LOWER RISK BOUNDS AND PROPERTIES OF CONFIDENCE SETS
FOR ILL-POSED ESTIMATION PROBLEMS WITH APPLICATIONS TO
SPECTRAL DENSITY AND PERSISTENCE ESTIMATION, UNIT ROOTS,
AND ESTIMATION OF LONG MEMORY PARAMETERS

Benedikt M. Pötscher
Department of Statistics
University of Vienna

Preliminary version: April 1998
First version: September 1999
This version: September 2000

1999 Meeting of the "Ökonometrischer Ausschuss des Vereins für Socialpolitik".
I thank H. Leeb, A. Samarov, the referees, and the co-editor A. Monfort for
helpful comments. I also thank A. Monfort for pointing out the reference
Bahadur and Savage (1956).
ABSTRACT

Important estimation problems in econometrics like estimating the value of a spectral density at frequency zero, which appears in the econometrics literature in the guises of heteroskedasticity and autocorrelation consistent variance estimation and long run variance estimation, are shown to be "ill-posed" estimation problems. A prototypical result obtained in the paper is that the minimax risk for estimating the value of the spectral density at frequency zero is infinite regardless of sample size, and that confidence sets are close to being uninformative. In this result the maximum risk is over commonly used specifications for the set of feasible data generating processes. The consequences for inference on unit roots and cointegration are discussed. Similar results for persistence estimation and estimation of the long memory parameter are given. All these results are obtained as special cases of a more general theory developed for abstract estimation problems, which readily also allows for the treatment of other ill-posed estimation problems such as, e.g., nonparametric regression or density estimation.

Key-words: Ill-posed problem, lower risk bound, minimax risk, confidence sets, nonparametric function estimation, spectral density estimation, persistence, unit root, cointegration, difference stationarity, trend stationarity, long memory, fractional differencing parameter
1. Introduction

A typical theoretical analysis of an estimation problem in the econometrics or statistics literature proceeds by first suggesting an estimator and then by establishing that this estimator is consistent or that some measure of risk like mean squared error converges to zero as sample size tends to infinity. A limiting distribution or at least a rate of convergence for the estimator or the associated risk-measure is often also supplied. Comparisons of estimators or claims of asymptotic optimality are then frequently based on these results. The history of the asymptotic efficiency concept for parametric models (Beran (1997)) clearly demonstrates the pitfalls plaguing such an approach. The Hájek-LeCam theory of asymptotic efficiency resolves these difficulties and shows that the asymptotic performance of an estimator should be judged not only in a "pointwise" manner, i.e., for each possible value of the true parameter individually, but rather "uniformly" over the parameter space (or at least "locally uniformly"). Unfortunately, however, this lesson does not seem to have been learned as much as it should, and "pointwise" asymptotic analyses as described at the beginning of this paragraph still dominate much of the econometrics and statistics literature. Looking beyond parametric problems, one of the aspects of the present paper is to ask if for typical estimation problems a "pointwise" asymptotic analysis can tell us anything relevant at all about the actual performance of an estimator or a confidence procedure. Not too surprisingly, it turns out that for many familiar estimation problems like, e.g., nonparametric density estimation, regression function estimation, spectral density estimation, estimation of persistence (and related questions of unit root inference), and long memory parameter estimation, the answer -- under realistic and commonly used assumptions -- unfortunately is "no". In fact, conclusions drawn from a "pointwise" asymptotic analysis can be highly misleading. A second, and perhaps more surprising and worrying conclusion from the paper is that -- in stark contrast to the case of regular parametric problems -- for these
estimation problems no estimators that are good in a uniform sense can exist, except under quite restrictive a priori assumptions regarding the set of feasible data generating processes (DGPs). We stress here that the phenomena discussed in the paper will occur regardless of how large sample size is and hence have nothing to do with small sample effects. Negative results like the ones just mentioned will first be established for a class of abstract estimation problems in Sections 2 and 3. These general results are then applied to some concrete and familiar estimation problems (spectral density and persistence estimation, estimation of long memory parameter) in Sections 4 and 5. A reader interested mainly in the applications may skip Sections 2 and 3 in a first reading.

To make the above discussion more concrete, consider an estimation problem where a (abstract) parameter \( h \in \mathcal{H} \) parameterizes the set of possible DGPs and where the real-valued functional \( \theta = \theta(h) \) is to be estimated. (A leading case is nonparametric function estimation where \( \mathcal{H} \) is a space of functions and \( \theta(h) = h(t_0) \) the value of the function \( h \) at a given point \( t_0 \).) A standard asymptotic analysis as sketched at the beginning of the previous paragraph would then typically deliver a result like

\[
R(\hat{\theta}_n, \theta(h)) \to 0 \text{ as } n \to \infty \text{ for every } h \in \mathcal{H},
\]

(1.1)

where \( R \) is some risk-measure like mean squared error, or a consistency result such as

\[
P_h(\hat{\theta}_n - \theta(h) > \varepsilon) \to 0 \text{ as } n \to \infty \text{ for every } h \in \mathcal{H}, \varepsilon > 0.
\]

(1.2)

Here \( \hat{\theta}_n \) denotes a sequence of estimators for \( \theta(h) \) and \( P_h \) stands for the probability measure of the DGP when \( h \) is the true parameter. A main result of the paper is to show that for a large class of estimation problems, so-called ill-posed problems, despite validity of (1.1)-(1.2), the following is true for any estimator \( \hat{\theta}_n \):

\[
\sup_{h \in \mathcal{H}} R(\hat{\theta}_n, \theta(h)) \geq c > 0 \quad \text{for all } n \geq 1,
\]

(1.3)

where the constant \( c \) can be described in terms of the estimation problem; in fact, often \( c = \infty \) holds. As a consequence, the minimax risk (i.e., the infimum
of the l.h.s. of (1.3) taken over all possible estimators) does not converge to zero as sample size tends to infinity. Using the maximum risk as in (1.3) to judge estimator performance could perhaps be criticized by some as being too pessimistic, who therefore might be inclined to discount the warning message sent by (1.3). It turns out, however, that for many ill-posed estimation problems a result much more worrying than (1.3) in fact holds: For any estimator \( \hat{\theta}_n \), any \( h_0 \in \mathcal{H} \), and any neighborhood \( U(h_0) \) of \( h_0 \) (in a suitable topology on \( \mathcal{H} \)) even

\[
\sup_{h \in U(h_0)} R(\hat{\theta}_n, \theta(h)) \geq c > 0 \quad \text{for all } n \geq 1, \tag{1.4}
\]

holds. While the standard asymptotic analysis leading to (1.1) tells us about the "pointwise" (i.e., for each fixed DGP) convergence to zero of the risk-measure, it does not tell us anything about the behaviour of the risk over arbitrarily small neighborhoods of a fixed DGP. For example, it is often the case that, although (1.1) holds (giving the impression that \( \hat{\theta}_n \) is a reasonable estimator with asymptotically vanishing risk), relation (1.4) with \( c = \infty \) is also true, implying that for any sample size there are DGPs arbitrarily close to the fixed DGP, such that the risk of \( \hat{\theta}_n \) at the former DGPs is arbitrarily large. The reason for this phenomenon is extreme local non-uniformity in the risk convergence described by (1.1). As can be shown (cf. Corollary 2.3), a further consequence of (1.3) or (1.4) is that no sequence of estimators for \( \theta(h) \) can be uniformly consistent, despite the fact that any estimator satisfying (1.1) will be consistent (provided a reasonable risk-measure like mean squared error is used). Recall that a sequence of estimators \( \hat{\theta}_n \) is uniformly consistent if for every \( \varepsilon > 0 \)

\[
\sup_{h \in \mathcal{H}} P_n(|\hat{\theta}_n - \theta(h)| > \varepsilon) \to 0 \quad \text{as } n \to \infty. \tag{1.5}
\]

Apart from the fact that in ill-posed estimation problems the standard "pointwise" asymptotic analysis can give a very misleading impression of the actual performance of an estimator, an even more serious consequence of (1.3)-(1.4) is that no reasonably good estimator exists for such estimation
problems at all as the (global or local) maximum risk of any estimator does not converge to zero with increasing sample size. Another aspect of this phenomenon is that confidence sets are necessarily very large and sometimes are completely uninformative with large probability. This implies that typical confidence sets based on "pointwise" asymptotic approximations have an actual minimal coverage probability much smaller than the nominal one, in fact, it is often zero. Cf. Gleser and Hwang (1987) and Dufour (1997) for related work on this aspect. Related is also Bahadur and Savage (1956), who consider nonparametric estimation of moments, which in the parlance of the present paper is an example of an ill-posed problem.

It is worth noting that, in contrast to ill-posed estimation problems, estimators with good "uniform" properties do exist for regular parametric estimation problems. For example, the usual consistency and asymptotic normality results for the maximum likelihood estimator (MLE) can quite easily be strengthened to (locally) uniform results and the (local) minimax risk of the MLE goes to zero (e.g., Ibragimov and Khasminskii (1981), Theorem 3.1.1). The need for a "uniform" asymptotic analysis in regular parametric models alluded to earlier only arises from the need to appropriately judge the performance of competitors that have a highly non-uniform limiting behaviour (e.g., superefficient estimators) and that in a standard "pointwise" analysis would appear to be superior to good estimators like the MLE.

An important estimation problem for which the results of the paper are relevant is nonparametric function estimation. Here the set $\mathcal{H}$ represents a set of functions like regression, density, or spectral density functions. The goal is estimation of the value $h(t_0)$ at a given point $t_0$. It turns out that the lower risk bound $c$ in (1.3) or (1.4) is typically related to the so-called oscillation of the map $h \mapsto h(t_0)$ at $h$, which is a measure of the (dis)continuity properties of this map at $h$. For many natural choices of the set $\mathcal{H}$ the map $h \mapsto h(t_0)$ turns out to be discontinuous (often at every $h \in \mathcal{H}$) implying a positive value of the oscillation measure, and hence of the lower
risk bounds in (1.3) or (1.4). Intuitively speaking, the positive lower risk bounds reflect the fact that a highly discontinuous functional of \( h \) cannot be estimated reliably (given \( h \) continuously parameterizes the set of DGPs). For precise results in the case of spectral density estimation see Section 4. Similar results can be given for regression or density function estimation.

Much of the literature on nonparametric function estimation (density estimation, regression function estimation, spectral density estimation, etc.) still employs a "pointwise" asymptotic point of view, although there is a growing number of papers taking a minimax point of view, e.g., Samarov (1977), Farrell (1979), Ibragimov and Khashminskii (1980, 1981), Stone (1982), Bentkus (1985), Efroimovich and Pinsker (1986), Donoho and Johnston (1994), Donoho et al. (1995), Lepski and Spokoiny (1997), to mention a few. A typical approach in this literature is to assume a uniform bound on the magnitude of certain derivatives (or of similar quantities) of the function to be estimated (uniform over the set of functions over which the minimax risk is to be taken), ruling out the kind of discontinuity mentioned above. Under this kind of assumptions the minimax risk can then be shown to converge to zero, its rate can be determined and asymptotically efficient estimators can be constructed. However, such assumptions — although indispensable for obtaining these results — are less than innocuous as they rule out many naturally occurring classes of DGPs as will transpire from the results in Sections 4 and 5 of the present paper.

An outline of the remainder of the paper is as follows: A general lower risk bound for abstract estimation problems including ill-posed ones is developed in Section 2, and properties of confidence procedures for such estimation problems are the subject of Section 3. These general results are applied to spectral density and persistence estimation in Section 4. For example, it is shown that the minimax risk for estimating the spectral density at frequency zero is infinite even for such small classes of spectral
densities as is, e.g., the class given by the ARMA(1,1) model. The implications for persistence estimation and unit root inference are also discussed in this section. Section 5 is concerned with estimation of the long memory parameter. Again, for typical classes of spectral densities the minimax risk is shown to be bounded away from zero as sample size increases and confidence intervals for the long memory parameter are shown to be necessarily uninformative with high probability. Concluding remarks are given in Section 6. All proofs and some auxiliary results are collected into appendices.

2. A General Lower Risk Bound and Ill-posed Problems

Let \((X,X)\) be a measurable space (the "sample space") and \(\mathcal{P}\) a nonempty set of probability measures on \(X\), describing a statistical experiment. We assume that the elements in \(\mathcal{P}\) are indexed by the elements of a set \(\mathcal{H}\) (the "parameter space"), i.e., there is a map \(h \mapsto P_h\) from \(\mathcal{H}\) onto \(\mathcal{P}\). E.g., \(\mathcal{H}\) could be a set of functions, like spectral density functions, probability densities, regression functions, etc., but clearly also parametric models are covered by the setup.

We want to estimate a function \(\theta: \mathcal{H} \to \mathcal{M}\), which describes the aspects of \(h\) we are interested in, where \((\mathcal{M},\mathcal{M})\) is a measurable space. The leading case is where \(\mathcal{M}\) is \(\mathbb{R}\) (or \(\mathbb{R}^m\)) and \(\theta\) is then a functional on \(\mathcal{H}\). For example, \(\theta(h)\) can be the value of \(h\) at a certain point in case \(\mathcal{H}\) is a class of functions. Given an estimator \(\hat{\theta}\), i.e., a measurable function \(\hat{\theta}:(X,X) \to (\mathcal{M},\mathcal{M})\), and a loss-function \(\ell: \mathcal{M} \times \mathcal{M} \to [0,\infty)\), we then compare estimators on the basis of the risk-function

\[
E_h \ell(\hat{\theta}, \theta(h)),
\]

where \(E_h\) denotes expectation w.r.t. \(P_h\). (As usual, we assign the risk the value \(+\infty\) if the loss is not \(P_h\)-integrable.) The minimal requirements we impose on the loss-function is that it is symmetric (i.e., \(\ell(a,b) = \ell(b,a)\) for all \(a, b \in \mathcal{M}\), \(\mathcal{M}\)-measurable in its first argument, and that it satisfies the generalized triangle inequality, i.e., there exists \(1 \leq \kappa < \infty\) such that

\[
\ell(a,b) \leq \kappa[\ell(a,c) + \ell(c,b)]
\]

holds for all \(a,b,c \in \mathcal{M}\). We shall call the loss-function "proper" if it
satisfies these requirements. For every proper loss-function $\ell$ there obviously exists a smallest $\kappa(\ell)\geq 1$ such that (2.1) holds. We also note for later use that the truncated loss-function $\ell_H(a,b) = \min(M,\ell(a,b))$, $M>0$, is proper if $\ell(a,b)$ has this property, and that $\kappa(\ell_H) = \kappa(\ell)$ holds. Moreover, $\kappa(\ell) \to \kappa(\ell)$ as $M \to \infty$ as is easily seen. Large classes of loss-functions are proper, including the leading case $\ell(a,b) = |a-b|^p$, $p>0$, when $M=\mathbb{R}$. For these loss-functions $\kappa(\ell) = 2^{p-1}$ for $p=1$ and $\kappa(\ell) = 1$ for $0<p<1$.

The next theorem provides a general lower bound for the "local" as well for the "global" maximum risk of an arbitrary estimator in the statistical experiment described above. The total variation distance between two probability measures on $(X,X)$ is defined as $\|P_1 - P_2\|_{TV} = \sup_{A \in \mathcal{X}} |P_1(A) - P_2(A)|$.

**Theorem 2.1:** Let $\mathcal{P}$ be a nonempty set of probability measures on the measurable space $(X,X)$. Assume that the elements of $\mathcal{P}$ are indexed by the elements of a set $\mathcal{H}$, i.e., there is a map $h \mapsto P_h$ from $\mathcal{H}$ onto $\mathcal{P}$. Assume further that there is a first countable topology $\tau$ on $\mathcal{H}$, such that the map $h \mapsto P_h$ from $\mathcal{H}$ onto $\mathcal{P}$ is continuous when $\mathcal{P}$ is endowed with the total variation distance.

Let $\theta: \mathcal{H} \to \mathcal{M}$ be a function, where $(\mathcal{M}, \mathcal{M})$ is a measurable space, and let $\ell$ be a proper loss-function. Define for each $h_0 \in \mathcal{H}$ the constants

$$c(\theta, \ell, \mathcal{H}, \tau, h_0) = (2\kappa(\ell))^{-1} \inf \{ \sup_{f,g \in \mathcal{W}} \ell(\theta(f), \theta(g)) : h_0 \in \mathcal{W}(\mathcal{H}), \ W \tau\text{-open} \}. \quad (2.2)$$

Then for every $h_0 \in \mathcal{H}$, every neighborhood $U(h_0)$ of $h_0$, and every estimator $\hat{\theta}$, i.e., every measurable function $\hat{\theta}:(X,X) \to (\mathcal{M}, \mathcal{M})$, the following lower bound holds:

$$\sup_{h \in U(h_0)} E_h \ell(\hat{\theta}(h)) \geq c(\theta, \ell, \mathcal{H}, \tau, h_0). \quad (2.3)$$

Consequently, the minimax risk for estimating $\theta(h)$ satisfies

$$\inf_{\theta} \sup_{h \in \mathcal{H}} E_h \ell(\hat{\theta}(h)) \geq c(\theta, \ell, \mathcal{H}, \tau), \quad (2.4)$$

where $c(\theta, \ell, \mathcal{H}, \tau) = \sup_{h_0 \in \mathcal{H}} c(\theta, \ell, \mathcal{H}, \tau, h_0)$ and where the infimum in (2.4) extends over all $X-M$-measurable $\hat{\theta}$. 

7
The significance of the above theorem derives from the following two observations: (i) In a typical application of the theorem the set $\mathcal{P}$ will be taken as $\mathcal{P}_n$, the family of distributions induced by the statistical model on the space of samples of size $n$, and hence will depend on sample size. However, typically there are natural choices for $\mathcal{H}$ and $\tau$ that will not depend on $n$. (Also $\Theta$, $(M,M)$ and $\ell$ will typically be independent of $n$.) In such a case, applying Theorem 2.1 for each $n$, entails a lower (local) global risk bound that is independent of sample size, since the lower bounds in (2.3) and (2.4) do not depend on $\mathcal{P}_n$. (ii) These bounds will turn out to be positive for many important estimation problems, including the ones discussed in Sections 4 and 5, showing that the (global or local) minimax risk in these problems does not vanish with increasing sample size. A consequence of positivity of these bounds is also nonexistence of uniformly consistent estimators, cf. Corollary 2.3. In the following we shall refer to an estimation problem, for which the lower bound in (2.4) is positive, as an "ill-posed" estimation problem. Also note that Theorem 2.1 allows for non-identifiable parameterizations and adapts easily to problems with nuisance parameters; see Remarks A.2 and A.3 in Appendix A.

As already alluded to in the Introduction, positivity of the lower risk bound in (2.3) is often equivalent to discontinuity of the functional $\Theta$ at $h_0$. General results of this sort are given in Theorem A.1 in Appendix A, which also provides simpler expressions or lower estimates for the "local" or "global" minimax risk bounds in (2.3) and (2.4). For the important special case where the function to be estimated, i.e., $\Theta(h)$, takes its values in Euclidean space, i.e., $\mathcal{M} = \mathbb{R}^m$, the subsequent result is an immediate corollary to Theorem A.1. Consider the loss-function $\ell_{p,\alpha}(a,b) = \min(\alpha,||a-b||_p)$, where $||a-b||_p = \sum_{i=1}^{m} |a_i-b_i|^p$, $p$ is a positive real number, and $\alpha$ is a positive real number or $+\infty$. These loss-functions are obviously proper with $\kappa(\ell_{p,\alpha}) = 1$ for $0<p<1$ and $\kappa(\ell_{p,\alpha}) = 2^{p-1}$ for $p \geq 1$. They are also continuous on $\mathcal{M} \times \mathcal{M}$. If $\alpha = \infty$, we have the usual $L^p$-loss, and if $\alpha < \infty$ we obtain a truncated version.
Corollary 2.2: Let the assumptions of Theorem 2.1 hold with \( M=\mathbb{R}^m \) and \( M=B(\mathbb{R}^m) \) and consider the loss-function \( \ell^{p,\alpha}(a,b), p>0, 0<\alpha \leq \infty \). Then the lower bound \( c(\theta, \ell^{p,\alpha}, \mathcal{H}, \tau, h_0) \) for the local minimax risk at \( h_0 \) is positive if and only if the function \( \theta \) is discontinuous at \( h_0 \). The lower bound \( c(\theta, \ell^{p,\alpha}, \mathcal{H}, \tau) \) for the global minimax risk is positive if and only if the function \( \theta \) is discontinuous at some \( h_0 \in \mathcal{H} \). Furthermore, for \( p \geq 1 \)
\[
c(\theta, \ell^{p,\alpha}, \mathcal{H}, \tau, h_0) = (1/2^p)\min(\alpha, [\text{osc}_p(\theta, h_0)])^p
\]
holds, where
\[
\text{osc}_p(\theta, h_0) = \inf\{\sup_{f,g \in W} ||\theta(f)-\theta(g)|| : h_0 \in W \subseteq \mathcal{H}, W \tau\text{-open}\}
\]
denotes the oscillation of the function \( \theta \) at \( h_0 \) w.r.t. the metric \( ||a-b||_p \).

The lower bound results derived in this section can be used to establish non-existence of uniformly consistent estimators in ill-posed problems. The following result is presented for functions \( \theta \) taking values in Euclidean space, but it is obvious how similar results can be obtained from Theorems 2.1 and A.1 in more general settings. Note that in typical applications of this result the set \( \mathcal{P}_n \) will correspond to the set of distributions induced by the statistical model on the space of samples of size \( n \). For a thorough investigation into the existence of uniformly consistent estimators in an i.i.d. framework see LeCam and Schwartz (1960).

Corollary 2.3: For each \( n \in \mathbb{N} \) let \( \mathcal{P}_n \) be a nonempty set of probability measures on the measurable space \( (X_n, \mathcal{X}_n) \). Assume that \( \mathcal{H} \) parameterizes the elements of \( \mathcal{P}_n \) for each \( n \in \mathbb{N} \), i.e., there are maps \( h \mapsto \mathcal{P}_{n,h} \) from \( \mathcal{H} \) onto \( \mathcal{P}_n \), and suppose that these maps are continuous w.r.t. a first countable topology \( \tau \) on \( \mathcal{H} \) and the total variation distance on \( \mathcal{P}_n \). Let \( \theta : \mathcal{H} \rightarrow \mathcal{H} \) be a function, where \( \mathcal{H}=\mathbb{R}^m \) and \( M=B(\mathbb{R}^m) \). If \( \theta \) is discontinuous at some \( h_0 \in \mathcal{H} \), no sequence of estimators \( \hat{\theta}_n \) (i.e., \( \hat{\theta}_n : X_n \rightarrow \mathbb{R}^m \) is \( X_n - B(\mathbb{R}^m) \)-measurable) can be uniformly consistent for \( \theta(h) \).
In fact, more is then true: there exists $\varepsilon_0 > 0$ and a constant $c(h_0) > 0$ such that for every sequence of estimators $\hat{\theta}_n$ and every $0 < \varepsilon < \varepsilon_0$

$$\inf_n \sup_{h \in U(h_0)} \mathbb{P}_n, h(||\hat{\theta}_n - \theta(h)||_1 > \varepsilon) \geq c(h_0) > 0$$

(2.5)

for every neighborhood $U(h_0)$ of $h_0$.

**Remark 2.1:** In fact it can be shown that for sufficiently small $\varepsilon$ the l.h.s. of (2.5) is not less than $1/2$ for all neighborhoods $U(h_0)$ of $h_0$.

**Remark 2.2:** (Resampling and Randomized Estimators) Theorem 2.1, and hence all results in the paper, are easily seen to hold also for estimators obtained through a resampling procedure. Suppose the resampling plan is given by the Markov-kernel $K(x,dz)$ from $(X,X)$ to a measurable space $(Z,Z)$ and $\hat{\theta}$ is now a measurable function on $(X \times Z, X \otimes Z)$. For each $P_h \in \mathcal{P}$ define $\tilde{P}_h$ via

$$\tilde{P}_h(A) = \int \mathbf{1}_A(x,z)K(x,dz)P_h(dx)$$

and let $\mathcal{P}$ denote the set of all such $\tilde{P}_h$. Observe that $||\tilde{P}_h - \tilde{P}_k||_{TV} = \sup_A |\tilde{P}_h(A) - \tilde{P}_k(A)| \leq \sup_x \sup_A |\mathbf{1}_A(x,z)K(x,dz)| \cdot ||P_h - P_k||_{TV} \leq ||P_h - P_k||_{TV}$.

Given the assumptions of Theorem 2.1 are satisfied for $\mathcal{P}$, it now immediately follows that the same is true for $\mathcal{P}$ (with $X \times Z$ replacing $X$), and hence Theorem 2.1 is seen to cover also randomized estimators.

3. **Properties of Confidence Sets and Ill-posed Problems**

In this section we provide results that relate the size of confidence sets to the size of certain nonrandom sets. For ill-posed estimation problems the latter sets can often be shown to be very large, implying that the same is also true for confidence sets. Let $(X,X)$, $\mathcal{P}$, $\mathcal{H}$, the parameterization map $h \rightarrow P_h$ from $\mathcal{H}$ onto $\mathcal{P}$, and the topology $\tau$ satisfy the assumptions of Theorem 2.1. Let again $\theta(h)$ denote the object of interest, i.e., the function of $h$ for which a confidence set is to be constructed. It will turn out to be convenient to assume only that $\theta$ is defined on some subset $\mathcal{H}_*$ of $\mathcal{H}$, i.e.,
\( \theta : \mathcal{H} \to \mathcal{M} \), where \( \mathcal{M} \) is a set (not necessarily carrying any further structure at this point). The subset \( \mathcal{H}_0 \) as well as the topology \( \tau \) will have to be chosen strategically in any particular application, cf. Remark C.1 in Appendix C. Typically, \( \mathcal{H}_0 \) will be a dense (w.r.t. \( \tau \)) subset of \( \mathcal{H} \) or coincide with \( \mathcal{H} \). The assumptions made above will be maintained throughout this section. For \( h_0 \in \mathcal{H} \) define the set
\[
\Theta(h_0, \mathcal{H}_0) = \cap \{ \Theta(\omega h_0 h_0) : W \text{ is neighborhood of } h_0 \}.
\]
Note that \( \Theta(h_0, \mathcal{H}_0) \) is the set of elements \( a \in \mathcal{M} \) that arise as the values of \( \Theta \) in arbitrary small neighborhoods of \( h_0 \) intersected with \( \mathcal{H}_0 \). Also define
\[
\mathcal{H}_0(h_0) = \{ h_0 \in \mathcal{H}_0 : \Theta(h_0) \in \Theta(h_0, \mathcal{H}_0) \},
\]
i.e., \( \mathcal{H}_0(h_0) = \Theta^{-1}(\Theta(h_0, \mathcal{H}_0)) \). Let \( C \) be a random set in \( \mathcal{M} \) in the sense that \( C(\omega) \) is a subset of \( \mathcal{M} \) for each \( \omega \in \mathcal{X} \) and the set \( \{ \omega \in \mathcal{X} : a \in C(\omega) \} \) is \( X \)-measurable for each \( a \in \mathcal{M} \). As usual, we shall write \( a \in C \) to denote the event \( \{ \omega \in \mathcal{X} : a \in C(\omega) \} \).

**Proposition 3.1:** Let \( C \) be a confidence set for \( \Theta(h_0), h_0 \in \mathcal{H}_0 \), with minimum coverage probability not less than \( 1-\alpha \), \( 0 \leq \alpha \leq 1 \), i.e., \( C \) is a random set in \( \mathcal{M} \) satisfying \( P_{h_0}(\Theta(h_0) \in C) \geq 1-\alpha \) for all \( h_0 \in \mathcal{H}_0 \). Then for every \( h_0 \in \mathcal{H} \)
\[
P_{h_0}(\Theta(h_0) \in C) \geq 1-\alpha \quad \text{for all } h_0 \in \mathcal{H}_0(h_0), \quad (3.1)
\]
and for every \( \varepsilon > 0 \) there is a neighborhood \( U(h_0) \) of \( h_0 \) in \( \mathcal{H} \) such that for all \( h \in \mathcal{U}(h_0) \)
\[
P_h(\Theta(h) \in C) \geq 1-\alpha-\varepsilon \quad \text{for all } h_0 \in \mathcal{H}_0(h_0). \quad (3.2)
\]
If even \( P_{h_0}(\Theta(h_0) \in C) = 1-\alpha \) holds for all \( h_0 \in \mathcal{H}_0 \), then equality holds in (3.1) and the 1.h.s. in (3.2) is also bounded from above by \( 1-\alpha+\varepsilon \).

Based on the above proposition, the next theorem provides lower bounds on the probability that a confidence set contains certain nonrandom sets related to \( \Theta(h_0, \mathcal{H}_0) \).

**Theorem 3.2:** Let \( C \) be as in Proposition 3.1. Then for every \( h_0 \in \mathcal{H} \):
(1) For any \( a_1, a_2, \ldots, a_p \) with \( a_i \in \Theta(h_0, H_\ast) \)
\[
P_{h_0}(a_1 \in C, \ldots, a_p \in C) \geq 1-\alpha.
\] (3.3)

(ii) Suppose \( M \) is equipped with a metric \( d \) and \( C \) is diameter-measurable, i.e., \( \text{diam}(C) \) is an \( X \)-measurable function. Then
\[
P_{h_0}(\text{diam}(C) \geq \text{diam}(\Theta(h_0, H_\ast))) \geq 1-2\alpha,
\] (3.4)

where \( \text{diam}(A) = \sup\{d(a, b) : a \in A, b \in A\} \) denotes the diameter of a set \( A \subseteq M \). If \( \text{diam}(\Theta(h_0, H_\ast)) = \infty \), then even
\[
P_{h_0}(\text{diam}(C) = \infty) \geq 1-\alpha
\] (3.5)
holds.\(^2\)

(iii) Suppose \( M \) is a subset of a vector space and \( C(\omega) \) is convex for all \( \omega \in X \). Then
\[
P_{h_0}(\text{conv}(a_1, \ldots, a_p) \subseteq C) \geq 1-\alpha
\] (3.6)

for any set of points \( a_1, a_2, \ldots, a_p \) with \( a_i \in \Theta(h_0, H_\ast) \). Here \( \text{conv}(.) \) denotes the convex hull of the points indicated. Consequently, for any set \( N \subseteq M \), which for some \( p \geq 1 \) can be written as a countable increasing union of sets of the form \( \text{conv}(a_1, \ldots, a_p) \) with \( a_i \in \Theta(h_0, H_\ast) \) we have
\[
P_{h_0}(N \subseteq C) \geq 1-\alpha.
\] (3.7)

(iv) Suppose \( M \) is a subset of a vector space, carries also a topology and \( C(\omega) \) is closed and convex for all \( \omega \in X \). Then
\[
P_{h_0}(\text{cl}(\text{conv}(a_1, \ldots, a_p)) \subseteq C) \geq 1-\alpha
\] (3.8)

for any set of points \( a_1, a_2, \ldots, a_p \) with \( a_i \in \Theta(h_0, H_\ast) \), where \( \text{cl}(A) \) denotes the closure of \( A \subseteq M \) w.r.t. the topology on \( M \). Consequently, for any set \( N \subseteq M \), which for some \( p \geq 1 \) can be written as a countable increasing union of sets of the form \( \text{conv}(a_1, \ldots, a_p) \) with \( a_i \in \Theta(h_0, H_\ast) \) we have
\[
P_{h_0}(\text{cl}(N) \subseteq C) \geq 1-\alpha.
\] (3.9)

(v) For every \( \varepsilon > 0 \) there is a neighborhood \( U(h_0) \) of \( h_0 \) in \( H \) such that for each \( h \in U(h_0) \) parts (i)-(iv) above continue to hold if \( P_h \) replaces \( P_{h_0} \) and \( \varepsilon \) is subtracted from the r.h.s. of each of these inequalities.
(vi) Suppose $P_h$ is absolutely continuous w.r.t. $P_{h_0}$ for all $h \in \mathcal{H}$. Then the inequalities (3.3)-(3.9), respectively, imply that the probabilities

$P_h(a_1 \in C, \ldots, a_p \in C), P_h(\text{conv}(a_1, \ldots, a_p) \subseteq \mathcal{C}), P_h(\mathcal{M}\subseteq \mathcal{C}), P_h(\text{cl}(\text{conv}(a_1, \ldots, a_p)) \subseteq \mathcal{C}),$

$P_h(\text{cl}(\mathcal{N}) \subseteq \mathcal{C})$ are positive for all $h \in \mathcal{H}$ provided $\alpha < 1/p$, that

$P_h(\text{diam}(C) \geq \text{diam}((h_0, \mathcal{H}_s)))$ is positive for all $h \in \mathcal{H}$ provided $\alpha < 1/2$, and that

$P_h(\text{diam}(C) = \infty)$ is positive for all $h \in \mathcal{H}$ provided $\alpha < 1$.

The significance of the above theorem is that it relates the "size" of the random set $C$ to the "size" of the nonrandom set $\Theta(h_0, \mathcal{H}_s)$. For example, if the latter set has a large (infinite) diameter, Theorem 3.2(ii) shows that the confidence set $C$ has to have large (infinite) diameter with probability at least $1-2\alpha (1-\alpha)$ under $P_{h_0}$; Theorem 3.2(v) furthermore shows that this is then true up to an error $\epsilon$ also for neighboring $P_h$'s, and under an absolute continuity assumption Theorem 3.2(vi) shows that the probability that $\text{diam}(C)$ is large (infinite) is positive even under all measures $P_h$. In many examples of ill-posed problems, cf. Sections 4 and 5, the set $\Theta(h_0, \mathcal{H}_s)$ will be sufficiently large such that the conditions on $N$ in Theorem 3.2(iii) are met by $N=M$. In this case Theorem 3.2(iii) then implies for convex confidence sets $C$ that $P_h(\mathcal{M}\subseteq \mathcal{C}) \geq 1-p\alpha$ for all $h_0 \in \mathcal{H}$, i.e., that convex confidence sets are completely noninformative with probability not less than $1-p\alpha$. (This covers the case of confidence intervals as a special case. Note that in this case $p=2$ and the lower bound $1-2\alpha$ is positive whenever $\alpha < 1/2$.)

Variants of the results in Theorem 3.2 concerning the diameter of $C$ have already been given in Gleser and Hwang (1987) and Dufour (1997). Theorem 3 in Gleser and Hwang (1987) can be obtained as a special case of Theorem 3.2(iv),(vi).

The results concerning the diameter of $C$ in Theorem 3.2 and in the above mentioned references do not convincingly show that $C$ is "large" (with high or positive probability), since -- without a further assumption like convexity or
connectedness -- a set can have large diameter, without being a "large" set (e.g. it could consist of isolated points separated by large distances). The following theorem shows that -- without imposing convexity or the like -- the confidence set \( C \) can in fact be shown to be "large" not only in terms of diameter but also in terms of volume of the set \( C \), where the volume of a set is expressed in terms of a nonnegative measure \( \mu \) on \( M \). If \( M = \mathbb{R}^n \) a natural candidate for \( \mu \) is Lebesgue-measure. We shall need to impose a measurability condition on \( C \), namely that

\[
(\omega, a) \to 1_C(\omega)(a) \text{ is } \mathcal{X} \odot \mathcal{M}-\text{measurable,} \tag{3.10}
\]

where \( \mathcal{M} \) is a given \( \sigma \)-field on \( M \). Note that (3.10) implies that \( C \) is a random set (in the sense that \( \{ \omega : a \in C(\omega) \} \in \mathcal{X} \) for all \( a \in \mathcal{M} \)) and that \( C(\omega) \in \mathcal{M} \) for all \( \omega \in \mathcal{X} \). We use the convention \( 0 \cdot \omega = 0 \) in (3.11) below if necessary.

**Theorem 3.3:** Let \( \mu \) be a \( \sigma \)-finite nonnegative measure on \((M, \mathcal{M})\) and let \( C \) be a confidence set for \( \Theta(h_0) \), \( h_0 \in \mathcal{H}_* \), with minimum coverage probability not less than \( 1-\alpha, 0 \leq \alpha \leq 1 \). Assume that (3.10) is satisfied. Then for every \( h_0 \in \mathcal{H} \) the expected volume \( E_{h_0} (\mu(C)) \) of the confidence set satisfies

\[
E_{h_0} (\mu(C)) \geq (1-\alpha)\mu(\Theta(h_0_0, \mathcal{H}_*)) \tag{3.11}
\]

provided that \( \Theta(h_0_0, \mathcal{H}_*) \) is an \( \mathcal{M} \)-measurable set. Consequently, for every real number \( \xi < (1-\alpha)\mu(\Theta(h_0_0, \mathcal{H}_*)) \) there is a neighborhood \( U(h_0) \) of \( h_0 \) in \( \mathcal{H} \) such that

\[
E_{h} (\mu(C)) > \xi \tag{3.12}
\]

for all \( h \in U(h_0) \). Furthermore, if \( P_{h_0} \) is absolutely continuous w.r.t. \( P_{h} \) for all \( h \in \mathcal{H} \), then \( E_{h_0} (\mu(C)) > 0 \) for all \( h \in \mathcal{H} \) if and only if \( E_{h_0} (\mu(C)) > 0 \).

In some applications, cf. Sections 4 and 5, \( \mu(\Theta(h_0, \mathcal{H}_*)) \) will not depend on \( h_0 \) and will equal \( c \), say. In such a case Theorem 3.3 provides a common lower bound of the form \((1-\alpha)c\) for the expected volume \( E_{h_0} (\mu(C)) \) for all \( h_0 \in \mathcal{H} \). The independence of \( \mu(\Theta(h_0, \mathcal{H}_*)) \) of \( h_0 \) will occur, e.g., if the set \( \Theta(h_0, \mathcal{H}_*) \)
itself does not depend on $h_0$ or, more generally, if the sets $\Theta(h_0, H_*)$ are translates of each other and $\mu$ is translation invariant.

To connect Theorem 3.3 with Theorem 3.2, observe that $E_{h_0} (\mu(C)) \geq d$ implies $P_{h_0} (\mu(C) \geq d) > 0$ for any $d \in \mathbb{R}$ (and the same is true for strict inequality). E.g., given $(1-\alpha)\mu(\Theta(h_0, H_*)) < \infty$, (3.11) implies $P_{h_0} (\mu(C) \geq (1-\alpha)\mu(\Theta(h_0, H_*))) > 0$. As in Theorem 3.2(v), (v1), this in turn implies

$$P_{h_0} (\mu(C) \geq (1-\alpha)\mu(\Theta(h_0, H_*))) > 0 \tag{3.13}$$

for all $h$ in a sufficiently small neighborhood $U(h_0)$, and it implies (3.13) for all $h \in H$ if $P_{h_0}$ is absolutely continuous w.r.t. $P_h$ for all $h \in H$.

4. Spectral Density Estimation and Estimation of Persistence

The general theory developed in the preceding sections will next be applied to the problem of estimating the value of the spectral density at frequency zero. This is a classical problem with a long history (Anderson (1971), Chp.9), which has regained importance in econometrics in recent years: It is intimately related to estimation of persistence, an important problem in macro-econometrics; see Hauser, Pötscher, and Reschenhofer (1999) for a discussion of the statistical issues involved. Furthermore, it is formally equivalent to heteroskedasticity and autocorrelation consistent variance estimation as well as to estimation of the long run variance involved in the computation of many unit root tests. In subsection 4.1 it is shown to be an ill-posed estimation problem. In fact, we show that for estimating the value of the spectral density at frequency zero the minimax risk over classes of DGP's as small as Gaussian ARMA(1,1) processes is infinite. A similar result is established for persistence estimation. Subsection 4.2 shows that in these contexts any confidence procedure is pretty much uninformative.

4.1 Lower Risk Bounds

Let $(y_t)$ be a Gaussian stationary process with zero mean and spectral
density $f$ belonging to a given set $\mathcal{F}$ of spectral densities that are finite everywhere, i.e., $\mathcal{F}$ is a subset of the set of all integrable and even functions from $[-\pi, \pi]$ to $[0, \infty)$. To exclude trivial cases, we always assume that no $f \in \mathcal{F}$ is zero a.e. If $\hat{\theta}_n$ is any real-valued estimator for $f(0)$ based on a sample $(y_1, \ldots, y_n)$ from the time series, we measure its risk by

$$E_{n,f} |\hat{\theta}_n - f(0)|^p, 1 \leq p < \infty.$$ Here $E_{n,f}$ stands for expectation w.r.t. $P_{n,f}$, the distribution of the sample $(y_1, \ldots, y_n)$ when $f$ is the true spectral density function. Let $\mathcal{F}_c$ denote the set of all spectral densities that are continuous on $[-\pi, \pi]$ and are not zero a.e. Let $\mathcal{F}_\infty$ denote the subset of $\mathcal{F}_c$ whose elements are infinitely often differentiable on $(-\pi, \pi)$. Furthermore, let $\mathcal{F}_{\ARMA(r,s)}$ denote the set of all spectral densities corresponding to an ARMA($r$,s)-model, i.e.,

$$\mathcal{F}_{\ARMA(r,s)} = \{ f_{a,b,\sigma} : 0 < \sigma^2 < \infty, a,b \text{ polynomials}, \deg(a) \leq r, \deg(b) \leq s, a(0)=b(0)=1, a(z)\neq0 \text{ for } |z|>1, b(z)\neq0 \text{ for } |z|<1 \},$$

where

$$f_{a,b,\sigma}(\lambda) = (\sigma^2/2\pi) |b(\exp(i\lambda))|^2 |a(\exp(i\lambda))|^{-2}.$$ We shall write $\mathcal{F}_{\ARMA(r)}$ for $\mathcal{F}_{\ARMA(r,0)}$ and $\mathcal{F}_{\ARMA(s)}$ for $\mathcal{F}_{\ARMA(0,s)}$. Furthermore, let $\mathcal{F}_{\ARMA}$ denote the union of all $\mathcal{F}_{\ARMA(r,s)}$, $r \geq 0$, $s \geq 0$, i.e., $\mathcal{F}_{\ARMA}$ is the set of all ARMA-spectral densities. Similarly, $\mathcal{F}_{\AR}$ denotes the union of all $\mathcal{F}_{\AR(r)}$, $r \geq 0$, i.e., the set of all finite-order autoregressive spectral densities, and $\mathcal{F}_{\MA}$ denotes the union of all $\mathcal{F}_{\MA(s)}$, $s \geq 0$. The $L_1$-(pseudo)distance is denoted by $d_1(f,g) = \frac{\pi}{-\pi} |f(\lambda) - g(\lambda)| d\lambda$. The $L_1$-ball in $\mathcal{F}$ with center $f_0$ and radius $\delta$, $0 < \delta < \infty$, is the set $U(f_0, \delta) = \{ f \in \mathcal{F} : d_1(f,f_0) < \delta \}$. For later use (cf. Theorem 4.1(b) below) we note that $\mathcal{F}_{\AR}, \mathcal{F}_{\MA}, \mathcal{F}_{\ARMA}$, and $\mathcal{F}_\infty$ are sup-norm dense in $\mathcal{F}_c$, see Lemma D.2 in Appendix D.

The following theorem shows in particular that the minimax risk is infinite (and hence does not approach zero as sample size increases) even for quite small classes $\mathcal{F}$ like $\mathcal{F}_{\ARMA(1,1)}$. We also stress that the result holds for any sample size $n \geq 1$. 

16
Theorem 4.1: Let $\hat{\theta}_n$ be any (real-valued) estimator for $f(0)$ based on a sample of size $n$ from the Gaussian process $(y_t)$, and let $1 \leq p < \infty$.

(a) If $\mathcal{F}$ contains $\mathcal{F}_{\text{ARMA}(1,1)}$, there exists $f_0 \in \mathcal{F}$ such that the "local" maximum risk of $\hat{\theta}_n$ over the $L_1$-balls $U(f_0, \delta)$ in $\mathcal{F}$, $\delta > 0$ arbitrary, satisfies for every $n \geq 1$

$$\sup_{f \in U(f_0, \delta)} P_{n,f} |\hat{\theta}_n - f(0)|^p = \infty. \quad (4.1)$$

(b) Suppose $\mathcal{F}$ is a sup-norm dense subset of $\mathcal{F}_c$ (e.g., $\mathcal{F} = \mathcal{F}_{\text{AR}}, \mathcal{F}_{\text{MA}}, \mathcal{F}_{\text{ARMA}}, \mathcal{F}_c$, or $\mathcal{F}_c$). Then (4.1) holds for all $f_0 \in \mathcal{F}$, all $\delta > 0$, and every $n \geq 1$, i.e., the "local" maximum risk of $\hat{\theta}_n$ over any $L_1$-ball in $\mathcal{F}$ is infinite.

(c) Under the assumptions of parts (a) or (b) the maximum risk of $\hat{\theta}_n$ satisfies for every $n \geq 1$

$$\sup_{f \in \mathcal{F}} P_{n,f} |\hat{\theta}_n - f(0)|^p = \infty, \quad (4.2)$$

i.e., the minimax risk for estimating $f(0)$ is infinite.

(d) Under the assumptions of parts (a) or (b) no sequence of uniformly consistent estimators $\hat{\theta}_n$ for $f(0)$ exists. In fact, there exists $\varepsilon_0 > 0$ such that for $0 < \varepsilon < \varepsilon_0$ and for every sequence of estimators $\hat{\theta}_n$

$$\inf_n \sup_{f \in \mathcal{F}} P_{n,f} (|\hat{\theta}_n - f(0)| > \varepsilon) > 0$$

holds. (Under the assumptions of part (b) the supremum over $\mathcal{F}$ can be replaced by the supremum over any $L_1$-ball $U(f_0, \delta)$ in $\mathcal{F}$ where then $\varepsilon_0$ depends on $f_0$.)

The proof of Theorem 4.1 proceeds by verifying the assumptions of Corollary 2.2. Inspection of that proof shows that the lower bound for the local minimax risk at $f_0$ obtained via Corollary 2.2 is positive if and only if the set $\mathcal{F}$ has the property that the evaluation map $f \mapsto f(0)$ is discontinuous (w.r.t. $L_1$-distance) at $f_0$; the corresponding lower bound for the global minimax risk is seen to be positive if and only if the evaluation map is discontinuous somewhere on $\mathcal{F}$. For the sets $\mathcal{F}$ covered by Theorem 4.1 the minimax risk is even infinite, since for those sets $\text{osc}(f(0), f_0)$ can not only be shown to be positive but even to be infinite (for some $f_0$ in part (a) and
all $f_0$ in part (b)). Loosely speaking, this means that one can find spectral densities $f$ and $g$, say, in $\mathcal{F}$ that are arbitrarily close w.r.t. $L_1$-distance (and hence imply sample distributions $P_{n,f}$ and $P_{n,g}$ that are arbitrarily close w.r.t. total variation distance for fixed $n$), but whose values $f(0)$ and $g(0)$ are arbitrarily far apart. In case $\mathcal{F}$ equals $\mathcal{F}_c$ or $\mathcal{F}_\infty$ this is quite obvious; that this is also possible for the smaller classes $\mathcal{F}_\text{AR}$, $\mathcal{F}_\text{MA}$, $\mathcal{F}_\text{ARMA}$, and even for the parametric class $\mathcal{F}_\text{ARMA(1,1)}$ is perhaps less obvious.

The above discussion suggests that in order for the minimax risk over a set $\mathcal{F}$ to converge to zero as $n \to \infty$, a condition that uniformly (over $\mathcal{F}$) limits the oscillatory behaviour of $f$ in a neighborhood around frequency zero is indispensable. A prototypical assumption of this kind could, e.g., limit the magnitude of the derivative of $f$ in a neighborhood of zero uniformly over $\mathcal{F}$. That is, one could require the existence of a finite constant $M$ and a positive $c$ such that

$$|f'(\lambda)| \leq M \text{ for all } |\lambda| < c \text{ and all } f \in \mathcal{F}.$$  

Indeed, under this type of assumptions on the class $\mathcal{F}$, convergence to zero of the minimax risk for estimating $f(0)$ and rates of convergence have been established, see Samarov (1977), Farrell (1979), and Bentkus (1985). Although assumptions on $\mathcal{F}$ like the one just given seem to be indispensable if one wants to establish convergence of the minimax risk to zero, Theorem 4.1 demonstrates that such assumptions are less than innocuous as they exclude standard classes of spectral densities like $\mathcal{F}_\text{AR}$, $\mathcal{F}_\text{MA}$, $\mathcal{F}_\text{ARMA}$, and $\mathcal{F}_\text{ARMA(r,s)}$, $r \geq 1$, $s \geq 1$.

Furthermore, even though the minimax risk over a class $\mathcal{F}$ satisfying the above assumption on the first derivative (with $M$ fixed) will converge to zero with increasing sample size, for fixed sample size the minimax risk will still be large if $M$ is large, since the minimax risk will go to infinity for $M \to \infty$.

An alternative assumption also limiting the oscillatory behaviour of $f$ uniformly in $f$, is to assume that $\mathcal{F}$ is generated by a sufficiently regular parametric time series model of fixed (finite) dimension. For example, consider the case where $\mathcal{F} = \mathcal{F}_{\text{MA}(s)}$ and $p = 2$, i.e., risk is mean squared error.
Then the minimax risk over \( \mathcal{F} \) indeed converges to zero at rate \( n^{-1} \) for each given model order \( s \), and the maximum risk of the MLE achieves this rate (Maliukeničius (1986)). However, note that for any given sample size \( n \) and any given estimator \( \hat{\theta}_n \), the maximum risk over \( \mathcal{F} \) will nevertheless be large when \( s \) is large, since it goes to infinity for \( s \to \infty \) by Theorem 4.1. Sometimes estimator sequences \( \tilde{\theta}_n \) for \( f(0) \) obtained via fitting an \( \text{MA}(s_n) \)-model with increasing model order \( s_n \) (or another parametric model whose dimension increases with \( n \)) are suggested. The rationale behind such procedures is that, while the true \( f \) is assumed to correspond to an \( \text{MA}(s) \)-process, the order \( s \) is unknown. But this implies \( \mathcal{F} = \bigcup_{s \geq 0} \mathcal{F}_{\text{MA}(s)} \), the union of all \( \mathcal{F}_{\text{MA}(s)} \), \( s \geq 0 \). Hence, Theorem 4.1 applies and shows that the maximum risk of \( \tilde{\theta}_n \) over \( \mathcal{F} \) is infinite. 4

Theorem 4.1(a) establishes the ill-posedness of the problem of estimating \( f(0) \) in the \( \text{ARMA}(r,s) \)-context, \( r \geq 1, s \geq 1 \). In particular, it shows the \( \text{ARMA}(r,s) \)-model, although being a parametric model, not to be sufficiently regular in the sense of the previous paragraph. Of course, considering sufficiently restricted subsets of the set \( \mathcal{F}_{\text{ARMA}(r,s)} \) of \( \text{ARMA}(r,s) \) spectral densities will lead to well-posed estimation problems. Such restrictions, however, are highly artificial and hard to defend on a priori grounds. (E.g., in the \( \text{ARMA}(1,1) \) case one such restricted subset is given by \( \mathcal{F}(\epsilon) = \{ f_{a,b,c} \in \mathcal{F}_{\text{ARMA}(1,1)} : \text{the AR and MA parameters differ by more than} \ \epsilon \} \) with \( \epsilon > 0 \).)

We turn to persistence estimation next. Define \( \mathcal{F}_{\text{reg}} \) to be the set of all spectral densities \( f: [-\pi, \pi] \to [0, \infty) \) that are not zero a.e. and correspond to linearly regular processes, i.e., satisfy \( \int_{-\pi}^{\pi} \log(f(\lambda))d\lambda < \infty \). Note that \( \mathcal{F}_{\text{MA}} \), \( \mathcal{F}_{\text{ARMA}} \), and \( \mathcal{F}_{\text{reg}} \) are subsets of \( \mathcal{F}_{\text{reg}} \), and hence they are also sup-norm dense in \( \mathcal{F}_{\text{MA}} \), \( \mathcal{F}_{\text{ARMA}} \), and \( \mathcal{F}_{\text{reg}} \) by Lemma D.2. Define \( \mathcal{F}_{\omega,\text{reg}} = \mathcal{F}_{\text{reg}} \cap \mathcal{F}_{\omega} \). Since \( \mathcal{F}_{\text{MA}} \subset \mathcal{F}_{\omega,\text{reg}} \), the latter set is also sup-norm dense in \( \mathcal{F}_{\omega,\text{reg}} \). For \( f \in \mathcal{F}_{\omega,\text{reg}} \) set \( \sigma^2(f) = 2\pi \exp((2\pi)^{-1} \int_{-\pi}^{\pi} \log(f(\lambda))d\lambda) \) and recall that \( \sigma^2(f) \) is then positive and is the variance of the innovations in the Wold-decomposition. Persistence of the process \( (Y_t) \), satisfying

\[ \Delta Y_t = \text{const} \cdot Y_t \]
with the spectral density \( f \) of \( (y_t) \) belonging to \( F_{\text{reg}} \), is defined as \( c_\infty(f) = (2\pi f(0)/\sigma^2(f))^{1/2} \).

**Theorem 4.2:** Let \( \hat{\theta}_n \) be any (real-valued) estimator for \( c_\infty(f) \) based on a sample of size \( n \) from the Gaussian process \( (y_t) \), and let \( 1 \leq p < \infty \).

(a) If \( F_{\text{reg}} \) contains \( F_{\text{ARMA}(1,1)} \), there exists \( f_0 \in F \) such that the "local" maximum risk of \( \hat{\theta}_n \) over the \( L_1 \)-balls \( U(f_0, \delta) \) in \( F \), \( \delta > 0 \) arbitrary, satisfies for every \( n \geq 1 \)

\[
\sup_{f \in U(f_0, \delta)} \mathbb{E}_n, f | \hat{\theta}_n - c_\infty(f) |^p = \infty. \tag{4.3}
\]

(b) Suppose \( F \) is a sup-norm dense subset of \( F_{\text{C, reg}} \) (e.g., \( F = F_{\text{AR}}, F_{\text{MA}}, F_{\text{ARMA}}, F_{\text{C, reg}} \) or \( F_{\text{ ARMA}}, F_{\text{ MA}}, F_{\text{ ARMA}}, F_{\text{C, reg}} \)). Then (4.3) holds for all \( f \in F \), all \( \delta > 0 \), and every \( n \geq 1 \), i.e., the "local" maximum risk of \( \hat{\theta}_n \) over any \( L_1 \)-ball in \( F \) is infinite.

(c) Under the assumptions of parts (a) or (b) the maximum risk of \( \hat{\theta}_n \) satisfies for every \( n \geq 1 \)

\[
\sup_{f \in F} \mathbb{E}_n, f | \hat{\theta}_n - c_\infty(f) |^p = \infty, \tag{4.4}
\]

i.e., the minimax risk for estimating \( c_\infty(f) \) is infinite.

(d) Under the assumptions of parts (a) or (b) no sequence of uniformly consistent estimators \( \hat{\theta}_n \) for \( c_\infty(f) \) exists. In fact, there exists \( \varepsilon_0 > 0 \) such that for \( 0 < \varepsilon < \varepsilon_0 \) and for every sequence of estimators \( \hat{\theta}_n \)

\[
\inf_n \sup_{f \in F} \mathbb{P}_n, f (| \hat{\theta}_n - c_\infty(f) | > \varepsilon) > 0
\]

holds. (Under the assumptions of part (b) the supremum over \( F \) can be replaced by the supremum over any \( L_1 \)-ball \( U(f_0, \delta) \) in \( F \) where then \( \varepsilon_0 \) depends on \( f_0 \).)

A discussion similar to the one following Theorem 4.1 also applies here and will not be repeated.

**Remark 4.1:** (Unknown Mean/Non-Gaussian Processes) (a) Suppose \((y_t)\) has mean \( \mu \) and spectral density \( f \) varying in \( F \). Let \( P_{n,f,\mu} \) denote the distribution of \((y_1, \ldots, y_n)\) and let \( E_{n,f,\mu} \) denote the corresponding expectation. Then Theorems 4.1 and 4.2 continue to hold if \( P_{n,f} \) and \( E_{n,f} \), respectively, are
replaced by $P_{n,f,\mu}$ and $E_{n,f,\mu}$ at every occurrence in those theorems. A fortiori all these results remain valid if the various suprema over $f \in \mathcal{F}$ are replaced by suprema over $f \in \mathcal{F}$ and $\mu \in \mathbb{R}$.

(b) The Gaussianity assumption is inessential in the sense that for any class of weakly stationary processes containing a class of Gaussian processes satisfying the assumptions of Theorems 4.1 or 4.2, corresponding results hold a fortiori; cf. the discussion in Remark A.3 (with the role of the nuisance parameter here being played by those aspects of the distribution of the process $y_t$ that are not already encoded in the spectral density).

**Remark 4.2:** (a) In the context of Theorem 4.1, suppose $\mathcal{F}$ is not contained in $\mathcal{F}_c$. Then (4.2) and the non-existence of uniformly consistent estimators for $f(0)$ continue to hold if $\mathcal{F} \cap \mathcal{F}_c$ is sup-norm dense in $\mathcal{F}_c$; if, additionally, $\mathcal{F}$ is the $L_1$-closure of $\mathcal{F} \cap \mathcal{F}_c$, then also (4.1) holds for any $f_0 \in \mathcal{F}$. A similar remark applies to Theorem 4.2.

(b) If the truncated loss $\ell^{p,\alpha}(a,b) = \min(\alpha, |a-b|^p)$ is used instead of $\ell^{p,\infty}(a,b) = |a-b|^p$, then in view of Corollary 2.2 the results in Theorems 4.1 and 4.2 continue to hold if "$=\omega$" is replaced by "$=2^p\alpha$" in (4.1)-(4.4).

(c) If a lower bound on the magnitude of $f$ is introduced, i.e., if the classes $\mathcal{F}$ in Theorems 4.1 and 4.2 are replaced by $\mathcal{F}(f: |f(\lambda)| \geq b > 0$ for all $\lambda \in [-\pi, \pi])$, both theorems continue to hold. If an upper bound is introduced, i.e., if the classes $\mathcal{F}$ in both theorems are replaced by $\mathcal{F}(f: |f(\lambda)| \leq b$ for all $\lambda \in [-\pi, \pi])$, the results continue to hold provided "$=\omega$" is replaced by "$\leq \mathrm{const} \geq 0$" in (4.1)-(4.3). The proofs of these results become more involved and are available on request.

(d) Results quite similar to those in this section can also be obtained for the problem of estimating $f(\lambda)$, $\lambda \neq 0$. This disproves the widely held belief that difficulties encountered in estimating persistence are a feature that is specific to estimation of long run behaviour.

21
4.2 Properties of Confidence Sets

Let \( C_n \) denote a confidence set for \( f(0) \) based on a sample of size \( n \) from the mean zero Gaussian process \( (y_t) \) with spectral density \( f \in \mathcal{F} \), where \( \mathcal{F} \) is as at the beginning of Section 4.1. I.e., \( C_n = C_n(y_1, \ldots, y_n) \) is a subset of \([0, \omega)\) for each \((y_1, \ldots, y_n) \in \mathbb{R}^n\) with the property that the set \( \{(y_1, \ldots, y_n) : a \in C_n\} \) is Borel-measurable for every \( a \in [0, \omega) \). We say that \( C_n \) has minimum coverage probability over \( \mathcal{F} \) not less than \( 1-\alpha, 0 \leq \alpha < 1 \), if \( P_{n,f}(f(0) \in C_n) = 1-\alpha \) for all \( f \in \mathcal{F} \).

We say that \( C_n \) is a (closed) confidence interval, if \( C_n(y_1, \ldots, y_n) \) is a (closed) interval of \([0, \omega)\) for all \((y_1, \ldots, y_n) \in \mathbb{R}^n\). We say that \( C_n \) is diameter-measurable if \( \text{diam}(C_n) \) is Borel-measurable. For parts of the results we shall need to assume "joint" measurability of \( C_n \), i.e.,

\[
(y_1, \ldots, y_n, a) \to 1_{C_n(y_1, \ldots, y_n)}(a) \text{ is Borel-measurable.} \quad (4.5)
\]

Theorem 4.3: Let \( C_n \) be a confidence set for \( f(0) \) with minimum coverage probability over \( \mathcal{F} \) not less than \( 1-\alpha, 0 \leq \alpha < 1 \), based on a sample of size \( n \) from the Gaussian process \( (y_t) \) with spectral density \( f \in \mathcal{F} \).

(a) Suppose \( \mathcal{F} \) contains \( \mathcal{F}_{\text{ARMA}}(1,1) \). Then for every \( f \in \mathcal{F} \) and \( n \geq 1 \) the following hold:

If \( C_n \) is diameter-measurable, then

\[
P_{n,f}(\text{diam}(C_n) = \omega) > 0. \tag{4.6}
\]

If \( C_n \) satisfies (4.5), then

\[
E_{n,f}(\text{length}(C_n)) > 0. \tag{4.7}
\]

where \( \text{length}(.) \) denotes Lebesgue-measure.

If \( C_n \) is a confidence interval and \( \alpha < 1/2 \), then

\[
P_{n,f}([0, \omega) \subseteq C_n) > 0. \tag{4.8}
\]

Furthermore, there exists \( f_0 \in \mathcal{F} \) such that the l.h.s. in (4.6) is not less than \( 1-\alpha \), the l.h.s. in (4.7) is infinite, and the l.h.s. in (4.8) is not less than \( 1-2\alpha \).

(b) Suppose \( \mathcal{F} \) is sup-norm dense in \( \mathcal{F}_c \) (e.g., \( \mathcal{F} = \mathcal{F}_{\text{AR}}, \mathcal{F}_{\text{MA}}, \mathcal{F}_{\text{ARMA}}, \mathcal{F}_\infty \), or \( \mathcal{F}_c \)).
Then for every \( f \in \mathcal{F} \) and \( n \geq 1 \) the following hold:

If \( C_n \) is diameter-measurable, then
\[
P_{n,f}(\text{diam}(C_n) = \infty) \geq 1 - \alpha. \tag{4.9}
\]

If \( C_n \) is a confidence interval, then
\[
P_{n,f}((0, \infty) \subseteq C_n) \geq 1 - 2\alpha. \tag{4.10}
\]

If \( C_n \) is a closed confidence interval, then
\[
P_{n,f}([0, \infty) \subseteq C_n) \geq 1 - 2\alpha. \tag{4.11}
\]

(c) Suppose \( \mathcal{F} = \mathcal{F}_N, \mathcal{F}_{\text{ARMA}}, \mathcal{F}_\infty \), or \( \mathcal{F}_C \). Then for every \( f \in \mathcal{F} \) and \( n \geq 1 \) in addition to (4.9) also the following hold:

If \( C_n \) is a confidence interval, then
\[
P_{n,f}((0, \infty) \subseteq C_n) \geq 1 - 2\alpha. \tag{4.12}
\]

If \( C_n \) satisfies (4.5), then
\[
E_{n,f}(\text{length}(C_n)) = \infty. \tag{4.13}
\]

where \( \text{length}(.) \) denotes Lebesgue-measure.

(d) Suppose \( \mathcal{F} = \mathcal{F}_{\text{AR}} \). Then for every \( f \in \mathcal{F} \) and \( n \geq 1 \) in addition to (4.9)-(4.11) also (4.13) holds.

Results similar to Theorem 4.3 can also be obtained for confidence sets for the persistence measure \( c_\infty(f) \). A consequence of (4.6) or (4.9) is that bounded confidence sets for \( f(0) \) necessarily must have coverage probability zero if \( \mathcal{F} \) is as above. (Faust (1999) sketches a similar result for bounded confidence intervals.\(^6\)) In particular, this applies to the conventional confidence intervals based on asymptotically normal spectral estimates.

4.3 Implications for Statistical and Econometric Practice

(1) Theorems 4.1-4.3 show that inference on \( f(0) \) (or on the persistence measure \( c_\infty(f) \)) is effectively only possible if very strong a priori assumptions on the set of feasible DGP's are made. Once this set is rich enough to contain all Gaussian ARMA(1,1) processes (all AR-, or all MA-processes), the problem is ill-posed and reliable inference is no longer

23
possible, as the minimax risk is infinite, no uniformly consistent estimator exists, and confidence sets become pretty much uninformative. One implication of these results is that inference on unit roots (and hence on cointegration) is also reliably possible only under very restrictive assumptions on the set of feasible DGP's since deciding on the presence of a unit root often amounts to determining whether or not \( f(0) \) is positive, where \( f \) is the spectral density of the process of first differences. In particular, whenever the model contains all ARMA(1,1) processes (all AR-, or all MA-processes) as possible candidates for the first differences, reliable unit root inference is not possible.

(11) Let \( Y_t \) be a process satisfying \( \Delta Y_t = \cdot \) and assume \( \mathcal{F} = \mathcal{F}^{\text{MA}} \). (A similar discussion applies for other sets like, e.g., \( \mathcal{F}^{\text{ARMA}} \) as well.) In this context difference-stationarity (DS) of \( Y_t \) amounts to \( f(0) > 0 \) and trend-stationarity (TS) to \( f(0) = 0 \). Let \( D \) be an arbitrary Borel-set in \( \mathbb{R}^n \), i.e., the critical region of an arbitrary test of the null hypothesis \( H_0: \text{"} Y_t \text{" is DS} \) versus \( H_1: \text{"} Y_t \text{" is TS} \) based on the data \( (y_1, \ldots, y_n) \). On the one hand, as shown in the proof of Theorem 4.1, the probability of rejection, i.e., \( P_{n,f}(D) \), depends continuously on \( f \in \mathcal{F} \) (w.r.t. \( L_1 \)-distance). On the other hand, it was shown in the proof of Lemma D.3 in Appendix D that every \( f \in \mathcal{F} \) has arbitrarily \( L_1 \)-close elements \( f' \in \mathcal{F} \) taking any given value for \( f'(0) \), from which follows that \( H_0 \) and \( H_1 \) are both dense in \( \mathcal{F} \). As a consequence power cannot exceed size; if the test given by \( D \) is also consistent against at least one alternative \( f \in H_1 \), then size approaches unity as sample size increases. The same result is also true if the roles of \( H_0 \) and \( H_1 \) are interchanged.\(^7\) (In fact, the above discussion applies if \( H_0 \) and \( H_1 \) are narrowed down to \( H_0: f(0) = c_0 \) and \( H_1: f(0) = c_1, c_0 \neq c_1 \).) The above discussion sharpens the results in Blough (1992) and Faust (1996) in that it allows for arbitrary tests. This is so since the proof of Theorem 4.1 establishes convergence in total variation norm of \( P_{n,f} \) when \( f \) converges, whereas the above references establish only weak convergence and therefore can handle only

\[ \text{end to the case where } y_t \text{ has non-zero mean, cf. Remark 4.1} \]
tests with critical regions that have a boundary that has probability zero under any $P_{f,n}^8$.

(iii) Motivated by the recently revived econometric interest in estimating $f(0)$, Ng and Perron (1996) computed exact formulae for the mean squared error of kernel estimators of $f(0)$, thus extending previous work by Næve (1971). The resulting expressions being rather unwieldy, Ng and Perron (1996) resort to numerical evaluation of these expressions for a collection of "representative" DGPs. Theorem 4.1, however, shows that the mean squared error is highly discontinuous (w.r.t. $L_1$-distance) at every spectral density (given $\mathcal{F}$ is rich enough). Hence, conclusions drawn from such "representative" DGPs may already be incorrect at arbitrarily small perturbations of these DGPs.

(iv) The difficulties arising in inference for $f(0)$ discussed above should actually not come as a surprise as it is not the value of the spectral density at frequency zero that should matter, but the area under the curve in a given band around that frequency. I.e., interest should focus, e.g., on $(2\pi)^{-1}\int_{-\epsilon}^{\epsilon} f(\lambda) d\lambda, \epsilon>0$, rather than $f(0)$. See Hauser, Pötscher, and Reschenhofer (1999), Section 2.1, for more discussion. Note that estimation of the former quantity is not an ill-posed problem. Of course, if $f$ is smooth, these two quantities are approximately the same for small $\epsilon$, but observe that the approximation error does not become small uniformly in $\mathcal{F}$ as $\epsilon \to 0$ (provided $\mathcal{F}$ is rich enough).

5. Estimation of Long Memory Parameter

Semiparametric estimation of the long memory parameter has been the focus of considerable research in recent years (e.g., Geweke and Porter-Hudak (1983), Robinson (1995a,b)). We next show that this estimation problem is ill-posed for commonly used specifications of the infinite-dimensional nuisance parameter.

Again let $(y_t)$ be a Gaussian stationary process with zero mean and
spectral density $f$ which is assumed to be of the form

$$f(\lambda) = \lambda^{-2d}g(\lambda)$$

(5.1)

where $g$ will be specified in a moment and where $d$, $-1/2 \leq d < 1/2$, is the long memory or fractional integration parameter which is the parameter of interest.

Define $\mathcal{F}_{\text{ARFI}}$, $\mathcal{F}_{\text{FIMA}}$, and $\mathcal{F}_{\text{ARIMA}}$ to be the set of all spectral densities $f$ of the form (5.1) where $g$ belongs to $\mathcal{F}_{\text{AR}}$, $\mathcal{F}_{\text{MA}}$, and $\mathcal{F}_{\text{ARMA}}$, respectively.

Furthermore, $\mathcal{F}_{C,FI}$ ($\mathcal{F}_{\infty,FI}$) denotes the set of all spectral densities of the form (5.1) where $g \in \mathcal{F}_{C}$ ($g \in \mathcal{F}_{\infty}$) and satisfies $g(0) > 0$. Note that for $f$ belonging to $\mathcal{F}_{\text{ARFI}}$, $\mathcal{F}_{\text{FIMA}}$, $\mathcal{F}_{\text{ARIMA}}$, $\mathcal{F}_{\infty,FI}$ or $\mathcal{F}_{C,FI}$ the parameter $d$ is uniquely defined and we shall sometimes write $d(f)$ for $d$.

Similarly as in Section 4, the proof of the following theorem proceeds by verifying the assumptions in Corollary 2.2 and by showing that $\text{osc}(d(f), f_0) = 1$ for every $f_0 \in \mathcal{F}$.

**Theorem 5.1:** Let $\hat{d}_n$ be any (real-valued) estimator for $d$ based on a sample of size $n$ from the Gaussian process $(y_t)$ with spectral density $f \in \mathcal{F}$, and let $1 \leq p < \infty$.

(a) Suppose $\mathcal{F} = \mathcal{F}_{\text{ARFI}}$, $\mathcal{F}_{\text{FIMA}}$, $\mathcal{F}_{\text{ARIMA}}$, $\mathcal{F}_{\infty,FI}$ or $\mathcal{F}_{C,FI}$. Then the maximum risk of $\hat{d}_n$ satisfies for every $n \geq 1$

$$\sup_{f \in \mathcal{F}} \mathbb{E}_n[f|\hat{d}_n - d|^p] \geq 1/2^p > 0,$$

(5.2)

i.e., the minimax risk for estimating $d$ is bounded from below by a positive constant independent of $n$. In fact, for every $f_0 \in \mathcal{F}$ the "local" maximum risk of $\hat{d}_n$ over any $L_1$-ball $U(f_0, \delta)$ in $\mathcal{F}$, $\delta > 0$ arbitrary, satisfies for every $n \geq 1$

$$\sup_{f \in U(f_0, \delta)} \mathbb{E}_n[f|\hat{d}_n - d|^p] \geq 1/2^p > 0.$$  

(5.3)

(b) Under the assumptions of part (a) no sequence of uniformly consistent estimators $\hat{d}_n$ for $d$ exists. In fact, there exists $\varepsilon_0 > 0$ such that for $0 < \varepsilon < \varepsilon_0$ and for every sequence of estimators $\hat{d}_n$

$$\inf_n \sup_{f \in \mathcal{F}} \mathbb{P}_n[f|\hat{d}_n - d| > \varepsilon] > 0$$

holds. (The supremum over $\mathcal{F}$ can in fact be replaced by the supremum over any
\[ L^1 \text{-ball } U(f_0, \delta) \text{ in } \mathcal{F} \text{ where then } \varepsilon_0 \text{ depends on } f_0. \]

For the sets \( \mathcal{F} \) in Theorem 5.1 the minimax risk does not go to zero with increasing sample size. In order to achieve convergence to zero, more restrictive assumptions on \( g \) are needed, e.g., uniform (w.r.t. \( g \)) restrictions on the behaviour of \( g \) or its derivatives near zero, see Giraitis, Robinson, and Samarov (1997). Turning to confidence sets, the following theorem establishes, e.g., that confidence intervals for \( d \) coincide with the entire parameter space for \( d \) (and hence are uninformative) with high probability.

**Theorem 5.2:** Let \( C_n \subseteq [-1/2,1/2] \) be a confidence set for the long memory parameter \( d \) with minimum coverage probability over \( \mathcal{F} \) not less than \( 1-\alpha \), \( 0 \leq \alpha \leq 1 \), based on a sample of size \( n \) from the Gaussian process \( (y_t) \) with spectral density \( f \in \mathcal{F} \). Suppose \( \mathcal{F} = \mathcal{F}_{\text{ARFI}}, \mathcal{F}_{\text{FIMA}}, \mathcal{F}_{\text{ARFIMA}}, \mathcal{F}_{\omega, \text{FI}} \) or \( \mathcal{F}_{\omega, \text{FI}} \). Then for every \( f \in \mathcal{F} \) and \( n \geq 1 \) the following hold:

If \( C_n \) is diameter-measurable, then

\[
P_{n, f}(\text{diam}(C_n) \geq 1) \geq 1 - 2\alpha. \quad (5.4)
\]

If \( C_n \) satisfies (4.5), then

\[
E_{n, f}(\text{length}(C_n)) \geq 1 - \alpha. \quad (5.5)
\]

where \( \text{length}(\cdot) \) denotes Lebesgue-measure.

If \( C_n \) is a confidence interval, then

\[
P_{n, f}(C_n = [-1/2,1/2]) \geq 1 - 2\alpha. \quad (5.6)
\]

Mutatis mutandis, most of the discussion in Section 4 also applies here and will not be repeated.

6. Conclusion

Important estimation problems in econometrics like (i) estimating the value of a spectral density at frequency zero, which appears in the econometric literature in various guises like heteroskedasticity and
autocorrelation consistent variance estimation or long run variance
estimation, (ii) the closely related problem of estimating persistence of a
time series, which is intimately connected to inference on unit roots and
cointegration, as well as (iii) the problem of estimating the long memory
parameter have been shown to fall into the category of ill-posed estimation
problems under commonly used specifications for the set of feasible DGPs.
These results have been obtained as special cases of a more general theory for
abstract estimation problems developed in Sections 2 and 3. The consequences
of ill-posedness are quite dramatic: the minimax risk does not vanish
asymptotically and may even be infinite, uniformly consistent estimators do
not exist, and confidence sets are close to being uninformative. This leads
to the conclusion that reliable inference is not possible in case of an
ill-posed estimation problem. Consider, for example, the problem of
estimating \( f(0) \), the value of the spectral density \( f \) at frequency zero. A
consequence of the results of the paper is that one either has to give up any
attempt at estimating \( f(0) \) (under the commonly used specifications for the set
of DGPs discussed in Section 4) and shift focus to a quantity like, e.g.,
\[
(2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda) d\lambda,
\]
which can be estimated more reliably, or one has to be
comfortable with quite strong a priori assumptions, i.e., with much more
restricted sets of DGPs (e.g., sufficiently regular parametric models of fixed
dimension or sets of spectral densities \( f \) that obey a given bound on the
derivative of \( f \) near zero uniformly in \( f \)). We finally note that the general
theory developed in Sections 2 and 3 can be applied to many other important
estimation problems including probability density and regression function
estimation; see Pötscher (1999) for a discussion of further examples.

Appendix A. Remarks on Section 2 and Some Generalizations

Remark A.1: (On Theorem 2.1) (a) The stronger the topology \( \tau \) on \( \mathcal{H} \) is chosen
in Theorem 2.1, the easier it is to obtain continuity of the map \( h \rightarrow \mathcal{P}_h \), but
the smaller are the resulting lower bounds in (2.3) and (2.4) as the infimum
in (2.2) is then taken over more sets. Hence, concerning (2.4), an optimal choice for \( \tau \) is \( \tau_w \), the weakest topology making the map \( h \rightarrow P_h \) continuous. (This topology is first countable as is easily seen.) However, sometimes using a stronger topology may be more convenient, e.g., by leading to more easily computed bounds. (In contrast to (2.4), the l.h.s. of (2.3) does depend on the choice of \( \tau \) and hence a similar statement on optimality of \( \tau_w \) does not apply.)

(b) In case the map \( h \rightarrow P_h \) is injective, one can w.l.o.g. set \( \mathcal{Ha} = \mathcal{P} \) (e.g., with the total variation topology which in this case coincides with \( \tau_w \)). However, the formulation of Theorem 2.1 as given has the advantage that it allows for non-identified parameterizations, as well as for a simple derivation of "automatically" sample size independent lower bounds when Theorem 2.1 is applied to a sequence of estimation problems as discussed after that theorem.

(c) The choice of the \( \sigma \)-field \( \mathcal{M} \) does not affect the size of the lower bounds, but only determines the class of estimators (and loss-functions) for which these bounds hold.

(d) Suppose \( \ell \) is a loss-function such that there exists a proper loss function \( l \) satisfying \( \ell(a, b) = l(a, b) \) for all \( a, b \in \mathcal{M} \). Then the lower risk bounds obtained for \( \ell \) obviously also hold for \( \ell \).

Remark A.2: (Non-identifiability) The formulation of Theorem 2.1 allows for non-identified parameterizations of \( \mathcal{P} \), since the map \( h \rightarrow P_h \) need not be injective. In order to obtain a large lower bound in (2.4) in case of a non-identified parameterization, a topology that does not separate observationally equivalent parameters (like \( \tau_w \)) is frequently a good choice. For such a topology the entire set of parameters \( h \) observationally equivalent to a given \( h_0 \) is contained in any neighborhood of \( h_0 \), which typically leads to a large lower bound in (2.4). We note that for two observationally equivalent points \( h_0 \) and \( h_1 \) the following rather trivial lower bound holds without any assumption on the map \( h \rightarrow P_h \): For any set \( A \subset \mathcal{M} \) containing \( h_0 \) and \( h_1 \) and any
estimator \( \hat{\theta} \)
\[
\sup_{h \in \mathcal{H}} E_{h}(\hat{\theta}, \theta(h)) \geq (2\kappa(\ell))^{-1} \ell(\theta(0), \theta(1)).
\]
This lower bound will typically be positive except if \( \theta \) is constant on the observational equivalence classes.

**Remark A.3:** (Nuisance Parameters) Theorem 2.1 assumes that the parameter \( h \) fully characterizes the corresponding probability measure \( P_h \). This is not so for estimation problems where (finite or infinite dimensional) nuisance parameters are present. One solution in this case is to redefine the parameter \( h \) such as to incorporate the nuisance parameter, thus bringing the estimation problem back under the scope of Theorem 2.1. Alternatively, if one can exhibit a subset \( \mathcal{P}^* \subseteq \mathcal{P} \), that is fully parameterized by the original parameter space \( \mathcal{H} \), Theorem 2.1 can be applied to \( \mathcal{P}^* \), which gives rise to bounds for the original estimation problem as follows:

(a) Suppose the set \( \mathcal{P} \) of probability measures on \((X, \mathcal{X})\) is parameterized by a parameter of interest \( h \in \mathcal{H} \) (a functional \( \theta \) of which is to be estimated) and a nuisance parameter \( r \in \mathcal{R} \), \( \mathcal{R} \) an arbitrary non-empty set. I.e., there is a map \( (h, r) \rightarrow P_{(h, r)} \) from \( D \times \mathcal{H} \times \mathcal{R} \) onto \( \mathcal{P} \). We define \( D_h = \{(r, h) \in D : \exists h \in \mathcal{H} \} \) and assume w.l.o.g. that \( D_h \neq \emptyset \) for all \( h \in \mathcal{H} \). Suppose further there exists a subset \( \mathcal{P}^* \) of \( \mathcal{P} \) such that \( \mathcal{P}^* = \{P_{(h, \sigma(h))} : h \in \mathcal{H} \} \) for some function \( \sigma : \mathcal{H} \rightarrow \mathcal{R} \) (e.g., \( \sigma(h) = r \) for some fixed \( r \in \mathcal{R} \)) and assume that there is a first countable topology \( \tau \) on \( \mathcal{H} \) such that the map \( h \rightarrow P_{(h, \sigma(h))} \) from \( \mathcal{H} \) to \( \mathcal{P}^* \) is continuous when \( \mathcal{P}^* \) is endowed with the total variation distance. Let \( \theta, \mathcal{H} \) and \( \ell \) be as in Theorem 2.1. Then this theorem implies that for every \( h_0 \in \mathcal{H} \) and every estimator \( \hat{\theta} \)
\[
\sup_{h \in U(h_0)} E_{(h, \sigma(h))}(\hat{\theta}, \theta(h)) \geq c(\theta, \ell, \mathcal{H}, \tau, h_0)
\]
and
\[
\sup_{h \in \mathcal{H}} E_{(h, \sigma(h))}(\hat{\theta}, \theta(h)) \geq c(\theta, \ell, \mathcal{H}, \tau)
\]
hold, where \( c(\theta, \ell, \mathcal{H}, \tau, h_0) \) and \( c(\theta, \ell, \mathcal{H}, \tau) \) are as in Theorem 2.1. It follows that for all \( h_0 \in \mathcal{H} \) and all estimators \( \theta \).
\[ \sup_{h \in U(h_0)} \sup_{r \in D_h} E_{(h, r)} \ell(\hat{\theta}, \theta(h)) \geq c(\theta, \ell, \mathcal{H}, \tau, h_0), \]  
(A.1)

and hence
\[ \sup_{(h, r) \in D_{(h, r)}} E_{(h, r)} \ell(\hat{\theta}, \theta(h)) \geq c(\theta, \ell, \mathcal{H}, \tau). \]  
(A.2)

(b) More generally, assuming the setup of part (a) but without the subset \( \mathcal{P}^* \), let \( \tau \) be a first countable topology on \( \mathcal{H} \) such that for all \( h_0 \in \mathcal{H} \) we have:

For any sequence \( h_1 \in \mathcal{H} \) with \( h_1 \to h_0 \) there exist \( r_1 \in \mathcal{R} \) and \( r_0 \in \mathcal{R} \) such that \( P(h_1, r_1) \) converges to \( P(h_0, r_0) \) in total variation distance.

Then, for all \( h_0 \in \mathcal{H} \) and all estimators \( \hat{\theta} \), inequalities (A.1) and (A.2) can again be shown to hold.

(c) Estimation problems, for which a map \( \psi: \mathcal{P} \to \mathcal{H} \), assigning parameters \( h = \psi(P) \) to the probability measures \( P \in \mathcal{P} \), is given and \( \theta(h) = (\theta \circ \psi)(P) \) is to be estimated, can be viewed as estimation problems where nuisance parameters are present, e.g., by choosing \( \mathcal{R} = \mathcal{P}, \mathcal{D}_h = \{ (h, P): \psi(P) = h \} \) and \( \mathcal{D} = \bigcup \mathcal{D}_h : h \in \mathcal{H} \). (Of course, the problem can also be brought under the framework of Theorem 2.1 by redefining \( \mathcal{H} \) so that it coincides with \( \mathcal{P} \).

**Remark A.4:** Suppose the assumptions of Theorem 2.1 are met except that the function \( \theta \) is only defined on a subset \( \mathcal{H}_* \) of \( \mathcal{H} \). Then Theorem 2.1 immediately applies with \( \mathcal{H}_* \) replacing \( \mathcal{H} \). Alternatively, the following simple extension of Theorem 2.1 is sometimes useful: Define for \( h_0 \in \mathcal{H} \)

\[ c_*(\theta, \ell, \mathcal{H}, \tau, h_0) = (2\kappa(\ell))^{-1}\inf\{ \sup_{f, g \in W_{H_0}} \ell(\theta(f), \theta(g)) ; h_0 \in W_{H_0}, W \tau\text{-open} \} \]

with the convention that the supremum over the empty set is zero. Also let

\[ c_*(\theta, \ell, \mathcal{H}, \tau) = \sup_{h_0 \in \mathcal{H}_*} c_*(\theta, \ell, \mathcal{H}, \tau, h_0). \]

Then for every \( h_0 \in \mathcal{H} \), every neighborhood \( U(h_0) \) of \( h_0 \) in \( \mathcal{H} \), and every estimator \( \hat{\theta} \), we have

\[ \sup_{h \in U(h_0) \cap \mathcal{H}} E_{(h, r)} \ell(\hat{\theta}, \theta(h)) \geq c_*(\theta, \ell, \mathcal{H}, \tau, h_0) \]

and

\[ \sup_{h \in \mathcal{H}_*} E_{(h, r)} \ell(\hat{\theta}, \theta(h)) \geq c_*(\theta, \ell, \mathcal{H}, \tau). \]
The proof is similar to the proof of Theorem 2.1.

The next theorem provides simpler expressions or lower estimates for the minimax risk bounds in (2.3) and (2.4). In particular, conditions are given under which these bounds are positive. Under appropriate conditions, positivity of the lower bound in (2.3) turns out to be equivalent to discontinuity of the functional $\theta$ at $h_0$.

**Theorem A.1:** Let the assumptions of Theorem 2.1 hold.

(1) Assume $M$ is endowed with a first countable topology. Given $\ell$ is continuous on $M \times M$ and

$$
\ell(a,a) = 0 \text{ for all } a \in M, \quad (A.3)
$$

the function $\theta$ is discontinuous at $h_0$ if the lower bound $c(\theta, \ell, \mathcal{H}, \tau, h_0)$ in (2.3) is positive. Given $\ell$ satisfies

$$
\inf_{a \in M \setminus U(b)} \ell(a, b) > 0 \text{ for all } b \in M \text{ and all neighborhoods } U(b), \quad (A.4)
$$

the lower bound $c(\theta, \ell, \mathcal{H}, \tau, h_0)$ in (2.3) is positive if the function $\theta$ is discontinuous at $h_0$. Consequently, given $\ell$ is continuous on $M \times M$ and satisfies (A.3) and (A.4), discontinuity of $\theta$ at $h_0$ is equivalent to positivity of $c(\theta, \ell, \mathcal{H}, \tau, h_0)$.

(11) Assume $M$ is endowed with a first countable topology and that $\ell$ is continuous on $M \times M$ and satisfies

$$
\ell(a, b) > 0 \text{ if } a \neq b. \quad (A.5)
$$

Assume further that there exists a neighborhood $W$ of $h_0$ such that its image $\theta(W)$ is contained in some compact set $K$. Then the lower bound $c(\theta, \ell, \mathcal{H}, \tau, h_0)$ in (2.3) is positive if the function $\theta$ is discontinuous at $h_0$.

(iii) Assume $M$ has the structure of a metric space with metric $d$, and the loss-function satisfies

$$
\ell(a, b) = s(d(a, b)) \text{ for all } a, b \in M, \quad (A.6)
$$

where $s: [0, \infty) \to [0, \infty]$ is continuous and $s(x)$ has a limit $s(\infty) \in [0, \infty]$ for $x \to \infty$.

Then
c(θ, ℓ, ℳ, τ, h₀) ≥ (2κ(ℓ))⁻¹s(osc(θ, h₀)),  \hspace{1cm} (A.7)

where the oscillation of θ at h₀ w.r.t. the metric d is defined as

\[
\text{osc}(θ, h₀) = \inf\{\sup_{f, g \in W} d(θ(f), θ(g)) : h₀ \in ℳ, W \text{ τ-open}\}.
\]

Hence, given s(x)>0 for x>0, the r.h.s. in (A.7), and hence c(θ, ℓ, ℳ, τ, h₀), is positive if θ is discontinuous at h₀; if θ is continuous at h₀ and s(0)=0, the r.h.s. in (A.7) is zero. Under the additional assumptions that equality holds in (A.6) and that the function s is nondecreasing, equality holds in (A.7). In this case, given s(0)=0 and s(x)>0 for x>0, the lower bound c(θ, ℓ, ℳ, τ, h₀) in (2.3) is positive if and only if θ is discontinuous at h₀.

(iv) Assume ℳ=ℝ and ℓ is continuous on ℝ×ℝ (w.r.t. Euclidean topology).

Define

\[
\sigma^*(θ, h₀) = \inf\{\sup_{h \in W} θ(h) : h₀ \notin W, W \text{ τ-open}\},
\]

\[
\sigma_*(θ, h₀) = \sup\{\inf_{h \in W} θ(h) : h₀ \notin W, W \text{ τ-open}\}
\]

and assume σ^*(θ, h₀) and σ_*(θ, h₀) are both finite. Then

\[
c(θ, ℓ, ℳ, τ, h₀) ≥ (2κ(ℓ))⁻¹ℓ(σ^*(θ, h₀), σ_*(θ, h₀)). \hspace{1cm} (A.8)
\]

Hence, given ℓ also satisfies (A.5), discontinuity of θ at h₀ implies positivity of the r.h.s. in (A.8), and hence of c(θ, ℓ, ℳ, τ, h₀). (Given ℓ satisfies (A.3), continuity of θ at h₀ implies that the r.h.s. of (A.8) is zero.)

From Theorem A.1 we can easily get similar results for the global lower minimax risk bound c(θ, ℓ, ℳ, τ). To illustrate, e.g., from Theorem A.1(i) we obtain: Given ℓ is continuous on ℳ×ℳ and satisfies (A.3) and (A.4), the lower bound c(θ, ℓ, ℳ, τ) in (2.4) is positive if and only if the functional θ is discontinuous at some h₀∈ℳ.

Remark A.5: (On Theorem A.1) (a) The σ-field ℳ used in Theorem 2.1 and the topology on ℳ used in Theorem A.1 need not necessarily be related in any way for Theorem A.1 to hold.
(b) Condition (A.4) implies (A.5) if the topology on \( M \) is Hausdorff.

(c) The assumption that the limit \( s(\omega) \) exists maintained in Theorem A.1(iii) can be dropped if \( \text{osc}(\theta, h_0) \) is finite.

(d) A more general formulation of Theorem A.1(iv) covering also the case where \( \sigma^*(\theta, h_0) \) and \( \sigma_*^*(\theta, h_0) \) are not both finite, requires that \( \ell \) is continuous on (or can be continuously extended to)

\[ D=\{(\sigma_*(\theta, h_0), \sigma^*(\theta, h_0)) | (\sigma_*(\theta, h_0), \sigma^*(\theta, h_0)) \setminus (-\infty, -\omega), (\omega, \omega)) \}. \]

(Note that \( -\infty = \sigma_*(\theta, h_0) = \theta(h_0) = \sigma^*(\theta, h_0) = \infty \), and hence \( \sigma_*(\theta, h_0) = \infty \) and \( \sigma^*(\theta, h_0) = -\infty \) hold.) Then (A.8) continues to hold, and the second claim in Theorem A.1(iv) continues to be true if (A.5) holds not only for all \((a, b) \in \mathcal{N}\), but for all \((a, b) \in D\). (With regard to the last claim in Theorem A.1(iv) note that under continuity of \( \theta \) at \( h_0 \), \( \sigma_*(\theta, h_0) = \sigma^*(\theta, h_0) = \theta(h_0) \) and hence both are finite.)

(e) If \( M=\mathbb{R} \), note that the difference \( \sigma^*(\theta, h_0) - \sigma_*(\theta, h_0) \) is always well-defined and equals \( \text{osc}(\theta, h_0) \), when \( \text{osc}(\theta, h_0) \) is taken w.r.t. the usual metric on \( M=\mathbb{R} \).

Hence, if in Theorem A.1(iv) or its generalization just discussed the loss-function \( \ell \) is a function of \(|a-b|\) only, i.e., \( \ell(a, b) = \ell(|a-b|) \), then the r.h.s. in (A.8) takes the form \( (2K(\ell))^{-1} \ell(\text{osc}(\theta, h_0)) \).

(f) Some of the assumptions in Theorem A.1 could be made "local". E.g., in Theorem A.1(i) continuity of \( \ell \) on all of \( \mathcal{M} \times \mathcal{M} \) could be replaced by continuity of \( x \mapsto \ell(x, \theta(h_0)) \) at \( x=\theta(h_0) \). Similarly, in Theorem A.1(ii) continuity of \( x \mapsto \ell(x, \theta(h_0)) \) on \( K \) suffices. Also, (A.3) could be weakened to \( \ell(\theta(h_0), \theta(h_0)) = 0 \) and (A.4) could be weakened to the "local" version

\[ \inf_{a \in \mathcal{M} \setminus U(b)} \ell(a, b) > 0 \]

for \( b=\theta(h_0) \) and all neighborhoods \( U(b) \). In a similar fashion, (A.5) can be weakened to \( \ell(a, \theta(h_0)) > 0 \) if \( a \neq \theta(h_0) \).

**Appendix B. Proofs for Section 2 and Appendix A**

**Proof of Theorem 2.1:** Consider an arbitrary \( h_0 \in \mathcal{M} \). Because \( \tau \) is first countable, we can find a neighborhood basis \( \{W_i : i \in \mathbb{N}\} \) of \( h_0 \) and we may assume w.l.o.g. that the sets \( W_i \) are decreasing. It then follows that
\[
c(\theta, \ell, \mathcal{H}, \tau, h_0) = (2\kappa(\ell))^{-1} \lim_{i \to \infty} \sup_{f, g \in \mathcal{W}_1} \ell(\theta(f), \theta(g)).
\]

Hence, for all \( i \in \mathbb{N} \), we can find elements \( f_i \in \mathcal{W}_1 \) and \( g_i \in \mathcal{W}_1 \) such that
\[
c(\theta, \ell, \mathcal{H}, \tau, h_0) = (2\kappa(\ell))^{-1} \lim_{i \to \infty} \ell(\theta(f_i), \theta(g_i))
\] (B.1)
holds. For every positive real number \( M \) define the truncated loss-function
\[
\ell_M(a,b) = \min(\max(M, \ell(a,b)))
\]
which is also proper as noted in Section 2.

Furthermore, \( g_i \to h_0 \) and \( f_i \to h_0 \) as \( i \to \infty \) by construction. By the assumed continuity of the map \( h \to P_h \) this in turn implies convergence of \( P_{g_i} \) and \( P_{f_i} \) to \( P_{h_0} \) in the total variation distance. Since \( \ell_M \) is bounded by \( M \), we obtain
\[
|E_{h_0, \mathcal{H}} \ell(h, \theta(f_i)) - E_{h_0, \mathcal{H}} \ell(h, \theta(f_1))| \leq M ||P_{f_i} - P_{h_0}||_{TV} \to 0 \quad \text{(B.2a)}
\]
and
\[
|E_{g_i, \mathcal{H}} \ell(g_i, \theta(g_i)) - E_{h_0, \mathcal{H}} \ell(g_i, \theta(g_1))| \leq M ||P_{g_i} - P_{h_0}||_{TV} \to 0 \quad \text{(B.2b)}
\]
as \( i \to \infty \). Applying the generalized triangle inequality to \( \ell_M(\theta(f_i), \theta(g_i)) \) and taking expectations w.r.t. \( E_{h_0} \) gives
\[
(2\kappa(\ell))^{-1} \ell_M(\theta(f_i), \theta(g_i)) \leq (1/2)\left[ E_{h_0, \mathcal{H}} \ell(h, \theta(f_i)) + E_{h_0, \mathcal{H}} \ell(h, \theta(g_i)) \right] \quad \text{(B.3)}
\]
where we have also made use of symmetry of \( \ell \) and of the fact that \( \kappa(\ell) \leq \kappa(\ell) \).

Combining (B.2) with (B.3) shows that for every \( \epsilon > 0 \) there exists an index \( i(\epsilon) \) such that for \( i \geq i(\epsilon) \) we have
\[
(2\kappa(\ell))^{-1} \ell_M(\theta(f_i), \theta(g_i)) - \epsilon < (1/2)\left[ E_{f_i, \mathcal{H}} \ell(h, \theta(f_i)) + E_{g_i, \mathcal{H}} \ell(h, \theta(g_i)) \right]
\]
and hence for \( i \geq i(\epsilon) \)
\[
\min((2\kappa(\ell))^{-1} \mathcal{M}, c(\theta, \ell, \mathcal{H}, \tau, h_0)) - 2\epsilon < (1/2)\left[ E_{f_i, \mathcal{H}} \ell(h, \theta(f_i)) + E_{g_i, \mathcal{H}} \ell(h, \theta(g_i)) \right].
\]

Consequently, for every neighborhood \( \mathcal{U}(h_0) \) of \( h_0 \) we obtain
\[
\min((2\kappa(\ell))^{-1} \mathcal{M}, c(\theta, \ell, \mathcal{H}, \tau, h_0)) \leq \sup_{h \in \mathcal{U}(h_0)} E_{h, \mathcal{H}} \ell(h, \theta(h)) \leq \sup_{h \in \mathcal{U}(h_0)} E_{h, \mathcal{H}} \ell(h, \theta(h))
\]
since \( \ell \leq \ell_M \). Sending \( M \) to infinity proves (2.3). The remaining claims are then simple consequences. \( \Box \)

**Proof of Theorem A.1:** (1) Let \( f_1 \) and \( g_1 \) be as in the proof of Theorem 1. If
\[ c(\theta, \ell, \mathcal{H}, \tau, h_0) \text{ is positive, the generalized triangle inequality and (B.1) imply}
\]
\[
\liminf_{t \to \infty} [\ell(\theta(f_1), \theta(h_0)) + \ell(\theta(g_1), \theta(h_0))] \geq \kappa(\ell)^{-1}\lim_{t \to \infty} \ell(\theta(f_1), \theta(g_1)) > 0.
\]

Consequently, at least one of the sequences \( \theta(f_1) \) and \( \theta(g_1) \) does not converge to \( \theta(h_0) \) in view of (A.3). Since \( f_1 \to h_0 \) and \( g_1 \to h_0 \) by construction, this establishes discontinuity of \( \theta \) at \( h_0 \). To prove the second claim observe that discontinuity of \( \theta \) at \( h_0 \) implies the existence of a sequence \( f_1 \) converging to \( h_0 \), such that \( \theta(f_1) \notin U \) for some neighborhood \( U \) of \( \theta(h_0) \). Condition (A.4) then implies that
\[
\ell(\theta(f_1), \theta(h_0)) \geq \delta > 0,
\]
which in turn implies \( c(\theta, \ell, \mathcal{H}, \tau, h_0) \geq (2\kappa(\ell))^{-1} \delta > 0 \). The third claim is now a trivial consequence.

(ii) Discontinuity of \( \theta \) at \( h_0 \) implies the existence of a sequence \( f_1 \in W \)
converging to \( h_0 \), such that \( \theta(f_1) \notin U \) for some neighborhood \( U \) of \( \theta(h_0) \). Since \( \theta(f_1) \notin \theta(W) \subseteq K \) and \( K \) is compact, the sequence has a cluster point, say \( a^* \), and \( a^* \neq \theta(h_0) \). Because of first countability, we may then choose a subsequence \( \theta(f_{1(j)}) \) that converges to \( a^* \). By continuity of \( \ell \) and (A.5) we obtain
\[
\ell(\theta(f_{1(j)}), \theta(h_0)) \to \ell(a^*, \theta(h_0)) > 0,
\]
which implies that \( \ell(\theta(f_{1(j)}), \theta(h_0)) \geq \delta > 0 \) for large \( j \). This in turn implies
\[
c(\theta, \ell, \mathcal{H}, \tau, h_0) \geq (2\kappa(\ell))^{-1} \delta > 0.
\]

(iii) Similarly as in the argument leading to (B.1) we can find \( h_{11} \in W_1 \) and
\( h_{21} \in W_1 \) such that \( \text{osc}(\theta, h_0) = \lim_{1 \to \infty} d(\theta(h_{11}), \theta(h_{21})) \). Hence,
\[
s(\text{osc}(\theta, h_0)) = s(\lim_{1 \to \infty} d(\theta(h_{11}), \theta(h_{21}))) = \lim_{1 \to \infty} s(d(\theta(h_{11}), \theta(h_{21}))),
\]
where we have made use of continuity of \( s \) and, if \( \text{osc}(\theta, h_0) = \infty \), of the fact that \( s(x) \) has a limit at \( x = \infty \). Because of (A.6), the r.h.s. in the above
chain of inequalities is not less than
\[
\lim_{1 \to \infty} \ell(\theta(h_{11}), \theta(h_{21})) \leq \lim_{1 \to \infty} \sup_{f, g \in W_1} \ell(\theta(f), \theta(g)) = (2\kappa(\ell))c(\theta, \ell, \mathcal{H}, \tau, h_0),
\]
which proves (A.7). Since \( \text{osc}(\theta, h_0) = \infty \) if and only if \( \theta \) is continuous at \( h_0 \),
also the second claim in (A.3) follows. Under the additional assumptions that
equality holds in (A.6) and that the function \( s \) is nondecreasing

\[
(2\kappa(\ell))c(\theta, \ell, \mathcal{H}, \tau, h_0) = \inf_{\mathcal{W}} \sup_{f, g \in \mathcal{W}} d(\theta(f), \theta(g)) = s(\text{osc}(\theta, h_0))
\]
clearly holds, completing the proof of the remaining claims in (iii).

(iv) We can find \( h_{11} \in \mathcal{W}_1 \) and \( h_{21} \in \mathcal{W}_1 \) such that \( \sigma^*(\theta, h_0) = \lim_{1 \to \infty} \theta(h_{11}) \) and \( \sigma_*(\theta, h_0) = \lim_{1 \to \infty} \theta(h_{21}) \). Because of continuity of \( \ell \),

\[
\ell(\sigma^*(\theta, h_0), \sigma_*(\theta, h_0)) = \lim_{1 \to \infty} \ell(\theta(h_{11}), \theta(h_{21})) \leq \lim_{1 \to \infty} \sup_{f, g \in \mathcal{W}_1} \ell(\theta(f), \theta(g)) \leq (2\kappa(\ell))c(\theta, \ell, \mathcal{H}, \tau, h_0),
\]

which establishes (A.8). The remaining claims are now obvious, observing that continuity of \( \theta \) at \( h_0 \) is equivalent to \( \sigma^*(\theta, h_0) = \sigma_*(\theta, h_0) \).

**Proof of Corollary 2.2:** The first two claims follow from Theorem A.1(i), and the third claim from Theorem A.1(iii).

**Proof of Corollary 2.3:** Consider the loss function \( \ell^\alpha(a, b) = \min(\alpha, ||a - b||_p^\alpha) \)
defined in Corollary 2.2 with \( p = \alpha = 1 \) and let \( \epsilon \) be positive. Then for each \( n \in \mathbb{N} \)

\[
E_n, h^\alpha, 1(\hat{\theta}_n - \theta(h)) = E_n, h^\alpha, 1(\theta(h))M(\hat{\theta}_n - \theta(h)||_1) + \epsilon
\]

\[
E_n, h^\alpha, 1(\theta(h))1(||\hat{\theta}_n - \theta(h)||_1 < \epsilon) \leq \min(1, \epsilon)P_n, h^\alpha, 1(\epsilon) + P_n, h^\alpha, 1(||\hat{\theta}_n - \theta(h)||_1 < \epsilon).
\]

Consequently,

\[
\sup_{h \in \mathcal{U}(h_0)} P_n, h^\alpha, 1(||\hat{\theta}_n - \theta(h)||_1 < \epsilon) \geq c(\theta, \ell^\alpha, 1, \mathcal{H}, \tau, h_0) - \min(1, \epsilon)
\]

and the r.h.s. in the above inequality is positive by Corollary 2.2, whenever \( \epsilon \) is small enough (namely whenever \( \epsilon \) is smaller than \( c(\theta, \ell^\alpha, 1, \mathcal{H}, \tau, h_0) \)).

**Appendix C. Proofs for Section 3**

**Lemma C.1:** Given the maintained assumptions of Section 3, let \( C \) be a random
set in \( \mathcal{H} \) and let \( h_0 \in \mathcal{H} \). Then for all \( h_1 \in \mathcal{H}_1(\epsilon) \)

\[
P_{h_0}(\theta(h_1) \in C) = \lim_{1 \to \infty} P_{h_1}(\theta(h_1) \in C)
\]

for some sequence \( h_1 \) satisfying \( h_1 \in \mathcal{H}_1 \) and \( \theta(h_1) = \theta(h_1) \). Furthermore, for every \( \epsilon > 0 \) there exists a neighborhood \( \mathcal{U}(h_0) \) in \( \mathcal{H} \) such that for every \( h_1 \in \mathcal{H}_1(h_0) \)

37
and every \( h \in U(h_0) \)

\[
|P_h(\Theta(h_0) \in C) - P_{h_0}(\Theta(h_0) \in C)| = |P_h(\Theta(h_0) \in C) - \lim_{1 \to \infty} P_{h_{*1}}(\Theta(h_{*1}) \in C)| \leq \varepsilon.
\]

In particular, if \( P_{h_{*1}}(\Theta(h_{*1}) \in C) \geq \varepsilon \) for all \( h_{*1} \in \mathcal{H}_*(h_0) \), then \( P_{h_0}(\Theta(h_0) \in C) \geq \varepsilon \) (\( = \varepsilon \)).

Proof: Let \( h_{*1} \in \mathcal{H}_*(h_0) \). By definition there exists a sequence \( h_{*1} \) satisfying \( h_{*1} \to h_0 \) and \( \Theta(h_{*1}) = \Theta(h_0) \). Then \( P_{h_{*1}}(\Theta(h_{*1}) \in C) = P_{h_{*1}}(\Theta(h_{*1}) \in C) \).

Since \( h_{*1} \to h_0 \), the corresponding probability measures converge in total variation distance and hence \( P_{h_{*1}}(\Theta(h_{*1}) \in C) \) converges to \( P_{h_0}(\Theta(h_0) \in C) \). This proves the first claim, and the second claim follows from the assumed continuity of the map \( h \to P_h \) when \( P \) carries the total variation distance. The last claim is a trivial consequence of the first claim.

Proof of Proposition 3.1: Immediate consequence of Lemma C.1.

Proof of Theorem 3.2: Part (i) follows from (3.1) and Bonferroni's inequality. Inequality (3.4) follows from part (i) with \( p=2 \) observing that for any \( c<\text{diam}(\Theta(h_0, \mathcal{H}_*)) \) one can find \( a_i \in \Theta(h_0, \mathcal{H}_*), i=1,2 \), with \( d(a_1, a_2) \geq c \) and that \( \{ \omega: a_i \in C(\omega), a_2 \in C(\omega) \} \) is contained in \( \{ \omega: \text{diam}(C(\omega)) \geq c \} \). We next prove (3.5). Let \( m \in M \) be arbitrary and define \( \rho(\omega) = \sup(d(a, m): a \in C(\omega)) \). Since \( \Theta(h_0, \mathcal{H}_*) \) has infinite diameter, we can find for every \( c \in \mathbb{R} \) an element \( a(c) \in \Theta(h_0, \mathcal{H}_*) \) with \( d(a(c), m) \geq c \). Hence,

\[
P_{h_0}(\rho(\omega) \geq c) \geq P_{h_0}(a(c) \in C) \geq 1-\alpha
\]

for every \( c \in \mathbb{R} \) in view of (i) with \( p=1 \), and thus \( P_{h_0}(\rho(\omega) = \omega) \geq 1-\alpha \). (If the events involving \( \rho \) are not \( X \)-measurable, the corresponding probabilities are to be interpreted as inner probabilities.) Now \( P_{h_0}(\text{diam}(C) = \omega) \geq 1-\alpha \) follows since \( \{ \omega: \rho(\omega) = \omega \} \subseteq \{ \omega: \text{diam}(C(\omega)) = \omega \} \). The first part of (iii) follows from part (i), since under convexity of \( C \) the event inside the probability on the l.h.s of (3.6) is identical to the event inside the probability on the l.h.s of
(3.3). Inequality (3.7) follows from (3.6) upon observing that the event $\mathcal{N}C$ is the intersection of a countable sequence of decreasing events of the form $\text{conv}(a_1, \ldots, a_p) \cap C$. Part (iv) follows from part (iii) observing that the event $\text{conv}(a_1, \ldots, a_p) \cap C$ is identical to the event $\text{cl}(\text{conv}(a_1, \ldots, a_p)) \cap C$ since $C$ is closed. Part (v) follows immediately from parts (i)-(iv) observing that $P_h(A)$ depends continuously on $h$ in view of the continuity of the map $h \mapsto P_h$ w.r.t. the variation distance on $\mathcal{P}$. Part (vi) is a trivial consequence of (i)-(iv) and absolute continuity. 

Proof of Theorem 3.3: Assumption (3.10) clearly implies that $C(\omega) \in \mathcal{M}$ for all $\omega \in \mathcal{X}$ and that $\mu(C(\omega)) = \int_{H} 1_{C(\omega)}(a) d\mu(da)$ is $\mathcal{X}$-measurable. Fubini's Theorem then implies

$$E_h(\mu(C)) = E_h\int_{H} 1_{C(\omega)}(a) d\mu(da) = \int_{H} E_h(1_{C(\omega)}(a)) d\mu(da) =$$

$$= \int_{H} P_h(\omega \in \mathcal{C}) d\mu(da) \geq \int_{\Theta(h_0, H^*)} P_h(\omega \in \mathcal{C}) d\mu(da) \geq (1-\alpha)\mu(\Theta(h_0, H^*)),$$

where the last inequality follows from Theorem 3.2(1) with $p=1$. This proves (3.11). Inequality (3.12) now follows since the map $h \mapsto E_h(Z)$ is lower semi-continuous for any nonnegative random variable as a consequence of the continuity of the map $h \mapsto P_h$ when $\mathcal{P}$ carries the total variation distance. The remaining claim is trivial. 

Remark C.1: (a) Similarly as in Section 2, the results in Section 3 allow for non-identifiable parameterizations of $\mathcal{P}$, cf. Remark A.2. In order to fully exploit the non-identifiable nature of such problems, a topology on $\mathcal{H}$ that does not separate observationally equivalent parameters (like the weak topology $\tau_w$, cf. Remark A.1) is a natural choice. If such a topology is used and if $H^* = H$, the entire set $G_h$ of parameters $h$ that are observationally equivalent to a given $h_0$ is contained in $H^*(h_0)$ and hence the entire image $\Theta(G_h)$ is contained in $\Theta(h_0, H^*)$. (Even if $H \neq H^*$, $\Theta(G_h) \subset \Theta(h_0, H^*)$ is still often the case.)
(b) We also note the following simple fact which holds without the assumption of continuity of the map \( h \mapsto P_h \): If \( C \) is a confidence set for \( \theta(h), h \in H \), with minimum coverage probability \( 1 - \alpha \), then \( P_h(\theta(h) \in C) \geq 1 - \alpha \) for all \( h \in G_0 \). From this fact results similar to Theorem 3.2 and Theorem 3.3 can be obtained; e.g., \( P_h(\text{diam}(C) \geq \text{diam}(G_0)) \leq 1 - 2\alpha \) as well as \( F_h(\mu(C)) \geq (1 - \alpha)\mu(G_0) \) hold (under the appropriate measurability assumptions).

(c) Requiring \( \theta \) to be defined only on a subset \( H_* \) of \( H \) in Section 3 allows for problems like the construction of confidence sets for ratios of parameters, a famous instance being the Fieller problem. In such a case \( H \) is typically Euclidean space, whereas \( H_* \) has to be chosen as the (dense) subset of \( H \) on which the ratio is well-defined, i.e., where the denominator does not vanish.

---

**Appendix D. Proofs for Section 4**

**Lemma D.1:** Let \( K, \phi, \) and \( \psi \) be real numbers satisfying \( K \geq 1, |\phi| < 1, |\psi| \leq 1 \). Then \( \max_{z \in C, |z| = 1} |1 - \phi z|^{-1} |1 - \psi z| \leq K \) if and only if \( (\phi, \psi) \in A_K \), where

\[
A_K = \{(\phi, \psi) \in \mathbb{R}^2 : |\phi| < 1, |\psi| \leq 1, \psi = (K-1)+K\phi, (1-K)+K(1) = K\phi = \psi\}.
\]

**Proof:** Define for \( -\pi \leq \lambda \leq \pi \) the function \( g(\lambda) = |1 - \phi \exp(i\lambda)|^{-2} |1 - \psi \exp(i\lambda)|^2 = (1 + \psi^2 - 2\psi \cos \lambda)/(1 + \phi^2 - 2\phi \cos \lambda) \). The derivative is given by \( g'(\lambda) = 2(\psi - \phi)(1 - \phi) + 2(\phi - \psi)^2(1 + \phi)^2 \sin \lambda \). By symmetry of \( g(\lambda) \) it suffices to consider the interval \( 0 \leq \lambda \leq \pi \) in order to find the maximum. Since the denominator is always positive (\( |\phi| < 1 \), since \( \sin \lambda \) is nonnegative on \( 0 \leq \lambda \leq \pi \), and since \( (1 - \phi \psi) \) is positive (\( |\phi| < 1, |\psi| \leq 1 \)), the derivative \( g'(\lambda) \) is nonnegative on \( 0 \leq \lambda \leq \pi \) if \( \psi \geq \phi \), and is nonpositive if \( \psi \leq \phi \). Therefore \( \max_{\lambda}(g(\lambda) : 0 \leq \lambda \leq \pi) = \max(g(0), g(\pi)) = \max((1 - \phi)^2(1 - \psi)^2, (1 + \phi)^2(1 + \psi)^2) \). This quantity is bounded by \( K^2 \) if and only if \( (\phi, \psi) \in A_K \) as is easily seen.

---

**Lemma D.2:** The sets \( \mathcal{F}_{\text{AR}}, \mathcal{F}_{\text{MA}}, \mathcal{F}_{\text{ARMA}}, \) and \( \mathcal{F}_\infty \) are sup-norm dense in \( \mathcal{F}_C \).

**Proof:** Since \( \mathcal{F}_{\text{AR}} \subset \mathcal{F}_{\text{MA}} \subset \mathcal{F}_{\text{ARMA}} \subset \mathcal{F}_\infty \), it suffices to prove the lemma for \( \mathcal{F}_{\text{AR}} \) and \( \mathcal{F}_{\text{MA}} \).

We make the following preparatory remark: Let \( \mathcal{F} \) be the algebra of all
trigonometric polynomials \( p(\lambda) = \sum_{k=-N}^{N} c_k \exp(ik\lambda) \) on \([0, \pi]\), where \( \mathbb{N} \cup \{0\} \), the \( c_k \)'s are real and satisfy \( c_k = c_{-k} \). This algebra contains the constant functions and separates points in \([0, \pi]\). By the Stone-Weierstrass Theorem its uniform closure is the set of continuous functions on \([0, \pi]\). Viewing \( \mathcal{F} \) now as a set of functions on \([-\pi, \pi]\), each \( p \in \mathcal{F} \) is obviously an even function and \( \mathcal{F} \) is uniformly dense in the set of even and continuous functions on \([-\pi, \pi]\). Next observe that every \( f \in \mathcal{F}_C \) can be uniformly approximated by positive elements \( f_k \in \mathcal{F}_C \) (e.g., \( f_k(\lambda) = f(\lambda) + 1/k \)). Hence, it suffices to show that each positive \( f \in \mathcal{F}_C \) is the uniform limit of elements in \( \mathcal{F}_M^A \) and \( \mathcal{F}_R^A \), respectively. Now given a positive \( f \in \mathcal{F}_C \), note that \( \inf_{-\pi \leq \lambda \leq \pi} f(\lambda) > 0 \) by continuity of \( f \) and compactness of \([-\pi, \pi]\). As shown in the preparatory remark, \( f \) is a uniform limit of a sequence \( p_k \in \mathcal{F} \). Hence, \( p_k \) is positive for \( k \geq 0 \), and therefore belongs to \( \mathcal{F}_M^A \). This proves the lemma for \( \mathcal{F}_M^A \). Since every positive element in \( \mathcal{F}_M^A \) can be uniformly approximated by elements in \( \mathcal{F}_R^A \) as is well-known, the result for \( \mathcal{F}_R^A \) also follows.

**Proof of Theorem 4.1:** We verify the assumptions of Corollary 2.2 with \( m = 1 \), \( p = 1 \), and \( \alpha = \omega \). For fixed \( n \geq 1 \) let \( X = \mathbb{R}^n \) and let \( \mathcal{P} \) denote the set of all \( n \)-dimensional normal distributions \( \mathcal{P} \). For each \( n \)-dimensional normal distribution \( \mathcal{P} \), let \( \Gamma_n(f) \) denote the \((1,1)\)-of distribution \( \mathcal{P} \), where the \((1,1)\)-of element of \( \Gamma_n(f) \) is given by \( \gamma(1,1), f = \int_{-\pi}^{\pi} \exp(\lambda f(\lambda))d\lambda \). Note that this matrix is positive definite except if \( f = 0 \) almost everywhere, a case that is excluded by assumption. Identify \( \mathcal{H} \) with \( \mathcal{F} \) and let \( \tau \) be the topology induced by the (pseudo)metric \( d_1 \). Since \( \sup_{1 \leq 0} |\gamma(i,f) - \gamma(i,g)| \leq d_1(f,g) \) holds, \( \Gamma_n(f) \) is continuous on \( \mathcal{F} \). Consequently, \( d_1(f,g) \to 0 \) for \( k \to \infty \) implies

\[
\int_{-\pi}^{\pi} \frac{d\mathcal{P}_n,f_k}{d\lambda} \to \int_{-\pi}^{\pi} \frac{d\mathcal{P}_n,f_k}{d\lambda} \to 0
\]

by Scheffé's Lemma, where \( \lambda_n \) denotes \( n \)-dimensional Lebesgue-measure. This shows that \( \mathcal{P}_n,f_k \to \mathcal{P}_n,f \) in total variation distance as \( k \to \infty \), and hence that the parameterization \( f \to \mathcal{P}_n,f \) is continuous. It remains to compute \( \text{osc}(f(0), f'_0) \)
for appropriate \( f_0 \). (Note that \( \text{osc}_p(\ldots, \ldots) \) does not depend on \( p \) since \( m=1 \).)

To prove part (a) define \( f_0(\lambda) = \sigma_0^2/2\pi \) for all \(-\pi \leq \lambda \leq \pi\) where \( \sigma_0^2 > 0 \) is arbitrary. Clearly \( f_0 \in \mathcal{F}_{\text{ARMA}(1,1)} \). Let \( M \) be a positive real number. Consider the family of spectral densities \( f(\lambda; \phi, \psi) = f_0(\lambda)g(\lambda; \phi, \psi) \) where

\[
g(\lambda; \phi, \psi) = |1-\phi \exp(i\lambda)|^{-2}|1-\psi \exp(i\lambda)|^2
\]

and the real numbers \( \phi \) and \( \psi \) satisfy \(|\phi| < 1, |\psi| \leq 1\). Note that any such \( f(\lambda; \phi, \psi) \) belongs to \( \mathcal{F}_{\text{ARMA}(1,1)} \). Choose a sequence \( \phi_k, |\phi_k| < 1 \), converging to unity. Define \( \psi_k = (1-M)+M\phi_k \) and \( \psi_k = 1 \) and observe that \(|\psi_k| \leq 1\) for large enough \( k \) and \( \psi_k \to 1 \). Observe that the elements \( (\phi_k, \psi_k) \) belong to \( A_K \) for \( K=M \)

if \( M=1 \) and to any \( A_K, K>1 \), if \( 0<M<1 \) whenever \( k=k_0(K) \). Similarly, \( (\phi_k, \psi_k) \) belongs to \( A_K \) for any \( K>1 \) whenever \( k=k_0(K) \). By Lemma D.1 the functions

\[
f_{k_1}(\lambda) = f(\lambda; \phi_k, \psi_k) \quad \text{and} \quad f_{k_2}(\lambda) = f(\lambda; \phi_k, \psi_k)
\]

are therefore eventually bounded by \( K^2 \sigma_0^2/2\pi \) on \([-\pi, \pi]\), and they converge to \( f_0(\lambda) \) for all \( \lambda \neq 0 \) as is easily seen. Hence, by dominated convergence \( f_{k_1} \) as well as \( f_{k_2} \) converge to \( f_0 \) in the \( L_1 \)-sense. Furthermore, \( f_{k_1}(0) = M^2 \sigma_0^2/2\pi \) and \( f_{k_2}(0) = 0 \). This shows that

\[
\text{osc}(f(0), f_0) \geq M^2 \sigma_0^2/2\pi \quad \text{for every } M > 0,
\]

which implies \( \text{osc}(f(0), f_0) = \infty \). Part (a) now follows from Corollary 2.2.

We next prove part (b). Let \( f_0 \in \mathcal{F} \) and \( M > 0 \) be arbitrary. Clearly, we can find sequences \( f_{k_1} \in \mathcal{F}_{c} \) and \( f_{k_2} \in \mathcal{F}_{c} \) such that \( f_{k_1} \) and \( f_{k_2} \) converge to \( f_0 \) in the \( L_1 \)-sense and satisfy \( f_{k_1}(0) = M \) and \( f_{k_2}(0) = 0 \). Since \( \mathcal{F} \) is sup-norm dense in \( \mathcal{F}_{c} \), we can find \( f^*_{k_1} \in \mathcal{F} \) and \( f^*_{k_2} \in \mathcal{F} \) such that the sup-norm of \( f^*_{k_1} - f_{k_1} \) converges to zero for \( i = 1, 2 \). Consequently, \( f^*_{k_1} \) and \( f^*_{k_2} \) converges to \( f_0 \) in the \( L_1 \)-sense, \( f^*_{k_1}(0) \to M \), and \( f^*_{k_2}(0) \to 0 \) for \( k \to \infty \). This shows \( \text{osc}(f(0), f_0) \geq M \) for every \( M \) and hence \( \text{osc}(f(0), f_0) = \infty \). Part (b) now follows from Corollary 2.2.

Part (c) is a trivial consequence of (a) and (b). Part (d) follows from parts (a) and (b) and Corollary 2.3.

\[ \]

Proof of Theorem 4.2: Verification of the assumptions of Corollary 2.2 proceeds as in the proof of Theorem 4.1. It remains to compute \( \text{osc}(c_{\infty}(f), f_0) \).

For part (a) we consider \( f_0, f_{k_1} \) and \( f_{k_2} \) constructed in the proof of Theorem
4.1(a). Note that \( \sigma^2(f_{k_1})=\sigma^2(f_{k_2})=\sigma^2_0 \). This shows \( c_\infty(f_{k_1})=M \) and \( c_\infty(f_{k_2})=0 \) which implies \( \text{osc}(c_\infty(f),f'_0)=\infty \). Part (a) now follows from Corollary 2.2.

For part (b) let \( f_0 \in \mathcal{F} \) and \( 0<\alpha<\infty \) be arbitrary real numbers. Clearly, we can find sequences \( f_{\alpha k_1} \in \mathcal{F}_{c_{\alpha k_1}} \), \( i=1,2 \), such that \( f_{\alpha k_1} \) and \( f_{\alpha k_2} \) converge to \( f_{+\alpha} \) in the \( L_1 \)-sense, satisfy \( \inf_{\lambda \in [-\pi,\pi]} f_{\alpha k_1}^{*}(\lambda)=\alpha \) and \( f_{\alpha k_1}(0)=M \), \( f_{\alpha k_2}(0)=\alpha \). Since \( \mathcal{F} \) is sup-norm dense in \( \mathcal{F}_{c_{\alpha}}, \) we can find \( f_{\alpha k_1}^{*} \in \mathcal{F}_{c_{\alpha k_1}} \) and \( f_{\alpha k_2}^{*} \in \mathcal{F}_{c_{\alpha k_2}} \) such that the sup-norm of \( f_{\alpha k_1}^{*}-f_{\alpha k_2}^{*} \) converges to zero for \( i=1,2 \). Consequently, \( f_{\alpha k_1}^{*} \) and \( f_{\alpha k_2}^{*} \) converges to \( f_{+\alpha} \) in the \( L_1 \)-sense, \( f_{\alpha k_1}^{*}(0)\rightarrow M \), and \( f_{\alpha k_2}^{*}(0)\rightarrow \alpha \) for \( k\rightarrow \infty \). Note that \( \inf_{\lambda \in [-\pi,\pi]} f_{\alpha k_1}^{*} \geq \alpha/2 \) for \( i=1,2 \) and large \( k \). Hence, \( \sigma^2(f_{\alpha k_1}) \) and \( \sigma^2(f_{\alpha k_2}) \) converge to \( \sigma^2(f_{+\alpha}) \) by dominated convergence, since \( |\log f_{\alpha k_1}(\lambda)| \leq \max(|\log(\alpha/2)|,f_{\alpha k_1}^{*}(\lambda)) \). It follows that \( c_\infty(f_{+\alpha})\rightarrow (2\pi M/\sigma^2(f_{+\alpha}))^{1/2} \) and \( c_\infty(f_{+\alpha})\rightarrow (2\pi M/\sigma^2(f_{0}+\alpha))^{1/2} \). For \( \alpha \) decreasing to zero \( \sigma^2(f_{0}+\alpha) \) converges to \( \sigma^2(f_{0})>0 \) by monotone convergence (note that \( \log(f_{0}+\alpha)\leq f_{0}+\alpha \) for \( \alpha \leq \alpha_0 \) and that \( f_{0}+\alpha_0 \) is integrable). Hence, by a diagonal sequence argument, we can find sequences \( f_{k_1}^{*} \in \mathcal{F} \), \( i=1,2 \), that converge to \( f_0 \) in the \( L_1 \)-sense and satisfy \( c_\infty(f_{k_1}^{*})\rightarrow (2\pi M/\sigma^2(f_{0}))^{1/2} \) and \( c_\infty(f_{k_1}^{*})\rightarrow 0 \). This shows \( \text{osc}(c_\infty(f),f'_0)=\infty \) since \( M \) was arbitrary. Part (b) now follows from Corollary 2.2.

Part (c) is a trivial consequence of (a) and (b). Part (d) follows from parts (a) and (b) and Corollary 2.3.

---

**Proof of Theorem 4.3:** Similar to the proof of Theorem 4.1 we set \( \mathcal{X}=\mathbb{R}^2 \), \( \mathcal{Y}=\mathbb{B}(\mathbb{R}^2) \), \( \mathcal{P}=\{ P_n, f \in \mathcal{F} \}, \mathcal{H}=\mathcal{H}_{f_0} \), \( M=[0,\infty) \) and \( \Theta(f)=f(0) \). The generally maintained assumptions of Section 3 are then seen to be satisfied, cf. the proof of Theorem 4.1. For part (a), note that \( \Theta(f_0)=[0,\infty) \) has been shown in the proof of Theorem 4.1(a) for the particular spectral density \( f_0 \in \mathcal{F}_{\text{ARMA}(1,1)} \) considered there. Part (a) then follows from Theorem 3.2(ii),(iii),(vi) (with \( p=2 \) and \( N=M \)) and Theorem 3.3. Part (b) follows from Theorem 3.2(ii)-(iv) (with \( p=2 \) and \( N=M \)), since \( \Theta(f_0) \) was shown to be dense in \( M=[0,\infty) \) for every \( f_0 \in \mathcal{F} \) in the proof of Theorem 4.1. As a consequence, (4.9)-(4.11) hold in particular for \( \mathcal{F}=\mathcal{F}_{\text{AR}}, \mathcal{F}_{\text{MA}}, \mathcal{F}_{\text{ARMA}}, \mathcal{F}_0 \), and \( \mathcal{F}_{c} \). The remaining claims in part
(c) follow from Theorem 3.2(iii) (with p=2 and N=M) and Theorem 3.3, since
\[ \Theta(f_0, \mathcal{F})=\{0, \omega\} \text{ for all } f_0 \in \mathcal{F} \] (where \( \mathcal{F} \) is as in part (c)) as shown in Lemma D.3 below. The remainder of part (d) follows from Theorem 3.3, since
\[ \Theta(f_0, \mathcal{F}_\infty) = \{0, \omega\} \text{ for all } f_0 \in \mathcal{F}_\infty \] as shown in Lemma D.3.

**Lemma D.3:** With the notation used in the proof of Theorem 4.3 we have
\[ \Theta(f_0, \mathcal{F}_M) = \{0, \omega\} \text{ for all } f_0 \in \mathcal{F}_M, \mathcal{F}_\text{ARMA}, \mathcal{F}_\infty, \text{ and } \mathcal{F}_c, \text{ and } \Theta(f_0, \mathcal{F}_\infty) = \{0, \omega\} \]
for all \( f_0 \in \mathcal{F}_\text{AR} \).

**Proof:** Recall that denseness of \( \Theta(f_0, \mathcal{F}) \) in \( [0, \omega] \) for every \( f_0 \in \mathcal{F} \) and for all the choices for \( \mathcal{F} \) appearing in the lemma has already been established in the proof of Theorem 4.1. Also, \( \Theta(f_0, \mathcal{F}_c) = \{0, \omega\} \) for all \( f_0 \in \mathcal{F}_c \) has already been shown in that proof. Consider \( \mathcal{F}_\text{ARMA} \) next and recall that this set, and hence any \( L_1 \)-ball \( U(f_0, \delta) \) in this set, is convex. Since \( \Theta(f) = f(0) \) is linear, it follows that \( \Theta(U(f_0, \delta)) \) is also convex. It is also dense in \( [0, \omega] \), since
\[ \Theta(f_0, \mathcal{F}_\text{ARMA}) \text{ has this property and } \Theta(f_0, \mathcal{F}_\text{ARMA}) \subseteq \Theta(U(f_0, \delta)) \]
holds. This implies that \( \Theta(U(f_0, \delta)) \) must contain \( (0, \omega) \) for every \( \delta > 0 \), and hence \( \Theta(f_0, \mathcal{F}_\text{ARMA}) \supseteq (0, \omega) \).

Furthermore, given \( f_0 \in \mathcal{F}_\text{ARMA} \), the sequence \( f_k(\lambda) = g(\lambda, 1-1/k, 1)f_0(\lambda) \) converges to \( f_0(\lambda) \) for every \( \lambda \neq 0 \), where \( g \) is the function defined in the proof of Theorem 4.1. Since the sequence \( f_k \) is uniformly bounded in view of Lemma D.1, it converges to \( f_0 \) also in the \( L_1 \)-sense. Since \( f_k \in \mathcal{F}_\text{ARMA} \) and \( f_k(0) = 0 \), also \( 0 \in \Theta(f_0, \mathcal{F}_\text{ARMA}) \) follows, thus establishing \( \Theta(f_0, \mathcal{F}_\text{ARMA}) = (0, \omega) \). Exactly the same proof shows that \( \Theta(f_0, \mathcal{F}_\infty) = (0, \omega) \). Since \( \mathcal{F}_M \) is well-known to be convex, the proof for \( \mathcal{F}_M \) proceeds identically until one arrives at \( \Theta(f_0, \mathcal{F}_M) \supseteq (0, \omega) \). It then remains to show that \( 0 \) also belongs \( \Theta(f_0, \mathcal{F}_M) \). Observe that for every \( k \geq 1 \) the function \( g(\lambda, 1-1/k, 1) \) can be approximated arbitrarily well in the sup-norm by moving average spectral densities of the form \( |1-\exp(i\lambda)|^2 p_k(\lambda) \), where \( p_k \) is also a moving average spectral density, cf. proof of Lemma D.2.

Defining now \( f_k(\lambda) = |1-\exp(i\lambda)|^2 p_k(\lambda)f_0(\lambda) \) one sees that \( f_k \) converges to \( f_0 \) in the \( L_1 \)-sense and that it belongs to \( \mathcal{F}_M \). Since \( f_k(0) = 0 \), the proof for \( \mathcal{F}_M \) is complete. Finally, consider \( \mathcal{F}_\text{AR} \) and let \( f_0 \) be an arbitrary element of \( \mathcal{F}_\text{AR} \).
Any $L_1$-ball $U(f_0, \delta)$, $\delta > 0$, in $\mathcal{F}_{AR}$ then satisfies $\Theta(U(f_0, \delta)) \Theta(f_0, \mathcal{F}_{AR})$, and hence is dense in $(0, \omega)$. Consider the family of functions $b(\lambda, \phi) = |1 - \phi \exp(i\lambda)|^{-2}$ for $0 \leq \phi < 1/2$, and observe that this family is uniformly bounded by 4. It is easy to see that the functions $b(\lambda, \phi)f(\lambda)$ belong to $U(f_0, \delta)$ for all $f \in U(f_0, \delta/8)$ and all $0 \leq \phi < \eta$ for appropriate $0 < \eta < 1/2$, where $\eta$ depends only on $f_0$ and $\delta$. The set of values of $b(0, \phi) = |1 - \phi|^{-2}$ for $0 \leq \phi < \eta$ is the interval $[1, |1 - \eta|^{-2}]$. Hence, with every $a \in \Theta(U(f_0, \delta))$ also the interval $[a, a|1 - \eta|^{-2}]$ is seen to be a subset of $\Theta(U(f_0, \delta))$. This shows that $\Theta(U(f_0, \delta))$ must coincide with $(0, \omega)$ for every $\delta > 0$, and hence $\Theta(f_0, \mathcal{F}_{AR}) = (0, \omega)$.

Appendix E. Proofs for Section 5

Proof of Theorem 5.1: Verification of the assumptions of Corollary 2.2

proceeds as in the proof of Theorem 4.1. To prove part (a) it remains to show that $\text{osc}(d(f), f_0) = 1$ for every $f_0 \in \mathcal{F}$. We first consider the case where

$\mathcal{F} = \mathcal{F}_{ARIMA}$. Define $\psi_0(\lambda; \xi) = |\lambda|^{-2\xi}$ for $|\lambda| \leq \delta$ and $\psi_0(\lambda; \xi) = |\delta|^{-2\xi}$ for $|\lambda| < \delta$, $\lambda \in [-\pi, \pi]$, where $\xi \in \mathbb{R}$ and $0 < \delta < \pi$. The function $\psi_0(\lambda; \xi)$ is continuous, even and positive on $[-\pi, \pi]$ and hence belongs to $\mathcal{F}_c$. Let $f_0(\lambda) = |\lambda|^{-2d(0)} g_0(\lambda)$ be an arbitrary element of $\mathcal{F}$. Set $f_1(\lambda) = |\lambda|^{-2(1)} \psi_0(\lambda; d(0) - d(1)) g_0(\lambda)$, where $-1/2 \leq d(1) < 1/2$. Note that $f_i$, $i = 0, 1$, are integrable over $[-\pi, \pi]$, since $d(1) < 1/2$. Hence, for every $c > 0$ we can find $\delta > 0$ (depending on $c$, $d(1)$, and $f_0$) such that $\int_{-\delta}^{\delta} f_1(\lambda)d\lambda < c/2$ for $i = 0, 1$. From the definition of $\psi_0(\lambda; d(0) - d(1))$ it follows that

$$\int_{-\pi}^{\pi} |f_0(\lambda) - f_1(\lambda)|d\lambda \leq \int_{-\delta}^{\delta} f_0(\lambda)d\lambda + \int_{-\delta}^{\delta} f_1(\lambda)d\lambda < c.$$

By Lemma D.2 we can find an element $\psi^* \in \mathcal{F}_{ARIMA}$ that is arbitrarily close to $\psi_0(\lambda; d(0) - d(1))$ in the sup-norm. Since $|\lambda|^{-2d(1)} g_0$ is integrable,

$f^*_1 = |\lambda|^{-2d(1)} \psi^* g_0$ is then arbitrarily close to $f_1$ in the $L_1$-sense, and hence $f^*_1 \in U(f_0, \epsilon)$ holds. Since $\mathcal{F}_{ARIMA}$ is closed under multiplication, $f^*_1$ belongs to $\mathcal{F}_{ARIMA}$. Clearly, $d(f^*_1) = d(1)$. Since $d(1) \epsilon [-1/2, 1/2]$ was arbitrary, this shows that in any $L_1$-neighborhood of $f_0$ in $\mathcal{F}$ we can find elements $f$ with arbitrary $d(f) \epsilon [-1/2, 1/2]$. As a consequence $\text{osc}(d(f), f_0) = 1$. The proofs for
\( \mathcal{F} = \mathcal{F}_{\text{ARFI}}, \mathcal{F}_{\text{FIMA}}, \mathcal{F}_{\omega, \mathcal{F}_1}, \text{ and } \mathcal{F}_{c, \mathcal{F}_1} \) are completely analogous. Part (b) follows from part (a) and Corollary 2.3.

**Proof of Theorem 5.2:** The results (5.4)-(5.6) follow immediately from Theorem 3.2 and Theorem 3.3 with \( \mathcal{H} = \mathcal{H}_* = \mathcal{F} \) and \( M = [-1/2, 1/2] \), upon observing that \( \Theta(f, \mathcal{F}) = M = [-1/2, 1/2] \) was established in the proof of Theorem 5.1.
REFERENCES


Gleser, L.J., and J.T. Hwang (1987): The Nonexistence of $100(1-\alpha)\%$


1 We note, however, that certain parametric estimation problems like the classical simultaneous equations model or the autoregressive moving average model, where identifiability issues arise, exhibit some features of ill-posed estimation problems, see Pötscher (1999).

2 Inspection of the proof shows that (3.4) and (3.5) in fact hold without the diameter-measurability assumption, provided the probabilities in these statements are interpreted as inner probabilities.

3 While (4.1) for $F=C$ could also be deduced from the development in Samarov (1977) and Farrell (1979), cf. the discussion surrounding (1.4) in the latter reference, this is not the case for the smaller classes $F_{AR}$, $F_{MA}$, $F_{ARMA}$, $F_{ARMA(r,n)}$.

4 We note that the behaviour of the maximum risk of $\tilde{\theta}_n$ over $F_n$ will depend on the behaviour of the sequence $s_n$. If $s_n$ increases slowly enough, this maximum risk will converge to zero, while it will diverge to infinity if $s_n$ increases too fast.

5 Hence, $E_n, f_n(length(C_n))$ is arbitrarily large in an $L_1-$ball in $F$ around $f_0$, cf. (3.12). Also, the l.h.s. of (4.6) exceeds $1-\alpha-\epsilon$, and the l.h.s. of (4.8) exceeds $1-2\alpha-\epsilon$ in an $L_1-$ball in $F$ around $f_0$, cf. Theorem 3.2(v).

6 Inspection of the proof of Proposition 2 in Faust (1999) reveals that a measurability condition on the confidence interval similar to diameter measurability is missing. The proof is further deficient as convergence of $P_n(A)$ to $P(A)$ for every Borel set $A$ is incorrectly deduced from weak convergence of $P_n$ to $P$ and absolute continuity of $P$. The same incorrect argument is also used in Faust (1996).

7 That power cannot exceed size in testing DS against TS is of course already true, e.g., for $F=F_{MA(1)}$. However, if $F=F_{MA(1)}$ the same is no longer true if the roles of DS and TS are interchanged.