \beta_0) \Rightarrow N(0, \Lambda_k)$ where $\Lambda_k^{-1} = E[\partial\psi_k(z,\beta)/\partial\beta]\Upsilon_k^{-1}E[\partial\psi_k(z,\beta)/\partial\beta']$. Chamberlain shows that $\Lambda_k \to \Lambda_0$ as $k \to \infty$ where the main idea of the proof is to approximate $D_0(x)$ by projecting it onto $\{g_j(x)\}$. In the time series context, Hansen [46] shows existence of an approximating martingale difference sequence for the optimal instruments.

Time series estimators based on such martingale approximations have been developed by Hayashi and Sims [51], Stoica, Söderström, and Friedlander [90], Hansen [46] and Hansen and Singelton [49]. For the case of independent innovations, frequency domain approximations of these procedures were obtained in Hannan [41] for the stationary case and for the nonstationary case by Phillips [78], [79].

Hansen [47] analyzes the relationship between the GMM variance lowerbound and the semiparametric lowerbound in the sense of Levit [62] and Chamberlain [19]. The setup is a linear vector time series with *iid* Gaussian innovations where conditional moment restrictions are available only for a subvector. The class of all linear processes satisfying these restrictions forms the class of distribution functions over which the semiparametric bound is formed. Hansen shows that the GMM lowerbound corresponds to the least informative likelihood specification. This establishes that GMM estimators attain the semiparametric lowerbound in the class of Gaussian processes.

Hansen and Singelton [49] consider the non Gaussian case but retain the independence assumption for the innovation process. The setup is again a linear vector time series where conditional moment restrictions are available only for a subvector. Hansen and Singelton obtain explicit expressions for the optimal instrument vector and show that a GMM

estimator based on an increasing number of lagged observations as instruments achieves the GMM lowerbound. This lowerbound corresponds to the asymptotic variance of full system estimators based on Gaussian likelihood functions.

The general case without independence assumptions is treated in Hansen [46] where the existence of a lower bound for the asymptotic covariance matrix of GMM estimators with an infinite set of $\mathcal{F}_{t-1}$ measurable instruments is established. The setup is general enough to include models with conditionally heterogeneous errors. Hansen, Heaton and Ogaki [48] show that the GMM efficiency bound for the conditionally heteroskedastic case can be achieved with a set of instruments based on the innovation sequence weighted by its conditional second moments. In fact this transformation reduces the innovations back to the case where $E(\varepsilon_t^2|\mathcal{F}_{t-1})$ is constant. This is the key assumption in Hannan's treatment. It also has to be emphasized that in general the conditional second moments are unknown functions depending on the entire past innovation sequence. The results in Hansen [46] show that this function can in principle be approximated by an infinite set of instruments. To this date however no feasible versions of such estimators have been constructed.

The results in this dissertation show, that efficiency gains for the conditionally heteroskedastic case can be achieved by restricting the instruments to the linear class. It can be shown that the IV estimators proposed here are asymptotically equivalent to GMM procedures based on a infinite set of instruments constructed from past observations y_{t-k}. Applying such a procedure literally is not feasible in any sample of reasonable size since it involves inversion of an $n \times n$ weight matrix. It is shown here how the problem can be transformed into a computationally efficient procedure.

IV estimators in the linear class do not attain the *GMM* lowerbound under conditional heterogeneity. Construction of feasible *IV* estimators attaining the bound remains therefore an open question for research. The justification for the use of a linear class of instruments lies in its simplicity. This aspect is especially important for generalizations of the approach to more general regression models and to multivariate contexts. Empirical applications include hypothesis tests and confidence interval estimation. One area where efficiency gains from accounting for conditional heteroskedasticity are important are rational expectations models for financial markets. The *IV* estimators proposed here offer efficiency gains without having to specify the functional form leading to heteroskedasticity. The simulation results in Part III indicate that estimators are sensitive to misspecification of these higher moment aspects of the data which underlines the need for robust semiparametric methods.

We turn to a discussion of alternative feasible estimators taking the conditional heteroskedasticity of the errors into account. An important class of procedures is fully parametric while more recently semiparametric and Bayesian alternatives have been suggested.

1.3. Feasible Procedures for Conditional Heteroskedasticity

Efficient estimation of regression parameters under stochastic conditional heteroskedasticity was first studied by Engle[26] in his influential paper introducing the ARCH model. Generalizations of ARCH include GARCH (Bollerslev[9]), ARCH-M (Engle, Lilian and Robins[27]) and EGARCH (Nelson[70]). ARCH specifications have been extended to multivariate models as for example in Bollerslev, Engle and Wooldbridge[10]. An asymptotic theory for ARCH models has been obtained by Weiss[94]. His results were extended in subsequent papers on GARCH(1,1) and IGARCH(1,1) specifications by Lumsdaine[64] and Lee and Hansen[61]. Yet as noted by Lee and Hansen[61], the asymptotic theory for the GARCH(p,q) class does not easily follow from their work on the special case of a GARCH(1,1) model.

Motivated by the study of continuous time option pricing models with changing volatility, an alternative formulation of conditionally heteroskedastic models has become popular. Stochastic volatility models represent the time changing variance as a separate stochastic process. This can be interpreted as a generalization of the ARCH model where the variance is restricted to be a function of the past errors in the measurement equation alone (see Andersen[1]). Models of this type are studied amongst others in Melino and Turnbull[67], Harvey, Ruiz and Shepard[50], Geweke[35] and Jacquier, Polson and Rossi[56]. Multivariate extensions have been considered by Harvey, Ruiz and Shepard[50] and Boudoukh[11].

Inference for stochastic volatility models is complicated by the unobservability of the conditional variance. Estimation was first carried out by moment matching techniques as in Chesney and Scott[21]. Melino and Turnbull[67] introduced GMM estimation, but note that selection of the appropriate moments is arbitrary to some degree.

A recent approach based on the marginal distribution of the errors has been proposed by several authors including Danielsson and Richard[23], Geweke[35], Jacquier, Polson and Rossi[56], Mahieu and Schotman[66] and Shephard[86],[87]. Despite distributional assumptions on the volatility process, the marginal distribution of the errors is generally not known analytically. Yet recent developments of simulation techniques, for example Pakes and Pollard[77] and Geweke[36], make a numerical evaluation of the marginal distribution possible. The strength of the simulations based approach is that it puts inference back into a likelihood context where optimality properties of the estimators should hold. The disadvantage of the procedure, however, lies in the complete distributional assumptions that have to be made about the conditional moments. Computational considerations might also be relevant especially for multivariate generalizations.

An alternative solution is proposed by Uhlig [93] where a Bayesian VAR with stochastic volatility is proposed. Conjugacy between Wishart and multivariate singular Beta distributions is exploited to determine the joint posterior analytically. The advantage of the analytical solution is that the posterior only depends on the final stage conditional variance leading to a significant reduction in the dimensionality of the numerical integration problem.

Semiparametric alternatives to quasi maximum likelihood inference for the ARCH class were introduced by Linton[63]. It is established that these adaptive procedures are local asymptotic minimax as defined by Hajek[38] and Fabian and Hannan[28]. Steigerwald[88] extends the analysis to GARCH, EGARCH and power ARCH cases. An alternative semiparametric approach is developed in Gallant and Nychka[32] and applied to financial data in Gallant and Tauchen[33] and Gallant and Long[34].

A common feature of the parametric and semiparametric approaches discussed so far is that they all specify generating mechanisms for the higher moment dependence of the error process. The focus of the analysis in this literature is often the volatility process itself. If, however, the investigator is only interested in estimating parameters in the regression equation, then choosing a particular parametric model for the statistical properties of the errors is often undesirable. In the *iid* case, we typically only impose restrictions on some moments of the errors, rather than specifying the distribution completely. This idea is extended here to the case where the errors are not independent. Inference is based solely on quantities that can be estimated by assuming that the model of interest, i.e., the regression equation, is correctly specified. As a consequence, the techniques introduced here differ in many ways from the literature on conditional heteroskedasticity. First, and most importantly, only estimation of the regression parameters is considered. The conditional volatility process is treated as a nuisance parameter which is handled by nonparametric techniques. Secondly, the inferential framework is not a likelihood approach, since we do not make enough assumptions to specify the data density or to estimate it nonparametrically. Neither do conditional moments enter the picture. Again, the assumptions made do not allow one to estimate them in a consistent way. Inference is therefore based on unconditional moments of the data and of estimated errors.

The procedures proposed here are similar in spirit to the semiparametric GLS and instrumental variables (IV) estimators of Robinson [81] and Newey [74], where no parametric assumptions about the form of conditional heteroskedasticity are made. However, in order to estimate the conditional variance, these authors have to assume independent errors. This assumption has precluded direct application of their techniques to the stochastic conditional variance case.

Hidalgo [52] relaxes the *iid* assumption for the errors but has to assume instead that the conditional variance is a smooth function of an independent stationary process. Hansen[44] treats the stochastic volatility model in a semiparametric GLS framework. He assumes that the conditional variance process converges to a Brownian motion in the limit. Sample path continuity of the limit process then allows for consistent kernel estimation of the conditional variance.

Semiparametric GLS estimators in general require consistent estimation of the conditional variance. Nelson[71] and Nelson and Foster[72] discuss conditions under which consistent estimation of the conditional variance is possible. An example where consistency fails is a mixed jump diffusion for the volatility. More generally, consistency is likely to fail if there exists no continuous approximation to the volatility process. Instead, the IV estimators proposed here do not rely on the consistent estimation of conditional moments. They are constructed uniquely from unconditional moments and thereby allow for a wide range of possible generating mechanisms.

1.4. Organization

The dissertation is organized as follows. Part I discusses efficiency bounds for the general case of univariate linear processes. Section 2 specifies the model and describes the parameter space. Section 3 analyzes the asymptotic distribution of the Gaussian Quasi Maximum Likelihood estimator. Section 4 derives a lower bound for the asymptotic covariance matrix of the Gaussian QMLE. Section 5 shows that a class of instrumental variables estimators with instruments that are linear in the innovation sequence has the same lower bound for the covariance matrix. This fact is then used to identify the optimal IV procedure in this linear class.

Part II is concerned with the implementation of the optimal IV procedure identified in the first part. This is done for the case when the linear time series model is an $AR(p)$

process. Section 6 introduces additional restrictions on the fourth order moments of the innovation process in order to simplify the construction of the optimal instrument. Section 7 obtains a time domain representation of the IV estimator. Section 8 discusses a frequency domain approximation and Section 9 constructs a feasible semiparametric estimator.

Part III analyzes the efficiency properties of the IV procedure for the case of an $AR(1)$ model. The IV procedure is compared to full maximum likelihood procedures and to misspecified likelihood procedures. Simulation results are reported for a variety of specifications regarding conditional heterogeneity of the error process.

Proofs of some important lemmas are given in Appendix A while the main results of Parts I and II are proved in Appendices B and C.

Part I

Efficiency Bounds

The first part of this dissertation analyzes the asymptotic distribution of estimators for the parameters of general univariate linear time series models. Special cases of these models are the $ARMA(p, q)$ class. The innovations driving the time series model are assumed to be martingale difference sequences. As a consequence, the linear specification correctly models the conditional mean of the data.

While the martingale difference assumption results in uncorrelatedness of the innovations, it generally does not imply independence. A consequence of the dependence in the errors is that fourth moment terms do not factor into a product of second moments. This has important consequences for the asymptotic distribution of linear time series models. The covariance matrix now is a function of fourth order cumulants and asymptotic normality depends on finiteness of these fourth moments.

In general, the asymptotic covariance matrix depends on the full trispectrum of the innovation sequence. Under the martingale difference assumption the trispectrum reduces to a bispectrum. This simplification is the key to an orthogonalization of the asymptotic covariance matrix.

Using the decomposition of the asymptotic covariance matrix, a matrix lower bound based on the Cauchy Schwartz inequality is obtained. The main result of this part is to show that this lower bound is a lower bound for the covariance matrix of a class of instrumental variables estimators. The class is defined by restricting the instruments to be linear functions of the innovation process. Such a restriction certainly involves a cost in terms of potential additional efficiency gains from nonlinear instruments. The justification of the procedure is based on practical considerations. It results in estimators which are easy to compute and weakly dominate Gaussian estimators in the sense that their asymptotic variance is bounded above by the variance of the Gaussian estimators on the whole parameter space. The asymptotic properties can be fully analyzed for parametric examples. This is done in Part III.

In this part, M-estimators based on orthogonality restrictions between current innovations and instruments are analyzed. The discussion is at a general level and the focus is on the efficiency properties of the estimators. The asymptotic theory is derived based on high level assumptions. The implementation of these estimators is left to Part II where a simplified case is treated in full detail.

2. Model Specification

We assume that we have a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathcal{F}_t$ of increasing σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}\ \forall t$. There is a doubly infinite sequence of random variables $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ generating the filtration $\mathcal{F}_t$. The innovations ε_t are assumed to be a martingale difference sequence. This has important consequences for the fourth order cumulants. Following Brillinger [13], the 4-th order cumulant is defined as

$$c_{\varepsilon..\varepsilon}(u_1, u_2, u_3, u_4) = \sum (-1)^{p-1} (p-1)! m_{\varepsilon..\varepsilon}(u_{\nu_1}) \cdots m_{\varepsilon..\varepsilon}(u_{v_k}) \tag{2.1}$$

where the sum is over all partitions $\nu_1, ..., \nu_k$ of the numbers $1, ..., 4$ and u_{ν_i} is a multi index of all the elements in ν_i. Strict stationarity implies that one time index u_i can always be normalized to zero without loss of generality. By the martingale difference property $c_{\varepsilon..\varepsilon}(u_1, ..., u_4)$ is zero if the largest index does not appear as a pair. These restrictions can be conveniently summarized by defining the following function

$$\sigma(s,r) = \begin{cases} E\left(\varepsilon_t^2 \varepsilon_{t-|s|} \varepsilon_{t-|r|}\right) & \text{for } s \neq r, \ sgn(r) = sgn(s) \\ E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) - \sigma^4 & \text{for } s = r, \ sgn(r) = sgn(s) \\ 0 & sgn(r) = -sgn(s) \end{cases} \quad \text{for } r, s \in \{0, \pm 1, \pm 2, ...\}. \tag{2.2}$$

Let $\alpha_{s,r} = \sigma(s,r)$ if $s \neq r$ and $\alpha_{r,r} = \sigma(r,r) + \sigma^4$. A detailed treatment of the form of the fourth order cumulant spectrum under the martingale difference sequence assumption is contained in the proof of Corollary (3.4) in Appendix B. The sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is assumed to satisfy the following assumptions.

Assumption A1. *(i) ε_t is strictly stationary and ergodic.*

(ii) $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ almost surely.

(iii) $E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) = \sigma_t^2$ almost surely.

(iv) $E\left(\varepsilon_t^2\right) = \sigma^2 < \infty$.

(v) $\sum_{k=1}^{\infty} \sum_{s=1}^{\infty} |k| \, |\sigma(s,r)| = B < \infty$ for $k = s$ and $k = r$.

(vi) $E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) > \underline{\alpha}$ some $\underline{\alpha} > 0$ for all s.

Remark 1. *Strict stationarity and ergodicity are assumed for convenience. The theory could be developed without these assumptions. The critical assumptions are the martin-*

gale difference sequence assumption A1(ii) and Assumption A1(iii) which states that the second moments are conditionally heterogeneous. A consequence is that terms of the form $E\left(\varepsilon_t^2\varepsilon_{t-s}\varepsilon_{t-r}\right)$ *are nonzero for* $s \neq r \neq 0$ *and depend on* s *for* $s = r \neq 0$. *Assumption (v) defines the fourth order cumulants of* ε_t *which reduce to the function* $\sigma(s, r)$ *due to the martingale difference assumption (i). Assumption (vi) is not restrictive. Its only purpose is to guarantee that the innovation distribution does not have all its mass concentrated at zero.*

By definition of the conditional expectation operator, σ_t is $\mathcal{F}_{t-1}$ measurable. Assumption (A1) implies that ε_t^2 is strictly stationary and ergodic and therefore covariance stationary. It should be emphasized that no assumptions about third moments are made. In particular this allows for skewness in the error process.

The econometrician does not observe the innovation sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ directly but has a finite stretch of data $\{y_t\}_{t=1}^{n}$ which is generated by the following mechanism

$$y_t = \sum_{j=0}^{\infty} c(\beta, j)\,\varepsilon_{t-j}. \tag{2.3}$$

with $\sum_{j=0}^{\infty} |c(\beta, j)|\,|j|^{1/2} < \infty$ for a given $\beta = \beta_0 \in \mathbb{R}^d$. We define the lag polynomial $C(\beta, z) = \sum_{j=0}^{\infty} c(\beta, j) z^j$ and impose an identifying condition $c(\beta, 0) = 1$.

For the special case of an $ARMA(p, q)$ process, the lag polynomial has the familiar rational form

$$C(\beta, z) = \frac{\theta(z)}{\phi(z)} \tag{2.4}$$

with $\theta(z) = 1-\theta_1 z-\ldots-\theta_q z^q$ and $\phi(z) = 1-\phi_1 z-\ldots-\phi_p z^p$ and $\beta' = (\phi_1, \ldots, \phi_p, \theta_1, \ldots \theta_q)$. Let $g_{yy}(\beta, \lambda) = |C(\beta, e^{i\lambda})|^2$. If the errors ε_t are weakly stationary and uncorrelated, as for example under Assumption (A1), then the spectrum of y_t is given by $f_{yy}(\beta, \lambda) = \frac{\sigma^2}{2\pi} g_{yy}(\beta, \lambda)$. It is an immediate consequence of the normalization assumption that

$$\int_{-\pi}^{\pi} \frac{\partial \ln f_{yy}(\beta, \lambda)}{\partial \beta} d\lambda = 0.$$

While it was shown by Hosoya [53] that such an assumption is not needed for consistent estimation of the model it will be maintained for convenience. The remaining assumptions of the problem are related to specifying the parameter space in a way to ensure stationarity of y_t. The assumptions basically require continuity of the spectral density f_{yy} to prove consistency and twice continuous differentiability for asymptotic normality. The necessary assumptions are discussed in Hannan [42], Dunsmuir and Hannan [25], and Deistler, Dunsmuir and Hannan[24]. As shown in these articles, a careful distinction between convergence of the parameters in $c(\beta, j)$ and the structural form parameters is needed. Consistency proofs typically establish convergence in the pointwise topology. An identification condition is then needed to obtain convergence in the quotient topology (see below). The following assumptions correspond to the assumptions in Hannan [42].

Assumption B1. *Let $C(\beta, z) = \sum_{j=0}^{\infty} c(\beta, j) z^j$ such that $c(\beta, j)$ is continuous in β for all j and $c(\beta, j) = 1$ and $\sum_{j=0}^{\infty} |c(\beta, j)|\, |j|^{1/2} < \infty$. The parameter space Θ is a subset of $\mathbb{R}^d$ defined by*

$$\Theta = \left\{ \beta \in \mathbb{R}^d \,|g_{yy}^{-1}(\beta, \lambda) \neq 0 \text{ for } |z| \leq 1 \right\}$$

with compact closure $\overline{\Theta}$.

Assumption B2. *Define* $\Theta_0 = \Theta \cap \{\beta \in \mathbb{R}^d \,|g_{yy}(\beta,\lambda) > 0 \text{ for } \lambda \in [-\pi,\pi]\}$ *and assume that* $\beta_0 \in \Theta_0$.

Assumption B3. $\forall\delta > 0$ $\{g_{yy}(\beta,\lambda)+\delta\}^{-1}$ *is continuous in* $(\lambda,\beta) \in [-\pi,\pi] \times \overline{\Theta}$.

Assumption B4. *For all* $\beta \in \overline{\Theta}$, $g_{yy}(\beta_0,\lambda) \neq g_{yy}(\beta,\lambda)$ *whenever* $\beta \neq \beta_0$.

Assumption B5. *For a neighbourhood* U *of* β_0, $U \subset \Theta_0$, $\partial^2 g_{yy}(\beta,\lambda)/\partial\beta\partial\beta'$ *is continuous in* $\lambda \in [-\pi,\pi]$ *and* $\beta \in U$.

Remark 2. *Assumption (B5) is needed for the asymptotic normality result. This is also true for the restriction* $\sum_{j=0}^{\infty} |c(\beta,j)|\,|j|^{1/2} < \infty$ *which can be relaxed to* $\sum_{j=0}^{\infty} |c(\beta,j)| < \infty$ *for the consistency proof.*

Remark 3. *Assumption (B4) guarantees that the limit of the criterion function of a certain Gaussian estimator is uniquely minimized at* β_0. *The assumption however only guarantees that this is true for a certain* β_0. *To ensure identification of the class of models indexed by* $\beta \in \Theta$, *stronger restrictions need to be imposed on the function* $C(\beta, e^{i\lambda})$. *These restrictions are discussed in general terms in the next remark and are later specialized to the case where* $C(\beta, e^{i\lambda})$ *corresponds to the class of ARMA models.*

Remark 4. *Additional restrictions on the class of functions* $C(\beta, e^{i\lambda})$ *are needed to guarantee identification of each member in the class. Let*

$$C_0 = C[-\pi,\pi] \cap \{g_{yy}(\beta,\lambda) \text{ satisfies B1-B4}\}.$$

Assumptions (B1-B4) guarantee that $g_{yy}(\beta,\lambda) \in C[-\pi,\pi]$ *for all* $\beta \in \Theta$*. A topology on* C_0 *can be defined in terms of pointwise convergence of the coefficients of* $c \in C_0$*. To be more precise and following Ash [5, pp. 376-378] let* $\{C_j\}_{j\in I}$ *be a collection of spaces of analytic functions of the form* $z(\beta,j) = c(\beta,j)e^{i\lambda j}$*. The product topology on* $C_0 = \prod_{j\in I} C_j$ *has as a base all sets of the from* $\{C(\beta,e^{i\lambda}) \in C_0 : c(\beta,j)e^{i\lambda j} \in C_j\}$ *and is also called the topology of pointwise convergence. It is the weakest topology making the projections* p_i *of* C_0 *onto* C_i *continuous. Then by Ash [5, theorem A3.2] for a net* $c^{(n)} \in C_0$ *and* $c \in C_0$*, then* $c^{(n)} \to c$ *iff* $c_j^{(n)} \to c_j$*. Also a map* f *from a topological space* Θ *onto* C_0 *is continuous iff* $p_i \circ f$ *is continuous. In our case* $p_i \circ f = c(\beta,j)$ *which is continuous in* β *by assumption (B1). If the inverse image of an open set* $U \subset C_0$*,* $f^{-1}(U)$ *is open in* Θ *then* f *is called an identification map of* Θ *onto* C_0*. If* f *also is injective then it is a homeomorphism and* Θ *and* C_0 *are homeomorphic.. Now let* R *be an equivalence relation on* Θ*. Let* Θ/R *be the set of equivalence classes such that the quotient space of* Θ *on* R *is* Θ/R*. The quotient topology on* Θ/R *is the strongest topology such that the canonical projection* $p : \Theta \to \Theta/R$ *is an identification. Let* $f : \Theta \to C_0$ *be an identification map and define the equivalence relation* R *on* Θ *by calling* θ_1 *and* θ_2 *equivalent iff* $f(\theta_1) = f(\theta_2)$*. Then* Θ/R *is homeomorphic to* C_0 *(see Ash [5, theorem A3.6]).*

Remark (4) shows that additional structure for the function $C(\beta, e^{i\lambda})$ is needed to construct an identification map between reduced form and structural parameters. An important class of functions where an identification map exists is the $ARMA$ class. The following assumptions are univariate versions of the assumptions in Dunsmuir and Hannan [25, p. 345] and Deistler, Dunsmuir and Hannan [24, p. 364]. To avoid confusion it should be emphasized that the notation is reversed relative to these articles, i.e., Θ stands for a subset of $\mathbb{R}^{p+q}$ containing the structural form ($ARMA$) parameters.

Assumption C1. *The parameter space Θ is defined by*

$$\Theta = \{\beta \in \mathbb{R}^{p+q} \,|\phi(\beta, z) \neq 0 \text{ for } |z| \leq 1, \theta(\beta, z) \neq 0 \text{ for } |z| < 1$$

$$\theta(\beta, z) \text{ and } \phi(\beta, z) \text{ have no common zeros}\}$$

Assumption C2. *$\overline{\Theta}$ is the closure of Θ. Let*

$$\Theta_1^* = \Theta \cap \{\beta \in \mathbb{R}^{p+q} \,|\beta_p \neq 0, \beta_{p+q} \neq 0\}$$

and $\Theta_1 = \Theta_1^* \cap \{\beta \in \mathbb{R}^{p+q} \,|\theta(\beta, z) \neq 0 \text{ for } |z| \leq 1\}$. The true parameter β_0 lies in Θ_1.

Remark 5. *Deistler, Dunsmuir and Hannan [24, theorem 3, lemma 1] show that Θ_1 is open in $\mathbb{R}^{p+q}$ and dense in Θ. They also state that Θ is homeomorphic to the quotient topology since all equivalence classes on Θ are singletons. It follows that the same property holds for Θ_1 since $\Theta_1 \subset \Theta$ and the canonical projection is the identity also on Θ_1. To establish that the quotient topology is homeomorphic to the product topology it is therefore enough to show that there is an injective identification map $f : \Theta_1 \to C_0$. For the $ARMA$ case the function f is $f(\beta, z) = \phi^{-1}(\beta, z)\,\theta(\beta, z)$ which is clearly continuous in $\beta \in \Theta$. Deistler, Dunsmuir and Hannan [24, p. 365] show that f has a continuous inverse such that $\Theta_1$and C_0 are homeomorphic.*

Properties of optimization estimators depend on the topological structure of the parameter space. While Θ_1 is not compact it can be shown to be locally compact. A natural way to proceed is to equip Θ_1 with the relative Euclidean topology. It can now be established that if $\beta_1, \beta_2 \in \Theta_1$ such that $\beta_1 \neq \beta_2$ then there exist open neighbourhoods N_1 and N_2 of β_1 and β_2 in the Euclidean topology (i.e. open spheres with constant radius) such that $N_1, N_2 \neq \emptyset$ and $N_1 \cap N_2 = \emptyset$ by the properties of $\mathbb{R}^d$. Continuity of $g_{yy}(\beta, \lambda)$ insures that N_1 and N_2 can be chosen such that $N_1, N_2 \subset \Theta_1$. This establishes that Θ_1 is a Hausdorff space. Since $\mathbb{R}^{p+q}$ with the Euclidean topology is locally compact it follows that Θ_1 is a locally compact Hausdorff space.

3. Properties of Gaussian Estimators

So far the focus of the discussion has been to specify the model to be estimated. Next a definition of the Gaussian estimator is given. We introduce discrete Fourier transforms of the data defined as $\omega_y(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t e^{-it\lambda}$. The periodogram is $I_{n,yy}(\lambda) = \omega_{n,y}(\lambda)\omega_{n,y}(-\lambda)$. Also let, $g_{yy}(\beta, \lambda) = 2\pi/\sigma^2 f_{yy}(\beta, \lambda)$. Then, as in Hannan [42], we consider the estimator $\hat{\beta}_n$ which minimizes

$$Q_n(\beta) = \int_{-\pi}^{\pi} \frac{I_{n,yy}(\lambda)}{g_{yy}(\beta, \lambda)} d\lambda.$$

It will turn out that the estimator is still consistent under Assumption (A1) but no longer efficient. Consistency still holds since it depends only on the uniform convergence of $Q_n(\beta)$ and the properties of its limit. We summarize these results in the following Lemma.

Lemma 3.1. *Under Assumption (A1) and either Assumptions (B1-B4) for the case of an unrestricted linear process or (C1-C2) for the case of an ARMA specification,* $Q_n(\beta) \overset{a.s.}{\rightarrow}$

$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{g_{yy}(\beta_0,\lambda)}{g_{yy}(\beta,\lambda)}d\lambda$ *for all* β *in* Θ *and for every* $\delta > 0$ $Q_n(\beta) \overset{a.s.}{\to} \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{g_{yy}(\beta_0,\lambda)}{g_{yy}(\beta,\lambda)+\delta}d\lambda$ *uniformly in* $\beta \in \overline{\Theta}$. *Also,* $\tilde{\beta}_n \overset{a.s.}{\to} \beta_0$.

Proof. The proof is the same as in Hannan [42], since y_t is strictly stationary and ergodic. ■

While the estimator is still consistent under the more general conditions it has a different limiting distribution. This has important consequences for statistical inference carried out on the basis of asymptotic approximations to the distribution of the estimators.

The following result shows how the limit distribution of the parameter estimates depends on the properties of the error process in the case of higher order dependence. In this sense, the result applies to a very general class of time series models under nonstandard assumptions. The form of the asymptotic covariance matrix has the typical form of the one for a maximum likelihood estimator under misspecification of the distribution. It reflects the fact that the expectation of the squared score function is not equal to the expectation of the Hessian matrix. At the same time it can also be represented in a form similar to the asymptotic covariance matrix of an inefficient GMM estimator. This last property will be exploited to derive a lower bound for the covariance matrix.

Introduce the notation $\dot{\eta}(\beta,\lambda) = \partial \ln g_{yy}(\beta,\lambda)/\partial\beta$ and $b_k = (2\pi)^{-1}\int \dot{\eta}(\beta,\lambda)e^{-ik\lambda}d\lambda$. For the $ARMA(p,q)$ model the derivatives of the log spectral density are given by

$$\begin{aligned}\frac{\partial \ln g_{yy}(\beta,\lambda)}{\partial\phi_l} &= -\frac{e^{-i\lambda l}}{\phi(e^{-i\lambda})} - \frac{e^{i\lambda l}}{\phi(e^{i\lambda})} \\ &= -\sum_{j=0}^{\infty}\psi_{\phi,j}\left(e^{-i\lambda(l+j)} + e^{i\lambda(l+j)}\right)\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \ln g_{yy}(\beta,\lambda)}{\partial \theta_l} &= -\frac{e^{-i\lambda l}}{\theta(e^{-i\lambda})} - \frac{e^{i\lambda l}}{\theta(e^{i\lambda})} \\
&= -\sum_{j=0}^{\infty} \psi_{\theta,j}\left(e^{-i\lambda(l+j)} + e^{i\lambda(l+j)}\right)
\end{aligned}$$

where $\phi^{-1}(z) = \sum_{j=0}^{\infty} \psi_{\phi,j} z^j$ and $\theta^{-1}(z) = \sum_{j=0}^{\infty} \psi_{\theta,j} z^j$. It follows immediately that $b_k = b_{-k}$ and $b_0 = 0$. The same can be established for the more general model with $g_{yy}(\beta,\lambda) = |C(\beta,\lambda)|^2$.

We now consider the asymptotic distribution of $n^{1/2}Q_n(\beta_0)$. This is done in the next proposition.

Proposition 3.2. *Under Assumption (A1) and either Assumptions (B1-B5) or Assumptions (C1-C2),* $n^{1/2}Q_n(\beta_0) \xrightarrow{d} N(0, B)$, *where*

$$B = \frac{1}{4\pi}\int_{-\pi}^{\pi} \dot{\eta}(\beta_0,\lambda)\dot{\eta}(\beta_0,\lambda)' d\lambda$$
$$+2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,\mu)\,\dot{\eta}(\beta_0,\lambda)\dot{\eta}(\beta_0,\mu)' d\mu d\lambda. \tag{3.1}$$

Here, $f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,\mu)$ *is defined as* $f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,\mu) = (2\pi)^{-2}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \sigma(k,l)\, e^{-i(\lambda k+\mu l)}$ *where* $\sigma(k,l)$ *is defined in (2.2).*

Proof. The proof follows immediately from Lemma (A.4) since $\sum \|b_k\| \, |k|^{1/2} < \infty$. ■

Next we turn to the proof of asymptotic normality of the estimator $\hat{\beta}_n$. Let $\|.\|$ denote the Euclidean norm. Then by the consistency result from Lemma 3.1 there exist neighbourhoods $N_\delta = \{\beta \in \Theta : \|\beta - \beta_0\| < \delta\}$ for $\delta > 0$ such that $P(\hat{\beta}_n \in N_\delta) \to 1$.

It is therefore possible to derive the asymptotic distribution of $\hat{\beta}_n$ by a component by component mean value expansion of the first order conditions around β_0. In particular, we consider as in Hannan [42]

$$n^{1/2}\frac{\partial Q_n(\hat{\beta}_n)}{\partial \beta} = 0 = n^{1/2}\frac{\partial Q_n(\beta_0)}{\partial \beta} + \frac{\partial^2 Q_n(\tilde{\beta}_n)}{\partial \beta \partial \beta'} n^{1/2}\left(\hat{\beta}_n - \beta_0\right)$$

where $\tilde{\beta}_n$ such that $\left\|\tilde{\beta}_n - \beta_0\right\| < \left\|\hat{\beta}_n - \beta_0\right\|$, which implies that $\tilde{\beta}_n \xrightarrow{p} \beta_0$ by the consistency result. Using Proposition (3.2), it remains to show that

$$\frac{\partial^2 Q_n(\tilde{\beta}_n)}{\partial \beta \partial \beta'} \xrightarrow{p} \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \dot{\eta}(\beta_0, \lambda)\dot{\eta}(\beta_0, \lambda)' d\lambda.$$

This follows directly from the consistency proof since

$$\begin{aligned} \frac{\partial^2 Q_n(\beta)}{\partial \beta \partial \beta'} &= \int I_{n,yy}(\lambda)\frac{\partial^2}{\partial \beta \partial \beta'} g_{yy}^{-1}(\beta, \lambda)d\lambda \\ &\rightarrow \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} g_{yy}(\beta_0, \lambda)\frac{\partial^2}{\partial \beta \partial \beta'} g_{yy}^{-1}(\beta, \lambda)d\lambda \end{aligned}$$

where the convergence is *a.s.* uniformly on a compact subset U of Θ. Choose U such that $\beta_0 \in U$. The desired result then follows by continuity of $\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} g_{yy}(\beta_0, \lambda)\partial^2 g_{yy}^{-1}(\beta, \lambda)/\partial \beta \partial \beta'$ and the fact that $\tilde{\beta}_n \xrightarrow{p} \beta_0$. Using these results we can now state the main result of this section.

Theorem 3.3. *Let ε_t satisfy Assumptions (A1) and let $y_t = C(\beta_0, L)\varepsilon_t$, where $C(\beta, L)$ either satisfies Assumptions (B1-B5) with $\beta \in \mathbb{R}^d$ or $C(\beta, z) = \theta(z)/\phi(z)$ with $\beta \in \mathbb{R}^{p+q}$*

and Assumption (C1-C2) holds. Define $\hat{\beta}_n = \arg\min_{\beta\in\Theta} Q_n(\beta)$ *and*

$$A = \frac{1}{4\pi}\int_{-\pi}^{\pi} \dot{\eta}(\beta_0,\lambda)\dot{\eta}(\beta_0,\lambda)'d\lambda. \tag{3.2}$$

Then, $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1})$.

The next corollary shows how the result in Theorem (3.3) relates to more general time series estimators as discussed in Hosoya and Taniguchi [54] and Keenan [59]. For this purpose, define the fourth order cumulant spectrum as

$$f_{\varepsilon..\varepsilon}(\lambda_1,\lambda_2,\lambda_3) = \frac{1}{(2\pi)^3}\sum_{u_1,u_2,u_3=-\infty}^{\infty} c_{\varepsilon..\varepsilon}(u_1,u_2,u_3)e^{-i\left\{\sum_{j=1}^{3} u_j\lambda_j\right\}}. \tag{3.3}$$

Following Brillinger [13], the k-th order cumulant is defined as

$$c_{\varepsilon..\varepsilon}(u_1,u_2,\ldots,u_k) = \sum(-1)^{p-1}(p-1)!m_{\varepsilon..\varepsilon}(u_{\nu_1})\cdots m_{\varepsilon..\varepsilon}(u_{\upsilon_k}), \tag{3.4}$$

where the summation is over all partitions of the numbers $1\ldots k$. For example, if $\upsilon_i = (1,2,5)$, then $t_{\nu_i} = (u_1,u_2,u_5)$ and $m_{\varepsilon\ldots\varepsilon}(u_{\nu_i}) = E(\varepsilon_{u_1},\varepsilon_{u_2},\varepsilon_{u_5})$. Stationarity implies that only the relative time difference matters so that one time index can always be normalized to zero.

Keenan[59, Corollary 3.5] derives the asymptotic distribution for spectral estimators without the martingale difference assumption for the errors but with additional summability assumptions for all higher order cumulants. Specializing his results to the linear process case leads to the asymptotic covariance matrix $\Omega = A^{-1}\tilde{B}A^{-1}$ for $\hat{\beta}$, where A is as defined in (3.2) and

$$\tilde{B} = 4\pi \int_{-\pi}^{\pi} \dot{\eta}(\beta_0, \lambda)\dot{\eta}(\beta_0, \lambda)' f_{\varepsilon\varepsilon}^2(\lambda)\, d\lambda + 2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \dot{\eta}(\beta_0, \lambda)\dot{\eta}(\beta_0, \mu)' f_{\varepsilon..\varepsilon}(\lambda, \mu, -\lambda)\, d\lambda d\mu. \tag{3.5}$$

The following corollary shows that the variance of the estimator in Theorem (3.3) is a special case of (3.5). The intuition behind this result is clear since under the martingale assumption, $f_{\varepsilon\varepsilon}^2(\lambda) = \sigma^4/(2\pi)^2$ and as will be shown in the proof of the corollary, $2\pi f_{\varepsilon..\varepsilon}(\lambda, \mu, -\lambda) = 1/2[f_{\varepsilon^2\varepsilon\varepsilon}(\lambda, -\lambda) + f_{\varepsilon^2\varepsilon\varepsilon}(\mu, -\mu)] + f_{\varepsilon^2\varepsilon\varepsilon}(\lambda, -\mu) + f_{\varepsilon^2\varepsilon\varepsilon}(\lambda, \mu)$. The fact that for the asymptotic variance only the $0 \bmod 2\pi$ submanifold $[\lambda, -\lambda, \mu, -\mu]$ of $[-\pi, \pi]^4$ matters is responsible for the appearance of the term $f_{\varepsilon^2\varepsilon\varepsilon}(\lambda, -\lambda)$.

Corollary 3.4. *Let B be defined as $\lim_{n\to\infty} var\left(\sqrt{n}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, \dot{\eta}(\beta_0, \lambda) d\lambda\right)$ under Assumption (A1) and $\dot{\eta}(\beta, \lambda)$ is a continuous even function on $[-\pi, \pi]$ with Fourier coefficients b_k such that $\sum_{k=1}^{\infty} |b_k|\, |k|^{1/2} < \infty$. Then, B can be written as*

$$B = 2\sigma^2 A + 2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon..\varepsilon}(\mu, \lambda, -\mu)\, \dot{\eta}(\beta_0, \lambda)\dot{\eta}(\beta_0, \mu)' d\lambda d\mu,$$

where $f_{\varepsilon..\varepsilon}(\mu, \lambda, -\mu)$ is the fourth order cumulant spectrum defined in (3.3).

Proof. See Appendix B ■

The corollary shows that the martingale assumption is the critical element in the reduction from the general case involving a fourth order cumulant spectrum to the case where the variance of the score process only depends on the spectral density of the squared errors.

It is the purpose of the next section to obtain a lower bound of this covariance matrix for a certain class of transformations. It is then shown that these transformations are related to the class of instrumental variables estimators with instruments that are linear filters of the innovation sequence.

4. Covariance Matrix Lowerbound

While the most compact representation of the covariance matrix $A^{-1}BA^{-1}$ is in terms of an integral of spectral density functions, more insight into the structure of the problem can be gained from a time domain representation of the matrix. Recall the definition of $b_k = (2\pi)^{-1} \int_{-\pi}^{\pi} \dot{\eta}(\beta, \lambda) e^{ik\lambda} d\lambda$, the vector of the k-th Fourier transform of the first derivative of the log spectral density. By Parseval's equality, we have $\sum_{k=1}^{\infty} b_k b_k' = \frac{1}{4\pi} \int_{-\pi}^{\pi} \dot{\eta}(\beta, \lambda) \dot{\eta}(\beta, \lambda)' d\lambda$. This also implies that $\sum_{k=1}^{\infty} b_k b_k' = A/(2\sigma^2)$. Further, define the matrix

$$P_m' = [b_1, ..., b_m].$$

Then, $\lim_{m \to \infty} P_m' P_m = \frac{1}{4\pi} \int_{-\pi}^{\pi} \dot{\eta}(\phi, \lambda) \dot{\eta}(\phi, \lambda)' d\lambda$. Next, introduce

$$\Omega_m = \begin{bmatrix} \sigma(1,1) + \sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m) + \sigma^4 \end{bmatrix}. \tag{4.1}$$

It follows that $B = 4 \lim_{m \to \infty} P_m' \Omega_m P_m$. Thus,

$$A^{-1}BA^{-1} = \sigma^4 \lim_{m \to \infty} (P_m' P_m)^{-1} (P_m' \Omega_m P_m) (P_m' P_m)^{-1}.$$

Before obtaining a matrix lower bound for this covariance matrix we need to investigate the properties of the fourth order cumulant matrix Ω_m. In particular, we need to establish that this matrix is invertible for all m. This is done in the next Lemma.

Lemma 4.1. *Let Ω_m be defined as in (4.1). Then, Ω_m^{-1} exists for all m.*

Proof. First, note that Ω_m is symmetric since $\sigma(k,l) = E(\varepsilon_t^2\varepsilon_{t-k}\varepsilon_{t-l}) = \sigma(l,k)$. Then. by the Shur decomposition (see Magnus and Neudecker [65, p. 16]) for all m there exists an orthogonal matrix S_m such that $S_m\Omega_m S_m = \Lambda_m$, where Λ_m is diagonal with elements λ_j^m, $j = 1, ..., m$. Now, for any $x_m \in \mathbb{R}^m$, $x_m \neq 0$, we have $x_m'\Omega_m x_m = E(\sum x_{i,m}\varepsilon_t\varepsilon_{t-i})^2 > 0$ where the inequality is strict by Assumption (A1). So Ω_m is positive definite such that $\lambda_j^m > 0\ \forall j, m$. This shows that Ω_m has full rank. ■

Invertibility of Ω_m for all m however is not enough to show that Ω_m is invertible in the limit. We briefly review the theory of invertible operators (see Gohberg and Goldberg [37, p. 65]). For two Hilbert spaces H_1 and H_2 denote the set of bounded linear operators mapping H_1 into H_2 by $L(H_1, H_2)$. Then $A \in L(H_1, H_2)$ is invertible if there exists an operator $A^{-1} \in L(H_2, H_1)$ such that $A^{-1}Ax = x$ for all $x \in H_1$ and $AA^{-1}y = y$ for all $y \in H_2$. Let $\ker A = \{x \in H_1 : Ax = 0\}$ and $\operatorname{Im} A = \{Ax : x \in H_1\}$. Then A is invertible if $\ker A = \{0\}$ and $\operatorname{Im} A = H_2$.

Following Hanani, Netanyahu and Reichaw [40] we now choose H_1, H_2 as linear spaces whose points are sequences of real numbers denoted by $x = \{x_1, x_2, ...\}$ and $y = \{y_1, y_2, ...\}$. Define the norms $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ and $\|x\|_\infty = \sup_i |x_i|$. Then H is the space of all bounded sequences, denoted by l^∞, if $\|x\|_\infty < \infty$ for all $x \in H$. Similarly let l^p be the space of all sequences that are bounded under the $\|.\|_p$ norm. An operator $A : l^p \mapsto l^p$,

$p \geq 1$, is now defined by the infinite dimensional matrix $A = (a_{i,j}), i, j = 1, 2, ...$ such that $y = Ax \in l^p$ for all $x \in l^p$. This can be written element by element as $y_i = \sum_j^\infty a_{i,j} x_j$ for all i. The operator A is invertible if the only solution to $Ax = 0$ is $x = \{0, 0, ...\}$ and $\operatorname{Im} A = l^p$.

Hanani, Netanyahu and Reichaw [40, Example 1] show that A may have an inverse on l^p for $p \geq 1$ but not on l^∞. Their example also shows that existence of an inverse in the subsystem of linear equations defined by $A^m = (a_{i,j}), i, j = 1, 2, ..., m$ does not imply the existence of an inverse for the infinite system. This follows from the fact that for finite sequences l^p and l^∞ are identical.

Invertibility of infinite dimensional matrices is analyzed in Hanani, Netanyahu and Reichaw [40], Gohberg and Goldberg [37, p. 65] and Farid [29], [30]. The conditions given by these authors do not readily apply to the matrix $\Omega = \lim_m \Omega_m$. We use arguments similar to the ones in the proof of Lemma (4.1) to establish invertibility.

Lemma 4.2. *Let Ω_m be defined as in (4.1) and let $\Omega = \lim_m \Omega_m$. Then $\Omega \in L(l^\infty, l^\infty)$ and Ω^{-1} exists.*

Proof. From Assumption (A1) it is clear that $\Omega x \in l^\infty$ for all $x \in l^\infty$. It remains to show that $\ker \Omega = 0$. Assume there is $x \in l^\infty$ such that $x \neq \{0, 0, ...\}$ and $\Omega x = 0$. Then also $x' \Omega x = 0$ which can be written as $E(\sum_{i=1}^\infty x_i \varepsilon_t \varepsilon_{t-i})^2 = 0$. But this is only possible if $\sum x_i \varepsilon_t \varepsilon_{t-i} = 0$ with probability one. Now $\sum x_i \varepsilon_t \varepsilon_{t-i} = 0$ *a.s.* if $\varepsilon_t \varepsilon_{t-i} = 0$ *a.s.* or the functions ε_{t-i} are linearly dependent *a.s.*

If ε_{t-i} are linearly dependent then $\exists \alpha \in l^\infty$, $\alpha \neq 0$ such that $\sum \alpha_i \varepsilon_{t-i} = 0$ *a.s.* Without loss of generality $\alpha_1 \neq 0$. If $\alpha_i = 0$ for all $i = 2, 3, ...$ then $\sum \alpha_i \varepsilon_{t-i} = 0$ *a.s.*

is trivially contradicted. Now assume $\alpha_i \neq 0$ for at least one $i = 2, 3, ...$ such that $\varepsilon_{t-1} = -\alpha_1^{-1} \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i}$ *a.s.* But then $E(\varepsilon_{t-1}|\mathcal{F}_{t-2}) = -\alpha_1^{-1} \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i}$ *a.s.* so that $E(\varepsilon_{t-1}|\mathcal{F}_{t-2}) \neq 0$ with positive probability. This contradicts the martingale difference assumption.

On the other hand if $\varepsilon_t \varepsilon_{t-i} = 0$ *a.s.* for all i then $\varepsilon_t^2 \varepsilon_{t-i}^2 = 0$ *a.s.* But then $E(\varepsilon_t^2 \varepsilon_{t-i}^2) = 0$ for all i which contradicts Assumption (A1). Therefore $\Omega x = 0$ can only hold if $x = 0$. Thus Ω is a positive definite bounded linear operator and therefore has an inverse (see Schmeidler [85, p. 62]) ■

Form Hanani, Netanyahu and Reichaw [40, Theorem X.4.2] it also follows that Ω^{-1} is bounded, i.e., $\|\Omega^{-1}\| = \sup_{\|x\| \leq 1} \|\Omega^{-1} x\| < \infty$. Thus $\sup_{i,j} |\omega_{i,j}| < \infty$ where $[\Omega^{-1}]_{i,j} = \omega_{i,j}$.

Next, we need to establish properties of the matrix Ω_m^{-1} as m tends to infinity. In particular we want to establish that the inverse Ω_m^{-1} approximates Ω^{-1} as $m \to \infty$. We define $\omega_{k,l}^m = [\Omega_m^{-1}]_{k,l}$ and analyze the properties of $\omega_{k,l}^m$ in the next Lemma.

Lemma 4.3. *Let Ω_m be as defined in (4.1). Define Ω_m^{-1} such that $\Omega_m^{-1} \Omega_m = I_m$ and $\Omega_m \Omega_m^{-1} = I_m$ $\forall m$. Let $\omega_{i,j}^m = [\Omega_m^{-1}]_{i,j}$ and $[\Omega^{-1}]_{i,j} = \omega_{i,j}$. Then $\omega_{i,j}^m \to \omega_{i,j}$ for all i, j as $m \to \infty$ and $\forall i \neq j$ $\omega_{i,j} \to 0$ as $j \to \infty$, $\forall j \neq i$ $\omega_{i,j} \to 0$ as $i \to \infty$, $\omega_{i,i} \to \sigma^{-4}$ as $i \to \infty$.*

Proof. By Assumption (A1) we know that $\sum \sum |\sigma(k,l)| < B$ thus $\sum_k |\sigma(k,l)| < B$ for any l. Therefore for any fixed l, $\sigma(k,l) \to 0$. This holds also if the roles of k and l are reversed. Also $\sum_k |\sigma(k,k)| < B$ such that $\sigma(k,k) \to 0$. Define the infinite dimensional matrices S_{12}^m, S_{21}^m and S_{22}^m according to the following partition

$$\Omega = \begin{bmatrix} \Omega_m & S_{12}^m \\ S_{21}^m & S_{22}^m \end{bmatrix}.$$

Then $tr(S_{12}^m S_{12}^{m'}) = \sum_{l=m+1}^{\infty} \sum_{k=1}^{m} |\sigma(k,l)|^2 \to 0$, $tr(S_{21}^m S_{21}^{m'}) \to 0$ and $tr(S_{22}^m - \sigma^4 I)(S_{22}^m - \sigma^4 I)' \to 0$ as $m \to \infty$. Then define the infinite dimensional approximation matrix

$$\Omega_m^* = \begin{bmatrix} \Omega_m & 0 \\ 0 & \sigma^4 I \end{bmatrix}.$$

Clearly Ω_m^{*-1} exists $\forall m$ by Lemma (4.1) and the partitioned inverse formula. We now have

$$(\Omega^{-1} - \Omega_m^{*-1}) = \Omega_m^{*-1}(\Omega - \Omega_m^*)\Omega^{-1}$$

such that

$$\|\Omega^{-1} - \Omega_m^{*-1}\| \leq \|\Omega_m^{*-1}\| \, \|\Omega - \Omega_m^*\| \, \|\Omega^{-1}\|.$$

We have shown in Lemma (4.1) that the smallest eigenvalue λ_1^m of Ω_m is nonzero. Then by a familiar inequality for all $x \in \mathbb{R}^m$ $x'\Omega_m^{-1}x/x'x \leq 1/\lambda_1^m < \infty$ $\forall m$. Then $\|\Omega_m^{*-1}\| = \sup_x x'\Omega_m^{-1}x/x'x + \sigma^{-4} < \infty$, $\|\Omega^{-1}\| < \infty$ and

$$\begin{aligned} \|\Omega - \Omega_m^*\| &= \sup_{\|x\| \leq 1} 2 \sum_{l=m+1}^{\infty} \sum_{k=1}^{m} |\sigma(k,l)| \, |x_k| \, |x_l| + \sum_{k=m+1}^{\infty} \sum_{l=m+1}^{\infty} |\sigma(k,l)| \, |x_k| \, |x_l| \\ &\leq 2 \sum_{l=m+1}^{\infty} \sum_{k=1}^{m} |\sigma(k,l)| + \sum_{k=m+1}^{\infty} \sum_{l=m+1}^{\infty} |\sigma(k,l)| \to 0 \text{ as } m \to \infty. \end{aligned}$$

Thus $\|\Omega^{-1} - \Omega_m^{*-1}\| \to 0$ as $m \to \infty$ ■

Using this notation we can state our next theorem, which establishes a lower bound for the covariance matrix.

Theorem 4.4. *Let the asymptotic covariance matrix of $\hat{\beta}$ be $A^{-1}BA^{-1}$, where A is defined in (3.2) and B in (3.1). Then, $\lim_{m\to\infty}\left(P_m'\Omega_m^{-1}P_m\right)^{-1}$ exists and*

$$A^{-1}BA^{-1} - \frac{1}{\sigma^4}\lim_{m\to\infty}\left(P_m'\Omega_m^{-1}P_m\right)^{-1} \geq 0,$$

where ≥ 0 stands for positive semi-definite.

Proof. See Appendix B ■

Define $\omega_{i,j}$ as the limits of the elements in the matrix Ω_m^{-1}. The optimal covariance matrix $\Xi = \lim_{m\to\infty}(P_m'\Omega_m^{-1}P_m)^{-1}$ can be evaluated to be

$$\Xi = \left[\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\omega_{k,l}\left(\int_{-\pi}^{\pi}\dot{\eta}(\phi,\lambda)\,e^{ik\lambda}d\lambda\right)\left(\int_{-\pi}^{\pi}\dot{\eta}(\phi,\lambda)'\,e^{il\lambda}d\lambda\right)\right]^{-1}. \tag{4.2}$$

In the next section we will show that the formal analogy between (4.2) and the lower bound for the standard GLS estimator can be exploited to construct an efficient instrumental variables estimator. The standard GLS estimator corrects for heteroskedasticity in the observations y_t. Here, however, y_t is strictly stationary and therefore has a homogeneous marginal distribution. Heterogeneity is introduced through a nonconstant autocovariance function of ε_t^2. The GLS type transformation needed in this case is thus to reweight each innovation by the correlation with ε_t^2. Heuristically speaking, a GLS transformation is performed on the innovation sequence rather than on the observations. This approach will be formalized next.

5. Optimal Instrumental Variables Estimators

The main result of this part is to relate the lower bound previously obtained to a class of instrumental variables estimators. The lower bound covariance matrix for the class of IV estimators will be shown to be equal to Ξ in (4.2).

The instruments z_t are restricted to be linear functions of the innovation process ε_t. Restricting the instruments to the linear class has implications for the efficiency properties of the estimators. It rules out conditional GLS transformations and ML estimators for parametric cases. Linearity, on the other hand, leads to a tractable theory. We define z_t as

$$z_t = \sum_{k=1}^{\infty} a_k \varepsilon_{t-k}$$

with $a_j \in \mathbb{R}^d$ and $\sum |j|^{1/2} \|a_j\| < \infty$. In particular we use the same number of instruments as parameters to estimate. The estimation problem for the time series model can be stated in terms of the orthogonality condition

$$E\left[\left(C(\beta_0, L)y_t\right) z_t\right] = 0 \tag{5.1}$$

with z_t an $\mathcal{F}_{t-1}$-measurable instrument.

Rather than offering a complete analysis of the consistency and asymptotic normality of estimators based on this orthogonality condition, we will make high level assumptions to that effect. The special case of an $AR(p)$ model will be treated in detail in the second part.

We follow Hansen [45] in working with a general criterion function $Q_n(\beta_n)$ such that the estimator β_n is defined as the solution to

$$\|Q_n(\beta_n)\|^2 /2 = 0.$$

$Q_n(\beta_n)$ will be assumed to behave asymptotically in a way to satisfy (5.1). Since typically $Q_n(\beta_0)$ is a function of $(C(\beta_0, L)y_t)z_t = \varepsilon_t z_t$ it is natural to require a martingale CLT to hold for $\sqrt{n}Q_n(\beta_0)$.

Consistency arguments are complicated by the fact that the parameter space for linear time series models usually is only locally compact. The consequences are far reaching. Jennrich's [57] lemma 2 and 3 are not applicable implying that one can not even ensure that β_n is a measurable function. Hosoya and Taniguchi [54], Kabaila [58], Taniguchi [92] are assuming compactness of the parameter space to avoid consistency problems. Such an assumption is not valid in the ARMA case.

Here we will proceed by adopting the Type B consistency proof in Huber [55]. The formulation there is in terms of low level assumptions on the criterion function and the data generating process which are not readily adaptable to the present situation. In a later development in the theory of nonlinear estimation, Wu [96] states the consistency proof in terms of high level assumptions on the criterion function. A similar approach is taken in Pakes and Pollard [77]. However, both treatments rely on compactness of the parameter space. Zaman [97] extends Wu's approach in several directions. In particular in his consistency proof [97, theorem 2] it is only assumed that the parameter space is

Hausdorff.

Here we will use Zaman's approach to force the estimator β_n into a compact subset of the parameter space with probability one. This result based on high level assumptions will correspond to Huber [55, lemma 2]. We start by making the following assumptions about $Q_n(\beta)$.

Assumption D1. *Let the sequence of estimators $\beta_n \in \mathbb{R}^d$ be defined such that*

$$\|Q_n(\beta_n)\|^2/2 \to 0 \text{ almost surely}$$

where $Q_n(.)$ is separable (see Huber [55, p.222]) and $\|.\|$ is the Euclidean norm.

Assumption D2. *$\beta_0 \in \Theta \subset \mathbb{R}^d$ where Θ is locally compact Hausdorff (LCH).*

Assumption D3. *The function $Q_n(\beta)$ is locally stochastically equicontinuous, i.e for every $\beta^* \in \Theta$ and for $\epsilon > 0$ $\exists \delta > 0$ and every open set $U \subset \Theta$ such that $\beta^* \in U$ it follows that*

$$\limsup_n P\left\{\sup_{\beta\in U}\sup_{\beta'\in B(\beta^*,\delta)}\left\|Q_n(\beta') - Q_n(\beta)\right\| > \epsilon\right\} < \epsilon.$$

Assumption D4. *Let $Q(\beta) = E\left[(C(\beta, L)y_t)z_t\right]$. Assume $EQ_n(\beta) = Q(\beta) < \infty$ exists for all $\beta \in \Theta$. For every $\beta \in \Theta$ $\|Q_n(\beta) - Q(\beta)\| \xrightarrow{p} 0$.*

Assumption D5. *Let the sets $B_k(\beta_0)$ for $k = 1, 2, ...$ form a local base around β_0. Then*

$$\inf_{\beta\in B_k(\beta_0)^C\cap\Theta}\|Q(\beta)\| > 0 \text{ for } k = 1, 2, ...$$

where $B_k(\beta_0)^C$ are the complements of $B_k(\beta_0)$. Moreover $\lim_n \|EQ_n(\beta_0)\| = 0$.

Assumption D6. *There exists a continuous function $b(\beta) > b_0 > 0$ such that the following hold almost surely*

i) $\liminf_n \inf_{\beta \in B_k(\beta_0)^c \cap \Theta} \|EQ_n(\beta)\| / b(\beta) \geq 1.$

ii) $\limsup_n \sup_{\beta \in B_k(\beta_0)^c \cap \Theta} \frac{\|Q_n(\beta) - EQ_n(\beta)\|}{b(\beta)} < 1.$

Remark 6. *A short discussion of the assumptions is in place to relate them to the existing literature. Assumption (D1) is the definition of the estimator and separability of the criterion function ensures that the supremum of the criterion function is measurable. Assumption (D2) insures that there are compact neighbourhoods around β_0 entirely contained in Θ. (D3) is a local stochastic equicontinuity condition. (D4) is a pointwise convergence condition. (D5) is a familiar identification condition which makes sure that the expectation of the criterion function is bounded away from zero outside a neighbourhood of the true parameter. The condition corresponds to the one in Zaman [97, theorem 2] and Pakes and Pollard [77, theorem 3.1, ii)] but is weaker than the one in Pakes and Pollard [77, corollary 3.2, ii)]. Also, the conditions of Wu [96, assumption A] can not be applied directly since we can not assume uniform convergence of the criterion function on Θ. (D6) corresponds to condition B4 in Huber's original article. It is also equivalent to Zaman's conditions (see [97, (6) and (7), p.276]). The conditions ensure that the criterion function. even though not converging uniformly, stays away from zero outside of a neighbourhood around β_0 with probability one. This then implies that an estimator satisfying (D1) has to converge eventually.*

Lemma 5.1. *Let β_n be defined as in Assumption (D1). Then under Assumptions (D2-D6) there exists a compact subset $C \subset \Theta$ such that $\beta_0 \in C$ and $P(\beta_n \in C) = 1$ for all*

$n > n_0$ *some* $n_0 < \infty$.

Proof. By local compactness there exists an open set $O \subset \Theta$ with $\beta_0 \in O$ such that the closure $\overline{O}$ is compact. Since $B_k(\beta_0)$ forms a local base $\exists$ k such that $B_k(\beta_0) \subset O$. Then $\overline{O}^c \subset B_k(\beta_0)^c$. Now let $\overline{O} = C$. The remainder of the proof follows immediately from Hubers argument. By assumption (D6ii) $\exists$ $\epsilon > 0$ and some $n_0 < \infty$

$$\sup_{\beta \notin C} \|Q_n(\beta) - EQ_n(\beta)\| / b(\beta) < 1 - 2\epsilon \tag{5.2}$$

for all $n > n_0$. Also by (D6i)

$$\inf_{\beta \notin C} \|EQ_n(\beta)\| / b(\beta) \geq 1 - \epsilon \tag{5.3}$$

Combining (5.2) and (5.3) then implies for all $\beta \notin C$ and n sufficiently large

$$\|Q_n(\beta) - EQ_n(\beta)\| \leq (1 - \epsilon) \|EQ_n(\beta)\|$$

or $\|Q_n(\beta)\| > \epsilon \|EQ_n(\beta)\| > \epsilon b_0$. The result now follows form the definition of β_n. ■

Consistency of β_n can be established by applying standard arguments for compact parameter spaces to the behaviour of $Q_n(\beta)$ on the set C.

Proposition 5.2. *Let* β_n *be defined as in assumption (D1). Then, under assumptions (D2-D6),* $\beta_n \xrightarrow{p} \beta_0$.

Proof. Restrict attention to the compact set C. Then for any open neighbourhood U of

β_0 $\inf_{\beta \in C \backslash U} \|Q(\beta)\| > \varepsilon$ by Assumption (D5) such that

$$P(\beta_n \in C\backslash U) \leq P\left(\sup_{\beta \in C} \|Q(\beta) - Q_n(\beta_n)\| + \|Q_n(\beta_n)\| > \epsilon\right).$$

For n sufficiently large this probability can be bounded by

$$P\left(\sup_{\beta \in C} \|Q(\beta) - Q_n(\beta)\| > \epsilon\right) + \epsilon$$

where we use $\|Q_n(\beta_n)\| = o_p(1)$. Then for every $\beta \in C$ choose a neighbourhood U_β such that for n large enough

$$P\left(\sup_{\beta' \in U_\beta} \left\|Q_n(\beta') - Q_n(\beta)\right\| > \epsilon\right) < \epsilon$$

by Assumption (D3). Select a finite subcover U_{β_s} $s = 1, ..., S$ of C. Now it follows that

$$\begin{aligned}
P\left(\sup_{\beta \in C} \|Q(\beta) - Q_n(\beta)\| > \epsilon\right) &= P\left(\max_{s \leq S} \sup_{\beta \in U_{\beta_s}} \|Q(\beta) - Q_n(\beta)\| > \epsilon\right) \\
&\leq P\left(\max_{s \leq S} \sup_{\beta \in U_{\beta_s}} \|Q_n(\beta_s) - Q_n(\beta)\| > \epsilon\right) \\
&\quad + P\left(\max_{s \leq S} \|Q_n(\beta_s) - Q(\beta_s)\| > \epsilon\right) \\
&\leq P\left(\sup_{\beta' \in C} \sup_{\beta \in U_{\beta'}} \left\|Q_n(\beta') - Q_n(\beta)\right\| > \epsilon\right) + \epsilon \\
&\leq \epsilon
\end{aligned}$$

where $P\left(\max_{s \leq S} \|Q_n(\beta_s) - Q(\beta_s)\| > \epsilon\right)$ goes to zero by Assumption (D4). This completes the proof ■

We now state assumptions that are enough to establish a result for the limiting distribution of $\sqrt{n}(\beta_n - \beta_0)$. We require that the criterion function has an asymptotic distribution which is the same as the sample analogue of the moment restrictions.

Assumption E1. $\sqrt{n}Q_n(\beta_n) = o_p(1)$.

Assumption E2. $\sqrt{n}(Q_n(\beta_0) - \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t z_t) = o_p(1)$.

For any $m \in \mathbb{N}$ define $A_m' = [a_1, ..., a_m]$ with $\sum \|a_k\| \, |k|^{1/2} < \infty$ and $z_t^m = \sum_{k=1}^{m} a_k \varepsilon_{t-k}$. Then, by the martingale CLT (A.1), it follows that

$$\sqrt{n}\sum_{t=1}^{n} \varepsilon_t z_t^m \xrightarrow{d} Y_m,$$

where $Y_m \sim N(0, A_m \Omega_m A_m')$ for any fixed m. Now Lemma (A.4) implies that $\sqrt{n}\sum_{t=1}^{n}\varepsilon_t z_t \xrightarrow{d} N(0, \lim_m A_m \Omega_m A_m')$. Next we want to expand the first order condition for β_n around β_0. We make the following assumption.

Assumption E3. *$Q_n(.)$ is twice continuously differentiable on a neighbourhood $N(\beta_0)$ of β_0. $E[(\frac{\partial}{\partial \beta}C(\beta_0, L)y_t)z_t] = M > 0$. Also,*

$$\sup_{\beta \in N(\beta_0)} \left\| \frac{\partial}{\partial \beta} Q_n(\beta) - E[(\frac{\partial}{\partial \beta}C(\beta, L)y_t)z_t] \right\| \xrightarrow{p} 0$$

and $\sup_{\beta \in N(\beta_0)} \frac{\partial^2}{\partial \beta \partial \beta'} Q_n(\beta) = O_p(1)$.

Since $y_t = C^{-1}(\beta_0, L)\varepsilon_t$, the expectation can also be expressed as

$$M = E[(\frac{\partial}{\partial \beta}\log C(\beta_0, L)\varepsilon_t)z_t].$$

Using the definition of $b_k = (2\pi)^{-1}\int_{-\pi}^{\pi}\dot{\eta}(\beta_0,\lambda)e^{i\lambda k}d\lambda$ and stacking the Fourier coefficients b_k in the matrix P_m we can also write $M = \lim_m P_m' A_m$. A familiar mean value expansion leads to

$$\begin{aligned} o_p(1) &= \frac{\partial}{\partial\beta}Q_n(\beta_n)\sqrt{n}Q_n(\beta_n) \\ &= (M + o_p(1))[Q_n(\beta_0) + \frac{\partial}{\partial\beta}Q_n(\beta_n^*)\sqrt{n}(\beta_n - \beta_0)]. \end{aligned}$$

The limiting distribution of the instrumental variables estimator is stated in the next theorem.

Theorem 5.3. *Let $z_t = \lim_{m\to\infty} A_m'\varepsilon_t^m$ with $A_m' = [a_1, ..., a_m]$ and $\sum \|a_k\| \, |k|^{1/2} < \infty$. Then the estimator based on $E[C(\beta, L)y_t z_t] = 0$ and defined by $\beta_n = \arg\min \|Q_n(\beta_n)\|^2$ has a limiting distribution given by*

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{d} N(0, M^{-1}(\lim_{m\to\infty} A_m'\Omega_m A_m)M^{-1})$$

The covariance matrix $M^{-1}(\lim_{m\to\infty} A_m'\Omega_m A_m)M^{-1}$ satisfies the matrix inequality

$$\lim_{m\to\infty}\left(P_m' A_m\right)^{-1} A_m'\Omega_m A_m \left(A_m' P_m\right)^{-1} \geq \Xi \qquad (5.4)$$

where Ξ is given in (4.2).

Proof. Use the same arguments as in the proof of Theorem 4.4 ■

Theorem (5.3) immediately leads to the construction of an efficient IV estimator. The matrix A_m has to be chosen in a way that (5.4) holds with equality. This is seen to be the case for $A_m = P'_m \Omega_m^{-1}$. Another way to characterize the lower bound in the GLS sense is to require that

$$Var\left[(C(\beta_0, L)y_t)\, z_t\right] = M,$$

where $Var\left[C(\beta_0, L)y_t z_t\right] = E\left[\varepsilon_t^2 z_t z_t'\right]$. The instrument z_t satisfying this condition is given by

$$z_t = \lim_{m \to \infty} P'_m \Omega_m^{-1} \varepsilon_t^m \tag{5.5}$$

where $\varepsilon_t^m = \left[\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}\right]'$. The expression for z_t shows that the optimal instrument balances two effects. It gives more weight to innovations which carry a strong signal measured by a high value in P_m. In fact, b_k measures the effect of ε_{t-k} on y_t. On the other hand, the contribution of ε_{t-k} is discounted if it is strongly linked to the error ε_t as measured by a high correlation between ε_t^2 and $\varepsilon_{t-k}\varepsilon_{t-l}$. The situation here differs from the case of independent errors ε_t. In that case, the innovation ε_{t-k} only affects y_t. With dependence in second moments the error not only generates the signal measured by y_t but also changes the error variance $E\left(\varepsilon_t^2 \left|\mathcal{F}_{t-1}\right.\right)$of the measurement equation. It is this second effect which results in the efficiency loss for the Gaussian QMLE.

We can now verify that an estimator based on z_t indeed attains the lower bound. The variance is

$$\begin{aligned} E\left[\varepsilon_t^2 z_t z_t'\right] &= E\left[\varepsilon_t^2 \lim_{m \to \infty} P'_m \Omega_m^{-1} \varepsilon_t^m \varepsilon_t^{m\prime} \Omega_m^{-1} P_m\right] \\ &= \lim_{m \to \infty} P'_m \Omega_m^{-1} P_m \end{aligned}$$

and

$$M = \lim_{m \to \infty} P_m' \Omega_m^{-1} P_m.$$

This establishes that the variance of the score process is equal to the expectation of the first derivative of the score process.

The analysis in this section is general regarding the functional form of the time series model. In particular, nonlinearities in the parameters are not excluded. In the next part, attention will be focused on the autoregressive case since this allows for explicit expressions of the estimators in terms of projection matrices.

Part II

Autoregressive Models

In this part the general setup from the previous analysis is specialized to the autoregressive case. For this case the instrumental variables estimator can be represented as a projection operator. The focus of this part is on developing a feasible semiparametric version of the IV estimator.

This is achieved by approximating the instruments in the frequency domain. The benefits of the frequency domain implementation are in a reduction of the algorithmic complexity from $O(n^2)$ to $O(n \log n)$. The resulting estimator is independent of a bandwidth choice which makes it attractive for applied work.

For expositional purposes, the dependence structure of the errors is somewhat simplified. The consequences of this simplifying assumption are discussed in the next section.

6. Model

We start by defining the stochastic environment of the model studied. Let $(\Omega, \mathcal{F}, P)$ be a general probability space and define a filtration $\mathcal{F}_t$ to be an increasing sequence of σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}\ \forall\ t$. If not stated otherwise, random variables indexed by t will be assumed to be adapted to the filtration $\mathcal{F}_t$. We assume that we have a sample of size n of a univariate time series y_t where $t = \{1, ..., n\}$. More specifically, we assume that y_t is generated by the following autoregressive model

$$\phi(L)y_t = \varepsilon_t \tag{6.1}$$

where ε_t is a martingale difference sequence. Here $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$. $\phi' = (\phi_1, ..., \phi_p)$ is the vector of parameters describing the mean equation of the model. It is assumed that $\phi(L)$ has all roots outside the unit circle. We are interested in estimating the parameter vector ϕ. The martingale difference assumption for ε_t implies absence of correlation between the errors. However, it is not assumed that the errors are independent. Rather we allow for dependence in higher than second moments to account for thick tails and conditional heteroskedasticity in the errors.

Assumption F1. *(i)* ε_t *is strictly stationary and ergodic,* $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2$, $E(\varepsilon_t^2) = \sigma^2 < \infty$.

(ii) $\phi(L)$ *has all roots outside the unit circle.*

(iii) $E\left((\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-s}^2 - \sigma^2)\right) = \sigma(s) < \infty$ *for* $s \geq 0$.

(iv) $E\left(\varepsilon_t^2 \varepsilon_{t-s} \varepsilon_{t-r}\right) = 0$ *for* $s \neq r$, $s, r > 0$.

(v) $\sum |s| |\sigma(s)| = B < \infty$, $E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) > \underline{\alpha}$ *some* $\underline{\alpha} > 0$ *for all* s.

Remark 7. *Assumption (iv) is somewhat restrictive as it rules out some nonsymmetric parametric examples such as EGARCH. While (iv) is not a critical element in the theory developed later, it is maintained to simplify the exposition. The IV estimators proposed in Section 7 are still consistent and asymptotically normal if (iv) fails. However, in this case they lose their optimality properties. Under these conditions the covariance matrix* *estimators (10.9) for the standard errors of the IV estimators are no longer consistent.*

By definition of the conditional expectation operator, σ_t is $\mathcal{F}_{t-1}$ measurable. Assumption (F1) implies that ε_t^2 is strictly stationary and ergodic and therefore covariance stationary. It should be emphasized that no assumptions about third moments are made. In particular this allows for skewness in the error process.

In this paper it is explicitly assumed that the parametric form generating the higher moment dependence is unknown. Nevertheless, we provide examples of widely used processes that exhibit features analyzed here, mainly to illustrate the relevance of the assumptions. A number of popular processes used mainly in financial econometrics satisfy Assumption (F1). Examples are provided next.

Example 6.1 (ARCH(1), Engle; 1982). *Let $\varepsilon_t = u_t h_t^{1/2}$, where u_t is $iid(0,1)$ with symmetric distribution and $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2$. Expanding leads to*

$$h_t = \gamma_0 \left(1 + \sum_{i=1}^{\infty} \gamma_1^i \left(\prod_{j=1}^{i} u_{t-j}^2 \right) \right).$$

In case of normal u_t, the conditional distribution of ε_t is also normal.

Example 6.2 (GARCH (1,1), Bollerslev; 1986). *Let $\varepsilon_t = u_t h_t^{1/2}$, where u_t is $iid(0,1)$ with symmetric distribution and $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. Substitution for ε_t^2 gives $h_t = \alpha_0 + \left(\alpha_1 u_{t-1}^2 + \beta_1\right) h_{t-1}$. Expansion again leads to*

$$h_t = \gamma_0 \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} \left(\gamma_1 u_{t-j}^2 + \beta_1\right) \right).$$

Example 6.3 (Stochastic Volatility). *Let* $\varepsilon_t = u_t \exp(h_t/2)$ *with* $h_t = \delta h_{t-1} + v_t$. $u_t \sim iid\ (0, \sigma_u)$, $v_t \sim iid(0, \sigma_v)$, u_t *and* v_t *independent of each other. Then,* $\varepsilon_t = u_t \exp(\sum_{i=0} \delta^i v_{t-i}) = u_t \exp(v_t) \exp(\sum_{i=1} \delta^i v_{t-i})$.

Example 6.4 (Linear Scale). *Let* $\varepsilon_t = u_t h_t$ *with* $h_t = \sum_{i=0} c(i, \theta) v_{t-i}$, $u_t \sim iid(0, \sigma_u)$, $v_t \sim iid(0, \sigma_v)$, u_t *and* v_t *independent of each other.*

Remark 8. *Nelson [69] obtains sufficient conditions for stationarity and ergodicity of 6.1 and 6.2. The martingale difference property follows immediately from independence of* u_t*. Assumption (F1iv) is shown to hold for the ARCH(p) case in Milhoj [68]. The reason is that odd moments appearing in the asymmetric fourth moment terms are zero by the symmetry assumption for the error density. The same argument extends to the* $GARCH(p, q)$ *case. For the Linear Scale and Stochastic Volatility examples, Assumption (F1iv) is satisfied because of independence between* u_t *and* v_t*. If* $u_t \sim N(0, 1)$*, then fourth moments are known to exist if* $3\gamma_1^2 + 2\gamma_1\beta_1 + \beta_1^2 < 1$*. This condition is valid for* $\beta = 0$ *and thus covers the ARCH case. In Milhoj [68] and Bollerslev [9], the autocorrelation structure* $\sigma(s)$ *is shown to be identical to the AR(p) and* $ARMA(\max(p, q), q)$ *case for* $ARCH(p)$ *and* $GARCH(p, q)$ *respectively. This implies that the summability condition holds if fourth moments exist. For (6.3) stationarity and ergodicity follow if* h_t *is stationary, i.e.* $|\delta| < 1$*. Assuming* $u_t \sim N(0, 1)$ *and* $v_t \sim N(0, 1)$*, it is shown by Breidt, Crato and deLima [12] that* $\sigma(s)/\sigma(0) = \exp(\delta^s/1 - \delta^2) - 1$ *or, using the expansion for the exponential, that* $\sigma(s)/\sigma(0) = \sum_{k=1}^{\infty} \frac{1}{k!} (\delta^s/1 - \delta^2)^k$*. Then we can bound* $\sum |s| \sigma(s) \leq \exp(1/1 - \delta^2)\, \sigma(0) \sum_{s=0}^{\infty} s\delta^s$*. A form similar to Example* 6.4 *was proposed in Hansen [43].*

Based on the results in Part I, we will now introduce the optimal instrumental variables estimator for the $AR(p)$ model. The estimator is constructed by reweighting the innovation sequence by the unconditional fourth moments $\sigma(k) + \sigma^4$ of the error process. Without parametric assumptions about the form of conditional heterogeneity these moments typically have to be estimated.

7. Instrumental Variables Estimator

The instrumental variables estimator defined by $E[(\phi(L)\, y_t)\, z_t] = 0$ can be written explicitly in the case of an autoregressive model as

$$\tilde{\phi} = \left(Z'Y_{-1}\right)^{-1} Z'Y, \tag{7.1}$$

where

$$\begin{aligned} Y' &= [y_{1+p}, \ldots, y_n] \\ Y'_{-1} &= \begin{bmatrix} y_p & ,\ldots, & y_{n-1} \\ \vdots & & \vdots \\ y_1 & ,\ldots, & y_{n-p} \end{bmatrix} \end{aligned}$$

Note that $Y = Y_{-1}\phi + \varepsilon$. The instrument z_t is assumed to be $\mathcal{F}_{t-1}$-measurable, strictly stationary and ergodic so that $\tilde{\phi}$ is consistent. Under additional moment conditions, the asymptotic distribution is

$$\sqrt{n}\left(\tilde{\phi} - \phi\right) \Rightarrow N\left(0, \lim_{n\to\infty} \left(En^{-1}Z'Y_{-1}\right)^{-1} En^{-1}Z'\varepsilon\varepsilon'Z \left(En^{-1}Z'Y_{-1}\right)^{-1}\right).$$

Using $\phi^{-1}(z) = \sum_{j=0}^{\infty} \psi_{\phi,j} z^j$, the time series y_t can be expressed as a linear filter of past shocks $y_t = \sum_{j=0}^{\infty} \psi_{\phi,j}\varepsilon_{t-j}$. Also, let

$$\alpha_k = E\left(\varepsilon_t^2 \varepsilon_{t-k}^2\right) = \sigma(k) + \sigma^4 \tag{7.2}$$

and $b_k' = (\psi_{\phi,k-1}, ..., \psi_{\phi,k-p})$ with $\psi_{\phi,k} = 0$ for $k < 0$. Then define $P_m' = [b_1, ..., b_m]$.

We have shown in Part I that for $z_t = \sum_{k=1}^{\infty} a_k \varepsilon_{t-k}$ the asymptotic covariance matrix of this estimator has the form $\lim_{m\to\infty}(P_m' A_m)^{-1} A_m' \Omega_m A_m (A_m' P_m)^{-1}$ with lower bound $\lim(\sigma^4 P_m' \Omega_m^{-1} P_m)^{-1}$. Under Assumption (F1), $\Omega_m = Diag(\alpha_1, ..., \alpha_m)$.

An optimal set of instruments Z constructed from $z_t = \lim_{m\to\infty} P_m' \Omega_m^{-1} \varepsilon_t^m$ as in (5.5) can be written explicitly for the $AR(p)$ case as

$$\begin{aligned} z_{t+1,1} &= \sum_{j=0}^{\infty} \frac{\psi_{\phi,j}}{\alpha_{j+1}} \varepsilon_{t-j} \\ &\vdots \\ z_{t+1,p} &= \sum_{j=0}^{\infty} \frac{\psi_{\phi,j}}{\alpha_{j+p}} \varepsilon_{t-j}. \end{aligned}$$

The instrument matrix Z is chosen in the following way

$$Z = \begin{bmatrix} z_{p,1} & ,\dots, & z_{n-1,1} \\ \vdots & & \vdots \\ z_{1,p} & ,\dots, & z_{n-p,p} \end{bmatrix},$$

where it should be emphasized that the k -th instrument, i.e. the instrument for parameter ϕ_k is lagged by k periods and has a convolution filter which is also shifted by k elements.

It is clear that the instrument matrix Z is not observable and the procedure is therefore infeasible. The main result of the paper will consist in the construction of an adaptive estimator for $\tilde{\phi}$. For the moment, however, the analysis is carried out under the assumption that Z is observable. The asymptotic distribution of $\tilde{\phi}$ is analyzed by considering a typical element of $En^{-1}Z'Y_{-1}$ which can be evaluated by direct calculation as

$$\begin{aligned}\lim_n \left[En^{-1}Z'Y_{-1}\right]_{k,l} &= \lim_n \left[n^{-1}\sum_{t=1+p}^{n} E\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{\psi_{\phi,j}\psi_{\phi,i}}{\alpha_{i+k\vee l}}\varepsilon_{t-k-i}\varepsilon_{t-l-j}\right] \\ &= \sigma^2\sum_{i=0}^{\infty}\frac{\psi_{\phi,i}\psi_{\phi,i+|k-l|}}{\alpha_{i+k\vee l}}.\end{aligned}$$

Also, a typical element of $En^{-1}Z'\varepsilon\varepsilon'Z$ is

$$\begin{aligned}\lim_n \left[En^{-1}Z'\varepsilon\varepsilon'Z\right]_{k,l} &= \lim_n \left[n^{-1}\sum_{t=1+p}^{n} E\varepsilon_t^2 z_t z_t'\right]_{k,l} \\ &= E\varepsilon_t^2\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{\psi_{\phi,j}\psi_{\phi,i}}{\alpha_{i+k}\alpha_{j+l}}\varepsilon_{t-k-i}\varepsilon_{t-l-j} \\ &= \sum_{i=0}^{\infty}\frac{\psi_{\phi,i+k-l}\psi_{\phi,i}}{\alpha_{i+k}^2}E\varepsilon_t^2\varepsilon_{t-k-i}^2 \\ &= \sum_{i=0}^{\infty}\frac{\psi_{\phi,i+k-l}\psi_{\phi,i}}{\alpha_{i+k}} = \sum_{i=0}^{\infty}\frac{\psi_{\phi,i}\psi_{\phi,i+|k-l|}}{\alpha_{i+k\vee l}}\end{aligned}$$

where the last equality follows from the fact that $\psi_{-k} = 0$ for $k > 0$. The asymptotic distribution of $\tilde{\phi}$ then is

$$\sqrt{n}\left(\tilde{\phi}-\phi\right) \Rightarrow N\left(0, \sigma^{-4}\left\{\left[\sum_{i=0}^{\infty}\frac{\psi_{\phi,i}\psi_{\phi,i+|k-l|}}{\alpha_{i+k\vee l}}\right]_{k,l}\right\}^{-1}\right)$$

and it is easy to check that $\left\{[\sum_{i=0}^{\infty}\frac{\psi_{\phi,i}\psi_{\phi,i+|k-l|}}{\alpha_{i+k\vee l}}]_{k,l}\right\}^{-1}$ equals Ξ in (4.2) for the autoregressive case.

As discussed before the instrument Z is not observable and needs to be replaced by a suitable estimate. While direct calculation of Z is possible, it is computationally inefficient, requiring $O(n^2)$ operations. A more natural way to formulate the estimator is in terms of discrete Fourier Transforms. Since the construction of the optimal instrument involves a convolution in the time domain, this is transformed into a simple multiplication in the frequency domain, thereby leading to a reduction in computing time.

Moreover, it turns out that direct calculation of the instruments can be avoided altogether in the frequency domain. It will be established in the next section that the instrumental variable estimator is asymptotically equivalent to an estimator based on the Whittle likelihood (8.2) where $g_{yy}^{-1}(\phi, \lambda)$ is replaced by an optimal filter.

8. IV Estimation in the Frequency Domain

In this section a frequency domain approximation to the optimal IV estimator is derived. To introduce notation and methodology, the Gaussian estimator for the $AR(p)$ model in the frequency domain is reviewed. It is then shown how the estimator can be transformed into an instrumental variables estimator by applying a linear filter to the data periodogram.

8.1. Spectral Representation of the AR(p) Model

We assume that starting values $y_{-p+1}, ..., y_0$ are drawn from the stationary distribution of y_t. Then, the Gaussian QMLE in the time domain can be approximated asymptotically by

$$\hat{\phi} = \left(Y'_{-1}Y_{-1}\right)^{-1} Y'_{-1}Y.$$

We derive a frequency domain analogue to $\hat{\phi}$ based on the periodogram of y_t alone. Consider the inverse of the spectral density for the *AR*(*p*) model $f_{yy}^{-1}(\lambda) = \frac{2\pi}{\sigma^2}\left|\phi\left(e^{i\lambda}\right)\right|^2$ and let $\left|\phi\left(e^{i\lambda}\right)\right|^2 = g_{yy}^{-1}(\phi, \lambda)$. The definition of the spectrum of the squared errors will be useful later and is

$$f_{\varepsilon^2\varepsilon^2}(\lambda) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty}\sigma(k)\,e^{-i\lambda k}. \tag{8.1}$$

Define the lag operator $a(\lambda) = \left[e^{i\lambda}, ..., e^{i\lambda p}\right]$ and denote the complex conjugate transpose by $a(\lambda)^*$. Also introduce the matrix $A(\lambda) = a(\lambda)^* a(\lambda)$. Then $g_{yy}^{-1}(\phi, \lambda)$ can be represented as

$$g_{yy}^{-1}(\phi, \lambda) = \left[1, \phi'\right]\begin{bmatrix} 1 & a(\lambda) \\ a(\lambda)^* & A(\lambda) \end{bmatrix}\begin{bmatrix} 1 \\ \phi \end{bmatrix}.$$

We introduce the discrete Fourier transform of the data as $\omega_y(\lambda) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t e^{-it\lambda}$ and the periodogram as $I_{n,yy}(\lambda) = \left|\omega_y(\lambda)\right|^2$. The Whittle likelihood, which is the spectral representation of the sum of squared errors, can now be written as

$$\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, g_{yy}^{-1}(\phi, \lambda) d\lambda = \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\left[1 + \phi' a(\lambda)^* + a(\lambda)\phi + \phi' A(\lambda)\phi\right] d\lambda. \tag{8.2}$$

For computational purposes (8.2) can be approximated by a discrete sum over the fundamental frequencies by Brillinger [14, theorem 5.10.2]. The QML estimator is defined as the value minimizing (8.2). The solution to the minimization problem is given by

$$\hat{\phi} = \left(\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\,\mathrm{Re}\left[A(\lambda)\right] d\lambda\right)^{-1}\int_{-\pi}^{\pi}\mathrm{Re}\left[a(\lambda)^*\right] I_{n,yy}(\lambda)\, d\lambda, \tag{8.3}$$

where the real part Re [.] is the orthogonal projection of the complex plane onto the real line. From Lemma (A.2) it follows that

$$\int_{-\pi}^{\pi} a(\lambda)^* I_{n,yy}(\lambda)\, d\lambda = \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, A(\lambda)\, \phi d\lambda + \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) \frac{a(\lambda)^*}{\left(1-\phi' a(\lambda)^*\right)} d\lambda + o_p\left(n^{-1/2}\right).$$

A similar expression can be obtained for $a(\lambda)' I_{n,yy}(\lambda)$. Next, noting that

$$\partial \ln g_{yy}(\phi,\lambda)/\partial\phi = \operatorname{Re}\left[a(\lambda)^* \left(1-\phi' a(\lambda)^*\right)^{-1}\right] \tag{8.4}$$

and introducing the notation $\dot{\eta}(\phi,\lambda) = \partial \ln g_{yy}(\phi,\lambda)/\partial\phi$ leads to

$$\begin{aligned} \int_{-\pi}^{\pi} \operatorname{Re}\left[a(\lambda)^* I_{n,yy}(\lambda)\right] d\lambda &= \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \operatorname{Re}\left[A(\lambda)\right] d\lambda \phi \\ &+ \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, \dot{\eta}(\phi,\lambda)\, d\lambda + o_p\left(n^{-1/2}\right). \end{aligned} \tag{8.5}$$

Substituting (8.5) back into (8.3) results in the following expression for the deviation of $\hat{\phi}$ from the true parameter vector ϕ

$$\sqrt{n}\left(\hat{\phi}-\phi\right) = \left(\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \operatorname{Re}\left[A(\lambda)\right] d\lambda\right)^{-1} \sqrt{n} \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, \dot{\eta}(\phi,\lambda)\, d\lambda + o_p(1). \tag{8.6}$$

Consistency of the estimator follows by the ergodic theorem from the fact that

$$E \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, \dot{\eta}(\phi,\lambda)\, d\lambda = \sigma^2 \int_{-\pi}^{\pi} \dot{\eta}(\phi,\lambda)\, d\lambda$$

and $\int_{-\pi}^{\pi} \dot{\eta}(\phi,\lambda)\, d\lambda = 0$. By the ergodic theorem, we also have $\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \operatorname{Re}\left[A(\lambda)\right] d\lambda \overset{a.s.}{\rightarrow}$

$\int_{-\pi}^{\pi} g_{yy}(\phi,\lambda)\,\mathrm{Re}\,[A(\lambda)]\,d\lambda$. It can be established that

$$2\int_{-\pi}^{\pi} g_{yy}(\phi,\lambda)\,\mathrm{Re}\,[A(\lambda)]\,d\lambda = \int_{-\pi}^{\pi} \dot{\eta}(\phi,\lambda)\,\dot{\eta}(\phi,\lambda)'\,d\lambda.$$

It follows from the results in Part I that $\sqrt{n}\left(\hat{\phi}-\phi\right) \Rightarrow N\left(0, A^{-1}BA^{-1}\right)$, where

$$A = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \dot{\eta}(\phi,\lambda)\,\dot{\eta}(\phi,\lambda)'\,d\lambda \tag{8.7}$$

and

$$B = 2\sigma^2 A + \pi^{-1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon^2}(\mu)\,\dot{\eta}(\phi,\mu-\lambda)\,\dot{\eta}(\phi,\lambda)'\,d\mu d\lambda. \tag{8.8}$$

The appearance of fourth order cumulant terms in the matrix B is caused by the dependence between errors and regressors. In the case of a constant conditional second moment $E\left(\varepsilon_t^2\,|\mathcal{F}_{t-1}\right) = \sigma^2$ the spectral density $f_{\varepsilon^2\varepsilon^2}(\mu)$ is equal to $\sigma^4/(2\pi)^2$. The second term in B is then zero and the covariance matrix reduces to $2\sigma^2 A^{-1}$, as in Hannan [42].

8.2. Frequency Domain Approximation

It will now be shown how the instrumental variables estimator introduced at the beginning can be implemented in the frequency domain. For the purpose of this and the next section, we introduce the spaces $L^k[-\pi,\pi]$ of functions $f : [-\pi,\pi] \to \mathbb{C}^p$ such that $\int |f|^k d\lambda < \infty$. Also, define the spaces $C^k[-\pi,\pi]$ of functions $f : [-\pi,\pi] \to \mathbb{R}^p$ such that f is k times continuously differentiable. Throughout, the function $R_n(\lambda)$ will denote a generic remainder term whose definition can change. We start by approximating the discrete

Fourier transform of the instrument variables $z_{t,j}$

$$
\begin{aligned}
\omega_{z_k}(\lambda) &= n^{-1/2}\sum_{j=0}^{\infty}\sum_{t=1}^{n}\frac{\psi_j}{\alpha_{k+j}}\varepsilon_{t-j}e^{-i\lambda t} \\
&= n^{-1/2}e^{i\lambda k}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda(j+k)}\sum_{t=1}^{n}\varepsilon_{t-j}e^{-i\lambda(t-j)} \\
&= l_{\psi,k}(\lambda)\, e^{i\lambda k}\omega_{\varepsilon}(\lambda) + n^{-1/2}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda j}U_{n,j}(\lambda) \\
&= l_{\psi,k}(\lambda)\, e^{i\lambda k}\phi\left(e^{-i\lambda}\right)\omega_y(\lambda) + R_n(\lambda) \qquad (8.9)
\end{aligned}
$$

where $l_{\psi,k}(\lambda) = \sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda(j+k)}$ and

$$
R_n(\lambda) = l_{\psi}(\lambda)\, n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{n,j}(\lambda) + n^{-1/2}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda j}U_{n,j}(\lambda)
$$

with $U_{n,j}(\lambda) = \sum_{t=1-j}^{n-j}\varepsilon_t e^{-i\lambda t} + \sum_{t=1}^{n}\varepsilon_t e^{-i\lambda t}$. In an analogous way we define

$$
\omega_{z_k}(-\lambda) = l_{\psi,k}(-\lambda)\, e^{-i\lambda k}\phi\left(e^{i\lambda}\right)\omega_y(-\lambda) + R_n(-\lambda). \qquad (8.10)
$$

Also, from Lemma (A.2), $\sqrt{n}\int I_{n,yz_k}(\lambda) - l_{\psi,k}(-\lambda)\, e^{-i\lambda k}\phi\left(e^{i\lambda}\right) I_{n,yy}(\lambda)\, d\lambda = o_p(1)$. Define $l_{\psi}(\lambda)$ as

$$
l_{\psi}(\lambda) = \begin{bmatrix} l_{\psi,1}(\lambda) \\ \vdots \\ l_{\psi,p}(\lambda) \end{bmatrix}. \qquad (8.11)
$$

The properties of $l_{\psi}(\lambda)$ determine the asymptotic distribution of the instrumental variables estimator. The next lemma gives a representation of $l_{\psi}(\lambda)$ in terms of convolution operators. This shows that the smoothness of $l_{\psi}(\lambda)$ is inherited from the smoothness of the $AR(p)$ spectrum.

Lemma 8.1. *Let* $l_{\psi,k}(\lambda) = \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{k+j}} e^{-i\lambda(j+k)}$ *with* $\alpha_k = E\left(\varepsilon_t^2 \varepsilon_{t-k}^2\right)$, ε_t *satisfying Assumption (F1) and* ψ_j *being the coefficients of the power series expansion of* $\phi(z)^{-1}$. *Also assume that* $\phi(z) = 1 - \phi' a(z)^*$ *has all characteristic roots outside the unit circle. Then* $l_\eta(\lambda) = l_\psi(\lambda) + l_\psi(-\lambda)$ *can be represented as*

$$l_\eta(\lambda) = \int_{-\pi}^{\pi} f_{\tilde{\alpha}}(\lambda - \xi)\, \dot{\eta}(\phi, \xi)\, d\xi + \frac{1}{\sigma^4} \dot{\eta}(\phi, \lambda)$$

where $f_{\tilde{\alpha}}(\lambda) = \sum_{j=-\infty}^{\infty} \tilde{\alpha}_j e^{-i\lambda j}$ *with* $\tilde{\alpha}_j = \left(\frac{1}{\alpha_j} - \frac{1}{\sigma^4}\right)$.

Proof. See Appendix B ■

Remark 9. *Using the convolution operator* $*$ *a compact notation for* $l_\eta(\lambda)$ *is* $l_\eta(\lambda) = (f_{\tilde{\alpha}} * \dot{\eta})(\lambda) + \frac{1}{\sigma^4} \dot{\eta}(\phi, \lambda)$. *The properties of* $l_\eta(\lambda)$ *can now be determined from those of* $f_{\tilde{\alpha}}$ *and* $\dot{\eta}$. *For* $f_{\tilde{\alpha}} \in L^1[-\pi, \pi]$ *and* $\dot{\eta} \in C^k[-\pi, \pi]$ *it follows from Folland [31, theorem 8.10] that* $l_\eta(\lambda) \in C^k[-\pi, \pi]$ *implying that* $\sum_{j=0}^{\infty} \left|\frac{\psi_j}{\alpha_{k+j}}\right| |j|^{1/2} < \infty$. *While* $f_{\tilde{\alpha}} \in L^1[-\pi, \pi]$ *is sufficient to obtain this result it is not necessary. Alternatively if* $\alpha_k > \underline{\alpha} > 0$ *for some* $\underline{\alpha}$ *and all* k *then* $\sup \left|\alpha_k^{-1}\right| < \underline{\alpha}^{-1} < \infty$ *and* $\sum_{j=0}^{\infty} \left|\frac{\psi_j}{\alpha_{k+j}}\right| |j|^{1/2} < \underline{\alpha}^{-1} \sum_{j=0}^{\infty} \left|\psi_j\right| |j|^{1/2} < \infty$ *if* $\dot{\eta} \in C^2[-\pi, \pi]$. *These arguments show that* $l_\eta(\lambda)$ *is sufficiently smooth to apply the central limit theorem in Appendix A.*

The representation of the discrete Fourier transforms of the instruments in terms of the DFT for the data allows to obtain a frequency domain version of $\tilde{\phi}$ without the need to go through an explicit calculation of the instruments in the time domain. The approximation relies on the fact that convolutions in the time domain are transformed into multiplications in the frequency domain and the fact that the residuals can be computed by a simple multiplication of $\omega_y(\lambda)$ by $\phi(e^{i\lambda})$.

The discrete Fourier transform of the instrument matrix is then obtained from (8.9) by lagging each $\omega_{z_k}(\lambda)$ by k periods leading to $e^{-i\lambda k}\omega_{z_k}(\lambda)$ and stacking the resulting transforms in the following way

$$\omega_z(\lambda) = \begin{bmatrix} e^{-i\lambda}\omega_{z_1}(\lambda) \\ \vdots \\ e^{-i\lambda p}\omega_{z_p}(\lambda) \end{bmatrix} = \phi\left(e^{-i\lambda}\right)\omega_y(\lambda)\, l_\psi(\lambda) + \tilde{R}_n(\lambda), \tag{8.12}$$

where $\tilde{R}_n(\lambda) = a(\lambda)^* \odot R_n(\lambda)$ and $\odot$ is the element by element product. The corresponding expression for the conjugate transpose of $\omega_z(\lambda)$ is

$$\omega_z(\lambda)^* = \begin{bmatrix} e^{i\lambda}\omega_{z_1}(-\lambda) \\ \vdots \\ e^{i\lambda p}\omega_{z_p}(-\lambda) \end{bmatrix} = \phi\left(e^{i\lambda}\right)\omega_y(\lambda)^*\, l_\psi(-\lambda) + \tilde{R}_n(-\lambda) \tag{8.13}$$

where the symmetry of $l_\psi(\lambda)$ is used. Using the notation $I_{n,yz}(\lambda) = \omega_y(\lambda)\,\omega_z(\lambda)^*$ and $I_{n,zy}(\lambda) = \omega_y(\lambda)^*\,\omega_z(\lambda)$ the frequency domain version of equation (7.1) can now be written as

$$\tilde{\phi} = \left[\int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,zy}(\lambda)\, a(\lambda)\right] d\lambda\right]^{-1} \int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,zy}(\lambda)\right] d\lambda \tag{8.14}$$

and using the fact that $\omega_y(\lambda)\,\omega_z(\lambda)^* = a(\lambda)^{*\prime}\,\phi\omega_y(\lambda)\,\omega_z(\lambda)^* + \omega_\varepsilon(\lambda)\,\omega_z(\lambda)^* + R_n(\lambda)$ where $\sqrt{n}\int R_n(\lambda)\, d\lambda = o_p(1)$. A similar expression holds for $\omega_y(\lambda)^*\,\omega_z(\lambda)$ reducing

(8.14) to

$$\tilde{\phi} - \phi = \left[\int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,zy}(\lambda)\, a(\lambda) \right] d\lambda \right]^{-1} \int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,z\varepsilon}(\lambda) \right] d\lambda + o_p\left(n^{-1/2} \right). \qquad (8.15)$$

The estimator $\tilde{\phi}$ corresponds to the instrumental variables estimator introduced in Section 7. It is possible to estimate directly the unobservable instruments. On the other hand, it is conceptually more convenient to separate the data and the unknown filters.

One additional approximation step produces an asymptotically equivalent estimator based on the periodogram of y_t and an unknown filter. It is convenient to define

$$h^{x}(\phi, \lambda) = \mathrm{Re}\left[l_{\psi}(-\lambda)\, \phi\left(e^{i\lambda} \right) a(-\lambda) \right]$$

and

$$h(\phi, \lambda) = \mathrm{Re}\left[l_{\psi}(-\lambda)\, \phi\left(e^{i\lambda} \right) \right].$$

By substituting for equations (8.12) and (8.13), (8.14) can be approximated by

$$\check{\phi} = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, h^{x}(\phi, \lambda)\, d\lambda \right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, h(\phi, \lambda) d\lambda. \qquad (8.16)$$

It is shown in the proof of Proposition (8.2) that $\tilde{\phi} - \check{\phi} = o_p\left(n^{-1/2} \right)$ and

$$\check{\phi} - \phi = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, h^{x}(\phi, \lambda) d\lambda \right]^{-1} \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, l_{\eta}(\lambda)\, d\lambda + o_p\left(n^{-1/2} \right)$$

such that consistency again follows from ergodicity and the fact that

$$E\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, l_{\eta}(\lambda)\, d\lambda = \sigma^2 \int_{-\pi}^{\pi} l_{\eta}(\lambda)\, d\lambda = 0.$$

It is transparent from equation (8.16) that $\breve{\phi}$ is infeasible as it stands, since it depends on knowledge of the true parameter values and the correlation structure of the squared errors. Feasible versions of $\tilde{\phi}$ and $\breve{\phi}$ will be discussed in Section 9 below.

Under the assumption that the weight matrix $\mathrm{Re}\left[l_{\psi}(-\lambda)\,\phi\left(e^{i\lambda}\right)\right]$ is known, the asymptotic distribution of $\breve{\phi}$ is now a straight forward consequence of Lemmas (A.2) and (A.4). The asymptotic distribution of $\tilde{\phi}$ and $\breve{\phi}$ can be analyzed by returning to equation (8.15). This is done in the next proposition.

Proposition 8.2. *Let $\phi(L)\, y_t = \varepsilon_t$ where all roots of $\phi(L)$ are outside the unit circle. If ε_t satisfies Assumption (F1) then for $\tilde{\phi}$ defined in (8.14) and $\breve{\phi}$ defined in (8.16) we have*

$$\sqrt{n}\left(\tilde{\phi} - \breve{\phi}\right) = o_p(1)$$

and

$$\sqrt{n}\left(\tilde{\phi} - \phi\right) \Rightarrow N\left(0, \sigma^{-4}\Xi\right)$$

where Ξ is defined in (4.2).

Proof. See Appendix B ■

The remainder of the paper will now be concerned with the construction of a semi-parametric estimator with the same distribution as $\tilde{\phi}$.

9. Adaptive Estimation

To develop an operationally efficient IV procedure, it has to be established that $h(\phi, \lambda) = \operatorname{Re}\left[l_{\psi}(\lambda)\, \phi\left(e^{-i\lambda}\right)\right]$ and $h^{x}(\phi, \lambda) = \operatorname{Re}\left[l_{\psi}(\lambda)\, \phi\left(e^{-i\lambda}\right) a(\lambda)\right]$ can be replaced by consistent estimates without affecting the limiting properties of the estimator. A semiparametric estimator having this property is called adaptive. No confusion should arise between this use of the terminology and the literature on feasible local minimax estimators such as Bickel [7], Kreiss [60], Linton [63] and Steigerwald [88]. The main difference, apart from efficiency issues, is the fact that here a nonparametric correction to the criterion function is made while the local minimax literature makes a nonparametric one step Newton Raphson improvement to a consistent first stage estimator.

Different approaches to prove adaptiveness are used in the semiparametric literature. Direct calculation is used in Robinson [81], [82] in the context of *iid* models and partially linear models and by Hidalgo [52] in the context of time series regression models. Newey [74] applies similar techniques as [81] to the instrumental variables case for *iid* data. Andrews [2] develops a general methodology based on stochastic equicontinuity arguments and applies it to the partially linear framework. Andrews' approach will be used here to break the proof into two parts. First, it is established that uniformly in a shrinking neighborhood of the true filter $h(\phi_0, \lambda)$ the distribution of an estimator is arbitrarily close to the distribution of the estimator based on the true filter. The second step shows that a nonparametric estimate $\hat{h}(\hat{\phi}, \lambda)$ converges to $h(\phi_0, \lambda)$ uniformly with probability one.

This argument will now be formalized. Let $l_{\psi} : [-\pi, \pi] \rightarrow \mathbb{C}^{p}$ and $\phi : [-\pi, \pi] \rightarrow \mathbb{C}$ and introduce a set of functions $\mathcal{H}$ defined as

$$\mathcal{H} = \left\{ h : [-\pi, \pi] \to \mathbb{R}^p \,\middle|\, h = \text{Re}\left[l_\psi(-\lambda)\, \phi\left(e^{i\lambda}\right) \right]; \text{Re}\left[l_\psi(-\lambda) \right], \text{Re}\left[\phi\left(e^{i\lambda}\right) \right] \in C^1[-\pi, \pi] \right\}. \tag{9.1}$$

We define the L^∞ Sobolev norm of order one as

$$\|f\|_1^s = \sup_{\lambda \in [-\pi,\pi]} \|f(\lambda)\| + \sup_{\lambda \in [-\pi,\pi]} \left\| \frac{\partial}{\partial \lambda} f(\lambda) \right\|$$

where $\|.\|$ is the Euclidean matrix norm defined by $\|A\| = (trAA^*)^{1/2}$. Introduce the metric on $\mathcal{H}$ as

$$\rho(h_1, h_2) = \|l_{\psi,1} - l_{\psi,2}\|_1^s + \|\phi_1 - \phi_2\|_1^s.$$

$(\mathcal{H}, \rho)$ is a complete metric space. If ϕ is given by (6.1) and l_ψ by (8.11) then it follows from Lemma (8.1) that $l_\eta(\phi, \lambda) \in C^k[-\pi, \pi]$. Therefore $h(\phi, \lambda) \in \mathcal{H}$.

We proceed by defining the estimator for $h(\phi, \lambda)$. We have established that we can obtain a consistent estimate $\hat{\phi}$ for example from $\hat{\phi} = (Y_{-1}' Y_{-1})^{-1} Y_{-1} Y$ or from its frequency domain analog introduced before. Residuals as a function of some fixed parameter value ϕ are obtained as in Kreiss [60] from

$$\varepsilon_t(\phi) = \varepsilon_t(\phi_0) + (\phi - \phi_0)' (y_{t-1}, \ldots, y_{t-p})$$

such that the estimated error $\varepsilon_t(\phi)$ can be decomposed into the true error and the $\mathcal{F}_{t-1}$

measurable part $(\phi - \phi_0)' (y_{t-1}, \ldots, y_{t-p})$. We form the following statistics.

$$\alpha_k^*(\hat{\phi}) = \frac{1}{n} \sum_{t=p+k+1}^{n} \varepsilon_t^2(\hat{\phi})\varepsilon_{t-k}^2(\hat{\phi})$$

$$\hat{\alpha}_k(\hat{\phi}) = \begin{cases} \frac{1}{n}\sum_{t=p+k+1}^{n} \varepsilon_t^2(\hat{\phi})\varepsilon_{t-k}^2(\hat{\phi}) & \text{if } \alpha_k^* > d_n \\ d_n & \text{else} \end{cases}$$

where the sequence $d_n > 0$ for all n with $d_n = O\left(n^{-1/2+\nu}\right)$ for some $0 < \nu < 1/2$. The truncation numbers d_n are used to avoid "too large" values for $\hat{\alpha}_k^{-1}(\hat{\phi})$. Truncation was introduced by Bickel [7] in the context of score estimation. More closely related to our context is Hidalgo's [52] semiparametric frequency domain estimator.

Next, an estimate for $b_k = (2\pi)^{-1} \int_{-\pi}^{\pi} \dot{\eta}(\phi, \lambda) e^{i\lambda k} d\lambda$ is needed. The vector b_k contains the impulse response function of the $AR(p)$ model evaluated at different points. Here we want to express b_k directly as a function of the underlying AR-parameters. From the definition of $\dot{\eta}(\phi, \lambda)$ in (8.4) and the expansion $\phi^{-1}(z) = \sum \psi_{\phi,j} z^j$ with $\psi_{\phi,j} = 0$ for $j < 0$, b_k can be written as

$$b_k = \begin{bmatrix} \psi_{\phi,k-1} \\ \vdots \\ \psi_{\phi,k-p} \end{bmatrix}$$

where the coefficients $\psi_{\phi,j}$ satisfy the recursion $\psi_{\phi,s} - \phi_1\psi_{\phi,s-1} - \cdots - \phi_p\psi_{\phi,s-p} = 0$ for all $s > 0$ and $\psi_{\phi,0} = 1$ (see Kreiss [60]). Let ψ_ϕ^p denote the vector of the first p coefficients of the polynomial expansion of $\phi^{-1}(z)$. This vector is the solution to

$$\begin{bmatrix} 1 & & & 0 \\ -\phi_1 & 1 & & \\ -\phi_2 & -\phi_1 & 1 & \\ . & . & . & . \end{bmatrix} \begin{bmatrix} \psi_{\phi,0} \\ \psi_{\phi,1} \\ \vdots \\ \psi_{\phi,p} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9.2}$$

which is denoted by $\psi_\phi^p = \Phi^{-1}e_1$ where e_1 is the first unit vector and Φ is the matrix defined in (9.2). Then, let b_{p+1} denote the vector of coefficients $\psi_{\phi,2}$ to $\psi_{\phi,p+1}$. Using a $p \times p+1$ selector matrix S_1 picking the last p elements from a $p+1 \times 1$ vector we have

$$b_{p+1} = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & & & 1 \\ \phi_1 & . & . & \phi_p \end{bmatrix} S_1\psi_\phi^p = T_\phi S_1 \Phi^{-1} e_1$$

where we define $T_\phi = \phi_1$ if $p = 1$. In a similar way we obtain $b_{p+s} = T_\phi^s S_1 \Phi^{-1} e_1$. The vectors $b_1 \ldots b_k$ can now be expressed as functions of the underlying parameters by

$$b_k = \left[S_{p,k} T_\phi^{(\max[0,k-p])} S_1 \Phi^{-1} e_1 \right]$$

where the convention $T_\phi^0 = I_p$ is assumed and the selector matrix $S_{p,k}$ is defined by

$$[S_{p,k}]_{i,j} = \{i = j\}\{j \leq k\}$$

with the indicator function $\{.\} = 1$ if the expression inside the bracket is true. b_k is continuous in the underlying parameters for all finite k and can therefore be consistently estimated from a consistent estimate $\hat{\phi}$.

A nonparametric estimate of $h(\phi, \lambda)$ is now defined as

$$\hat{l}_\psi(\lambda) = \sum_{j=1}^{n-p-1} \hat{\alpha}_j^{-1} \hat{b}_j e^{-i\lambda j}$$

and

$$\hat{h}_n\left(\hat{\phi}, \lambda\right) = \operatorname{Re}\left[\hat{l}_\psi(\lambda)\,\hat{\phi}\left(e^{-i\lambda}\right)\right]. \tag{9.3}$$

No additional kernel smoothing is needed. The reason is, that $h(\phi, \lambda)$ is already a convolution between a bounded sequence and a twice continuously differentiable function. In fact, the $\hat{b}_k$ implicitly contain a bandwidth since for every ϕ inside the stationary region they will decay to zero quickly.

Nevertheless, the implied bandwidth might not be optimal in finite samples such that one might want to introduce an additional lag window $k(\frac{j}{M})$ with bandwidth parameter M. $k(\frac{j}{M})$ is the j-th Fourier coefficient of a spectral Window $K(\lambda)$ satisfying $\int |K(\lambda)|\, d\lambda < \infty$, $\int K(\lambda)\, d\lambda = 1$. Robinson [83] derives cross validation methods for spectral density estimates where a goodness of fit measure for the density is minimized with respect to the bandwidth parameter. As pointed out by Newey [74], however, such a procedure does not necessarily lead to improved finite sample properties of the semiparametric estimator which should be the final objective.

We will also need the following matrix $\hat{h}_n^x(\hat{\phi}, \lambda)$, whose elements are continuous functions of $\hat{h}_n(\hat{\phi}, \lambda)$ and which is defined by

$$\hat{h}_n^x\left(\hat{\phi}, \lambda\right) = \mathrm{Re}\left[\hat{l}_\psi(\lambda)\,\hat{\phi}\left(e^{-i\lambda}\right) a(\lambda)\right].$$

The success of a semiparametric estimator depends on the ability to uniformly estimate the weights α_j^{-1}. Additional assumptions about the moments of the driving error process are needed to assure this. Since $\hat{\alpha}_j$ depends on fourth moments such conditions necessarily involve higher than fourth moments. Here we prove uniform convergence by a mean square argument which necessitates summability assumptions on eighth moments. The following assumption is sufficient to prove the main result.

Assumption G1. *Let $c_{\varepsilon\ldots\varepsilon}(t_1, \ldots, t_7)$ be the eighth order cumulant of the error process ε_t. Then*

$$\sum_{t_1} \cdots \sum_{t_7} |1 + |t_j||\, |c_{\varepsilon\ldots\varepsilon}(t_1, \ldots, t_7)| < \infty, \text{ for all } j = 1, ..., 7$$

Assumption (G1) implies that higher order cumulant spectra of order eight exist. This assumption enables us to state the following result.

Proposition 9.1. *Let $\hat{h}_n(\hat{\phi}_n, \lambda)$ be as defined in (9.3), let Assumptions (F1, G1) hold and assume that $\hat{\phi}_n \to \phi_0$ in probability or almost surely. Then*

$$\sup_{\lambda \in [-\pi, \pi]} \left\| \hat{h}_n\left(\hat{\phi}_n, \lambda\right) - h(\phi_0, \lambda) \right\| = o_p(1)$$

as $n \to \infty$. Also $P\left(\rho\left(\hat{h}_n(\hat{\phi}_n, \lambda), h(\phi_0, \lambda)\right) > \delta\right) \to 0$ for any $\delta > 0$ as $n \to \infty$ and $P\left(\hat{h}_n\left(\hat{\phi}_n, \lambda\right) \in \mathcal{H}\right) \to 1$ as $n \to \infty$.

Proof. See Appendix C ■

We proceed to define the semiparametric estimator $\bar{\phi}_n(\hat{h}_n)$ by replacing $h_0 = h(\phi_0, \lambda)$ with a nonparametric estimate (9.3). We will establish that

$$\sqrt{n}\left(\bar{\phi}_n(\hat{h}_n) - \bar{\phi}_n(h_0)\right) = o_p(1). \tag{9.4}$$

By applying Lemma (A.2) it can be shown that for $h \in \mathcal{H}$

$$I_{n,yy}(\lambda) h(\phi, \lambda) = I_{n,yy}(\lambda) \operatorname{Re}\left[l_\psi(\lambda) \phi\left(e^{-i\lambda}\right) a(\lambda)\right] \phi \tag{9.5}$$

$$+ I_{n,\varepsilon\varepsilon}(\lambda) h_{\phi_0}(\phi, \lambda) + h(\phi, \lambda) R_n(\lambda) \tag{9.6}$$

where the remainder term $R_n(\lambda)$ is such that $\sqrt{n} \int R_n(\lambda) \varsigma(\lambda) d\lambda = o_p(1)$ for any continuous function $\varsigma(\lambda)$ with absolutely summable Fourier coefficients. Let

$$h_{\phi_0}(\phi, \lambda) = \operatorname{Re}\left[l_\psi(-\lambda) \phi\left(e^{i\lambda}\right) \phi_0^{-1}\left(e^{i\lambda}\right)\right]$$

such that $h_{\phi_0}(\phi_0, \lambda) = \operatorname{Re}[l_\psi(-\lambda)] = l_\eta(\lambda)$ and

$$\hat{h}_{\phi_0}\left(\hat{\phi}, \lambda\right) = \operatorname{Re}\left[\hat{l}_\psi(-\lambda) \hat{\phi}\left(e^{i\lambda}\right) \phi_0^{-1}\left(e^{i\lambda}\right)\right].$$

(9.4) then follows if

$$\left\|\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \left(\hat{h}_n^x\left(\hat{\phi}_n, \lambda\right) - h^x(\phi_0, \lambda)\right) d\lambda\right\| = o_p(1) \tag{9.7}$$

and

$$\sqrt{n}\left\|\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(\hat{h}_{\phi_0}\left(\hat{\phi}_n,\lambda\right)-l_\eta(\lambda)\right)d\lambda\right\| = o_p(1) \tag{9.8}$$

$$\sqrt{n}\left\|\int_{-\pi}^{\pi} R_n(\lambda)\left(\hat{h}_n\left(\hat{\phi}_n,\lambda\right)-h(\phi_0,\lambda)\right)d\lambda\right\| = o_p(1). \tag{9.9}$$

(9.7) can be established easily with the help of proposition (9.1) by the following argument

$$\begin{aligned}
&\left\|\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\left(\hat{h}_n^x\left(\hat{\phi}_n,\lambda\right)-h^x(\phi_0,\lambda)\right)d\lambda\right\| \\
\leq\ & \sup_{\lambda\in[-\pi,\pi]}\left\|\hat{h}_n^x\left(\hat{\phi}_n,\lambda\right)-h^x(\phi_0,\lambda)\right\|\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\,d\lambda \\
\leq\ & 2\sup_{\lambda\in[-\pi,\pi]}\left\|\hat{l}_\psi(-\lambda)\hat{\phi}\left(e^{i\lambda}\right)-l_{\psi,0}(-\lambda)\phi_0\left(e^{i\lambda}\right)\right\|\sup_{\lambda\in[-\pi,\pi]}\|a(\lambda)\|\,\hat{\gamma}_{yy}(0)\to 0
\end{aligned}$$

where the first inequality uses the fact, that $I_{n,yy}(\lambda)$ is a positive scalar and the second inequality uses $tr(ab'ba') = (a'a)(b'b)$ where a and b are two conformable vectors. The last expression goes to zero by (9.1) and the fact that $\sup_{\lambda\in[-\pi,\pi]}\|a(\lambda)\|$ is bounded. To prove (9.8) we work with the metric space $(\mathcal{H},\rho)$ defined in (9.1). Also let $h_0 = h(\phi_0,\lambda)$, $\hat{h} = \hat{h}_n(\hat{\phi}_n,\lambda)$, $\phi^\lambda = \phi\left(e^{i\lambda}\right)$, $\phi_0^\lambda = \phi_0\left(e^{i\lambda}\right)$ and

$$v_n(h) = \sqrt{n}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(h_{\phi_0}(\phi,\lambda)-l_{\eta,0}\right)+R_n(\lambda)(h-h_0)\,d\lambda$$

for $h\in\mathcal{H}$. Following Andrews [2], (9.8) follows if for any given $\vartheta,\epsilon>0$ there exists a $\delta>0$ such that

$$\limsup_{n\to\infty} P\left(\left\|v_n\left(\hat{h}_n\right)-v_n(h_0)\right\|>\vartheta\right)$$

$$\leq \limsup_{n\to\infty} P\left(\left\|v_n\left(\hat{h}_n\right) - v_n\left(h_0\right)\right\| > \vartheta, \hat{h}_n \in \mathcal{H}, \rho\left(\hat{h}_n, h_0\right) < \delta\right)$$
$$+ \limsup_{n\to\infty} P\left(\hat{h}_n \notin \mathcal{H} \text{ or } \rho\left(\hat{h}_n, h_0\right) > \delta\right)$$
$$\leq \limsup_{n\to\infty} P\left(\sup_{h\in\mathcal{H}, \rho\left(\hat{h}_n, h_0\right)<\delta} \left\|v_n\left(h\right) - v_n\left(h_0\right)\right\| > \vartheta\right) \leq \varepsilon$$

since we have established in proposition (9.1) that

$$\limsup_{n\to\infty} P\left(\hat{h}_n \notin \mathcal{H} \text{ or } \rho\left(\hat{h}_n, h_0\right) > \delta\right) = 0.$$

Therefore if

$$\limsup_{n\to\infty} P\left(\sup_{h\in\mathcal{H}, \rho\left(\hat{h}_n, h_0\right)<\delta} \left\|v_n\left(h\right) - v_n\left(h_0\right)\right\| > \vartheta\right) \leq \varepsilon \tag{9.10}$$

then the following theorem can be established.

Theorem 9.2. *Let $\hat{h}_n(\hat{\phi}_n, \lambda)$ as defined in (9.3). Let assumption (F1) hold and let $\hat{\phi}_n$ be a previous estimator for which $\hat{\phi}_n \to \phi_0$ in probability or almost surely. Then, the semiparametric estimator $\tilde{\phi}(\hat{h}_n)$ defined by*

$$\tilde{\phi}\left(\hat{h}_n\right) = \left[\int_{-\pi}^{\pi} I_{n,yy}\left(\lambda\right) \hat{h}_n^{x}(\hat{\phi}_n, \lambda) d\lambda\right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}\left(\lambda\right) \hat{h}_n(\hat{\phi}_n, \lambda) d\lambda$$

has a limiting distribution characterized by

$$\sqrt{n}\left(\tilde{\phi}(\hat{h}_n) - \phi_0\right) \Rightarrow N\left(0, \sigma^{-4}\Xi\right)$$

Proof. See Appendix C ■.

This result establishes the feasibility of a semiparametric estimator that improves on the efficiency of the conventional Gaussian estimator in the presence of higher order dependence. While time domain versions of this estimator could certainly be obtained, the frequency domain version developed here seems most natural in the present framework. The frequency domain representation allows to avoid estimating the instruments for each observation in the sample. Instead an optimal filter applied to the periodogram of the data leads to an asymptotically equivalent procedure. Moreover, the fact that the optimal filter itself is a convolution integral in the frequency domain solves the problem of truncating the approximation of the optimal instrument at a given lag in a natural and elegant way.

Part III

Simulation Results

10. Monte Carlo Simulations

In this section a small Monte Carlo experiment is conducted. To keep the exposition as simple as possible we focus on an $AR(1)$ model. We consider what the efficiency gains/losses of the IV estimator are relative to a correctly specified likelihood procedure and relative to a misspecified ML estimator. We also consider estimation of covariance matrices for the IV estimator.

10.1. Relative Efficiency

The following questions are of interest: Under what circumstances does the optimal IV estimator achieve efficiency gains, how big are they relative to the Gaussian QMLE and how much is lost by not specifying the true likelihood. These questions are analyzed for the case where the true generating mechanism is an $ARCH(1)$ process.

We generate samples of size $n = 256$, $n = 512$ and $n = 1024$ from the following model

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \qquad (10.1) \\ \varepsilon_t &= u_t h_t^{1/2} \\ h_t &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 \\ & \quad u_t \tilde{} N(0,1). \end{aligned}$$

Starting values are $y_0 = 0$ and $\varepsilon_0 = 0$. Small sample properties of three different estimators to be defined below are evaluated for different values of ϕ, $\gamma_1 \in [0,1)$. It is clear from Milhoj [68] that asymptotic normality established in previous chapters only obtains for values of $\gamma_1 \in [0, \sqrt{1/3})$. Nevertheless, simulation results are reported for parametrizations outside this interval in order to analyze the robustness of the proposed IV procedure to departures from the assumptions. The parameter γ_0 is fixed at .1 for all experiments.

The parameter ϕ is estimated by three different estimators. The least squares estimator is denoted by $\hat{\phi}_n^{OLS} = \sum_{t=2}^n y_t y_{t-1} / \sum_{t=2}^n y_{t-1}^2$. The optimal instrumental variables estimator is obtained from the consistent first stage estimator $\hat{\phi}_{OLS}$ as

$$\hat{\phi}_n^{IV} = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \hat{h}^x \left(\hat{\phi}_{OLS}, \lambda \right) d\lambda \right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \hat{h}(\hat{\phi}_{OLS}, \lambda) d\lambda$$

where $\hat{h}^x(\phi, \lambda)$ and $\hat{h}(\phi, \lambda)$ are computed as explained in Section 9. If the data are generated by (10.1), the likelihood estimator $\hat{\phi}_n^{ML}$ is obtained from maximizing

$$l(\phi, \gamma_0, \gamma_1; Y) = -\frac{1}{2} \sum_{t=3}^{n} \left(\ln h_t + \frac{\varepsilon_t^2}{h_t} \right) \tag{10.2}$$

with $\varepsilon_t = y_t - \phi y_{t-1}$ and $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2$. We use the $BHHH$ algorithm described in Engle [26] to maximize the likelihood. Figure 10.1 shows the potential efficiency gains of the IV estimator relative to the Gaussian QMLE as a function of the autoregressive parameter ϕ. The efficiency gains are computed from the asymptotic covariance matrix when the generating mechanism is (10.1). More explicitly, the asymptotic covariance

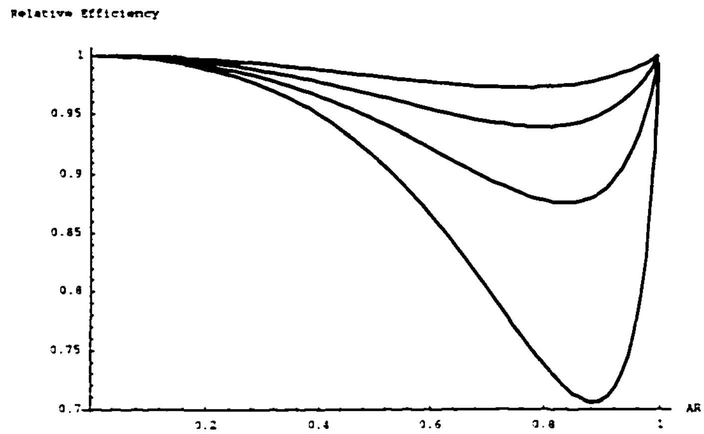

Figure 10.1: *Asymptotic efficiency of OLS relative to the IV estimator as a function of the parameter ϕ. Generating mechanisms considered are from bottom to top: $\gamma_1 = .5$, $\gamma_1 = .4$, $\gamma_1 = .3$ and $\gamma_1 = .2$.*

matrix of $\hat{\phi}_n^{OLS}$ can be expressed as

$$\sigma^2_{OLS}(\phi, \gamma_0, \gamma_1) = \frac{(1-\phi)^2}{\sigma^4} \sum_{i=0}^{\infty} \phi^{2i} a_{i+1} \tag{10.3}$$

where $\sigma^4 = (\gamma_0/1 - \gamma_1)^2$ and $a_{i+1} = 2\gamma_0^2\gamma_1^{i+1}/[(1-\gamma_1)^2(1-3\gamma_1^2)] + \sigma^4$. The asymptotic covariance matrix for the optimal IV estimator can be obtained from (4.2). It is given by

$$\sigma^2_{IV}(\phi, \gamma_0, \gamma_1) = \left[\sigma^4 \sum_{i=0}^{\infty} \phi^{2i} a_{i+1}^{-1}\right]^{-1}. \tag{10.4}$$

Figure 10.1 plots $\sigma^2_{IV}(\phi, .1, \gamma_1)/\sigma^2_{OLS}(\phi, .1, \gamma_1)$ for $\phi \in [0, 1)$ and different values of γ_1. These theoretical gains are contrasted to the empirical efficiency of the estimators $\hat{\phi}_n^{OLS}$, $\hat{\phi}_n^{IV}$ and $\hat{\phi}_n^{ML}$ based on 3000 replications for sample sizes 256, 512 and 1024. The results are summarized in Table 10.1.

As expected, gains for the *IV* estimator are achieved for models where the autoregres-

Table 10.1: Relative efficiency of OLS for ARCH(1) innovations

Model:		$y_t = \phi y_{t-1} + \varepsilon_t$		$\varepsilon_t = u_t h_t^{1/2}$		$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$	
		$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$
ϕ	γ_1	$n = 256$		$n = 512$		$n = 1024$	
0.0	0.0	0.9969	1.0294	0.9966	1.0126	0.9980	1.0107
		(0.0257)	(0.0266)	(0.0257)	(0.0261)	(0.0258)	(0.0261)
0.0	0.5	0.9918	0.5158	0.9947	0.4144	0.9984	0.3801
		(0.0534)	(0.0278)	(0.0535)	(0.0223)	(0.0258)	(0.0098)
0.0	0.9	0.9502	0.4432	0.9619	0.4458	0.9631	0.3978
		(0.0272)	(0.0127)	(0.0276)	(0.0128)	(0.0276)	(0.0114)
0.5	0.0	0.9937	1.0272	0.9985	1.0087	0.9971	1.0106
		(0.0268)	(0.0277)	(0.0270)	(0.0272)	(0.0269)	(0.0273)
0.5	0.5	0.9388	0.4761	0.9375	0.4515	0.9616	0.3658
		(0.0469)	(0.0238)	(0.0469)	(0.0226)	(0.0248)	(0.0094)
0.5	0.9	0.9361	0.4083	0.9171	0.4478	0.9106	0.4759
		(0.0262)	(0.0114)	(0.0257)	(0.0125)	(0.0255)	(0.0133)
0.7	0.0	1.0046	1.0190	1.0046	1.0096	1.0012	1.0025
		(0.0259)	(0.0263)	(0.0259)	(0.0261)	(0.0259)	(0.0259)
0.7	0.5	0.9071	0.5084	0.8958	0.4416	0.8590	0.3715
		(0.0436)	(0.0244)	(0.0431)	(0.0212)	(0.0222)	(0.0096)
0.7	0.9	5.5958	0.3998	0.8274	0.4470	0.8274	0.4470
		(0.1445)	(0.0103)	(0.0214)	(0.0115)	(0.0214)	(0.0115)
0.9	0.0	1.0391	1.0097	1.0152	1.0101	1.0085	1.0022
		(0.0272)	(0.0264)	(0.0265)	(0.0264)	(0.0264)	(0.0262)
0.9	0.5	0.9209	0.5420	0.8522	0.4570	0.8308	0.4663
		(0.0430)	(0.0253)	(0.0398)	(0.0213)	(0.0215)	(0.0120)
0.9	0.9	0.7281	0.4352	0.5951	0.3486	0.4878	0.4247
		(0.0207)	(0.0124)	(0.0169)	(0.0099)	(0.0126)	(0.0110)

Relative Efficiency is defined as S_{IV}^2/S_{OLS}^2 or S_{ML}^2/S_{OLS}^2, respectively, where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parenthesis are asymptotic standard deviations of the variance ratio. Results are based on 3000 replications.

sive parameter is above .5. This conforms with the theoretical analysis based on asymptotic approximations. For the sample sizes considered here, the theoretical efficiency gains are not achieved completely. The table shows that the relative efficiency of the *IV* estimator improves with the sample size. The most significant increase takes place from size 256 to 512. It is also interesting to note that the *IV* procedure maintains its properties even for values of $\gamma_1 > \sqrt{1/3}$. In fact the gains are strongest when both autocorrelation and dependence in the conditional variance are strong. The reason is that in this case the dependence between the regressors and the errors is largest.

Figure 10.2 shows the empirical densities of the three estimators $\hat{\phi}_n^{OLS}$, $\hat{\phi}_n^{IV}$ and $\hat{\phi}_n^{ML}$ when no ARCH effects are present. The graph confirms the information summarized in the tables: The three estimators are identical under *iid* conditions. Figure 10.3 shows the empirical distributions of $\hat{\phi}_n^{OLS}$, $\hat{\phi}_n^{IV}$ and $\hat{\phi}_n^{ML}$ for a sample size of 1024 when $\phi = .9$ and $\gamma_1 = .9$.

Here $\hat{\phi}_n^{ML}$ clearly dominates the two other estimators in terms of efficiency and mean and median unbiasedness. The *IV* estimator has surprisingly good properties even though the asymptotic theory used for its construction does not hold for this set of parameter values.

Table 10.2 contains the means and medians for $\hat{\phi}_n^{OLS}$, $\hat{\phi}_n^{IV}$ and $\hat{\phi}_n^{ML}$ when $n = 512$ based on 3000 replications. The bias tends to be largest for the IV estimator, but the difference between $\hat{\phi}_n^{OLS}$ and $\hat{\phi}_n^{IV}$ is smaller than the difference of the former with $\hat{\phi}_n^{ML}$. The bias for $\hat{\phi}_n^{OLS}$ and $\hat{\phi}_n^{IV}$ increases with ϕ. For a fixed ϕ, it is largest when $\gamma_1 = .5$. The bias of the *ML* estimator, on the other hand, is little affected by the parametrization of

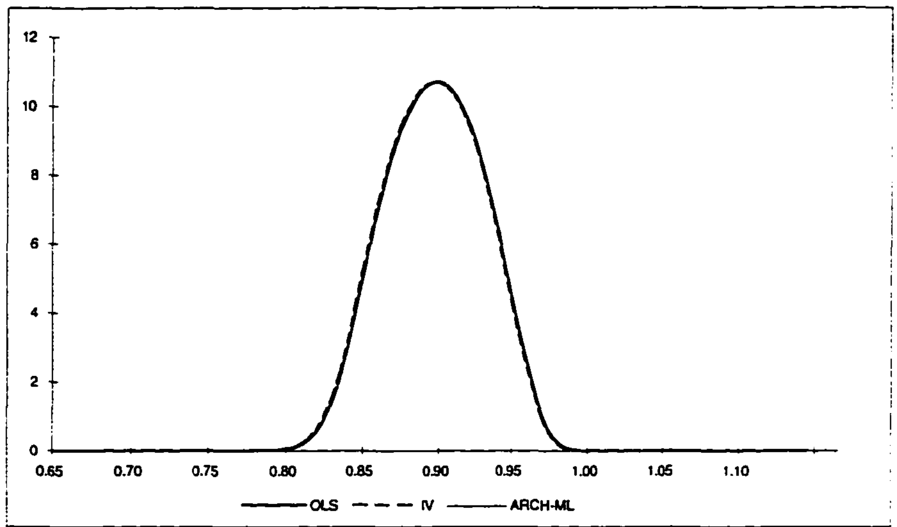

Figure 10.2: *Empirical density of parameter estimates for an AR(1) model with* $\phi = .9$ *when the errors have no ARCH effects*

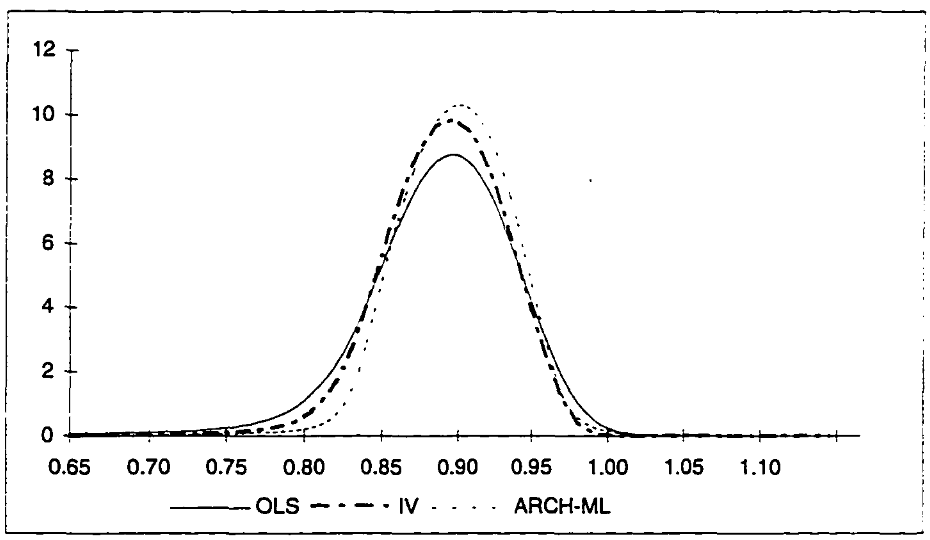

Figure 10.3: *Empirical densities of estimated AR parameters when* $\phi = .9$ *and* $\gamma_1 = .9$

Table 10.2: Means and Medians

Model:		$y_t = \phi y_{t-1} + \varepsilon_t$		$\varepsilon_t = u_t h_t^{1/2}$		$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$	
		$\hat{\phi}_n^{OLS}$		$\hat{\phi}_n^{IV}$		$\hat{\phi}_n^{ML}$	
ϕ	γ_1	Mean	Median	Mean	Median	Mean	Median
0.0	0	-0.0007	-0.0006	-0.0006	-0.0005	-0.0007	-0.0007
0.0	0.5	-0.0025	-0.0014	-0.0024	-0.0016	-0.0009	-0.0009
0.0	0.9	0.0006	0.0000	0.0009	-0.0002	0.0011	0.0009
0.5	0	0.4978	0.4994	0.4969	0.4980	0.4977	0.4992
0.5	0.5	0.4927	0.4935	0.4908	0.4919	0.4987	0.4995
0.5	0.9	0.4770	0.4830	0.4694	0.4780	0.4945	0.4984
0.7	0	0.6962	0.6969	0.6948	0.6957	0.6962	0.6970
0.7	0.5	0.6933	0.6950	0.6913	0.6933	0.6979	0.7000
0.7	0.9	0.6779	0.6888	0.6733	0.6847	0.6945	0.6992
0.9	0	0.8965	0.8981	0.8947	0.8962	0.8965	0.8982
0.9	0.5	0.8948	0.8977	0.8933	0.8957	0.8982	0.8992
0.9	0.9	0.8850	0.8925	0.8860	0.8921	0.8984	0.8999

Sample size is 512. Results are based on 3000 replications.

the model.

10.2. Covariance Matrix Estimation

Estimation of covariance matrices for conditionally heteroskedastic errors is considered in the literature by White[95], Newey and West [76], Andrews[3] and Andrews and Monahan[4].

In our case, the covariance of $\hat{\phi}_n^{OLS}$ is given by

$$var\left(\sqrt{n}(\hat{\phi}_n^{OLS} - \phi_0\right) = \left(\frac{1}{n}\sum_{t=1}^{n} y_t^2\right)^{-2} \left(\frac{1}{n}\sum_{t=2}^{n}\sum_{s=2}^{n} E\varepsilon_t y_{t-1}\varepsilon_s y_{s-1}\right) \tag{10.5}$$

Estimation of $\frac{1}{n}\sum_{t=2}^{n}\sum_{s=2}^{n} E\varepsilon_t y_{t-1}\varepsilon_s y_{s-1}$ is carried out in White by

$$\hat{V}_W = \sum_{t=2}^{n} \hat{\varepsilon}_t^2 y_{t-1}^2 \tag{10.6}$$

Table 10.3: Standard Deviations, Mean Absolute Errors and Mean Squared Errors

Model:		$y_t = \phi y_{t-1} + \varepsilon_t$			$\varepsilon_t = u_t h_t^{1/2}$			$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$		
			$\hat{\phi}_n^{OLS}$			$\hat{\phi}_n^{IV}$			$\hat{\phi}_n^{ML}$	
ϕ	γ_1	SDV	MAE	MSE	SDV	MAE	MSE	SDV	MAE	MSE
0.0	0	0.0442	0.0351	0.0020	0.0441	0.0350	0.0019	0.0445	0.0353	0.0020
0.0	0.5	0.0759	0.0591	0.0058	0.0757	0.0589	0.0057	0.0489	0.0391	0.0024
0.0	0.9	0.1429	0.1070	0.0204	0.1402	0.1045	0.0196	0.0954	0.0605	0.0091
0.5	0	0.0384	0.0308	0.0015	0.0384	0.0308	0.0015	0.0386	0.0309	0.0015
0.5	0.5	0.0609	0.0485	0.0038	0.0590	0.0473	0.0036	0.0409	0.0324	0.0017
0.5	0.9	0.1213	0.0893	0.0152	0.1162	0.0855	0.0144	0.0812	0.0499	0.0066
0.7	0	0.0321	0.0257	0.0010	0.0321	0.0259	0.0011	0.0322	0.0259	0.0011
0.7	0.5	0.0473	0.0371	0.0023	0.0448	0.0356	0.0021	0.0314	0.0248	0.0010
0.7	0.9	0.0958	0.0687	0.0097	0.0871	0.0616	0.0083	0.0640	0.0366	0.0041
0.9	0	0.0198	0.0158	0.0004	0.0199	0.0161	0.0004	0.0199	0.0158	0.0004
0.9	0.5	0.0259	0.0201	0.0007	0.0239	0.0192	0.0006	0.0175	0.0136	0.0003
0.9	0.9	0.0487	0.0342	0.0026	0.0376	0.0264	0.0016	0.0287	0.0176	0.0008

Sample size is 512. Results are based on 3000 replications.

Table 10.4: Coverage Probabilities 90%

Model:		$y_t = \phi y_{t-1} + \varepsilon_t$		$\varepsilon_t = u_t h_t^{1/2}$		$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$		
		W	NW	NW	NW	OLS	P-OLS	P-IV
ϕ	γ_1		$n^{1/5}$	$n^{1/4}$	$n^{1/2}$			
0.0	0	0.8997	0.8897	0.8880	0.8763	0.9007	0.9590	0.9590
0.0	0.5	0.8890	0.8780	0.8757	0.8510	0.6790	0.8663	0.8660
0.0	0.9	0.8677	0.8280	0.8177	0.7517	0.4503	0.7773	0.7760
0.5	0	0.9010	0.8883	0.8887	0.8703	0.9043	0.9587	0.9587
0.5	0.5	0.8903	0.8827	0.8800	0.8540	0.7147	0.8827	0.8790
0.5	0.9	0.8710	0.8293	0.8173	0.7540	0.4793	0.7953	0.7840
0.7	0	0.8997	0.8907	0.8900	0.8707	0.9003	0.9560	0.9560
0.7	0.5	0.8850	0.8770	0.8707	0.8400	0.7460	0.8933	0.8890
0.7	0.9	0.8677	0.8420	0.8247	0.7527	0.5140	0.8157	0.7840
0.9	0	0.9030	0.8980	0.8957	0.8693	0.9040	0.9583	0.9583
0.9	0.5	0.8900	0.8863	0.8830	0.8463	0.8117	0.9243	0.9167
0.9	0.9	0.8790	0.8797	0.8710	0.7830	0.6190	0.8690	0.8163

Sample size is 512. Results are based on 3000 replications.

Table 10.5: Coverage Probabilities 95%

Model:	$y_t = \phi y_{t-1} + \varepsilon_t$		$\varepsilon_t = u_t h_t^{1/2}$		$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$			
		White	NW	NW	NW	OLS	P-OLS	P-IV
ϕ	γ_1		$n^{1/5}$	$n^{1/4}$	$n^{1/2}$			
0.0	0	0.9507	0.9473	0.9460	0.9340	0.9530	0.9803	0.9803
0.0	0.5	0.9400	0.9380	0.9357	0.9170	0.7630	0.9210	0.9210
0.0	0.9	0.9227	0.8927	0.8843	0.8323	0.5230	0.8483	0.8447
0.5	0	0.9500	0.9457	0.9457	0.9313	0.9513	0.9833	0.9833
0.5	0.5	0.9373	0.9343	0.9303	0.9147	0.7960	0.9353	0.9303
0.5	0.9	0.9293	0.8990	0.8873	0.8203	0.5493	0.8660	0.8543
0.7	0	0.9463	0.9413	0.9390	0.9253	0.9517	0.9830	0.9830
0.7	0.5	0.9390	0.9357	0.9323	0.9087	0.8247	0.9477	0.9460
0.7	0.9	0.9260	0.9180	0.9043	0.8377	0.5823	0.8863	0.8550
0.9	0	0.9510	0.9483	0.9450	0.9220	0.9517	0.9857	0.9857
0.9	0.5	0.9440	0.9380	0.9363	0.9137	0.8803	0.9627	0.9563
0.9	0.9	0.9407	0.9450	0.9413	0.8623	0.7007	0.9253	0.8850

Sample size is 512. Results are based on 3000 replications.

where $\hat{\varepsilon}_t = y_t - \hat{\phi}_n^{OLS} y_{t-1}$. This estimator accounts for conditional heteroskedasticity but not for autocorrelation in the errors. However, $\hat{V}_W$ is consistent under the martingale difference assumption for ε_t. Newey and West[76] account for autocorrelation by using

$$\hat{V}_{NW} = \Gamma_0 + 2\sum_{j=1}^{m} k\left(j/m\right)\Gamma_j,\ \Gamma_j = \frac{1}{n}\sum_{t=j+2}^{n} \hat{\varepsilon}_t y_{t-1}\hat{\varepsilon}_{t-j} y_{t-j-1} \tag{10.7}$$

where $k\left(j/m\right)$ is the Bartlett kernel

$$k\left(x\right) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Andrews[3] obtains optimal choices for the bandwidth parameter m. The optimal bandwidth is zero in the case of no autocorrelation in the errors. In this case, $\hat{V}_{NW} = \Gamma_0$, which yields $\hat{V}_{NW} = \hat{V}_W$. We examine whether this choice is optimal in our case by reporting

$\hat{V}_{NW}$ for a bandwidth $m = n^{1/2}, n^{1/4}, n^{1/5}$.

Instead of estimating (10.5) nonparametrically we can use the results from (10.3) to construct a parametric estimator of the covariance matrix based on consistent estimates of ϕ and a_i. In particular, we define

$$\hat{V}_{P-OLS} = \frac{\left(1-\hat{\phi}\right)^2}{\hat{\sigma}^4} \sum_{i=0}^{n-2} \hat{\phi}^{2i} \hat{a}_{i+1}, \tag{10.8}$$

where estimation of $\hat{a}_{i+1}$ is discussed in Section 9. In the same way, we also obtain a covariance estimator for the IV procedure as

$$\hat{V}_{P-IV} = \left[\hat{\sigma}^4 \sum_{i=0}^{n-2} \hat{\phi}^{2i} \hat{a}_{i+1}^{-1}\right]^{-1}. \tag{10.9}$$

Tables 10.4, 10.5 and 10.6, respectively, contain empirical levels of a t-test of the two sided hypothesis $\hat{\phi} = \phi_0$. These levels can also be interpreted as coverage probabilities of an interval constructed around the estimated parameter. The coverage probabilities represent the empirical frequency of the event $\left|\hat{\phi}_n - \phi_0\right| / \sqrt{var(\hat{\phi}_n)} > Z_{1-\alpha/2}$ for $\alpha = 0.1$, 0.05 and 0.01 respectively where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. The variance of the estimator is estimated by four different procedures when $\hat{\phi}_n = \hat{\phi}_n^{OLS}$, namely, $\hat{V}_W, \hat{V}_{NW}$ and $\hat{V}_{P-OLS}$. For comparative purposes, we also report $\hat{V}_{OLS} = \hat{\sigma}^2(\frac{1}{n}\sum_{t=2}^{n} y_{t-1}^2)^{-1}$, which is inconsistent. When $\hat{\phi}_n = \hat{\phi}_n^{IV}$ the covariance estimator is $\hat{V}_{P-IV}$.

The results in Tables 10.4, 10.5 and 10.6 confirm what is expected by theoretical arguments. $\hat{V}_{OLS}$ is roughly equivalent to $\hat{V}_W$ when there is no conditional heteroskedasticity,

Table 10.6: Coverage Probabilities 99%

Model:	$y_t = \phi y_{t-1} + \varepsilon_t$		$\varepsilon_t = u_t h_t^{1/2}$		$h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$			
		White	NW	NW	NW	OLS	P-OLS	P-IV
ϕ	γ_1		$n^{1/5}$	$n^{1/4}$	$n^{1/2}$			
0.0	0	0.9767	0.9760	0.9747	0.9663	0.9787	0.9933	0.9933
0.0	0.5	0.9787	0.9767	0.9750	0.9567	0.8380	0.9613	0.9610
0.0	0.9	0.9607	0.9497	0.9420	0.8967	0.5907	0.9033	0.9007
0.5	0	0.9787	0.9773	0.9757	0.9667	0.9807	0.9953	0.9953
0.5	0.5	0.9747	0.9703	0.9690	0.9540	0.8660	0.9730	0.9710
0.5	0.9	0.9617	0.9523	0.9420	0.8853	0.6257	0.9197	0.9093
0.7	0	0.9777	0.9760	0.9760	0.9597	0.9810	0.9967	0.9967
0.7	0.5	0.9757	0.9723	0.9713	0.9573	0.8897	0.9750	0.9730
0.7	0.9	0.9637	0.9670	0.9590	0.8993	0.6690	0.9317	0.9130
0.9	0	0.9823	0.9803	0.9767	0.9590	0.9850	0.9953	0.9953
0.9	0.5	0.9733	0.9737	0.9720	0.9573	0.9340	0.9837	0.9813
0.9	0.9	0.9787	0.9807	0.9790	0.9310	0.7700	0.9593	0.9347

Sample size is 512. Results are based on 3000 replications.

but deteriorates dramatically as γ_1 increases. Overall $\hat{V}_W$ has the most accurate coverage probabilities. The quality of $\hat{V}_{NW}$ deteriorates with increasing bandwidth. These two results confirm that $m = 0$ is the optimal bandwidth choice.

The performance of $\hat{V}_{P-OLS}$ is best when $\gamma_1 = .5$. However, even in this case, it is not as accurate as $\hat{V}_W$. For values of $\gamma_1 > \sqrt{1/3}$, the asymptotic approximation leading to $\hat{V}_{P-OLS}$ does not exist. This is reflected in biased coverage probabilities. Somewhat surprisingly, bias also exists for $\gamma_1 = 0$. The same characterizations also apply to $\hat{V}_{P-IV}$.

10.3. Relative Efficiency under Misspecification

In this section, we analyze the finite sample efficiency of the three estimators under alternative generating mechanisms for the conditional heteroskedasticity. We consider ARCH(2), GARCH(1,1) and stochastic volatility models. While the *IV* estimator is still valid under these circumstances, the *ARCH-ML* estimator now is misspecified. In particular the es-

Table 10.7: Relative efficiency of OLS with GARCH(1,1) innovations

Model: $y_t = \phi y_{t-1} + \varepsilon_t \varepsilon_t = u_t h_t^{1/2} h_t = 0.1 + 0.3\varepsilon_{t-1}^2 + 0.6h_{t-1}$				
	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$
ϕ	n=512		n=1024	
0.925	0.9501	1.1253	0.9368	1.1253
	(0.0470)	(0.0557)	(0.0242)	(0.0291)
0.950	0.9518	1.0040	0.9091	1.1370
	(0.0457)	(0.0482)	(0.0235)	(0.0294)
0.975	1.4346	0.8818	0.9650	1.0303
	(0.0668)	(0.0410)	(0.0249)	(0.0266)
0.990	1.0561	0.8979	0.9770	0.8689
	(0.0481)	(0.0409)	(0.0252)	(0.0224)

timated conditional variances h_t are inconsistent. This inconsistency is particularly severe for the Stochastic Volatility process where h_t is independent of the u_t sequence.

Table 10.7 summarizes the results for the case when the generating mechanism is

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \qquad (10.10) \\ \varepsilon_t &= u_t h_t^{1/2} \\ h_t &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1} \\ & u_t \backsim N(0,1). \end{aligned}$$

Starting values are $y_0 = 0$, $h_0 = 0$ and $\varepsilon_0 = 0$.We focus on values for $\gamma_1 = .3$ and $\gamma_2 = .6$. In general, the volatility process for the GARCH specification is much smoother than for the ARCH case. This is the reason why the potential gains for the IV procedure are smaller than in the ARCH case. Not surprisingly, the $ARCH\text{-}ML$ estimator (10.2) loses its efficiency properties. It is even less efficient than simple OLS when the autocorrelation coefficient is not too large, i.e., $\phi < .975$.

In Table 10.8 we report the results for the case when the true generating mechanism is an ARCH(2) process. The data are now generated by (10.11) where the parameters γ_1 and γ_2 are chosen to reflect moderate and strong conditional heteroskedasticity.

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \\ \varepsilon_t &= u_t h_t^{1/2} \\ h_t &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 \\ & \quad u_t \backsim N(0,1). \end{aligned} \tag{10.11}$$

Starting values are $y_0 = 0$ and $\varepsilon_{-1}, \varepsilon_0 = 0$. In particular, the parametrization $(\gamma_1, \gamma_2) = (.5, .4)$ implies infinite fourth moments for ε_t. In this case the asymptotic distribution of $\hat{\phi}_n^{OLS}$ is not normal and the formal derivation of the IV procedure is not applicable to this case. Nevertheless the simulation results show strong efficiency gains for the IV procedure for this case. The misspecified ARCH estimator is now less efficient than simple OLS in three out of four cases. The empirical densities for $n = 1024$ and $(\phi, \gamma_1, \gamma_2) = (.9, .5, .4)$ are shown in Figure 10.4.

Finally, Table 10.9 contains the simulation results when the true generating mechanism is the following stochastic volatility model

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \\ \varepsilon_t &= u_t \exp(h_t/2) \end{aligned} \tag{10.12}$$

Table 10.8: Relative efficiency of OLS with ARCH(2) innovations

Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t h_t^{1/2} \quad h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2$						
			$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$
ϕ	γ_1	γ_2	n=512		n=1024	
0.5	0.3	0.2	0.9897	1.0215	0.9781	1.0843
			(0.0486)	(0.0501)	(0.0253)	(0.0280)
0.5	0.5	0.4	0.9680	1.4400	0.9585	1.5088
			(0.0491)	(0.0731)	(0.0247)	(0.0390)
0.9	0.3	0.2	0.9632	0.8404	0.9366	0.8690
			(0.0462)	(0.0403)	(0.0242)	(0.0224)
0.9	0.5	0.4	0.8837	1.2000	0.7843	1.2963
			(0.0418)	(0.0568)	(0.0203)	(0.0335)

Relative Efficiency is defined as S_{IV}^2/S_{OLS}^2 or S_{ML}^2/S_{OLS}^2 respectively where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parenthesis are asymptotic standard deviations of the variance ratio. Results are based on 3000 replications.

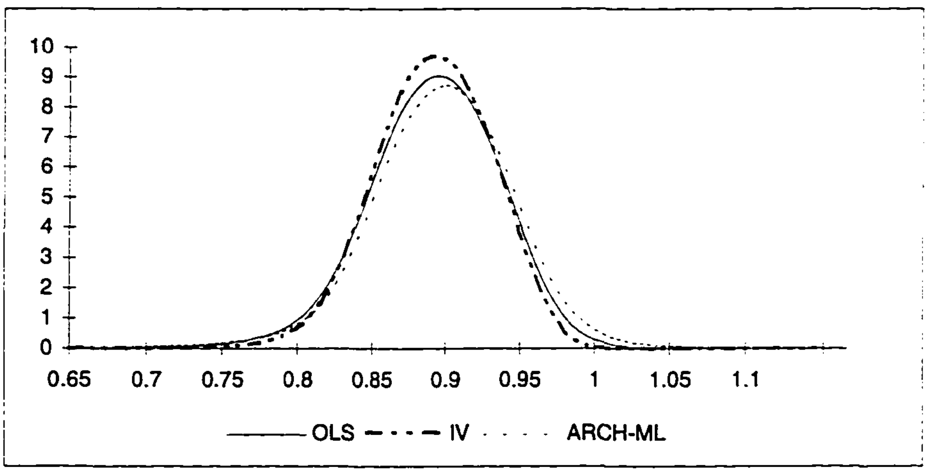

Figure 10.4: *Empirical densities for estimators of the autoregressive parameter when the errors are generated by an ARCH(2) model with $\phi = .9$, $\gamma_1 = .5$ and $\gamma_2 = .4$.*

Table 10.9: Relative efficiency of OLS with stochastic volatility innovations

Model: $y_t = \phi y_{t-1} + \varepsilon_t$	$\varepsilon_t = u_t \exp(h_t/2)$	$h_t = \gamma_1 h_{t-1} + v_t$			
		$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$	$\hat{\phi}_n^{IV}$	$\hat{\phi}_n^{ML}$
ϕ	γ_1	$n = 512$		$n = 1024$	
0.7	0.50	0.9501	1.4299	0.9461	1.4926
		(0.0357)	(0.0537)	(0.0244)	(0.0385)
0.9	0.50	0.9447	1.2109	0.9628	1.2125
		(0.0375)	(0.0480)	(0.0249)	(0.0313)
0.5	0.90	0.9557	7.2483	0.9557	7.2483
		(0.0247)	(0.1872)	(0.0247)	(0.1872)
0.7	0.90	0.8315	7.4975	0.8154	9.2688
		(0.0215)	(0.1936)	(0.0211)	(0.2393)
0.9	0.90	0.7062	5.8414	0.5970	8.6569
		(0.0182)	(0.1508)	(0.0154)	(0.2235)
0.5	0.95	0.9475	8.9600	0.9475	8.9600
		(0.0245)	(0.2313)	(0.0245)	(0.2313)
0.7	0.95	0.9019	8.9705	0.9146	8.9516
		(0.0233)	(0.2316)	(0.0236)	(0.2311)
0.9	0.95	1.1910	7.3298	0.7370	9.9652
		(0.0308)	(0.1893)	(0.0190)	(0.2573)

Relative Efficiency is defined as S^2_{IV}/S^2_{OLS} or S^2_{ML}/S^2_{OLS} respectively where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parenthesis are asymptotic standard deviations of the variance ratio. Results are based on 3000 replications.

$$h_t = \gamma_1 h_{t-1} + v_t$$

$$u_t \backsim N(0,1), v_t \backsim N(0,1).$$

Starting values are $y_0 = 0$ and $h_0 = 0$. The misspecified ARCH estimator now is less efficient than OLS in all cases considered. When the dependence in conditional variances as measured by γ_1 is strong, the inefficiency of the ARCH estimator is extremely large with an eight to nine fold increase in the variance relative to OLS. The IV estimator on the other hand has properties similar to the ARCH and GARCH cases. The breakdown of the ARCH estimator is documented in Figure 10.5.

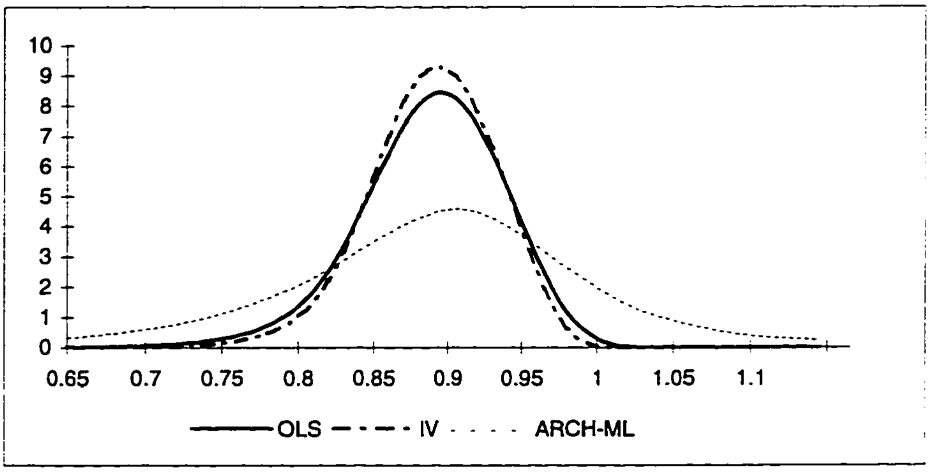

Figure 10.5: *Empirical densities for estimators of the autoregressive parameter when the errors are generated by a Stochastic Volatility model with* $\phi = .9$, $\gamma_1 = .9$.

11. Conclusions

This dissertation analyzes the consequences of higher moment dependence between the errors and the regressors in the context of the univariate linear time series model. Prominent parametric examples of such models are mainly encountered in financial econometrics. They include ARCH, GARCH and stochastic volatility models. Here the focus is on the estimation of the parameters of the linear time series model. No parametric assumptions about the conditional variance process are made. Instead, higher moment dependence is treated as an unobservable nuisance parameter which is estimated nonparametrically.

It is shown that the asymptotic covariance matrix of estimators based on Gaussian criterion functions contains fourth order cumulant terms. This result is a special case of models with a more general dependence structure of the errors. The assumption of martingale difference errors maintained in this paper allows us to decompose the fourth order cumulant term appearing in the covariance matrix of the score function. The decomposition in turn is used to derive a lower bound for the asymptotic covariance matrix of Gaussian estimators. It is shown that the lower bound is related to the class of instrumental variables estimators with instruments which are linear filters of the innovation sequence.

The lower bound covariance matrix is used to identify the optimal instrument. The optimal instrument is constructed from reweighting the innovation sequence in a way to minimize the asymptotic variance of the parameter estimates. Intuitively the innovation sequence is weighted in a way to balance the signal of the innovation as measured by the impulse response function and the noise generated through the dependence of squared errors.

Unobservability of the optimal instrument necessitates a semiparametric approach. It is shown that the optimal filter can be consistently estimated from fourth order cumulant terms of consistent first stage regression residuals. Computational efficiency is achieved by formulating the instrumental variables estimator in the frequency domain. This allows for the use of Fast Fourier Transform algorithms to compute the optimal filter and the optimal instrument. The complexity of the algorithm can thus be kept at $O(n \log n)$.

The paper offers an alternative way to treat conditionally heterogeneous processes. Parametric approaches decompose the process into independent sources of randomness by assuming a certain generating mechanism. Here the focus is not on the conditional but on the unconditional distribution of the errors. It is shown that instrumental variables techniques can be used to correct the statistical properties of the score process in a way that accounts for the omitted higher moment dependence. The advantage of the instrumental variables approach lies in the fact that it remains valid in situations where the form of the conditional density is unknown and too complex to be estimated nonparametrically. It is clear that the ideas developed here can be extended to related problems. These include multivariate extensions, nonlinear models and regression models. The author is currently working on these extensions.

A. Appendix - Lemmas

In this appendix the limiting normal distribution for the Whittle likelihood is derived under the conditions of Assumption (A1).

First a CLT will be stated for the error process ε_t.

Lemma A.1. *Under Assumption (A1) the following statements hold:*

i) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t \Rightarrow N(0,\sigma^2)$

ii) for each $m \in \mathbb{N}^+ \setminus \{1\}$, *m fixed, the vector* $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\varepsilon_t\varepsilon_{t-1},...,\varepsilon_t\varepsilon_{t-m}] \Rightarrow N(0,\Omega)$ *with*

$$\Omega = \begin{bmatrix} \sigma(1,1)+\sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m)+\sigma^4 \end{bmatrix}.$$

Proof. For i) we note that since $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ by assumption 1.ii) ε_t is a martingale difference sequence. Then by the martingale CLT (see Hall and Heyde[39], Theorem 3.2, p.58) we define $\varepsilon_{nt} = s_n^{-1}\varepsilon_t$ where $s_n = \sqrt{var(\sum_{t=1}^{n}\varepsilon_t)} = \sqrt{n}\sigma$. Now

$$\begin{aligned} & P\left[\left|\sum_t E\left(\varepsilon_t^2 1\left\{|\varepsilon_t| > \lambda\sqrt{n}\sigma\right\} \mid \mathcal{F}_{t-1}\right)\right| > \eta n\sigma^2\right] \\ \leq & \frac{E\left|\sum_t E\left(\varepsilon_t^2 1\left\{|\varepsilon_t| > \lambda\sqrt{n}\sigma\right\} \mid \mathcal{F}_{t-1}\right)\right|}{\eta\sigma^2 n} \\ \leq & \frac{E\sum_t E\left|\left(\varepsilon_t^2 1\left\{|\varepsilon_t| > \lambda\sqrt{n}\sigma\right\} \mid \mathcal{F}_{t-1}\right)\right|}{\eta\sigma^2 n} \\ = & \frac{\sum_t E\left(\varepsilon_t^2 1\left\{|\varepsilon_t| > \lambda\sqrt{n}\sigma\right\}\right)}{\eta\sigma^2 n} \\ \leq & \frac{E\left(\varepsilon_1^2 1\left\{|\varepsilon_t| > \lambda\sqrt{n}\sigma\right\}\right)}{\eta\sigma^2} \underset{n\to\infty}{\to} 0 \text{ for all } \eta,\lambda > 0 \end{aligned}$$

where the first inequality follows from Markov's inequality and the third inequality follows from the fact that ε_t is strictly stationary. Convergence to zero is then a consequence of the finite unconditional variance. Next

$$\begin{aligned}
& P\left[\left|\sum_t E\left(\varepsilon_{nt}^2 \mid \mathcal{F}_{t-1}\right) - 1\right| > \eta\right] \\
= \; & P\left[\left|\frac{1}{n}\sum_t \left(E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) - \sigma^2\right)\right| > \eta\right] \\
\leq \; & \frac{\sum_t \sum_s E\left(E\left(\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right) - \sigma^2\right)\left(E\left(\varepsilon_s^2 \mid \mathcal{F}_{s-1}\right) - \sigma^2\right)}{n^2\eta^2} \\
\leq \; & \frac{\sum_s \sigma(s,s)}{n\eta^2} \leq \frac{\sum_s |\sigma(s,s)|}{n\eta^2} \underset{n\to\infty}{\to} 0
\end{aligned}$$

where the first inequality follows from Chebychev's inequality. This establishes the conditions of the martingale CLT.

For part ii) we note that individually all the terms $\varepsilon_t\varepsilon_{t-k}$ with $k \geq 1$ are martingale differences. Now define $Y_t' = [\varepsilon_t\varepsilon_{t-1}, ..., \varepsilon_t\varepsilon_{t-m}]$. Then also $E(Y_t \mid \mathcal{F}_{t-1}) = 0$ so that Y_t is a vector martingale difference sequence. To show that $\frac{1}{\sqrt{n}}\sum Y_t \Rightarrow N(0,\Omega)$ it is enough to show that for all $\ell \in \mathbb{R}^m$ such that $\ell'\ell = 1$ we have $\frac{1}{\sqrt{n}}\sum \ell'\tilde{Y}_t \Rightarrow N(0,1)$ where now $\tilde{Y}_t = \Omega^{-1/2}Y_t$ and $\Omega = EY_tY_t'$. This is easily evaluated to be

$$\Omega = E\begin{bmatrix} \varepsilon_t^2\varepsilon_{t-1}^2 & \cdots & \varepsilon_t^2\varepsilon_{t-1}\varepsilon_{t-m} \\ & \ddots & \\ \varepsilon_t^2\varepsilon_{t-1}\varepsilon_{t-m} & \cdots & \varepsilon_t^2\varepsilon_{t-m}^2 \end{bmatrix} = \begin{bmatrix} \sigma(1,1)+\sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m)+\sigma^4 \end{bmatrix}.$$

Next we note that for any $\ell \in \mathbb{R}^m$ such that $\ell'\ell = 1$, ℓ fixed, $\ell'\tilde{Y}_t$ is a martingale by linearity of the conditional expectation and the fact that m is fixed and finite. We can therefore apply the martingale CLT used in part i) for the variable $Y_{nt} = \frac{1}{\sqrt{n}}\tilde{Y}_t$. Again we have to check the conditional Lindeberg condition. In particular we consider

$$
\begin{aligned}
& P\left[\left|\sum_t E\left(Y_{nt}^2 1\{|Y_{nt}|>\lambda\} \mid \mathcal{F}_{t-1}\right)\right| > \eta\right] \\
\leq\ & \frac{E\left|\sum_t E\left(\left(\ell'\tilde{Y}_t\right)^2 1\{|Y_{nt}|>\lambda\} \mid \mathcal{F}_{t-1}\right)\right|}{\eta n} \\
\leq\ & \frac{E\sum_t E\left|\left(\left(\ell'\tilde{Y}_t\right)^2 1\left\{\left|\ell'\tilde{Y}_t\right| > \lambda\sqrt{n}\right\} \mid \mathcal{F}_{t-1}\right)\right|}{\eta n} \\
=\ & \frac{\sum_t E\left(\left(\ell'\tilde{Y}_t\right)^2 1\left\{\left|\ell'\tilde{Y}_t\right| > \lambda\sqrt{n}\right\}\right)}{\eta n} \\
\leq\ & \frac{E\left(\left(\ell'\tilde{Y}_1\right)^2 1\left\{\left|\ell'\tilde{Y}_t\right| > \lambda\sqrt{n}\right\}\right)}{\eta} \underset{n\to\infty}{\to} 0 \text{ for all } \eta, \lambda > 0
\end{aligned}
$$

where again we make use of the fact that $\ell'\tilde{Y}_t$ is strictly stationary and that

$$E\left(\ell'\tilde{Y}_t\right)^2 = E\left(\ell'\Omega^{-1/2}Y_tY_t'\Omega^{-1/2}\ell\right) = 1$$

which is clearly finite so that the tail bound in the expectation of the last inequality converges to zero for all $\eta, \lambda > 0$. The second condition of the martingale CLT requires that $\sum E(Y_{nt}^2 \mid \mathcal{F}_{t-1}) \xrightarrow{p} 1$. Since for all of the examples provided it is the case that $E(Y_{nt} \mid \mathcal{F}_{t-1})$ is strictly stationary, this convergence will also hold *a.s.* which, however, is not required for the proof of the CLT, so the fact will not be further exploited. To prove the convergence condition we use

$$\begin{aligned}
&\left|\sum E\left(Y_{nt}^2 \mid \mathcal{F}_{t-1}\right) - 1\right| \\
\leq\ & \left|\sum E\left(Y_{nt}^2 \mid \mathcal{F}_{t-1}\right) - Y_{nt}^2\right| + \left|\sum Y_{nt}^2 - 1\right| \\
=\ & \left|\frac{1}{n}\sum\left[E\left(\left(\ell'\tilde{Y}_t\right)^2 \mid \mathcal{F}_{t-1}\right) - \left(\ell'\tilde{Y}_t\right)^2\right]\right| + \left|\frac{1}{n}\sum\left[\left(\ell'\tilde{Y}_t\right)^2 - 1\right]\right|.
\end{aligned}$$

Now since measurable functions of ergodic random variables are ergodic and $E(\ell'\tilde{Y}_t)^2 = 1$, we have from the ergodic theorem $\frac{1}{n}\sum(\ell'\tilde{Y}_t)^2 - 1 \overset{a.s.}{\to} 0$. Next we note that $(\ell'\tilde{Y}_t)^2$ is $\mathcal{F}_t$ measurable by assumption. Also $(\ell'\tilde{Y}_t)^2$ is strictly stationary such that

$$P\left[\left(\ell'\tilde{Y}_t\right)^2 > x\right] = P\left[\left(\ell'\tilde{Y}_1\right)^2 > x\right] < cP\left[\left(\ell'\tilde{Y}_1\right)^2 > x\right]$$

for all $x \geq 0$, $t \geq 1$ and $c > 1$. Then by Theorem 2.19, Hall and Heyde[39] it follows that

$$\frac{1}{n}\sum\left[E\left(\left(\ell'\tilde{Y}_t\right)^2 \mid \mathcal{F}_{t-1}\right) - \left(\ell'\tilde{Y}_t\right)^2\right] \overset{p}{\to} 0.$$

This now establishes that $\sum E(Y_{nt}^2 \mid \mathcal{F}_{t-1}) - 1 = o_p(1)$ as required. The proof of the CLT is complete since ℓ was fixed arbitrarily. ■

Next the distribution of the estimators is derived in the following lemmas. The proof follows the standard argument of Hannan [42] with necessary modifications made where required. The details of the proof also follow closely the exposition of Brockwell and Davis[15].

Lemma A.2. *Let $I_{n,yy}(\lambda)$ be the periodogram of $\{y_1, \ldots, y_n\}$ and $I_{n,\varepsilon\varepsilon}(\lambda)$ is the peri-*

odogram of $\{\varepsilon_1, \ldots, \varepsilon_n\}$. *Assume* ε_t *satisfy Assumption* $(A1)$ *and that* $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *with spectral density* $\frac{\sigma^2}{2\pi} g_{yy}(\beta_0, \lambda)$ *such that* $\sum_{j=0}^{\infty} |\psi_j| |j|^{1/2} < \infty$. *Let* $\varsigma(.)$ *be any continuous even function on* $[-\pi, \pi] \rightarrow \mathbb{R}$ *with absolutely summable Fourier coefficients* $\{z_k, -\infty < k < \infty\}$, *then for any* $\eta, \epsilon > 0$

$$P\left(\sqrt{n}\left|\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \varsigma(\lambda)\, d\lambda - \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, g_{yy}(\beta_0, \lambda) \varsigma(\lambda)\, d\lambda\right| > \eta\right) < \epsilon$$

as $n \longrightarrow \infty$.

Proof. First an expression for $R_n(\lambda) = I_{n,yy}(\lambda) - \frac{\sigma^2}{2\pi} I_{n,\varepsilon\varepsilon}(\lambda)\, g_{yy}(\beta_0, \lambda)$ is obtained. Let $\omega_y(\lambda) = n^{-1/2} \sum_{t=1}^{n} y_t e^{i\lambda t}$ be the discrete Fourier transform of the data. Then

$$\begin{aligned}
\omega_y(\lambda) &= n^{-1/2} \sum_{j=0}^{\infty} \sum_{t=1}^{n} \psi_j \varepsilon_{t-j} e^{-i\lambda t} \\
&= n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} \sum_{t=1}^{n} \varepsilon_{t-j} e^{-i\lambda(t-j)} \\
&= n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{i\lambda j} \sum_{t=1-j}^{n-j} \varepsilon_t e^{i\lambda t} \\
&= n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} \left(\sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t} + \sum_{t=1-j}^{n-j} \varepsilon_t e^{-i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t} \right) \\
&= \psi\left(e^{-i\lambda}\right) \omega_\varepsilon(\lambda) + n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} U_{nj}(\lambda) \qquad \text{(A.1)}
\end{aligned}$$

such that

$$\begin{aligned}
& I_{n,yy}(\lambda) - \left|\psi\left(e^{i\lambda}\right)\right|^2 I_{n,\varepsilon\varepsilon}(\lambda) \\
= \;& \psi\left(e^{-i\lambda}\right) \omega_\varepsilon(\lambda)\, n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} U_{nj}(\lambda)
\end{aligned}$$

$$+\psi\left(e^{i\lambda}\right)\omega_{\varepsilon}(-\lambda)\,n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{i\lambda j}U_{nj}(-\lambda)+n^{-1}\left|\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)\right|^2$$

Then using the Markov inequality we have

$$\begin{aligned}
& P\left(\sqrt{n}\left|\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\varsigma(\lambda)\,d\lambda-\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\,g_{yy}(\beta_0,\lambda)\varsigma(\lambda)\,d\lambda\right|>\eta\right)\\
< \; & E\frac{\sqrt{n}}{\eta}\left|\int_{-\pi}^{\pi} R_n(\lambda)\varsigma(\lambda)\,d\lambda\right|
\end{aligned}$$

so that it is enough to show that $E\sqrt{n}\left|\int_{-\pi}^{\pi} R_n(\lambda)\varsigma(\lambda)\,d\lambda\right|\to 0$. First consider

$$\begin{aligned}
& E\sqrt{n}\left|\int_{-\pi}^{\pi}\psi\left(e^{-i\lambda}\right)\omega_{\varepsilon}(\lambda)\,n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)\varsigma(\lambda)\,d\lambda\right|\\
= \; & En^{-1/2}\left|\int_{-\pi}^{\pi}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=1}^{l}\sum_{m=-\infty}^{\infty}\sum_{t=1}^{n}\psi_k\psi_l z_m\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)e^{-i\lambda(k+m-r+t)}d\lambda\right|.
\end{aligned}$$

Then for k, l, m, and r fixed such that $1\leq r-k+m\leq n$

$$\begin{aligned}
& E\left|\int_{-\pi}^{\pi}\sum_{t=1}^{n}\psi_k\psi_l z_m\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)e^{-i\lambda(k+m-r+t)}d\lambda\right|\\
= \; & E\left|\sum_{t=1}^{n}\psi_k\psi_l z_m\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)\int_{-\pi}^{\pi} e^{-i\lambda(k+m-r+t)}d\lambda\right|\\
\leq \; & \left|\psi_k\psi_l z_m\right| E\left|\varepsilon_{r-k-m}\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)\right|\\
\leq \; & \left|\psi_k\psi_l z_m\right|\left(E\left|\varepsilon_{r-k-m}\varepsilon_{r-l}\right|+E\left|\varepsilon_{r-k-m}\varepsilon_{n-l+r}\right|\right)\\
\leq \; & \left|\psi_k\psi_l z_m\right|\left(\alpha_{l-k-m}^{1/2}+\alpha_{n-l-k-m}^{1/2}\right).
\end{aligned}$$

where α_k is short hand notation for $\alpha_{k,k}$. Now summing over k, l, m gives

$$\begin{aligned} & E\sqrt{n}\left|\int_{-\pi}^{\pi}\psi\left(e^{-i\lambda}\right)\omega_{\varepsilon}(\lambda)\,n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)\,\varsigma(\lambda)\,d\lambda\right| \\ \leq\; & n^{-1/2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty}|\psi_k\psi_l z_m|\,|l|\left(\alpha_{l-k-j}^{1/2}+\alpha_{n-l-k-j}^{1/2}\right) \\ \leq\; & 2\sup_k \alpha_k^{1/2} n^{-1/2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty}|\psi_k\psi_l z_m|\,|l| \to 0. \end{aligned}$$

Next consider the term $n^{-1}\left|\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}\right|^2$. First a bound for the expected value is obtained by

$$\begin{aligned} En^{-1}\left|\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}\right|^2 &= En^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_j\psi_k e^{-i\lambda(j-k)}U_{nj}(\lambda)\,U_{nk}(-\lambda) \\ &\leq n^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\psi_j|\,|\psi_k|\,E\,|U_{nj}(\lambda)|\,|U_{nk}(\lambda)| \\ &\leq n^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\psi_j|\,|\psi_k|\left(E\,|U_{nj}(\lambda)|^2\right)^{1/2}\left(E\,|U_{nk}(\lambda)|^2\right)^{1/2} \\ &\leq \left(n^{-1/2}\sum_{j=0}^{\infty}|\psi_j|\left(E\,|U_{nj}|^2\right)^{1/2}\right)^2 \end{aligned}$$

where the Cauchy-Schwarz inequality was used in the second inequality. Now we use the martingale property of ε_t to bound $E\,|U_{nj}|^2$.

$$E\,|U_{nj}|^2 = E\left|\sum_{t=1-j}^{n-j}\varepsilon_t e^{i\lambda t}-\sum_{t=1}^{n}\varepsilon_t e^{i\lambda t}\right|^2$$

$$
\begin{aligned}
&= \begin{cases} E\left|\sum_{t=1-j}^{0} \varepsilon_t e^{-i\lambda t} - \sum_{t=n-j+1}^{n} \varepsilon_t e^{-i\lambda t}\right|^2 & 1 < j < n \\ E\left|\sum_{t=1-j}^{n-j} \varepsilon_t e^{-i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t}\right|^2 & j \geq n \\ 0 & j = 0 \end{cases} \\
&\leq \begin{cases} E\sum_{s\in I}\sum_{t\in I} \varepsilon_t\varepsilon_s e^{-i\lambda(t-s)} & I = \{1-j,\ldots,0\} \cup \{n-j+1,\ldots,n\} \\ E\sum_{s\in I}\sum_{t\in I} \varepsilon_t\varepsilon_s e^{-i\lambda(t-s)} & I = \{1-j,\ldots,n-j\} \cup \{1,\ldots,n\} \\ 0 & j = 0 \end{cases} \\
&\leq \begin{cases} 2j\sigma^2 & 1 < j < n \\ 2n\sigma^2 & j \geq n \\ 0 & j = 0 \end{cases} \\
&= 2\min(j,n)\sigma^2.
\end{aligned}
$$

This implies

$$
\left(n^{-1/2}\sum_{j=0}^{\infty} |\psi_j| \left(E\left|U_{nj}\right|^2\right)^{1/2}\right)^2 \leq 2\sigma^2\left(n^{-1/2}\sum_{j=0}^{\infty} |\psi_j| \min(j,n)^{1/2}\right)^2.
$$

Now

$$
n^{-1/2}\sum_{j=0}^{\infty} |\psi_j| \min(j,n)^{1/2} \leq n^{-1/2}\sum_{j=0}^{n} |\psi_j|\,|j|^{1/2} + \sum_{j=n+1}^{\infty} |\psi_j|.
$$

Since $\sum_{j=0}^{\infty} |\psi_j|\,|j|^{1/2} < \infty$ it follows that

$$
\begin{aligned}
\overline{\lim_{n\to\infty}} \sum_{j=0}^{\infty} |\psi_j| \min(j,n)^{1/2} &\leq \overline{\lim_{n\to\infty}} \left(\sum_{j=0}^{n} |\psi_j|\,|j|^{1/2} + n^{1/2}\sum_{j=n+1}^{\infty} |\psi_j|\right) \\
&\leq \sum_{j=0}^{\infty} |\psi_j|\,|j|^{1/2} < \infty.
\end{aligned}
$$

Pulling these results together gives

$$E\sqrt{n}\left|\int_{-\pi}^{\pi} n^{-1}\left|\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}\right|^2 \varsigma(\lambda)\,d\lambda\right|$$

$$\leq \; 4\pi\sigma^2\left(n^{-1/2}\sum_{j=0}^{\infty}|\psi_j|\,|j|^{1/2}\right)^2 \sup_{\lambda\in[-\pi,\pi]}\varsigma(\lambda) \to 0.$$

Together these results imply

$$E\sqrt{n}\left|\int_{-\pi}^{\pi} R_n(\lambda)\,\varsigma(\lambda)\,d\lambda\right| \to 0$$

as had to be shown ■

Next a Lemma is stated which allows us to use the finite dimensional CLT proved in lemma A.1 to approximate the limit distribution of a countably infinite dimensional vector of random variables.

Lemma A.3. *(Billingsley) Let X_{mn}, Y_n be random variables defined on $(\Omega, \mathcal{F}, P)$. Suppose that for each m $X_{mn} \xrightarrow{d} X_m$ as $n \to \infty$ and that $X_m \xrightarrow{d} X$ as $m \to \infty$. Suppose further that*

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\{|X_{mn} - Y_n| \geq \epsilon\} = 0$$

for each positive ϵ. Then $Y_n \xrightarrow{d} X$ as $n \to \infty$.

Now a general CLT for criterion functions based on the periodogram and involving sufficiently well behaved spectral weighting functions is presented.

Lemma A.4. *Let $I_{n,yy}(\lambda)$ be the periodogram of $\{y_1, \ldots, y_n\}$ and $I_{n,\varepsilon\varepsilon}(\lambda)$ is the peri-*

odogram of $\{\varepsilon_1, \ldots, \varepsilon_n\}$ *. Suppose the* ε_t *satisfy Assumption* $(A1)$ *and that* $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *with spectral density* $\frac{\sigma^2}{2\pi} g_{yy}(\beta_0, \lambda)$ *such that* $\sum_{j=0}^{\infty} |\psi_j| |j|^{1/2} < \infty$*. Let* $\varsigma(.)$ *be any continuous even function on* $[-\pi, \pi] \to \mathbb{R}$ *with Fourier coefficients* $\{z_k, -\infty < k < \infty\}$ *such that*

$$\sum_{k=1}^{\infty} |z_k| |k|^{1/2} < \infty$$

and $\int_{-\pi}^{\pi} \varsigma(\lambda) g_{yy}(\beta_0, \lambda) d\lambda = 0$*, then*

$$n^{1/2} \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \varsigma(\lambda) d\lambda \xrightarrow{d} N\left(0, 4 \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \alpha_{k,l} b_k b_l\right)$$

with $b_k = \sum_{j=-\infty}^{\infty} \gamma_{yy}(k-j) z_j$*.*

Proof. It has already been established that

$$n^{1/2} \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \varsigma(\lambda) d\lambda - n^{1/2} \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) g_{yy}(\beta_0, \lambda) \varsigma(\lambda) d\lambda = o_p(1)$$

so that it remains to show

$$n^{1/2} \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) g_{yy}(\beta_0, \lambda) \varsigma(\lambda) d\lambda \xrightarrow{d} N\left(0, 4 \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \alpha_{k,l} b_k b_l\right).$$

Let $\chi(\lambda) = g(\beta_0, \lambda) \varsigma(\lambda)$. Using the Fourier approximation

$$\chi(\lambda) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \gamma_{yy}(k-j) z_j e^{i\lambda k}$$

such that $(2\pi)^{-1}\int_{-\pi}^{\pi}\chi(\lambda)e^{i\lambda k}d\lambda = \sum_{j=-\infty}^{\infty}\gamma_{yy}(k-j)z_j$. It follows that

$$\begin{aligned}
\sum_{k=1}^{\infty}|b_k||k|^{1/2} &= \sum_{k=1}^{\infty}\left|\sum_{j=-\infty}^{\infty}\gamma_{yy}(k-j)z_j\right||k|^{1/2} \\
&\leq \sum_{k=1}^{\infty}\sum_{j=-\infty}^{\infty}|\gamma_{yy}(k-j)||z_j||k|^{1/2} \\
&\leq \sum_{j=-\infty}^{\infty}|z_j||j|^{1/2}\sum_{k=1}^{\infty}|\gamma_{yy}(k-|j|)|\left(\frac{|k|}{\max|j|,1}\right)^{1/2} \\
&= \sum_{j=-\infty}^{\infty}|z_j||j|^{1/2}\frac{1}{(\max|j|,1)^{1/2}}\sum_{k=1+|j|}^{\infty}|\gamma_{yy}(k)||k-|j||^{1/2} \\
&\leq \sum_{j=-\infty}^{\infty}|z_j||j|^{1/2}\sum_{k=1}^{\infty}|\gamma_{yy}(k)||k|^{1/2} < \infty.
\end{aligned}$$

Then define $\chi_m(\lambda) = \sum_{|k|<m} b_k e^{i\lambda k}$. $\chi_m(\lambda)$ converges uniformly to $\chi(\lambda)$. Using Lemma (A.3) it has to be shown that for all $\epsilon > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left\{\left|n^{1/2}\int_{-\pi}^{\pi}I_{n,\varepsilon\varepsilon}(\lambda)(\chi_m(\lambda)-\chi(\lambda))d\lambda\right| \geq \epsilon\right\} = 0$$

where

$$n^{1/2}\int_{-\pi}^{\pi}I_{n,\varepsilon\varepsilon}(\lambda)(\chi_m(\lambda)-\chi(\lambda))d\lambda = n^{1/2}\sum_{|k|>m}^{n}\tilde{\gamma}_{\varepsilon\varepsilon}(k)b_k$$

with $\tilde{\gamma}_{\varepsilon\varepsilon}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}\varepsilon_t\varepsilon_{t+|k|}$. It follows that $E\tilde{\gamma}_{\varepsilon\varepsilon}(k) = 0$,

$$\begin{aligned}
E\tilde{\gamma}_{\varepsilon}(k)\tilde{\gamma}_{\varepsilon}(j) &= \frac{1}{n^2}\sum_{t=1}^{n-|k|}\sum_{s=1}^{n-|k|}E\varepsilon_t\varepsilon_s\varepsilon_{t+|k|}\varepsilon_{s+|j|} \\
&= \frac{1}{n^2}\sum_{t=1}^{n-|k|}E\varepsilon_t^2\varepsilon_{t+|k|}\varepsilon_{t+|j|} = \frac{n-|k|}{n^2}\alpha_{k,j}.
\end{aligned}$$

where the second equality follows from the martingale difference sequence property of the errors. Using the Markov inequality we have

$$
\begin{aligned}
& P\left\{\left|n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(\chi_m(\lambda)-\chi(\lambda)\right)d\lambda\right|\geq\epsilon\right\} \\
\leq\ & \frac{var\left(n^{1/2}\sum_{|k|>m}\tilde{\gamma}_{\varepsilon\varepsilon}(k)\,b_k\right)}{\epsilon^2} \\
=\ & \frac{n\sum_{|k|>m}^{n}\sum_{|l|>m}^{n}\frac{n-|k|}{n^2}\alpha_{k,l}b_kb_l}{\epsilon^2} \\
\leq\ & \frac{\sup_{k,l}|\alpha_{k,l}|}{\epsilon^2}\left(\sum_{|k|>m}^{n}|b_k|\right)^2
\end{aligned}
$$

which in turn implies

$$
\limsup_{n\to\infty} P\left\{\left|n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(\chi_m(\lambda)-\chi(\lambda)\right)d\lambda\right|\geq\epsilon\right\}\leq\frac{C}{\epsilon^2}\left(\sum_{|k|>m}^{\infty}|b_k|\right)^2
$$

where $C=\sup_{k,l}|\alpha_{k,l}|<\infty$ so that the result follows form

$$
\lim_{m\to\infty}\left(\sum_{|k|>m}^{\infty}|b_k|\right)^2=0.
$$

From Lemma (A.1) we have

$$
\begin{aligned}
n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\,\chi_m(\lambda)\,d\lambda &= 2n^{1/2}\sum_{k=1}^{m}\tilde{\gamma}_{\varepsilon\varepsilon}(k)\,b_k \\
&\xrightarrow{d} N\left(0,4\sum_{l=1}^{m}\sum_{k=1}^{m}\alpha_{k,l}b_kb_l\right).
\end{aligned}
$$

Letting $X_m\sim N\left(0,4\sum_{l=1}^{m}\sum_{k=1}^{m}\alpha_{k,l}b_kb_l\right)$ it remains to show that $X_m\xrightarrow{d}X$ where $X\sim$

$N\left(0, 4\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\alpha_{k,l}b_kb_l\right)$. By Billingsley [8], Theorem 7.6 it is enough to show

$$E\left(e^{itX_m}\right) \to E\left(e^{itX}\right)$$

for all t. Since

$$E\left(e^{itX_m}\right) = e^{-\frac{1}{2}t^2 4\sum_{l=1}^{m}\sum_{k=1}^{m}\alpha_{k,l}b_kb_l} \to e^{-\frac{1}{2}t^2 4\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\alpha_{k,l}b_kb_l}$$

if $\sum_{l=1}^{m}\sum_{k=1}^{m}\alpha_{k,l}b_kb_l \to \sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\alpha_{k,l}b_kb_l < \infty$. This follows by absolute convergence from

$$\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}|\alpha_{k,l}|\,|b_k|\,|b_l| \leq \sup_{k,l}|\alpha_{k,l}|\left(\sum_{k=1}^{\infty}|b_k|\right)^2 < \infty$$

the proof is completed ■

B. Appendix - Proofs of Part I

Proof of Theorem 3.3 Since $\dot{\eta}(\beta,\lambda) \in C^2[-\pi,\pi]$ it follows from Lemma (A.4) by setting $\varsigma(\lambda) = \ell' \dot{\eta}(\beta,\lambda)$ for all $\ell \in \mathbb{R}^p$ with $\ell'\ell = 1$ that

$$\sqrt{n}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\,\dot{\eta}(\beta,\lambda)\,d\lambda \Rightarrow N(0,B)$$

with $B = 4\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\alpha_{k,l}\left\langle \dot{\eta}(\beta,\lambda), e^{ik\lambda}\right\rangle \left\langle \dot{\eta}(\beta,\lambda), e^{il\lambda}\right\rangle'$. Next it is shown that the variance of the score process can be expressed as

$$B = 2\sigma^2 A + 2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda)\,\dot{\eta}(\beta,\mu)\,\dot{\eta}(\beta,\lambda)\,d\mu d\lambda.$$

Redefine the vector of Fourier coefficients as

$$b_k = \left\langle \dot{\eta}(\beta,\lambda), e^{ik\lambda}\right\rangle$$

with j-th element

$$b_{j,k} = \left\langle \dot{\eta}_j(\beta,\lambda), e^{ik\lambda}\right\rangle.$$

B can now be written more compactly as

$$4\sigma^4\sum_{k=1}^{\infty} b_k b_k' + 4\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} b_k b_l' \sigma(k,l).$$

By Parseval's identity we have $\sum_{k=1}^{\infty} b_k b_k' = \frac{1}{4\pi}\int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\,\dot{\eta}(\beta,\lambda)'\,d\lambda$. Considering the derivatives of $\ln g_{yy}(\beta,\lambda)$ implies immediately that $b_k = b_{-k}$ and $b_0 = 0$. Using the definition of $f_{\varepsilon^2 \varepsilon\varepsilon}(\mu, \lambda)$ we note that $\sigma(k,l) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,\mu)\, e^{i(k\lambda+\mu l)} d\lambda d\mu$. By definition $\sigma(k,l) = \sigma(-k,-l)$. Looking at a typical element of $\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} b_k b_l' \sigma(k,l)$ we first note

$$2\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} b_{j,k} b_{m,l}\sigma(k,l) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_{j,k} b_{m,l}\sigma(k,l)$$

since $\sigma(k,-l) = \sigma(-k,l) = 0\ \forall l,k > 0$ by (2.2) and $b_0 = 0$. Then

$$\begin{aligned}
\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_{j,k} b_{m,l}\sigma(k,l) &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} b_{j,k}e^{-i\lambda k} b_{l,m} e^{-i\lambda l} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda) d\mu d\lambda \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{k=-\infty}^{\infty} b_{j,k} e^{-i\lambda k}\sum_{m=-\infty}^{\infty} b_{l,m} e^{-i\mu l} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda) d\mu d\lambda \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \dot{\eta}_j(\beta,\mu)\,\dot{\eta}_l(\beta,\lambda)\, f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda) d\mu d\lambda
\end{aligned}$$

where we used dominated convergence to exchange integration and summation in the second but last line. Now the result follows from building up the matrix B from its individual elements ■

Proof of Corollary 3.4 The fourth order cumulant is defined in general as

$$f_{\varepsilon..\varepsilon}(\lambda_1,...,\lambda_4) = \frac{1}{(2\pi)^3}\sum_{u_1,...,u_4=-\infty}^{\infty} c_{\varepsilon..\varepsilon}(u_1,...,u_4) e^{-i\{\sum_{j=1}^{4} u_j\lambda_j\}}.$$

Stationarity implies that $c_{\varepsilon..\varepsilon}(u_1,...,u_4) = c_{\varepsilon..\varepsilon}(u_1+t,...,u_4+t)$ for any t. Therefore $f_{\varepsilon..\varepsilon}(\lambda_1,...,\lambda_4) = 0$ for $\sum_{j=1}^{4}\lambda_j \neq 0 \bmod 2\pi$. We can thus choose $u_1 = 0$ without loss of generality. The fourth order cumulant spectrum $f_{\varepsilon..\varepsilon}(\lambda_1,...,\lambda_4)$ then reduces to (3.3). From the martingale difference property $c_{\varepsilon..\varepsilon}(0,u_2,...,u_4)$ is non zero only if at least one

$u_i = 0$ and all $u_j < 0$ for $j \neq i$ or $u_i = u_j > 0$ and $u_k < u_i$ for $k \neq i, j$. It follows that $c_{\varepsilon..\varepsilon}(0, u_2, ..., u_4)$ can be replaced by $\sigma(u, s)$ defined in (2.2) for $u, s \in \{0, \pm 1, \pm 2, ...\}$. This is the symmetry property of cumulants for stationary time series. The trispectrum (3.3) can be written by taking the restrictions for the *mds* case into account

$$
\begin{aligned}
&(2\pi)^3 f_{\varepsilon..\varepsilon}(\lambda, \mu, -\lambda) \\
= &\sum_{s=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} c_{\varepsilon..\varepsilon}(q, r, s) e^{-i\{(q-s)\lambda + r\mu\}} \\
= &\sum_{s=1}^{\infty} \sum_{r=-\infty}^{s-1} c_{\varepsilon..\varepsilon}(s, r, s) e^{-ir\mu} + \sum_{s=-\infty}^{0} \sum_{q=-\infty}^{0} c_{\varepsilon..\varepsilon}(q, 0, s) e^{-i(q-s)\lambda} \\
&+ \sum_{q=1}^{\infty} \sum_{s=-\infty}^{q-1} c_{\varepsilon..\varepsilon}(q, q, s) e^{-i(q\mu + (q-s)\lambda)} + \sum_{q=-\infty}^{0} \sum_{r=-\infty}^{0} c_{\varepsilon..\varepsilon}(q, r, 0) e^{-i(q\lambda + r\mu)} \\
&+ \sum_{s=1}^{\infty} \sum_{q=-\infty}^{s-1} c_{\varepsilon..\varepsilon}(q, s, s) e^{-i(s\lambda + (q-s)\lambda)} + \sum_{s=-\infty}^{0} \sum_{r=-\infty}^{0} c_{\varepsilon..\varepsilon}(0, r, s) e^{-i(r\mu - s\lambda)}.
\end{aligned}
$$

This is a list of all nonzero parts of the triple sum. Now for $s > 0$ and $r < s$ we have $c_{\varepsilon..\varepsilon}(s, r, s) = \sigma(s, s-r)$ and using $k = s-r$ we can rewrite $\sum_{s=1}^{\infty} \sum_{r=-\infty}^{s} c_{\varepsilon..\varepsilon}(s, r, s) e^{-ir\mu} = \sum_{s=1}^{\infty} \sum_{k=1}^{\infty} \sigma(s, k) e^{-i(s-k)\mu}$. In the same way $c_{\varepsilon..\varepsilon}(q, q, s) = \sigma(q, q-s)$ such that

$$\sum_{q=1}^{\infty} \sum_{s=-\infty}^{q-1} c_{\varepsilon..\varepsilon}(q, q, s) e^{-i(q\mu + (q-s)\lambda)} = \sum_{q=1}^{\infty} \sum_{k=1}^{\infty} \sigma(q, k) e^{-i(q\mu + k\lambda)}$$

and

$$\sum_{s=1}^{\infty} \sum_{q=-\infty}^{s-1} c_{\varepsilon..\varepsilon}(q, s, s) e^{-i(q\mu + (q-s)\lambda)} = \sum_{s=1}^{\infty} \sum_{k=1}^{\infty} \sigma(s, k) e^{-i(s\mu - k\lambda)}.$$

Also for $q, s < 0$ we have $c_{\varepsilon..\varepsilon}(q, 0, s) = \sigma(s, q)$ and the same holds for $c_{\varepsilon..\varepsilon}(0, q, s)$ and $c_{\varepsilon..\varepsilon}(q, s, 0)$. Then $\sum_{s=-\infty}^{0} \sum_{q=-\infty}^{0} c_{\varepsilon..\varepsilon}(q, 0, s) e^{-i(q-s)\lambda} = \sum_{s=-\infty}^{0} \sum_{q=-\infty}^{0} \sigma(s, q) e^{-i(q-s)\lambda}$

and the same is true for the other two double sums. Using the fact that $\sigma(s,-q) = \sigma(-s,q) = 0$ for all $s,q > 0$ we can now write

$$\begin{aligned}(2\pi)\, f_{\varepsilon..\varepsilon}(\lambda,\mu,-\lambda) &= \frac{1}{2}(f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,-\lambda) + f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\mu)) \\ &\quad + f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,\mu) + f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,-\mu)\end{aligned}$$

Substituting into the fourth order cumulant part of the asymptotic covariance matrix now leads to

$$\begin{aligned}&(2\pi)\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\varepsilon..\varepsilon}(\mu,\lambda,-\mu)\,\dot{\eta}(\phi,\mu)\,\dot{\eta}(\phi,\lambda)\,d\lambda d\mu \\ &= \int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda)\,\dot{\eta}(\phi,\mu)\,d\mu d\lambda \\ &\quad + \int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\phi,\mu)\,d\mu d\lambda.\end{aligned}$$

where we use $\int_{-\pi}^{\pi}\dot{\eta}(\beta,\mu)\,d\mu = 0$ such that $\int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\,d\lambda\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\mu)\,\dot{\eta}(\phi,\mu)\,d\mu = 0$ and the same holds for the term with $f_{\varepsilon^2\varepsilon\varepsilon}(\lambda,-\lambda)$. Define

$$(f_{\varepsilon^2\varepsilon\varepsilon} * \dot{\eta})(\lambda) = \int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\phi,\mu)\,d\mu.$$

Using $\dot{\eta}(\beta,\lambda) = \dot{\eta}(\beta,-\lambda)$ the result follows if $(f_{\varepsilon^2\varepsilon\varepsilon} * \dot{\eta})(\lambda)$ is symmetric as well, since then

$$\begin{aligned}&\int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda)\,\dot{\eta}(\beta,\mu)\,d\mu d\lambda \\ &= \int_{-\pi}^{\pi}\dot{\eta}(\beta,\lambda)\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\beta,\mu)\,d\mu d\lambda.\end{aligned}$$

But since

$$\begin{aligned}
& \int_{-\pi}^{\pi} \dot{\eta}(\beta,\lambda) \int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\beta,\mu)\,d\mu d\lambda \\
= & \int_{-\pi}^{\pi} \dot{\eta}(\beta,\mu) \int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\beta,\lambda)\,d\lambda d\mu \\
= & \left(f_{\varepsilon^2\varepsilon\varepsilon} * \dot{\eta}\right)(\mu)
\end{aligned}$$

and $\int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,-\lambda)\,\dot{\eta}(\beta,\lambda)\,d\lambda = \int_{-\pi}^{\pi} f_{\varepsilon^2\varepsilon\varepsilon}(\mu,\lambda)\,\dot{\eta}(\beta,-\lambda)\,d\lambda$ by symmetry of $\dot{\eta}(\beta,\lambda)$. This completes the proof ■

Proof of Theorem 4.4 For m fixed it follows from standard results that for any non singular matrix C_m

$$\left(P_m' C_m P_m\right)^{-1}\left(P_m' C_m \Omega_m C_m P_m\right)\left(P_m' C_m P_m\right)^{-1} - \frac{1}{\sigma^4}\left(P_m' \Omega_m^{-1} P_m\right)^{-1} \geq 0$$

which after pre and postmultiplying by $(P_m' C_m P_m)$ is

$$\left(P_m' C_m \Omega_m C_m P_m\right) - \frac{1}{\sigma^4}\left(P_m' C_m P_m\right)\left(P_m' \Omega_m^{-1} P_m\right)^{-1}\left(P_m' C_m P_m\right) \geq 0$$

where ≥ 0 here signifies matrix positive semi definiteness. While this inequality holds for all positive definite matrices C_m as long as m is finite, further restrictions need to be imposed to guarantee that the inequality also holds in the limit. It is enough to require that for all m the elements of C_m denoted by $c_{j,k}^m$ are such that $\sum_{k=1}^m \left|c_{j,k}^m\right| \leq M < \infty$ and $\sum_{k=1}^m \left|c_{k,j}^m\right| \leq M < \infty$, i.e. rows and columns are absolutely summable. Using the matrix norm $\|A\| = \left(trA'A\right)^{1/2}$ it then follows from the matrix version of the triangle inequality that

$$\begin{aligned}\left\|P_m' C_m P_m\right\| &= \left\|\sum_{k=1}^{m}\sum_{l=1}^{m} b_k b_l' c_{k,l}^m\right\| \\ &\leq \sum_{k=1}^{m}\sum_{l=1}^{m}\left\|b_k b_l'\right\| \left|c_{k,l}^m\right| \\ &\leq M\left(\sum_{k=1}^{m}\left(b_k' b_k\right)^{1/2}\right)^2 < \infty\end{aligned}$$

for all m. This then implies that $\lim_m P_m' C_m P_m$ exists and has bounded elements. By definition all elements of Ω_m are positive and so $P_m' C_m \Omega_m C_m P_m$ is always positive definite. As before

$$\begin{aligned}\left\|P_m' C_m \Omega_m C_m P_m\right\| &\leq \sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|b_i c_{i,l}^m b_j' c_{k,j}^m\right\| \left|\alpha_{k,l}\right| \\ &\leq \sup_{k,l}\left|\alpha_{k,l}\right| \sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|b_i c_{i,l}^m b_j' c_{l,j}^m\right\| \\ &\leq \sup_{k}\left|\alpha_k\right| \sum_{i=1}^{m}\sum_{j=1}^{m}\left\|b_i b_j'\right\| \sum_{k=1}^{m}\sum_{l=1}^{m}\left|c_{i,l}^m\right| \left|c_{k,j}^m\right| \\ &\leq \sup_{k}\left|\alpha_k\right| M^2 \sum_{i=1}^{m}\sum_{j=1}^{m}\left\|b_i b_j'\right\| < \infty\end{aligned}$$

Then $\lim_m P_m' C_m \Omega_m C_m P_m$ exists and has bounded elements. Finally $\lim_{m\to\infty}\left(P_m' \Omega_m^{-1} P_m\right)$ exists since all elements of $\Omega_m^{-1}$are bounded by Lemma (4.3). Now clearly $C_m = I_m$ satisfies the summability assumptions and $C_m = \Omega_m^{-1}$ is the optimal transformation. Taking limits then establishes the result. ■

C. Appendix - Proofs of Part II

Proof of Lemma 8.1 First we show that $f_{\tilde{\alpha}}(\lambda) \in L_1[-\pi,\pi]$ which follows from $\int_{-\pi}^{\pi} |f_{\tilde{\alpha}}(\lambda)| d\lambda = \sum_{j=-\infty}^{\infty} \left| \frac{1}{\alpha_j} - \frac{1}{\sigma^4} \right| < \infty$. Next note for a typical element k

$$
\begin{aligned}
& \int_{-\pi}^{\pi} f_{\tilde{\alpha}}(\lambda-\xi) [\dot{\eta}(\phi,\xi)]_k d\xi \\
= & \int_{-\pi}^{\pi} f_{\tilde{\alpha}}(\lambda-\xi) \left(\sum_{j=0}^{\infty} \psi_j e^{-i\xi(j+k)} + \sum_{j=0}^{\infty} \psi_j e^{i\xi(j+k)} \right) d\xi \\
= & \int_{-\pi}^{\pi} \sum_{l=-\infty}^{\infty} \tilde{\alpha}_l e^{-i(\lambda-\xi)l} \left(\sum_{j=0}^{\infty} \psi_j e^{-i\xi(j+k)} + \sum_{j=0}^{\infty} \psi_j e^{i\xi(j+k)} \right) d\xi \\
= & \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{\alpha}_l \psi_j e^{-i\xi(j+k)} e^{-i(\lambda-\xi)l} + \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \tilde{\alpha}_l e^{-i(\lambda-\xi)l} e^{i\xi(j+k)} d\xi \\
= & \sum_{j=0}^{\infty} \tilde{\alpha}_{j+k} \psi_j e^{-i\lambda(j+k)} + \sum_{j=0}^{\infty} \tilde{\alpha}_{-j-k} \psi_j e^{i\lambda(j+k)} \\
= & \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} e^{-i\lambda(j+k)} + \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} e^{i\lambda(j+k)} - \frac{1}{\sigma^4} \left(\sum_{j=0}^{\infty} \psi_j e^{-i\lambda(j+k)} + \sum_{j=0}^{\infty} \psi_j e^{i\lambda(j+k)} \right) \\
= & l_{\psi,k}(\lambda) + l_{\psi,k}(-\lambda) - \frac{1}{\sigma^4} \dot{\eta}(\phi,\lambda)
\end{aligned}
$$

such that the result follows. ■

Proof of Proposition 8.2 We start by defining two remainder terms $R_n^1(\lambda)$ and $R_n^2(\lambda)$ where

$$
\begin{aligned}
R_n^1(\lambda) &= n^{-1/2} \sum_{j=0}^{n} \psi_j e^{-i\lambda j} U_{n,j}(\lambda) \\
R_n^2(\lambda) &= l_{\psi}(\lambda) \phi\left(e^{-i\lambda}\right) n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} U_{n,j}(\lambda)
\end{aligned}
$$

$$+n^{-1/2}a(\lambda)^{*}\odot\sum_{j=0}^{\infty}\tilde{\psi}_{j}e^{-i\lambda j}U_{n,j}(\lambda)$$

with $U_{n,j}(\lambda)=\sum_{t=1-j}^{n-j}\varepsilon_t e^{-i\lambda t}+\sum_{t=1}^{n}\varepsilon_t e^{-i\lambda t}$ and

$$\tilde{\psi}_j=\begin{bmatrix}\psi_j a_{j+1}^{-1}\\ \vdots\\ \psi_j a_{j+p}^{-1}\end{bmatrix},$$

such that $R_n^1(\lambda)$ is a scalar and $R_n^2(\lambda)$ is a $p\times 1$ vector. Then, using equations (A.1) and (8.9), $I_{n,zy}(\lambda)$ can be written as

$$I_{n,zy}(\lambda)=l_{\psi}(\lambda)\,I_{n,yy}(\lambda)\,\phi\left(e^{-i\lambda}\right)a(\lambda)\,\phi+l_{\psi}(\lambda)\,I_{n,\varepsilon\varepsilon}(\lambda)+R_n(\lambda)$$

and in the same way

$$I_{n,yz}(\lambda)=l_{\psi}(\lambda)\,I_{n,yy}(\lambda)\,\phi\left(e^{i\lambda}\right)a(-\lambda)\,\phi+l_{\psi}(\lambda)\,I_{n,\varepsilon\varepsilon}(\lambda)+R_n(-\lambda)$$

where

$$\begin{aligned}R_n(\lambda) &= \phi^{-1}\left(e^{i\lambda}\right)\omega_{\varepsilon}(\lambda)\,R_n^2(\lambda)+R_n^1(-\lambda)\,R_n^2(\lambda)\\ &\quad +l_{\psi}(\lambda)\,\omega_{\varepsilon}(\lambda)\,R_n^1(\lambda)+l_{\psi}(\lambda)\left|R_n^1(\lambda)\right|^2\end{aligned}$$

such that

$$\begin{aligned}
& \int_{-\pi}^{\pi} \left(I_{n,zy}(\lambda) + I_{n,yz}(\lambda)\right) d\lambda \\
= & \left[2\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \operatorname{Re}\left[l_{\psi}(-\lambda)\,\phi\left(e^{i\lambda}\right) a(-\lambda)\right] d\lambda\right] \phi \\
& +2\int_{-\pi}^{\pi} l_{\eta}(\lambda)\, I_{n,\varepsilon\varepsilon}(\lambda)\, d\lambda + \int_{-\pi}^{\pi} R_n(\lambda) + R_n(-\lambda)\, d\lambda. \qquad \text{(C.1)}
\end{aligned}$$

Alternatively one has also

$$I_{n,yz}(\lambda)\, a(-\lambda) = I_{n,yy}(\lambda)\, l_{\psi}(-\lambda)\,\phi\left(e^{i\lambda}\right) a(-\lambda) + R_n^p(\lambda)$$

with

$$R_n^p(\lambda) = \phi^{-1}\left(e^{i\lambda}\right)\omega_{\varepsilon}(\lambda)\, R_n^2(\lambda) + R_n^1(-\lambda)\, R_n^2(\lambda)$$

such that $\int I_{n,yz}(\lambda) \operatorname{Re}[a(-\lambda)]\, d\lambda = \int I_{n,yy}(\lambda) \operatorname{Re}\left[l_{\psi}(-\lambda)\,\phi\left(e^{i\lambda}\right) a(-\lambda)\right] d\lambda + \int \operatorname{Re}\left[R_n^p(\lambda)\right] d\lambda$. Now if $\int \operatorname{Re}\left[R_n^p(\lambda)\right] d\lambda = o_p\left(n^{-1/2}\right)$ it follows that

$$\begin{aligned}
\sqrt{n}\left(\tilde{\phi} - \phi\right) = & \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda) \operatorname{Re}\left[l_{\psi}(-\lambda)\,\phi\left(e^{i\lambda}\right) a(-\lambda)\right] d\lambda\right]^{-1} \\
& \times\sqrt{n}\left[\int_{-\pi}^{\pi} l_{\eta}(\lambda)\, I_{n,\varepsilon\varepsilon}(\lambda)\, d\lambda + \int_{-\pi}^{\pi} R_n(\lambda) + R_n(-\lambda)\, d\lambda\right].
\end{aligned}$$

Since $l_{\psi}(\lambda) \in C^2[-\pi,\pi]$ and $\phi\left(e^{i\lambda}\right) \in C^2[-\pi,\pi]$ it follows from Lemma (A.4) that

$$\sqrt{n}\int_{-\pi}^{\pi} l_{\eta}(\lambda)\, I_{n,\varepsilon\varepsilon}(\lambda)\, d\lambda \Rightarrow N\left(0, \sum_{j=1}^{\infty} \alpha_j \left\langle l_{\eta}(\lambda), e^{i\lambda j}\right\rangle \left\langle l_{\eta}(\lambda), e^{i\lambda j}\right\rangle'\right),$$

where the matrix $\sum_{l=1}^{\infty} \alpha_l \left\langle l_\eta(\lambda), e^{i\lambda l}\right\rangle \left\langle l_\eta(\lambda), e^{i\lambda l}\right\rangle'$ has typical element

$$\left[\sum_{l=1}^{\infty} \alpha_l \left\langle l_\eta(\lambda), e^{i\lambda l}\right\rangle \left\langle l_\eta(\lambda), e^{i\lambda l}\right\rangle'\right]_{k,l} = \sum_{j=0}^{\infty} \alpha_{j+k\vee l}^{-1} \psi_{\phi,j} \psi_{\phi,j+|k-l|}.$$

Using lemma (8.1) it is easy to show that

$$\begin{aligned} &\sum_{j=1}^{\infty} \alpha_j \left\langle l_\eta(\lambda), e^{i\lambda j}\right\rangle \left\langle l_\eta(\lambda), e^{i\lambda j}\right\rangle' \\ = &\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left(f_{\tilde{\alpha}}(\lambda-\xi)\, \dot{\eta}(\phi,\xi) + \frac{1}{\sigma^4}\dot{\eta}(\phi,\lambda)\right) \dot{\eta}(\phi,\lambda)'\, d\xi d\lambda. \end{aligned}$$

The result then follows if $\sqrt{n}\int_{-\pi}^{\pi} R_n(\lambda)\, d\lambda = o_p(1)$. We discuss the four different types of remainder terms separately. Since $l_\psi(\lambda) \in C^2[-\pi,\pi]$ the proof of Lemma (A.2) can be applied to $l_\psi(\lambda)\,\omega_\varepsilon(-\lambda)\, R_n^1(\lambda)$ and $l_\psi(\lambda)\left|R_n^1(\lambda)\right|^2$. To show that $\sqrt{n}\int_{-\pi}^{\pi} l_\psi(\lambda)\,\omega_\varepsilon(-\lambda)\, R_n^2(\lambda)\, d\lambda = o_p(1)$ it is enough to show that $E\sqrt{n}\left|\int_{-\pi}^{\pi} l_\psi(\lambda)\,\omega_\varepsilon(-\lambda)\, R_n^2(\lambda)\, d\lambda\right| \to 0$. Using the definition of $R_n^2(\lambda)$ one has

$$\begin{aligned} & E\sqrt{n}\left|\int_{-\pi}^{\pi} l_\psi(\lambda)\,\omega_\varepsilon(-\lambda)\, R_n^2(\lambda)\, d\lambda\right| \\ \leq\; & E\sqrt{n}\left|\int_{-\pi}^{\pi} l_\psi(\lambda)\,\phi\left(e^{-i\lambda}\right)\phi^{-1}\left(e^{i\lambda}\right)\omega_\varepsilon(-\lambda)\, R_n^1(\lambda)\, d\lambda\right| \\ & +E\sqrt{n}\left|\int_{-\pi}^{\pi} \phi^{-1}\left(e^{i\lambda}\right)\omega_\varepsilon(-\lambda)\, n^{-1/2} a(\lambda)^* \odot \sum_{j=0}^{\infty} \tilde{\psi}_j e^{-i\lambda(j+k)} U_{n,j}(\lambda)\, d\lambda\right|. \end{aligned}$$

The first term can be analyzed by setting $\varsigma(\lambda) = l_\psi(\lambda)\,\phi\left(e^{-i\lambda}\right) \in L_1[-\pi,\pi]$ and applying the proof of Lemma (A.2). Next look at a typical element k of the second term

$$
\begin{aligned}
& E\sqrt{n}\left|\int_{-\pi}^{\pi}\phi^{-1}\left(e^{i\lambda}\right)\omega_{\varepsilon}(-\lambda)\, n^{-1/2}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda(j+k)}U_{n,j}(\lambda)\, d\lambda\right| \\
= \;& En^{-1/2}\left|\int_{-\pi}^{\pi}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=1}^{l}\sum_{t=1}^{n}\frac{\psi_j\psi_l}{\alpha_{k+j}}\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)e^{-i\lambda(k+j-r+t)}d\lambda\right|.
\end{aligned}
$$

Then for j, l, m, and r fixed such that $1 \leq r-k+m \leq n$

$$
\begin{aligned}
& E\left|\int_{-\pi}^{\pi}\sum_{t=1}^{n}\frac{\psi_j\psi_l}{\alpha_{k+j}}\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)e^{-i\lambda(k+j-r+t)}d\lambda\right| \\
= \;& E\left|\sum_{t=1}^{n}\frac{\psi_j\psi_l}{\alpha_{k+j}}\varepsilon_t\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)\int_{-\pi}^{\pi}e^{-i\lambda(k+j-r+t)}d\lambda\right| \\
\leq \;& \left|\frac{\psi_j\psi_l}{\alpha_{k+j}}\right| E\left|\varepsilon_{r-k-j}\left(\varepsilon_{r-l}-\varepsilon_{n-l+r}\right)\right| \\
\leq \;& \left|\frac{\psi_j\psi_l}{\alpha_{k+j}}\right|\left(E\left|\varepsilon_{r-k-j}\varepsilon_{r-l}\right| + E\left|\varepsilon_{r-k-j}\varepsilon_{n-l+r}\right|\right) \\
\leq \;& \left|\frac{\psi_j\psi_l}{\alpha_{k+j}}\right|\left(\alpha_{l-k-j}^{1/2}+\alpha_{n-l-k-j}^{1/2}\right).
\end{aligned}
$$

Now summing over j, l, m gives

$$
\begin{aligned}
& E\sqrt{n}\left|\int_{-\pi}^{\pi}\psi\left(e^{-i\lambda}\right)\omega_{\varepsilon}(\lambda)\, n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)\,\varsigma(\lambda)\, d\lambda\right| \\
\leq \;& n^{-1/2}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\left|\frac{\psi_k\psi_l}{\alpha_{k+j}}\right||l|\left(\alpha_{l-k-j}^{1/2}+\alpha_{n-l-k-j}^{1/2}\right) \\
\leq \;& \sup_k 2\alpha_k^{1/2} n^{-1/2}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\left|\frac{\psi_j\psi_l}{\alpha_{k+j}}\right||l| \to 0.
\end{aligned}
$$

This holds true for all $k = 1, ..., p$. Next consider

$$\begin{aligned}
& E\sqrt{n}\left|\int_{-\pi}^{\pi} R_n^1(-\lambda) R_n^2(\lambda) d\lambda\right| \\
= \; & E\sqrt{n}\left|\int_{-\pi}^{\pi} l_\psi(\lambda)\phi\left(e^{-i\lambda}\right) R_n^1(-\lambda) R_n^1(\lambda) d\lambda\right| \\
& + E\sqrt{n}\left|\int_{-\pi}^{\pi} R_n^1(-\lambda) n^{-1/2}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{j+k}} e^{-i\lambda(j+k)} U_{n,j}(\lambda) d\lambda\right|
\end{aligned}$$

such that the first term goes to zero by the proof of Lemma (A.2). The second term then is

$$\begin{aligned}
& E\sqrt{n}\left|\int_{-\pi}^{\pi} R_n^1(-\lambda) n^{-1/2}\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{j+k}} e^{-i\lambda(j+k)} U_{n,j}(\lambda) d\lambda\right| \\
\leq \; & E n^{-1/2}\int_{-\pi}^{\pi}\left|\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\frac{\psi_j\psi_l}{\alpha_{j+k}} e^{-i\lambda(j+k)} U_{n,l}(-\lambda) U_{n,j}(\lambda)\right| d\lambda \\
\leq \; & n^{-1/2}\int_{-\pi}^{\pi}\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\left|\frac{\psi_j\psi_l}{\alpha_{j+k}}\right| E\left|U_{n,l}(-\lambda)\right|\left|U_{n,j}(\lambda)\right| d\lambda \\
\leq \; & n^{-1/2}\int_{-\pi}^{\pi}\sum_{l=0}^{\infty}|\psi_l|\left(E\left|U_{n,l}(\lambda)\right|^2\right)^{1/2}\sum_{j=0}^{\infty}\left|\frac{\psi_j}{\alpha_{j+k}}\right|\left(E\left|U_{n,j}(\lambda)\right|^2\right)^{1/2} d\lambda \\
\leq \; & n^{-1/2}\int_{-\pi}^{\pi}\sum_{l=0}^{\infty}|\psi_l|\min(l,n)^{1/2}\sum_{j=0}^{\infty}\left|\frac{\psi_j}{\alpha_{j+k}}\right|\min(j,n)^{1/2} d\lambda \to 0.
\end{aligned}$$

The remaining terms can be shown to go to zero by the same arguments and the proofs are omitted. This completes the proof of the proposition ■

Next we turn to the proof of adaptiveness. Before proving the main results a few useful lemmas are obtained

Lemma C.1. *$P\left(\left|\min\hat{\alpha}_j(\phi)\min\alpha_j(\phi_0)\right| > cn^{-1/2+\nu}\right) = 1$ for some $0 < \nu < 1/2$ and some $c > 0$.*

Proof. From Assumption (F1) we know that $\min a_j(\phi_0) > \underline{\alpha} > 0$. Also by definition $\hat{\alpha}_j(\phi) \geq d_n$ for all $j = 1, ..., n$. The result follows from the definition of d_n. ■

Lemma C.2. *Let $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with ε_t satisfying assumptions (F1, G1) and y_t observed for $t = 1, ..., n$. Define $T = 2n - 1$, $\lambda_{s_j} = \frac{2\pi s_j}{T}$ where the integers s_j are indexed by an index j, $s_j \in \{0, 1, ..., T\}$ and*

$$\hat{y}_t = \begin{cases} y_t & \text{for } t = 1, ..., n \\ 0 & \text{for } t = n+1, ..., T \end{cases}.$$

Then

$$T^{\frac{-k+1}{2}} \sum_{s_1=1}^{T} \cdots \sum_{s_{k-1}=1}^{T} \omega_{\hat{y}}(\lambda_{s_1}) \cdots \omega_{\hat{y}}(\lambda_{s_{k-1}}) \overline{\omega}_{\hat{y}}\left(\sum \lambda_{s_j}\right) \exp\left(i \sum \lambda_{s_j} t_j\right) \qquad \text{(C.2)}$$

is bounded in probability.

Proof. By Chebychev's inequality it is enough to show that the variance of (C.2) is bounded.

$$\begin{aligned} & var\left[T^{\frac{-k+1}{2}} \sum_{s_1=1}^{T} \cdots \sum_{s_{k-1}=1}^{T} \omega_{\hat{y}}(\lambda_{s_1}) \cdots \overline{\omega}_{\hat{y}}\left(\sum \lambda_{s_j}\right) \exp\left(i \sum \lambda_{s_j} t_j\right)\right] \\ = \; & T^{-2k+1} n^k \sum_{s_1=1}^{T} \cdots \sum_{s_{2k-2}=1}^{T} \exp\left(i \sum \lambda_{s_j} t_j\right) \exp\left(-i \sum \lambda_{s_{j+k-1}} t_j\right) \\ & \times cov\left[\omega_y(\lambda_{s_1}) \cdots \overline{\omega}_y\left(\sum \lambda_{s_j}\right), \omega_y(\lambda_{s_k}) \cdots \overline{\omega}_y\left(\sum \lambda_{s_{j+k-1}}\right)\right] \qquad \text{(C.3)} \end{aligned}$$

where we used $\omega_y(\lambda) = (T/n)^{1/2} \omega_{\tilde{y}}(\lambda_{s_1})$. The covariance term is evaluated as

$$\sum_{v} \prod_{v_j \in v} cum\left(\omega_y\left(\lambda_{s_j}\right); j \in v_i\right)$$

where the sum is over all indecomposable partitions of the table

$$\begin{array}{cccc} \omega_y(\lambda_{s_1}) & \cdots & \omega_y\left(\lambda_{s_{k-1}}\right) & \overline{\omega}_y\left(\sum \lambda_{s_j}\right) \\ \omega_y(\lambda_{s_k}) & \cdots & \omega_y\left(\lambda_{s_{2k-2}}\right) & \overline{\omega}_y\left(\sum \lambda_{s_{j+k-1}}\right) \end{array}$$

and $j \in v_i$ if $\omega_y(\lambda_{s_i})$ is an element of the set v_i of partition v. The notation $cum\left(\omega_y\left(\lambda_{s_j}\right); j \in v_i\right)$ is short hand for the cumulant of all the elements in set v_i of partition v. (See Brillinger [14, p. 20] for details). From Brillinger [14, theorem 2.8.1 and 4.3.1] we have

$$\begin{aligned} & cum\left(\omega_y\left(\lambda_{s_j}\right); j \in v_i\right) \\ = \; & (2\pi)^{\#v_i - 1} n^{-\#v_i/2} \Delta^{(T)}\left(\sum \lambda_{s_j} \{j \in v_i\}\right) f_{y \ldots y}\left(\lambda_{s_j}; j \in v_i\right) + O\left(n^{\frac{-\#v_i}{2}}\right) \end{aligned}$$

where $\#$ stands for cardinality and $\Delta^{(T)}(\lambda) = \sum e^{i\lambda t}$. Substituting back into (C.3) leads to

$$\begin{aligned} & T^{-2k+1} \sum_{s_1=1}^{T} \cdots \sum_{s_{2k-2}=1}^{T} \exp\left(i \sum \lambda_{s_j} t_j\right) \exp\left(-i \sum \lambda_{s_{j+k-1}} t_j\right) \\ & \times \sum_{v} \prod_{v_j \in v} \left[(2\pi)^{\#v_i - 1} \Delta^{(T)}\left(\sum \lambda_{s_j} \{j \in v_i\}\right) f_{y \ldots y}\left(\lambda_{s_j}; j \in v_i\right) + O(1)\right]. \end{aligned} \tag{C.4}$$

Since the $\lambda_{s_j} \neq 0 \bmod T$ the largest number of combinations $\sum \lambda_{s_j} \{j \in v_i\} = 0 \bmod T$ is k. The error term is at most of order $O(n^{k-1})$ in which case the number of sums is reduced to $k-1$. Therefore

$$\prod_{v_j \in v} \left[(2\pi)^{\#v_i - 1} \Delta^{(T)} \left(\sum \lambda_{s_j} \{ j \in v_i \} \right) f_{y \dots y} \left(\lambda_{s_j}; j \in v_i \right) + O(1) \right] = O(T^k) \qquad \text{(C.5)}$$

for all partitions v and $\left| f_{y \dots y} \left(\lambda_{s_j}; j \in v_i \right) \right|$ is bounded because of assumption (G1). If (C.5) is of order T^k then the $i\lambda_{s_j}$ appear in complex pairs reducing the number of outer summations by half such that for these terms (C.4) is $O(1)$. If (C.5) is of order T^{k-l} then the summation runs over $k-2+l$ terms such that the order of (C.4) is $O(T^{-1})$. Together these results imply that (C.3) is $O(1)$ ■

Lemma C.3. *Let* $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *with* ε_t *satisfying assumptions (F1, G1) then for any integers* $t_1 \dots t_{k-1} \in \{1, ..., n-1\}$

$$\begin{aligned} & \sum_{t=\max(t_1,\dots,t_{k-1})}^{n} y_t y_{t-t_1} \cdots y_{t-t_k} \\ = \; & T^{\frac{-k}{2}+1} \sum_{s_1=1}^{T} \cdots \sum_{s_{k-1}=1}^{T} \omega_{\tilde{y}} \left(\lambda_{s_1} \right) \cdots \omega_{\tilde{y}} \left(\lambda_{s_{k-1}} \right) \overline{\omega}_{\tilde{y}} \left(\sum \lambda_{s_j} \right) \\ & \times \exp \left(i \sum \lambda_{s_{j+k-1}} t_j \right) + O_p \left(n^{-1/2} \right) \end{aligned}$$

where the error is uniform in $t_1, ..., t_{k-1}$.

Proof. From Brillinger [14, lemma 3.6.2] it follows that

$$\begin{aligned} & \sum_{t=\max(t_1,\dots,t_{k-1})}^{n} y_t y_{t-t_1} \cdots y_{t-t_{k-1}} \\ = \; & T^{\frac{-k}{2}+1} \sum_{s_1=0}^{T} \cdots \sum_{s_{k-1}=0}^{T} \omega_{\tilde{y}} \left(\lambda_{s_1} \right) \cdots \overline{\omega}_{\tilde{y}} \left(\sum \lambda_{s_j} \right) \exp \left(i \sum \lambda_{s_j} t_j \right) \end{aligned}$$

so it remains to be shown that

$$T^{\frac{-k}{2}+1}\sum_{l=1}^{k}\sum_{s_m=0,m\neq l}^{T}\cdots\sum^{T}\omega_{\hat{y}}\left(\lambda_{s_1}\right)\cdots\omega_{\hat{y}}\left(\lambda_{s_{l-1}}\right) \tag{C.6}$$

$$\times\omega_{\hat{y}}\left(\lambda_{s_{l+1}}\right)\cdots\overline{\omega}_{\hat{y}}\left(\sum\lambda_{s_j}\right)\omega_{\hat{y}}\left(0\right)\exp\left(i\sum\lambda_{s_{j-1}}t_j\right)$$

$$= O_p\left(n^{-1/2}\right).$$

This argument can be applied recursively to obtain expressions where $1, 2, ...k$ different λ_{s_j} are zero. Here only the first step of the recursion is discussed since the others follow from the same argument. The recursion stops at k where (C.6) reduces to

$$T^{-k+1}\sum_{t_1}\cdots\sum_{t_k}y_{t_1}\cdots y_{t_k} = O_p\left(T^{-\frac{k}{2}+1}\right).$$

Now for the first step (C.6) can be written as

$$T^{-1}\sum\hat{y}_t T^{\frac{-k+1}{2}}\sum_{l=1}^{k}\sum_{s_m=0,m\neq l}^{T}\cdots\sum^{T}\omega_{\hat{y}}\left(\lambda_{s_1}\right)\cdots\omega_{\hat{y}}\left(\lambda_{s_{l-1}}\right)$$

$$\times\omega_{\hat{y}}\left(\lambda_{s_{l+1}}\right)\cdots\overline{\omega}_{\hat{y}}\left(\sum\lambda_{s_j}\right)\exp\left(i\sum\lambda_{s_{j-1}}t_j\right).$$

The inner summation is again split into a part which involves only $\lambda_{s_m} > 0$ and a part with zero terms. Then by Lemma C.2 and because the summation is now only $k-1$ fold

$$T^{\frac{-k+1}{2}}\sum_{s_m=1,m\neq l}^{T}\cdots\sum^{T}\omega_{\hat{y}}\left(\lambda_{s_1}\right)\cdots\omega_{\hat{y}}\left(\lambda_{s_{l-1}}\right)$$

$$\times\omega_{\hat{y}}\left(\lambda_{s_{l+1}}\right)\cdots\overline{\omega}_{\hat{y}}\left(\sum\lambda_{s_j}\right)\exp\left(i\sum\lambda_{s_{j-1}}t_j\right)$$

$$= O_p(n^{-1/2}).$$

Also by a standard CLT $T^{-1}\sum \hat{y}_t = O_p\left(n^{-1/2}\right)$ so that all terms with only one $\lambda_{s_j} = 0$ are $O_p\left(n^{-1/2}\right)$. All other terms are of lower order. The argument in the proof of C.4 can then be used to show that the result holds uniformly over $t_1, ..., t_{k-1}$. ■

The next lemma establishes uniform convergence of the estimates $\hat{\alpha}_l(\phi)$.

Lemma C.4. *For any $\varepsilon > 0$ there exists some $\delta, \eta > 0$ such that*

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\phi\in N_\delta(\phi_0)} \max_l n^{1/2} \left|\hat{\alpha}_l(\phi) - \hat{\alpha}_l(\phi_0)\right| > \eta\right) < \varepsilon$$

Proof. We first show that the result only needs to be established for the untruncated estimator $\alpha_l^*(\phi)$. Define the set $A_n = \left\{\omega \in \Omega \left| \sup_{\phi\in N_\delta(\phi_0)} \min_l \left|\alpha_l^*(\phi)\right| > d_n \right.\right\}$ and let A_n^c be the complement of A_n

$$\begin{aligned}
& \overline{\lim_{n\to\infty}} P\left(\sup_{\phi\in N_\delta(\phi_0)} \max_l n^{1/2}\left|\hat{\alpha}_l(\phi) - \alpha_l(\phi_0)\right| > \eta\right) \\
\leq\ & \overline{\lim_{n\to\infty}} P\left(\sup_{\phi\in N_\delta(\phi_0)} \max_l n^{1/2}\left|\hat{\alpha}_l(\phi) - \alpha_l(\phi_0)\right| > \eta,\ A_n\right) \\
& + \overline{\lim_{n\to\infty}} P\left(\sup_{\phi\in N_\delta(\phi_0)} \max_l n^{1/2}\left|\hat{\alpha}_l(\phi) - \alpha_l(\phi_0)\right| > \eta,\ A_n^c\right) \\
\leq\ & \overline{\lim_{n\to\infty}} P\left(\sup_{\phi\in N_\delta(\phi_0)} \max_l n^{1/2}\left|\alpha_l^*(\phi) - \alpha_l(\phi_0)\right| > \eta\right) + \overline{\lim_{n\to\infty}} P\left(A_n^c\right)
\end{aligned}$$

where the second inequality uses the fact, that $\hat{\alpha}_l(\phi) = \alpha_l^*(\phi)$ on the set where $\min_l \left|\alpha_l^*(\phi)\right| > d_n$. Yet $|\delta| \leq \left|a_l(\phi_0)\right| \leq \left|a_l^*(\phi) - a_l(\phi_0)\right| + \left|a_l^*(\phi)\right|$ so that $\overline{\lim_{n\to\infty}} P(A_n^c) \to 0$ follows from $\max_l \left|a_l^*(\phi) - a_l(\phi_0)\right| \xrightarrow{p} 0$. Define $y_{-1,t} = (y_{t-1}, \ldots, y_{t-p})$. Then

$$\sup_{\phi\in N_\delta(\phi_0)} \max_l \left|\alpha_l^*(\phi) - \alpha_l(\phi_0)\right| = \sup_{\phi\in N_\delta(\phi_0)} \max_l \left|\frac{1}{n}\sum \varepsilon_t(\phi)^2 \varepsilon_{t-l}(\phi)^2 - \alpha_l(\phi_0)\right|.$$

Using $\varepsilon_t(\phi) = \varepsilon_t(\phi_0) + (\phi - \phi_0)'(y_{t-1}, ..., y_{t-p})$ and collecting terms leads to

$$n^{1/2} \sup_{\phi \in N_\delta(\phi_0)} \max_l \left| \frac{1}{n} \sum \varepsilon_t^2(\phi_0)\, \varepsilon_{t-l}^2(\phi_0) + 2\varepsilon_t(\phi_0)\, \varepsilon_{t-l}^2(\phi_0)(\phi - \phi_0)' y_{-1,t} \right.$$
$$+ \left[(\phi - \phi_0)' y_{-1,t}\right]^2 \varepsilon_{t-l}^2(\phi_0) + 2(\phi - \phi_0)' y_{-1,t-l}\varepsilon_{t-l}(\phi_0)\, \varepsilon_t^2(\phi_0)$$
$$+4\left[(\phi - \phi_0)' y_{-1,t}\right]^2 \varepsilon_t(\phi_0)\, \varepsilon_{t-l}(\phi_0)$$
$$+2(\phi - \phi_0)' y_{-1,t-j}\varepsilon_{t-l}(\phi_0)\left[(\phi - \phi_0)' y_{-1,t}\right]^2 + \left[(\phi - \phi_0)' y_{-1,t-l}\right]^2 \varepsilon_t^2(\phi_0) +$$
$$+2(\phi - \phi_0)' y_{-1,t}\varepsilon_t(\phi_0)\left[(\phi - \phi_0)' y_{-1,t-l}\right]^2$$
$$\left. + \left[(\phi - \phi_0)' y_{-1,t}\right]^2 \left[(\phi - \phi_0)' y_{-1,t-l}\right]^2 - \alpha_l(\phi_0) \right|.$$

Now from the Cauchy-Schwartz inequality

$$\left[(\phi - \phi_0)' y_{-1,t}\right]^2 \leq (\phi - \phi_0)'(\phi - \phi_0)\, y_{-1,t}' y_{-1,t}$$

and the fact that

$$\sup_{\phi \in N_\delta(\phi_0)} (\phi - \phi_0)'(\phi - \phi_0) = \delta^2$$

the above expression is dominated by

$$\leq \max_l \left| n^{-1/2} \sum \varepsilon_t^2(\phi_0)\, \varepsilon_{t-l}^2(\phi_0) - \alpha_l(\phi_0) \right| \quad \text{(C.7)}$$
$$+\delta^2 \max_j \left| n^{-1/2} \sum y_{-1,t}' y_{-1,t}\varepsilon_{t-l}^2(\phi_0) + y_{-1,t-l}' y_{-1,t-l}\varepsilon_t^2(\phi_0) \right.$$
$$\left. +4y_{-1,t}' y_{-1,t-l}\varepsilon_t(\phi_0)\, \varepsilon_{t-l}(\phi_0) \right|$$
$$+\delta^2 \sup_{\phi \in N_\delta(\phi_0)} \left| n^{-1/2} \sum 2(\phi - \phi_0)' y_{-1,t-l}\varepsilon_{t-l}(\phi_0)\, y_{-1,t}' y_{-1,t} \right.$$
$$+2\varepsilon_t(\phi_0)\, \varepsilon_{t-l}^2(\phi_0)(\phi - \phi_0)' y_{-1,t} + 2(\phi - \phi_0)' y_{-1,t-l}\varepsilon_{t-l}(\phi_0)\, \varepsilon_t^2(\phi_0)$$

$$+2\left(\phi-\phi_0\right)' y_{-1,t}\varepsilon_t\left(\phi_0\right) y_{-1,t-l}' y_{-1,t-l}\Big|$$

$$+\delta^2 \max_j \left| n^{-1/2} \sum y_{-1,t}' y_{-1,t} y_{-1,t-l}' y_{-1,t-l} \right|.$$

From

$$\begin{aligned}
& \sup_{\phi \in N_\delta(\phi_0)} \left| n^{-1/2} \sum_t^n \left(\phi-\phi_0\right)' y_{-1,t-j}\varepsilon_{t-l}\left(\phi_0\right) y_{-1,t}' y_{-1,t} \right| \\
= & \sup_{\phi \in N_\delta(\phi_0)} \left| n^{-1/2} \sum_t^n \sum_i^p \left(\phi_i-\phi_{i,0}\right) y_{t-l-i}\varepsilon_{t-l}\left(\phi_0\right) y_{-1,t}' y_{-1,t} \right| \\
\leq & \ \delta \sum_i^p \left| n^{-1/2} \sum_t^n y_{t-l-i}\varepsilon_{t-l}\left(\phi_0\right) y_{-1,t}' y_{-1,t} \right|
\end{aligned}$$

all the terms in (C.7) multiplied by δ^2 are of the form $\left|n^{-1/2}\sum_t y_{t-t_1}y_{t-t_2}y_{t-t_3}\varepsilon_t\right|$ with the number of ε_t terms varying from one to three. The indices t_1, t_2, t_3 are of general form $t_i(l) = l + c_i$ or $t_i(l) = c_i$ where the dependence on l will be dropped for notational convenience. It is important to point out that in all the fourth moment expressions there is always at least one term which does not depend on l. It remains to establish that these terms are $O_p(1)$ uniformly in t_1, t_2, t_3 as a function of l. From Lemma (C.3) and Brillinger [14, theorem 2.8.1 and 4.3.1] it follows that

$$\begin{aligned}
& n^{-1/2} \sum_{t=1}^n y_{t-t_1}y_{t-t_2}y_{t-t_3}\varepsilon_t \\
= & \ n^{-1/2}T^{-1} \sum_{s_1=1}^T \cdots \sum_{s_3=1}^T \omega_{\hat{y}}\left(\lambda_{s_1}\right)\cdots\omega_{\hat{y}}\left(\lambda_{s_3}\right)\omega_{\hat{\varepsilon}}\left(\sum \lambda_{s_j}\right) \\
& \times \exp\left(i\sum \lambda_{s_j} t_j\right) + O_p\left(n^{-1}\right)
\end{aligned}$$

where the error is uniform in t_1, t_2, t_3. Then by the arguments in the proof of Lemma (C.2)

$$\begin{aligned}
& var\left(n^{-1/2}\sum_{t=1}^{n} y_{t-t_1}y_{t-t_2}y_{t-t_3}\varepsilon_t\right) \\
= \; & n^{-1}T^{-6}\sum_{s_1=1}^{T}\cdots\sum_{s_{2k-2}=1}^{T}\exp\left(i\sum\lambda_{s_j}t_j\right)\exp\left(-i\sum\lambda_{s_{j+3}}t_j\right) \\
& \times\sum_{v}\prod_{v_j\in v}\left[(2\pi)^{q-1}\Delta^{(T)}\left(\sum\lambda_{s_j}\{j\in v_i\}\right)f_{x_1\ldots x_q}\left(\lambda_{s_j};j\in v_i\right)+O(1)\right]
\end{aligned} \tag{C.8}$$

where the summation runs over all indecomposable partitions of the table

$$\begin{array}{cccc}
\omega_y(\lambda_{s_1}) & \omega_y(\lambda_{s_2}) & \omega_y(\lambda_{s_3}) & \overline{\omega}_\varepsilon\left(\sum\lambda_{s_j}\right) \\
\omega_y(\lambda_{s_4}) & \omega_y(\lambda_{s_5}) & \omega_y(\lambda_{s_6}) & \overline{\omega}_\varepsilon\left(\sum\lambda_{s_{j+3}}\right)
\end{array}$$

and $f_{x_1\ldots x_q}\left(\lambda_{s_j};j\in v_i\right)$ is the q-th order cumulant of the variables x_j. x_j denotes the variable associated with entry j in the above table if $j\in v_i$. From Brillinger [14, theorem 2.8.1] $f_{x_1\ldots x_q}\left(\lambda_{s_j};j\in v_i\right)$ can be expressed in terms of the cumulant spectrum of ε_t. Let $C(\lambda)=\sum\psi_j e^{-i\lambda j}$ and assume w.l.g. that the y-variables are numbered $1,\ldots,6$ then

$$f_{x_1\ldots x_8}\left(\lambda_{s_j};j\in v_i\right)=f_{\varepsilon\ldots\varepsilon}\left(\lambda_{s_1},\ldots,\lambda_{s_7}\right)\prod_{i=1}^{6}C\left(\lambda_{s_j}\right)\{j\in v_i\}.$$

From

$$\begin{aligned}
& \overline{\lim_{n\to\infty}}\,P\left(\max_l n^{-1/2}\sup\left|a_l^*(\phi)-\alpha_l(\phi_0)\right|>\eta\right) \\
\leq \; & \overline{\lim_{n\to\infty}}\sum_{l=1}^{n}P\left(n^{-1/2}\sup\left|a_l^*(\phi)-\alpha_l(\phi_0)\right|>\eta\right) \\
\leq \; & \overline{\lim_{n\to\infty}}\sum_{l=1}^{n}var\left(n^{-1/2}\sum\left(\varepsilon_t^2(\phi_0)\varepsilon_{t-l}^2(\phi_0)-\alpha_l(\phi_0)\right)\right)
\end{aligned}$$

$$+\overline{\lim_{n\to\infty}}\sum_{l}^{n}\sum_{i} var\left(n^{-1/2}\delta^{2}\sum_{t=1}^{n}\prod_{m=1}^{4}C^{m,i}(L)\,\varepsilon_{t-t_m}\right)$$

where the sum in the last line is over all terms appearing in $\alpha_l^*(\phi)$ except $\varepsilon_t^2(\phi_0)\,\varepsilon_{t-j}^2(\phi_0)$ and $C^{m,i}(L) = C(L)$ if the m-th term is y_{t-t_m} and $C^{m,i}(L) = 1$ if the m-th term is ε_{t-t_m}. The result now follows if

$$\overline{\lim_{n\to\infty}}\sum_{l}^{n} var\left(n^{-1/2}\sum_{t=1}^{n}\prod_{m=1}^{4}C^{m,i}(L)\,\varepsilon_{t-t_m}\right) < \infty,\ \forall i$$

and

$$\overline{\lim_{n\to\infty}}\sum_{l=1}^{n} var\left(n^{-1/2}\sum\left(\varepsilon_t^2(\phi_0)\,\varepsilon_{t-l}^2(\phi_0) - \alpha_l(\phi_0)\right)\right) < \infty.$$

However from (C.8) we have

$$\begin{aligned}
&\sum_{l}^{n} var\left(n^{-1/2}\sum_{t=1}^{n}\prod_{m=1}^{4}C^{m,i}(L)\,\varepsilon_{t-t_m}\right)\\
=\ & n^{-1}T^{-6}\sum_{l=1}^{n}\sum_{s_1=1}^{T}\cdots\sum_{s_{2k-2}=1}^{T}\exp\left(i\sum\lambda_{s_j}t_j\right)\exp\left(-i\sum\lambda_{s_{j+3}}t_j\right) \qquad \text{(C.9)}\\
&\times\sum_{v}\prod_{v_j\in v}\left[(2\pi)^{q-1}\Delta^{(T)}\left(\sum\lambda_{s_j}\{j\in v_i\}\right)f_{x_1\ldots x_q}\left(\lambda_{s_j};j\in v_i\right)+O(1)\right].
\end{aligned}$$

The higher order cumulant spectra are defined in terms of higher order cumulants in the following way

$$f_{x_1\ldots x_q}\left(\lambda_{s_j};j\in v_i\right) = (2\pi)^{-q+1}\sum_{u_1,\ldots,u_{q-1}} c_{x_1,\ldots,x_q}(u_1,\ldots,u_{q-1})\exp\left(-i\sum u_r\lambda_r\right)$$

where in $\sum u_r \lambda_r$ we assign each u_i to one λ_{s_j}. Now substituting back intoC.9 leads to

$$\begin{aligned} &= n^{-1}T^{-6}\sum_{l=1}^{n}\sum_{v}\left\{\prod_{v_i\in v}\left[\sum_{u_1,\ldots,u_q} c_{x_1,\ldots,x_q}\left(u_1,\ldots,u_{q-1}\right)\sum_{s_j;j\in v_i}\Delta^{(T)}\left(\sum\lambda_{s_j}\right)\right.\right. \\ &\quad \left.\left.\times\exp\left(-i\sum\left[\left(u_r-t_r\right)\{r\le 3\}+\left(u_r+t_r\right)\{r>3\}\right]\lambda_{s_r}\right)\right]+O\left(T^6\right)\right\} \\ &= \sum_{l=1}^{n}\sum_{v}\prod_{v_j\in v} c_{x_1,\ldots,x_q}\left(t_j;j\in v_i\right)+O\left(1\right). \end{aligned}$$

The last line follows from the fact that each $\Delta^{(T)}\left(\sum\lambda_{s_j}\right)$ which is nonzero puts an additional restriction on the λ_{s_j}. The maximal number of sets in a partition is k. If in the largest partition all $k-1$ restrictions are valid then the k-th restriction is automatically satisfied as well since always $\sum_{j=1}^{k-1}\lambda_{s_j}=\lambda_{s_k}$. As a consequence the number of summations is reduced to $k-1$ while k terms of the form $\Delta^{(T)}\left(\sum\lambda_{s_j}\right)$ are equal to T. Therefore the leading term of the product is $O\left(T^{2k-1}\right)$. The $O\left(T^6\right)$ error term comes from multiplying out the product in (C.9). Keeping in mind the dependence of t_j on l we can now state that

$$\begin{aligned} &\sum_{l=1}^{n}\sum_{v}\prod_{v_j\in v}\left|c_{x_1,\ldots,x_q}\left(t_j;j\in v_i\right)\right| \\ &\le \sum_{v}\prod_{v_j\in v}\sum_{l=1}^{n}\left|c_{x_1,\ldots,x_q}\left(t_j;j\in v_i\right)\right|\le\sum_{v}\prod_{v_j\in v}\sum_{l=1}^{\infty}\left|c_{x_1,\ldots,x_q}\left(t_j;j\in v_i\right)\right|<\infty \end{aligned}$$

The same argument can be applied to the remaining terms in (C.7). Also

$$var\left(n^{-1/2}\sum\varepsilon_t^2\left(\phi_0\right)\varepsilon_{t-l}^2\left(\phi_0\right)-\alpha_l\left(\phi_0\right)\right)=O\left(1\right)$$

uniformly in l by the same argument. Then again the above argument can be applied to show that we can sum over l ■

Proof of Proposition 9.1 For the first part of the proposition we have to show that for any $\eta, \epsilon > 0$

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\lambda\in[-\pi,\pi]} \left\|\hat{h}_n\left(\hat{\phi}_n, \lambda\right) - h\left(\phi_0, \lambda\right)\right\| > \eta\right) < \varepsilon.$$

This holds if there is a δ and a neighborhood $N_\delta\left(\phi_0\right)$ of ϕ_0 such that $N_\delta\left(\phi_0\right)$ is contained in the interior of the stationary region of the parameter space and

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\lambda\in[-\pi,\pi]} \sup_{\phi\in N_\delta(\phi_0)} \left\|\hat{h}_n\left(\phi, \lambda\right) - h\left(\phi_0, \lambda\right)\right\| > \eta\right) + \overline{\lim_{n\to\infty}} P\left(\hat{\phi} \notin N_\delta\left(\phi_0\right)\right) \leq \epsilon.$$

Then consistency implies $P\left(\hat{\phi} \in N_\delta\left(\phi_0\right)\right) \to 1$ so that only the first term needs to be considered. From

$$\begin{aligned}
& \left\|\hat{h}_n\left(\phi, \lambda\right) - h\left(\phi_0, \lambda\right)\right\| \\
= & \left\|\mathrm{Re}\left[\hat{l}_\psi\left(\lambda\right)\phi\left(e^{-i\lambda}\right)\right] - \mathrm{Re}\left[l_{\psi,0}\left(\lambda\right)\phi_0\left(e^{-i\lambda}\right)\right]\right\| \\
\leq & \frac{1}{2}\left\|\hat{l}_\psi\left(\lambda\right)\phi\left(e^{-i\lambda}\right) - l_{\psi,0}\left(\lambda\right)\phi_0\left(e^{-i\lambda}\right)\right\| + \frac{1}{2}\left\|\hat{l}_\psi\left(-\lambda\right)\phi\left(e^{i\lambda}\right) - l_{\psi,0}\left(-\lambda\right)\phi_0\left(e^{i\lambda}\right)\right\|
\end{aligned}$$

and

$$\begin{aligned}
& \left\|\hat{l}_\psi\left(\lambda\right)\phi\left(e^{-i\lambda}\right) - l_{\psi,0}\left(\lambda\right)\phi_0\left(e^{-i\lambda}\right)\right\| \\
\leq & \left\|\hat{l}_\psi\left(\lambda\right) - l_{\psi,0}\left(\lambda\right)\right\| \left|\phi\left(e^{-i\lambda}\right)\right| + \left\|l_{\psi,0}\left(\lambda\right)\right\| \left|\phi\left(e^{-i\lambda}\right) - \phi_0\left(e^{-i\lambda}\right)\right|
\end{aligned}$$

it is enough to show that $\sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \hat{l}_\psi(\lambda) - l_{\psi,0}(\lambda) \right\| = o_p(1)$, while $\left|\phi\left(e^{i\lambda}\right) - \phi_0\left(e^{i\lambda}\right)\right| < \eta/2$ on $N_\delta(\phi_0)$ by uniform continuity of $\phi_0\left(e^{i\lambda}\right)$ on $[-\pi, \pi]$. Now

$$
\begin{aligned}
& \sup_{\lambda \in [-\pi,\pi]} \sup_{\phi \in N_\delta(\phi_0)} \left\| \hat{l}_\psi(\lambda) - l_{\psi,0}(\lambda) \right\| \\
= & \sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \sum_{j=1}^{n-p-1} \hat{\alpha}_j(\phi)^{-1} b_j(\phi) e^{-i\lambda j} - \sum_{j=1}^{\infty} \alpha_j(\phi_0)^{-1} b_j(\phi_0) e^{-i\lambda j} \right\| \\
\leq & \sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \sum_{j=1}^{n-p-1} \hat{\alpha}_j(\phi)^{-1} b_j(\phi) e^{-i\lambda j} - \sum_{j=1}^{n-p-1} \alpha_j(\phi_0)^{-1} b_j(\phi_0) e^{-i\lambda j} \right\| \\
& + \sup_\lambda \left\| \sum_{j \geq n-p}^{\infty} \alpha_j(\phi_0)^{-1} b_j(\phi_0) e^{-i\lambda j} \right\|.
\end{aligned}
$$

We note that $\sup_\lambda \left\| \sum_{j \geq n-p}^{\infty} \alpha_j(\phi_0)^{-1} b_j(\phi_0) e^{-i\lambda j} \right\| \leq \sup_j \alpha_j(\phi_0)^{-1} n^{-1/2} \sum_{j \geq n-p}^{\infty} j^{1/2} \left\| b_j(\phi_0) \right\| = o(n^{-1/2})$. Next

$$
\begin{aligned}
& \sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \sum_{j=1}^{n-p-1} \hat{\alpha}_j(\phi)^{-1} b(\phi)_j e^{-i\lambda j} - \sum_{j=1}^{n-p-1} \alpha_j(\phi_0)^{-1} b_j(\phi_0) e^{-i\lambda j} \right\| \\
\leq & \sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \sum_{j=1}^{n-p-1} \left(\hat{\alpha}_j(\phi)^{-1} - \alpha_j(\phi_0)^{-1} \right) b_j(\phi) e^{-i\lambda j} \right\| \qquad \text{(C.10)} \\
& + \sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \sum_{j=1}^{n-p-1} \alpha_j(\phi_0)^{-1} \left(b_j(\phi) - b_j(\phi_0) \right) e^{-i\lambda j} \right\|. \qquad \text{(C.11)}
\end{aligned}
$$

It is therefore enough to show that (C.10) and (C.11) go to zero in probability uniformly on $N_\delta(\phi_0)$ as $\delta \to 0$. First consider (C.11). Let $b_{j,k}(\phi)$ denote the k-th element of $b_j(\phi)$. Then the norm can be bounded by

$$
\left\| \sum_{j=1}^{n-p-1} \alpha_j(\phi_0)^{-1} \left(b_j(\phi) - b_j(\phi_0) \right) e^{-i\lambda j} \right\|
$$

$$
\begin{aligned}
&= \left(\sum_{k=1}^{p} \left| \sum_{j=1}^{n-p-1} \alpha_j (\phi_0)^{-1} (b_{j,k}(\phi) - b_{j,k}(\phi_0)) e^{-i\lambda j} \right|^2 \right)^{1/2} \\
&\leq \sum_{k=1}^{p} \left| \sum_{j=1}^{n-p-1} \alpha_j (\phi_0)^{-1} (b_{j,k}(\phi) - b_{j,k}(\phi_0)) e^{-i\lambda j} \right| .
\end{aligned}
$$

For the k -th element of the sum we have

$$
\begin{aligned}
& \sup_{\lambda \in [-\pi,\pi]} \left| \int_{-\pi}^{\pi} \sum_{j=1}^{n-p-1} \alpha_j (\phi_0)^{-1} e^{-i(\lambda-\mu)j} (\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi,\mu)) \, d\mu \right| \\
\leq & \sup_{\lambda \in [-\pi,\pi]} \left| \int_{-\pi}^{\pi} \sum_{j=1}^{n-p-1} \left(\alpha_j (\phi_0)^{-1} - \frac{1}{\sigma^4} \right) e^{-i(\lambda-\mu)j} (\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi,\mu)) \, d\mu \right| \\
& + \frac{1}{\sigma^4} \sup_{\lambda \in [-\pi,\pi]} \left| \sum_{j=1}^{n-p-1} (b_{j,k}(\phi) - b_{j,k}(\phi_0)) e^{-i\lambda j} \right| \\
\leq & \sup_{\lambda \in [-\pi,\pi]} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-p-1} \tilde{\alpha}_j (\phi_0)^{-1} e^{-i(\lambda-\mu)j} \right| |\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi,\mu)| \, d\mu \\
& + \frac{1}{\sigma^4} \sup_{\lambda \in [-\pi,\pi]} \left| \sum_{j=1}^{n-p-1} (b_{j,k}(\phi) - b_{j,k}(\phi_0)) e^{-i\lambda j} \right| \\
\leq & \sum_{j=-n+p+1}^{n-p-1} \left| \tilde{\alpha}_j (\phi_0)^{-1} \right| \int_{-\pi}^{\pi} |\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi,\mu)| \, d\mu \\
& + \frac{1}{\sigma^4} \sup_{\lambda \in [-\pi,\pi]} \left| \sum_{j=1}^{n-p-1} (b_{j,k}(\phi) - b_{j,k}(\phi_0)) e^{-i\lambda j} \right| \\
< & \; C\epsilon(\delta) + \epsilon(n) .
\end{aligned}
$$

Then $\sum_{j=1}^{n-p-1} \left| \tilde{\alpha}_j (\phi_0)^{-1} \right| < \sum_{j=1}^{\infty} \left| \tilde{\alpha}_j (\phi_0)^{-1} \right| < \infty$. By uniform continuity of $\dot{\eta}(\phi,\mu)$ on $[-\pi,\pi] \times N_\delta(\phi_0)$ we have $|\dot{\eta}(\phi,\mu) - \dot{\eta}(\phi_0,\mu)|_k < \epsilon$ for some $\delta > 0$ from which it follows that $\int_{-\pi}^{\pi} |\dot{\eta}(\phi,\mu) - \dot{\eta}(\phi_0,\mu)|_k \, d\mu < 2\pi\epsilon$. Also for $\phi \in N_\delta(\phi_0)$ the finite Fourier approximation of $\dot{\eta}(\phi,\mu)$ converges uniformly

$$\begin{aligned}
&\sup_{\lambda\in[-\pi,\pi]}\left|\sum_{j=1}^{n-p-1}\left(b_{j,k}(\phi)-b_{j,k}(\phi_0)\right)e^{-i\lambda j}\right| \\
=&\sup_{\lambda\in[-\pi,\pi]}\left|\sum_{j=n-p}^{\infty}\left(b_{j,k}(\phi)-b_{j,k}(\phi_0)\right)e^{-i\lambda j}+\left(\dot{\eta}(\phi,\lambda)-\dot{\eta}(\phi_0,\lambda)\right)_k\right| \\
\leq&\sup_{\lambda\in[-\pi,\pi]}\left|\dot{\eta}(\phi,\lambda)-\dot{\eta}(\phi_0,\lambda)\right|_k+2\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=n-p}^{\infty}\left|b_{j,k}(\phi)\right| \\
\leq&\ \epsilon+2\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=n-p}^{\infty}\left|b_{j,k}(\phi)\right|
\end{aligned}$$

so that the constant C can be bounded by $\left(2\pi\sum_{j=1}^{\infty}\left|\tilde{a}_j(\phi_0)^{-1}\right|+\frac{1}{\sigma^4}\right)$ independent of n and $\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=n-p}^{\infty}\left|b_{j,k}(\phi)\right|$ goes to zero as $n\to\infty$. Now turn to (C.10) which can be analyzed element by element as before

$$\begin{aligned}
&\sup_{\lambda\in[-\pi,\pi]}\sup_{\phi\in N_\delta(\phi_0)}\left|\sum_{j=1}^{n-p-1}\left(\hat{\alpha}_j(\phi)^{-1}-\alpha_j(\phi_0)^{-1}\right)b_{j,k}(\phi)e^{-i\lambda j}\right| \\
\leq&\sup_{\lambda\in[-\pi,\pi]}\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=1}^{n-p-1}\left|\hat{\alpha}_j(\phi)^{-1}-\alpha_j(\phi_0)^{-1}\right|\left|b_{j,k}(\phi)\right| \\
\leq&\sup_{\phi\in N_\delta(\phi_0)}\sup_{j}\left(\left|\hat{\alpha}_j(\phi)^{-1}-\alpha_j(\phi_0)^{-1}\right|\right)\sum_{j=1}^{n-p-1}\left|b_{j,k}(\phi)\right|.
\end{aligned}$$

Now $\sup_j\left(\left|\hat{\alpha}_j(\phi)^{-1}-\alpha_j(\phi_0)^{-1}\right|\right)\leq\left(\min\hat{\alpha}_l(\phi)\min\alpha_l(\phi_0)\right)^{-1}\max\left|\hat{\alpha}_l(\phi)-\alpha_l(\phi_0)\right|$ and $\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=1}^{n-p-1}\left|b_{j,k}(\phi)\right|<\infty$ for all n. The result then follows if for any $\epsilon>0$ there exists $\eta,\delta>0$ such that

$$\lim_{n\to\infty}P\left(\sup_{\phi\in N_\delta(\phi_0)}\left(\min\hat{\alpha}_l(\phi)\min\alpha_l(\phi_0)\right)^{-1}\max\left|\hat{\alpha}_l(\phi)-\alpha_l(\phi_0)\right|>\eta\right)<\epsilon.$$

The probability can be bounded by

$$P\left(\sup_{\phi\in N_\delta(\phi_0)} n^{1/2-\nu}\max_l|\hat{\alpha}_l(\phi)-\alpha_l(\phi_0)|>\eta\right)$$
$$+P\left(\sup_{\phi\in N_\delta(\phi_0)}\left(\min\hat{\alpha}_l(\phi)\min\alpha_l(\phi_0)\right)^{-1}<cn^{-1/2+\nu}\right)$$

The second term is zero by Lemma (C.1) and the first term tends to zero by Lemma (C.4).

It remains to show that $P\left(\rho\left(\hat{h}_n(\hat{\phi}_n,\lambda),h(\phi_0,\lambda)\right)>\delta\right)\to 0$. Using the result from the first part it remains to be shown that $\sup_{\lambda\in[-\pi,\pi]}\left\|\partial\hat{h}_n\left(\hat{\phi}_n,\lambda\right)/\partial\lambda-\partial h(\phi_0,\lambda)/\partial\lambda\right\|=o_p(1)$. Since

$$\frac{\partial^h\dot{\eta}(\phi,\lambda)_k}{\partial\lambda^h}=(ik)^h\,\dot{\eta}(\phi,\lambda)_k$$

implies that $\dot{\eta}(\phi,\lambda)_k\in C^h[-\pi,\pi]\ \forall h<\infty$ and $\partial^h\dot{\eta}(\phi,-\pi)_k/\partial\lambda^h=\partial^h\dot{\eta}(\phi,\pi)_k/\partial\lambda^h$ such that by the Riemann-Lebesgue lemma $\sum j^{h-3/2}|b_j|<\infty$ or in particular $\sum j^{3/2}|b_j|<\infty$. Also $(ij)b_j(\phi)=\int\left(\frac{\partial}{\partial\lambda}\dot{\eta}(\phi,\lambda)\right)e^{-i\lambda j}d\lambda$ such that by the Weierstrass Theorem

$$\sum(ij)b_j(\phi)\,e^{i\lambda j}\to\frac{\partial}{\partial\lambda}\dot{\eta}(\phi,\lambda)$$

uniformly on $[-\pi,\pi]$. Using these facts the proof of the first part can be applied to the first derivative.

Since $(\mathcal{H},\rho)$ is a complete metric space $P\left(\rho\left(\hat{h}_n(\hat{\phi}_n,\lambda),h(\phi_0,\lambda)\right)>\delta\right)\to 0$ implies that $P\left(\hat{h}_n\in\mathcal{H}\right)\to 1$. ■

Proof of Theorem 9.2 We start by obtaining an expression for $R_n(\lambda)$ in (9.5). From the proof of Lemma (A.2) we have

$$\omega_y(\lambda) = \phi^{-1}\left(e^{-i\lambda}\right)\omega_\varepsilon(\lambda) + n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)$$

Letting $R_n^1(\lambda) = n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)$ we can write

$$\begin{aligned}
I_{n,yy}(\lambda) &= \omega_y(\lambda)\omega_y(-\lambda) \\
&= \omega_y(\lambda)\omega_y(-\lambda)a(\lambda)\phi + \omega_y(\lambda)\omega_\varepsilon(-\lambda) + \omega_y(\lambda)\phi\left(e^{i\lambda}\right)R_n^1(-\lambda) \\
&= I_{n,yy}(\lambda)a(\lambda)\phi + I_{n,\varepsilon\varepsilon}(\lambda)\phi^{-1}\left(e^{-i\lambda}\right) + \omega_\varepsilon(-\lambda)R_n^1(\lambda) \\
&\quad + \omega_\varepsilon(\lambda)\frac{\phi\left(e^{i\lambda}\right)}{\phi\left(e^{-i\lambda}\right)}R_n^1(-\lambda) + \left|R_n^1(-\lambda)\right|^2
\end{aligned}$$

In the same way we also have

$$\begin{aligned}
I_{n,yy}(\lambda) &= I_{n,yy}(\lambda)a(-\lambda)\phi + I_{n,\varepsilon\varepsilon}(\lambda)\phi^{-1}\left(e^{i\lambda}\right) + \omega_\varepsilon(\lambda)R_n^1(-\lambda) \\
&\quad + \omega_\varepsilon(-\lambda)\frac{\phi\left(e^{-i\lambda}\right)}{\phi\left(e^{i\lambda}\right)}R_n^1(\lambda) + \left|R_n^1(-\lambda)\right|^2
\end{aligned}$$

Let

$$R_n(\lambda) = \mathrm{Re}\left[\omega_\varepsilon(\lambda)R_n^1(-\lambda) + \omega_\varepsilon(-\lambda)\frac{\phi\left(e^{-i\lambda}\right)}{\phi\left(e^{i\lambda}\right)}R_n^1(\lambda) + \left|R_n^1(-\lambda)\right|^2\right].$$

Next write $I_{n,yy}(\lambda)h_0(\phi_0,\lambda)$ for the filter based on the true parameter ϕ_0 as

$$\begin{aligned}
I_{n,yy}(\lambda)h_0(\phi_0,\lambda) &= I_{n,yy}(\lambda)\mathrm{Re}\left[l_{\psi,0}(-\lambda)\phi_0\left(e^{i\lambda}\right)a(-\lambda)\right]\phi \\
&\quad + I_{n,\varepsilon\varepsilon}(\lambda)l_{\eta,0}(\lambda) + h_0(\phi_0,\lambda)R_n(\lambda)
\end{aligned}$$

while a filter based on a different parameter ϕ is expressed as

$$\begin{aligned} I_{n,yy}(\lambda) h(\phi,\lambda) &= I_{n,yy}(\lambda) \operatorname{Re}\left[l_{\psi}(-\lambda) \phi\left(e^{i\lambda}\right) a(-\lambda)\right] \phi \\ &\quad + I_{n,\varepsilon\varepsilon}(\lambda) h_{\phi_0}(\phi,\lambda) + h(\phi,\lambda) R_n(\lambda) \end{aligned}$$

with $h_{\phi_0}(\phi,\lambda) = \operatorname{Re}\left[l_{\psi}(-\lambda)\phi\left(e^{i\lambda}\right)\phi_0^{-1}\left(e^{i\lambda}\right)\right]$. Then by the proof of Lemma $(A.2)$ it follows that $\sqrt{n}\int R_n(\lambda)\varsigma(\lambda)\,d\lambda = o_p(1)$ for all continuous $\varsigma(\lambda)$ with absolutely summable Fourier coefficients. Next (9.10) is established if for all $\vartheta, \epsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{h\in\mathcal{H},\rho(h,h_0)<\delta} \|v_n(h) - v_n(h_0)\| > \vartheta\right) \leq \varepsilon$$

which in turn follows from

$$\limsup_{n\to\infty} P\left(\sup_{h\in\mathcal{H},\rho(h,h_0)<\delta} \sqrt{n}\left\|\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda\right\|\right) \leq \varepsilon \qquad \text{(C.12)}$$

and

$$\limsup_{n\to\infty} P\left(\sup_{h\in\mathcal{H},\rho(h,h_0)<\delta} \sqrt{n}\left\|\int_{-\pi}^{\pi} R_n(\lambda)\left[h - h_0\right] d\lambda\right\|\right) \leq \varepsilon. \qquad \text{(C.13)}$$

For (C.13) define the open neighborhood

$$\mathcal{H}_\delta = \left\{h : [-\pi,\pi] \to \mathbb{R}^p \,|\, h \in \mathcal{H}, \rho(h,h_0) < \delta\right\}.$$

Now for all $h \in \mathcal{H}_\delta$ it is the case that $\sup_{\lambda\in[-\pi,\pi]}|h(\lambda)| < \infty$ which in turn implies absolute summability of h. In turn $h - h_0$ is also continuous and uniformly bounded such that (C.13) follows from the proof of Lemma (A.2).

Next turn to (C.12). The main idea of the proof is taken from Robinson [83] where integration by parts is used to separate h from $I_{n,\varepsilon\varepsilon}(\lambda)$.

$$\begin{aligned} & \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda \right\| \\ \leq & \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} \left(I_{n,\varepsilon\varepsilon}(\lambda) - EI_{n,\varepsilon\varepsilon}(\lambda)\right) \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda \right\| \\ & + \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} EI_{n,\varepsilon\varepsilon}(\lambda) \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda \right\| \end{aligned}$$

Since $EI_{n,\varepsilon\varepsilon}(\lambda) = \sigma^2$ it follows that the last term is $\sigma^2 \int_{-\pi}^{\pi} \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda$. Now note that $\int_{-\pi}^{\pi} l_{\eta,0}(\lambda)\, d\lambda = 0$ and $\int_{-\pi}^{\pi} h_{\phi_0}(\phi,\lambda)\, d\lambda = 0$. Next use integration by parts

$$\begin{aligned} & \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} \left(I_{n,\varepsilon\varepsilon}(\lambda) - EI_{n,\varepsilon\varepsilon}(\lambda)\right) \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) d\lambda \right\| \\ \leq & \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} \frac{\partial}{\partial\lambda} \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) \int_{-\pi}^{\lambda} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\mu d\lambda \right\| \\ & + \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \left(h_{\phi_0}(\phi,\pi) - l_{\eta,0}(\pi)\right) \int_{-\pi}^{\pi} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\lambda \right\| \end{aligned}$$

Now both $\left\|\frac{\partial}{\partial\lambda}\left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right)\right\|$ and $\left\|h_{\phi_0}(\phi,\pi) - l_{\eta,0}(\pi)\right\|$ are uniformly bounded in $\mathcal{H}_\delta$ such that for some constant $C < \infty$

$$\begin{aligned} & \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} \frac{\partial}{\partial\lambda} \left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right) \int_{-\pi}^{\lambda} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\mu d\lambda \right\| \\ \leq & \; C\delta\sqrt{n} \int_{-\pi}^{\pi} \left| \int_{-\pi}^{\lambda} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\mu \right| d\lambda. \end{aligned}$$

It remains to show that $\sqrt{n}\sup_\lambda \left|\int_{-\pi}^{\lambda} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\mu\right|$ is bounded in probability.

Using Markov's inequality we look at $\sup_\lambda nE\left|\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu\right|^2$ which is analyzed by looking at

$$
\begin{aligned}
& nE\left|\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu\right|^2 \\
= \; & n\int_{-\pi}^{\lambda_1}\int_{-\pi}^{\lambda_2} cov\left(I_{n,\varepsilon\varepsilon}(\mu_1), I_{n,\varepsilon\varepsilon}(\mu_2)\right) d\mu_1 d\mu_2.
\end{aligned}
$$

From Brillinger [14, theorem 5.10.1]

$$
\begin{aligned}
& \int_{-\pi}^{\lambda_1}\int_{-\pi}^{\lambda_2} cov\left(I_{n,\varepsilon\varepsilon}(\mu_1), I_{n,\varepsilon\varepsilon}(\mu_2)\right) d\mu_1 d\mu_2 \\
= \; & n^{-1}\int_{-\pi}^{\lambda} f_{\varepsilon\varepsilon}^2(\mu)\,d\mu \\
& +n^{-1}\int_{-\pi}^{\lambda_1}\int_{-\pi}^{\lambda_2} f_{\varepsilon..\varepsilon}(\mu_1,\mu_2,-\mu_1)\,d\mu_1 d\mu_2 + O\left(n^{-2}\right)
\end{aligned}
$$

where the error is uniform in λ. Then $f_{\varepsilon\varepsilon}^2(\lambda)=\sigma^4$ and $f_{\varepsilon..\varepsilon}(\mu_1,\mu_2,-\mu_1)$ is uniformly bounded under assumption (F1). So

$$
\begin{aligned}
& \sup_\lambda nE\left|\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu\right|^2 \\
= \; & \sup_\lambda\int_{-\pi}^{\lambda} f_{\varepsilon\varepsilon}^2(\mu)\,d\mu + \sup_{\lambda_1,\lambda_2}\int_{-\pi}^{\lambda_1}\int_{-\pi}^{\lambda_2} f_{\varepsilon..\varepsilon}(\mu_1,\mu_2,-\mu_1)\,d\mu_1 d\mu_2 + O\left(n^{-1}\right)
\end{aligned}
$$

is bounded in the limit as $n\to\infty$.

By the same argument we also prove that $nE\left|\int_{-\pi}^{\pi}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\lambda\right|^2$ is bounded. This completes the proof. ■

References

[1] Andersen, T. G. Stochastic autoregressive volatility: A framework for volatility modeling. *Mathematical Finance*, 4:75–102, 1994.

[2] Andrews, D. W. K. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica*, 62:43–72, 1994.

[3] Andrews, D. W. K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

[4] Andrews, D. W. K. and J. C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966, 1992.

[5] Ash, R. B. *Real Analysis and Probability*. Academic Press, Inc., 1972.

[6] Beran, R. Adaptive estimates for autoregressive processes. *Annals of the Institute of Statistical Mathematics*, 26:77–89, 1976.

[7] Bickel, P. J. On adaptive estimation. *The Annals of Statistics*, 10(3):647–671, 1982.

[8] Billingsley, P. *Convergence of Probability Measures*. John Wiley and Sons, Inc., 1968.

[9] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, pages 307–327, 1986.

[10] Bollerslev, T., R. F. Engle, and J. M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.

[11] Boudoukh, J. An equilibrium model of nominal bond prices with inflation-output correlation and stochastic volatility. *Journal of Money, Credit and Banking*, 25(3):636–663, August 1993.

[12] Breidt, F. J., N. Crato, and P. de Lima. On the detection and estimation of long memory in stochastic volatility. manuscript.

[13] Brillinger, D. R. An introduction to polyspectra. *The Annals of Mathematical Statistics*, pages 1351–1374, 1965.

[14] Brillinger, D. R. *Time Series, Data Analysis and Theory, Expanded Edition.* Holden-Day, Inc., 1981.

[15] Brockwell, P. J. and R. A. Davis. *Time Series: Theory and Methods.* Springer Verlag-New York, Inc., 1987.

[16] Le Cam, L. *Locally Asymptotically Normal Families of Distributions*, volume 3, pages 27–98. University of California Publications in Statistics, 1960.

[17] Le Cam, L. *Asymptotic Methods in Statical Theory.* Springer-Verlag New York Inc., 1986.

[18] Le Cam, L. and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag New York Inc., 1990.

[19] Chamberlain, G. Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32:189–218, 1986.

[20] Chamberlain, G. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334, 1987.

[21] Chesney, M. and L. Scott. Pricing european currency options: A comparison of the modified black scholes model and a random variance model. *Journal of Financial and Quantitative Analysis*, 24(3):267–284, September 1989.

[22] Cragg, J.G. More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica*, 51(3):751–763, 1983.

[23] Danielsson, J. and J.-F. Richard. Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8:S153–S173, 1993.

[24] Deistler, M. , W. Dunsmuir, and E.J. Hannan. Vector linear time series models: Corrections and extensions. *Adv. Appl. Prob.*, 10:360–372, 1978.

[25] Dunsmuir,W. and E.J. Hannan. Vector linear time series models. *Adv. Appl. Prob.*, 8:339–364, 1976.

[26] Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages 987–1007, 1982.

[27] Engle, R. F., David M. Lilien, and Russell P. Robins. Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, 55(2):391–407. March 1987.

[28] Fabian, V. and J. Hannan. On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59:459–478, 1982.

[29] Farid, F.O. Spectral properties of perturbed linear operators and their application to infinite matrices. *Proceedings of the American Mathematical Society*, 112:1013–1022. 1991.

[30] Farid, F.O. Criteria for invertibility of diagonally dominant matrices. *Linear Algebra and its Applications*, 215:63–93, 1995.

[31] Folland, G. B. *Real Analysis - Modern Techniques and Their Applications.* John Wiley and Sons, 1984.

[32] Gallant, R. A., and D. Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55:363–390, 1987.

[33] Gallant, R. A. and G. Tauchen. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, 57(5):1091–1120, 1989.

[34] Gallant, R. A. and J. R. Long. Estimating stochastic differential equations efficiently by minimum chi square. Technical report, Department of Economics, University of North Carolina, 1995.

[35] Geweke, J. F. Bayesian comparison of econometric models. *Federal Reseverve Bank of Minneapolis, Working Paper 532*, 1994.

[36] Geweke, J. F. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339, November 1989.

[37] Gohberg, I. and S. Goldberg. *Basic Operator Theory.* BirkhĎuser, 1991.

[38] Hajek, J. Local asymptotic minimax and admissiblity in estimation. *Proceedings of the Sixth Berkeley Symposium in Mathematical Statistics 1972*, pages 175–194, 1972.

[39] Hall, P. and C.C. Heyde. *Martingale Limit Theory and its Application.* Academic Press, 1980.

[40] Hanani, H., E. Netanyahu, and M. Reichaw. Eigenvalues of infinite matrices. *Colloquium Mathematicum*, 19:89–101, 1968.

[41] Hannan, E.J. Nonlinear time series regression. *Journal of Applied Probability*, 8:767–780, 1971.

[42] Hannan, E.J. The asymptotic theory of linear time series models. *Journal of Applied Probability*, 10:130–145, 1973.

[43] Hansen, B.E. Heteroskedastic cointegration. *Journal of Econometrics*, 54:139–158, 1992.

[44] Hansen, B.E. Regression with nonstationary volatility. *Econometrica*, 63(5):1113–1132, 1995.

[45] Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1053, 1982.

[46] Hansen, L. P. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30:203–238, 1985.

[47] Hansen, L. P. Semi-parametric efficiency bounds of linear time series models. In P.C.B. Phillips and V.B. Hall, editors, *The Bergstrom Festschrift*, pages 253–271. Oxford: Basil Blackwell, 1990.

[48] Hansen, L. P., J. C. Heaton, and M. Ogaki. Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association*, 83:863–871, 1988.

[49] Hansen, L. P. and K. J. Singleton. Computing semiparametric efficiency bounds for linear time series models. In William A. Barnett, James Powell, and George Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pages 387–412, 1991.

[50] Harvey, A., E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

[51] Hayashi, F. and C. Sims. Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. *Econometrica*, 51(3):783–798, 1983.

[52] Hidalgo, J. Adaptive estimation in time series regression models with heteroskedasticity of unknown form. *Econometric Theory*, 8:161–187, 1992.

[53] Hosoya, Y. High-order efficiency in the estimation of linear processes. *The Annals of Statistics*, 7:516–530, 1979.

[54] Hosoya, Y. and M. Taniguchi. A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10:132–153, 1982.

[55] Huber, P. J. The behaviour of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233. University of California Press, 1967.

[56] Jacquier, E., N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 1994.

[57] Jennrich, R I. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, pages 633–643, 1969.

[58] Kabaila, P. V. An optimality property of the least squares estimate of the parameter of the spectrum of a purely nondeterministic time series. *The Annals of Statistics*, 8:1082–1092, 1980.

[59] Keenan, D. M. Asymptotic properties of minimization estimators of time series parameters. *Annals of Statistics*, 13:369–382, 1985.

[60] Kreiss, J. On adaptive estimation in stationary ARMA processes. *The Annals of Statistics*, 15(1):112–113, 1987.

[61] Lee, S. and B. E. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52, March 1994.

[62] Levit, B. Y. On the efficiency of a class of non-parametric estimates. *Theory of Probability and its Applications*, 20:723–740, 1975.

[63] Linton, O. Adaptive estimation in ARCH models. *Econometric Theory*, 9(4):539–569, 1993.

[64] Lumsdaine, R.L. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, 64(3):575–596, 1996.

[65] Magnus, J.R. and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons, 1988.

[66] Mahieu, R. and P. Schotman. Stochastic volatility and the distribution of exchange rate news. Discussion paper 96, Institute for Empirical Macroecnomics, 1994.

[67] Melino, A. and S. M. Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239–265, 1990.

[68] Milhoj, A. The moment structure of ARCH processes. *Scandinavian Journal of Statistics*, 12:281–292, 1985.

[69] Nelson, D. B. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990.

[70] Nelson, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, March 1991.

[71] Nelson, D. B. Filtering and forecasting with misspecified ARCH models I. *Journal of Econometrics*, 52:61–90, 1992.

[72] Nelson, D. B. and D. P. Foster. Filtering and forecasting with misspecified ARCH models II. Making the right forecast with the wrong model. *Journal of Econometrics*, 67:303–335, 1995.

[73] Newey, W. K. Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics*, 38:301–339, 1988.

[74] Newey, W. K. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837, 1990.

[75] Newey, W. K. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5:99–135, 1990.

[76] Newey, W. K. and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

[77] Pakes, A. and D. Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057, September 1989.

[78] Phillips, P.C.B. Error correction and long-run equilibrium in continuous time. *Econometrica*, 59(4):967–980, 1991.

[79] Phillips, P.C.B. Spectral regression for cointegrated time series. In J. Powell W. Barnett and G. Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pages 417–435. Cambridge University Press, 1991.

[80] PPhillips, P.C.B. Econometric model determination. *Econometrica*, 64(4):763–812, 1996.

[81] Robinson, P.M. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*, 55:875–891, 1987.

[82] Robinson, P.M. Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954, 1988.

[83] Robinson, P.M. Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica*, 59:1329–1363, 1991.

[84] Robinson, P.M. Best nonlinear three-stage least squares estimation of certain econometric models. *Econometrica*, 59(3):755–786, 1991.

[85] Schmeidler, W. *Linear Operators in Hilbert Space.* Academic Press, 1954.

[86] Shephard, N. Fitting nonlinear time series models with applications to stochastic variance models. *Journal of Applied Econometrics*, 8:S135–S152, 1993.

[87] Shephard, N. Partial Non-Gaussian state space. *Biometrika*, 81(1):115–131, 1994.

[88] Steigerwald, D. G. Efficient estimation of models with conditional heteroscedasticity. Technical report, Department of Economics, University of California Santa Barbara, 1994.

[89] Stein, C. Efficient nonparametric testing and estimation. In *Proc. Third Berkeley Symp. Math. Statist. Prob.*, pages 187–196. University of California Press, 1956.

[90] Stoica, P., T. Söderström, and B. Friedlander. Optimal instrumental variable estimates of the ar parameters of an ARMA process. *IEEE Transactions on Automatic Control*, 30:1066–1074, 1985.

[91] Swensen, A. R. The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16:54–70, 1985.

[92] Taniguchi, M. On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *The Annals of Statistics*, 11:157–169, 1983.

[93] Uhlig, H. Baesian vector autoregressions with stochastic volatility. *Econometrica*. 65(1):59–73, 1997.

[94] Weiss, A. A. Asymptotic theory for arch models: Estimation and testing. *Econometric Theory*, 2:107–131, 1986.

[95] White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

[96] Wu, C. F. Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, pages 501–513, 1981.

[97] Zaman, A. Consistency via type 2 inequalities: A generalization of Wu's theorem. *Econometric Theory*, pages 272–286, 1989.