C O L L O Q U I U M   M A T H E M A T I C U M

VOL. 100 2004 NO. 1

CONSISTENCY OF THE LSE IN LINEAR REGRESSION WITH STATIONARY NOISE

BY

GUY COHEN (Beer-Sheva), MICHAEL LIN (Beer-Sheva) and
ARKADY TEMPELMAN (University Park, PA)

Abstract. We obtain conditions for L_2 and strong consistency of the least square estimators of the coefficients in a multi-linear regression model with a stationary random noise. For given non-random regressors, we obtain conditions which ensure L_2-consistency for all wide sense stationary noise sequences with spectral measure in a given class. The condition for the class of all noises with continuous (i.e., atomless) spectral measures yields also L_p-consistency when the noise is strict sense stationary with continuous spectrum and finite absolute pth moment, $p \geq 1$ (even without finite variance).

When the spectral measure of the noise is not continuous, we assume that the non-random regressors are Hartman almost periodic, and obtain a spectral condition for L_2-consistency. An additional assumption on the regressors yields strong consistency for strictly stationary noise sequences.

We also treat the case when the regressors are random sequences, with trends having some good averaging properties and with additive stationary ergodic random fluctuations independent of the noise. When the noise and the fluctuations have disjoint point spectra and the noise is strict sense stationary, we obtain strong consistency of the LSE.

The results are applied to amplitude estimation in sums of harmonic signals with known frequencies.

1. Introduction. We consider the multiple linear regression model

$$Y_n = \sum_{j=1}^{r} \beta^{(j)} (a_n^{(j)} + X_n^{(j)}) + Z_n, \qquad n = 1, 2, \ldots, \tag{1.1}$$

where $\beta^{(1)}, \ldots, \beta^{(r)}$ are the unknown regression coefficients, the random variable Y_n is the "observed signal" (or "response") corresponding to the random observable regressors $a_n^{(1)} + X_n^{(1)}, \ldots, a_n^{(r)} + X_n^{(j)}$, with random unobservable "noise" (or error) Z_n; the noise is assumed to be independent of the regressors. We study the L_2-consistency and strong consistency of the least square coefficient estimators $\widehat{\beta}_m^{(1)}, \ldots, \widehat{\beta}_m^{(r)}$ obtained from the first m observations of the signal and the regressors.

2000 *Mathematics Subject Classification*: Primary 37A30, 37A50, 62J05; Secondary 28D05, 47A35.

This problem has received wide attention for a long time. Many authors (Grenander and Rosenblatt (1957), Rozanov (1969), Kholevo (1969, 1971), Tempelman (1970, 1973), Drygas (1971, 1976), Anderson and Taylor (1976), Lai and Robbins (1977), Lai *et al.* (1978, 1979), Hannan (1978), Chen *et al.* (1981), Solo (1981), Novikov (1985), Gaposhkin (1979, 1985), Le Breton and Musiela (1987), Yajima (1988), Savichev and Tempelman (1990), Chen and Wu (1993)) have investigated the consistency in the case of non-random regressors (i.e., $X_n^{(j)} \equiv 0$ for every $1 \le j \le r$ and $n \ge 1$), imposing various conditions on the noise sequence $\mathbf{Z} := \{Z_n\}$, trying to assume minimal conditions on the regressor sequences $\mathbf{a}^{(j)} := \{a_n^{(j)}\}_{n\ge1}$, $1 \le j \le r$. Of course, for the least square estimator (LSE) $\widehat{\beta}_m := (\widehat{\beta}_m^{(1)}, \dots, \widehat{\beta}_m^{(r)})$ to be well defined (for non-random regressors), the $r \times r$ symmetric matrix $A_m := [\sum_{n=1}^m a_n^{(i)} a_n^{(j)}]_{1\le i,j\le r}$ must be invertible (e.g., Rao and Toutenbourg (1994), Stapleton (1995)). If A_{m_0} is invertible, so is A_m for $m \ge m_0$. The following statement is a special case of Theorem 5.7 of Tempelman (1973), announced in Tempelman (1970).

THEOREM A. *Assume $X_n^{(j)} \equiv 0$ for $1 \le j \le r$ and $n \ge 1$, and A_{m_0} is invertible. Then*

$$\sum_{n=1}^{\infty} \Big| \sum_{j=1}^{r} \alpha_j a_n^{(j)} \Big|^2 = \infty \qquad \forall\ r\text{-tuples } \vec{\alpha} = (\alpha_1, \dots, \alpha_r) \ne \vec{0} \tag{1.2}$$

is a necessary and sufficient condition for the L_2-consistency of the LSE for every uncorrelated noise sequence $\{Z_n\}$ with zero means and common variance $\sigma^2 > 0$.

Apparently unaware of Tempelman's result, Lai *et al.* (1979) obtained Theorem A with (1.2) replaced by

$$\lim_{m\to\infty} A_m^{-1} = \mathbf{0}, \tag{1.3}$$

deducing the necessity from Theorem 4.1(iii) of Drygas (1976). By symmetry of A_m, an equivalent reformulation of (1.3) is

$$\lambda_{\min}(A_m) \to \infty \tag{1.3a}$$

(where $\lambda_{\min}(A_m)$ is the minimal eigenvalue of A_m). The equivalence of (1.2) and (1.3a) is not hard—see Proposition 2.3.

Clearly, (1.2) leads to the necessary (but not sufficient) condition

$$\sum_{n=1}^{\infty} |a_n^{(j)}|^2 = \infty \qquad \forall 1 \le j \le r. \tag{1.4}$$

One of the most general conditions imposed on the noise is that the series $\sum_{n=1}^{\infty} c_n Z_n$ converges almost surely for every sequence $\{c_n\} \in l_2$ (i.e., such that $\sum_{n=1}^{\infty} c_n^2 < \infty$). Lai *et al.* (1979), who introduced this condition, proved that together with (1.3), it yields strong consistency of the LSE for non-random regressors. It follows that the LSE is strongly consistent when (1.3) is satisfied and $\{Z_n\}$ is an L_2-bounded martingale difference sequence; see also Lai (1991) and Buldygin and Koval' (2000) (where the vector-valued case is considered). Conditions for strong consistency for $\{Z_n\}$ orthogonal were given by Chen *et al.* (1981), and Gaposhkin (1985). Jin and Chen (1996, 1999) have recently found conditions for strong consistency of the LSE when the Z_n's are i.i.d. and possess only a moment of order less than 2. Stringent conditions on systems satisfying (1.3), which imply strong consistency for every $\{Z_n\}$ with $\sup_n \mathbf{E}(|Z_n|) < \infty$, were given by Slabospitskii (1994).

It is known (see Petersen (1983), p. 94) that for every ergodic probability preserving transformation τ on a non-atomic space there exists f bounded measurable with 0 expectation, such that $\sum_{n=1}^{\infty} f \circ \tau^n / n$ is a.e. divergent. Thus, $Z_n = f \circ \tau^n$ is a *bounded* strict sense stationary centered ergodic sequence, which does not satisfy the above condition of Lai *et al.* (1979) for strong consistency; if τ is weakly mixing (Krengel (1985), §2.3), then $\{Z_n\}$ has continuous (atomless) spectral measure. Thus, the condition of Lai *et al.* does not apply to bounded stationary noises, even with good spectral and ergodic properties, and different conditions are needed. This is done in Sections 3 and 4.

In this paper, we study the consistency of the LSE when the noise sequence $\{Z_n\}$ is a wide sense stationary sequence, or a strict sense stationary sequence with certain finite moments. Note that if the spectral measure of the wide sense stationary noise $\mathbf{Z}$ is *absolutely continuous with bounded spectral density* f, then condition (1.3) is sufficient for L_2-consistency of the LSE's (and when $f > 0$ a.e., this condition is also necessary); see §7.0 of Grenander and Rosenblatt (1957) for $r = 1$, and Kholevo (1969), Rozanov (1969), and Tempelman (1973) for $r \geq 1$. Theorem 5.2 in Tempelman (1973) shows that in general condition (1.3) is *not sufficient* for consistency when the spectral measure of the noise is absolutely continuous with *unbounded* density; Theorem 5.6 therein asserts that in that case, it is sufficient for L_2-consistency to reinforce (1.3) by adding the condition

$$\sup_{m \geq m_0} \left\{ |c_m^{(kj)}| \sum_{n=1}^{m} |a_n^{(j)}| \right\} < \infty, \quad 1 \leq k, j \leq r, \tag{1.5}$$

where A_{m_0} is invertible, and $C_m := [c_m^{(kj)}] = A_m^{-1}$ for $m \geq m_0$.

In this paper we do not necessarily assume any absolute continuity of the spectral measure of the stationary noise. In our model we assume that the regressors are known, and look for conditions on the *given* regressors which yield consistency of the LSE for *all* stationary noise sequences with continuous (i.e., atomless), possibly singular, spectral measure; the point of view of Grenander and Rosenblatt (1957) and Kholevo (1969) is different—they want to find *all* regressors which yield L_2-consistency for a *given* stationary noise (with bounded spectral density). For strong consistency, we consider here systems of regressors for which $\lim_{m\to\infty} A_m/m$ exists. For such systems, condition (1.5) is implied by the "asymptotic non-degeneracy" condition: the matrix $\lim_{m\to\infty} A_m/m$ is not singular; this condition is often used in econometrics, e.g., Green (2003), §5.2.1. In §2 we will see that, when the regressors are non-random, the condition $\liminf_m \lambda_{\min}(A_m)/m > 0$ (which is implied by asymptotic non-degeneracy) is sufficient for L_2-consistency for all wide sense stationary noises with continuous (atomless) spectral measure.

For random regressors, we will make the plausible assumptions that the random r-dimensional sequence $\{(X_n^{(1)}, \dots, X_n^{(r)})\}_{n\geq 1}$ is ergodic and strict sense stationary with zero expectation and finite variance, that it is stochastically independent of the noise $\{Z_n\}$, and that the covariance matrix of the random vector $(X_1^{(1)}, \dots, X_1^{(r)})$ is invertible. Under different assumptions, the problem of consistency for random regressors has been treated, among others, by Christopeit and Helms (1980), Lai and Wei (1982), and Wei (1985).

To illustrate the connection between conditions on the regressors and different classes of stationary noise sequences for which we want consistency of the LSE, consider the simplest case: the linear regression with $r = 1$ and non-random regressors, i.e., $Y_n = \beta a_n + Z_n$, $n = 1, 2, \dots$. The LSE for the regression coefficient based on the first m coordinates is

$$\widehat{\beta}_m = \frac{\sum_{n=1}^m a_n Y_n}{\sum_{n=1}^m |a_n|^2} = \beta + \frac{\sum_{n=1}^m a_n Z_n}{\sum_{n=1}^m |a_n|^2}. \tag{1.6}$$

When $\{Z_n\}$ are uncorrelated with zero means and common finite variance, L_2-consistency, i.e., norm convergence to zero of the last ratio, is equivalent to $\sum_{n=1}^\infty |a_n|^2 = \infty$ (this is a direct proof of Theorem A in this simple case); under this assumption, Drygas (1976) has shown that if the $\{Z_n\}$ are i.i.d., we also have a.s. convergence in (1.6). We show in §2 that (1.5) for $r = 1$, when added to $\sum_{n=1}^\infty |a_n|^2 = \infty$, yields L_2-consistency for every mixing noise sequence, but in general these conditions do not imply the L_2-consistency for every noise with continuous spectral measure. We are therefore led to a stronger assumption on the regressors, namely $\liminf_m m^{-1} \sum_{n=1}^m |a_n|^2 > 0$, which is shown to imply L_2-consistency for every stationary noise with continuous (atomless), even singular, spectral measure.

We remark that *if* $m^{-1} \sum_{n=1}^m |a_n|^2 \to \infty$*, then the LSE sequence is* L_2*-consistent for every stationary* $\mathbf{Z}$. Indeed, let $\|Z_n\| = \sigma$. Then

$$\text{(1.7)} \qquad \frac{\| \sum_{n=1}^{m} a_n Z_n \|}{\sum_{n=1}^{m} |a_n|^2} \le \frac{\sigma \sum_{n=1}^{m} |a_n|}{\sum_{n=1}^{m} |a_n|^2} \le \frac{\sigma \sqrt{\sum_{n=1}^{m} |a_n|^2} \sqrt{m}}{\sum_{n=1}^{m} |a_n|^2} = \frac{\sigma}{\sqrt{m^{-1} \sum_{n=1}^{m} |a_n|^2}} \to 0.$$

The asymptotic non-degeneracy assumption, for $r = 1$, means that $m^{-1} \sum_{n=1}^{m} |a_n|^2$ converges to a non-zero finite limit. Note that this assumption implies (1.5), by the second inequality in (1.7). With the non-degeneracy assumption, L_2 (strong) consistency of the LSE with a centered wide (strict) sense stationary noise sequence $\{Z_n\}$ is equivalent to the L_2 (almost sure) convergence to 0 of $m^{-1} \sum_{n=1}^{m} a_n Z_n$. Limit theorems of this type are called *modulated ergodic theorems.*

The paper is organized as follows:

In §2 we study the L_2-consistency for non-random regressors. We present a simple example which shows that in general (1.3) is not sufficient for L_2-consistency (even if the noise has absolutely continuous spectral measure). Under various (additional) assumptions on the regressor system, we obtain the L_2-consistency for various classes of wide sense stationary noises, defined in terms of properties of the spectral measure of the noise sequence. The application of our results when $r = 1$ is summarized in a table at the end of the section.

In §3 we prove, for non-random regressors with the non-degeneracy assumption, L_p-consistency for all strictly stationary noises with finite pth moment and continuous spectrum, even for $p \in [1, 2)$ without finite variance. A necessary and sufficient condition for L_p-consistency when the noise has non-empty point spectrum is given. An application to amplitude estimation in sums of harmonic signals with known frequencies and unknown constant phases is presented.

In §4 we use the previous material in order to obtain sufficient conditions for L_p and strong consistency of the LS regression estimators in the presence of stationary "noise" and additive ergodic stationary "random perturbations" of the regressors. These conditions are satisfied if all sequences have continuous spectral measures. An application to amplitude estimation in sums of harmonic signals with known frequencies and stationary random phases is presented.

In Appendix A we show that if in the model (1.1) the random regressors are ergodic strictly stationary with finite second moments, then for almost every realization the regressors have the property that $m^{-1} A_m$ converges; the limiting matrix (which does not depend on the realization) is invertible if and only if $X_0^{(1)}, \dots, X_0^{(r)}$ are linearly independent in L_2.

In Appendix B we state the modulated ergodic theorems used in the paper.

2. L_2-consistency of the least square estimators of the regression coefficients

2.1. *Description of the model.* In this section we consider the multiple linear regression model with non-random (complex-valued) regressors

$$Y_n = \sum_{j=1}^{r} \beta^{(j)} a_n^{(j)} + Z_n, \qquad n = 1, 2, \dots. \tag{2.1}$$

We assume that the r numerical (complex) regressor sequences $\mathbf{a}^{(j)}$ are known, the "regression coefficients" $\beta^{(j)}$, $1 \le j \le r$, are unknown constants, and the "pure signal" sequence $\{v_n\} = \mathbf{v} := \sum_{j=1}^{r} \beta^{(j)} \mathbf{a}^{(j)}$ is observed with some additive (complex-valued) unobservable random "noise" $\mathbf{Z} = \{Z_n\}$, so the random "response" (or "output") sequence $\mathbf{Y} = \mathbf{Z} + \mathbf{v}$ is observed. We assume that $\{Z_n\}$ is wide sense stationary defined on the probability space $(\mathcal{Z}, \nu)$, so the response Y_n is a random variable in $L_2(\mathcal{Z}, \nu)$. Our goal is to estimate the regression coefficients.

The *least square estimator* (LSE) for the regression coefficients is obtained from the random vector $(\widetilde{v}_1^{(m)}, \dots, \widetilde{v}_m^{(m)})$, the orthogonal projection (in $\mathbb{C}^m$) of $(Y_1, \dots, Y_m)$ on the subspace generated by the set of r constant vectors $\mathcal{S}_m := \{\mathbf{a}^{(j,m)} = (a_1^{(j)}, \dots, a_m^{(j)}) : 1 \le j \le r\}$. When the vectors in $\mathcal{S}_m$ are linearly independent (possible only if $m \ge r$), there is a *unique* representation $(\widetilde{v}_1^{(m)} \dots, \widetilde{v}_m^{(m)}) = \sum_{j=1}^{r} \widehat{\beta}_m^{(j)} \mathbf{a}^{(j,m)}$, and we take $\widehat{\beta}_m^{(j)}$ as the mth estimator for $\beta^{(j)}$. Note that if $\mathcal{S}_{m_0}$ is linearly independent, so is $\mathcal{S}_m$ for every $m \ge m_0$.

We now describe the actual computation of the LSE, which is valid also for the complex case (standard texts in statistics usually give proofs valid only for the real-valued case). In order to deal with the complex case, the matrix A_m of the introduction should be defined by $A_m = [\sum_{n=1}^{m} a_n^{(i)} \overline{a_n^{(j)}}]$ (which is now symmetric over $\mathbb{C}$). When we assume that the Gram matrix A_m is non-singular (for $m \ge m_0$), which is equivalent to the linear independence of $\mathcal{S}_m$ (e.g., Davis (1975), p. 178), we use its inverse $C_m = [c_m^{(ij)}] = A_m^{-1}$ to obtain (e.g., Davis (1975), p. 176)

$$\widehat{\beta}_m^{(k)} = \sum_{j=1}^{r} c_m^{(jk)} \langle \mathbf{Y}, \mathbf{a}^{(j)} \rangle = \sum_{j=1}^{r} c_m^{(jk)} \sum_{n=1}^{m} Y_n \overline{a_n^{(j)}}. \tag{2.2}$$

Thus, computation of the LSE requires invertibility of A_m, and we obtain

$$\widehat{\beta}_m^{(k)} - \beta^{(k)} = \sum_{j=1}^{r} c_m^{(jk)} \sum_{n=1}^{m} Z_n \overline{a_n^{(j)}} = \sum_{n=1}^{m} \Big(\sum_{j=1}^{r} c_m^{(jk)} \overline{a_n^{(j)}} \Big) Z_n. \tag{2.3}$$

Since in our model the regressors are *known*, the problem is to find for which wide sense stationary noise sequences $\{Z_n\}$ we have L_2-consistency of the LSE (i.e., the $L_2(\mathcal{Z}, \nu)$-convergence of $\widehat{\beta}_m^{(k)}$ to $\beta^{(k)}$, for $1 \le k \le r$).

We mention that Hannan (1978) and Gaposhkin (1979) obtained conditions for strong consistency (i.e., almost sure convergence) when the noise is centered wide sense stationary with bounded spectral density. Condition (1.3) is always assumed.

2.2. *L_2-consistency of the LSE.* We now obtain sufficient conditions on the *given* sequence $\{A_m\}$, which ensure L_2-consistency of the LSE for the class of *all* wide sense stationary noise sequences with a certain property of the spectral measure. We deal with the following classes (listed in increasing order): absolutely continuous spectral measure with density in L_p ($p > 1$), absolutely continuous spectral measure (density in L_1), mixing spectral measure (i.e., its Fourier coefficients tend to 0), continuous (atomless), even *singular*, spectral measure. Of course, consistency for a larger class of noises requires stronger assumptions on $\{A_m\}$. Except for the class of all stationary noises with bounded density (Theorem 2.1), which is discussed for the sake of completeness, our results are new. We give some examples which show the limitations of the sufficient conditions for consistency when the class of noises is enlarged. The results when $r = 1$ are tabulated at the end of this section.

Recall (e.g., Doob (1953), §X.1; Krengel (1985), p. 32) that to any complex-valued wide sense stationary process $\{Z_n\}_{n\geq 0}$ we can associate an isometry U (the shift isometry) defined on the closed subspace generated by $\{Z_n\}$, such that $Z_n = U^n Z_0$. Recall also (e.g., Doob (1953), §X.3; Krengel (1985), p. 95) that there exists a unique finite measure μ on the Borel sets of $[-\pi, \pi)$, called *the spectral measure of* $\{Z_n\}$, such that

$$\widehat{\mu}(n) := \int_{-\pi}^{\pi} e^{int}\, d\mu(t) = \mathbf{E}(Z_n \overline{Z}_0), \qquad n = 0, 1, \ldots.$$

If $\{Z_n\}$ is centered, $\{\widehat{\mu}(n)\}$ is its covariance sequence. An easy computation shows that we have

$$\Big\| \sum_{j=0}^{N} b_j Z_j \Big\|^2 = \int_{-\pi}^{\pi} \Big| \sum_{j=0}^{N} b_j e^{ijt} \Big|^2 d\mu(t). \tag{2.4}$$

For $t_0 \in [-\pi, \pi)$ this yields, by Lebesgue's bounded convergence theorem,

$$\Big\| \frac{1}{N} \sum_{j=0}^{N-1} e^{-ijt_0} Z_j \Big\|^2 = \int_{-\pi}^{\pi} \Big| \frac{1}{N} \sum_{j=0}^{N-1} e^{-ijt_0} e^{ijt} \Big|^2 d\mu(t) \xrightarrow[N\to\infty]{} \mu(\{t_0\}).$$

THEOREM 2.1. *The LSE sequences given by* (2.2) *are L_2-consistent for every wide sense stationary $\{Z_n\}$ with absolutely continuous spectral measure with bounded density if and only if* (1.3) *holds.*

Proof. The sufficiency of (1.3) was proved by Kholevo (1969). To prove the necessity, take $\{Z_n\}$ centered i.i.d.; the spectral measure is a multiple of Lebesgue's measure, so by the consistency we have convergence to 0 in (2.3), which by the orthogonality and Lemma 2.2 below yields (1.3) (we do not need here the second part of Kholevo's theorem, since we are free to use *any* noise sequence from the class of noises with bounded density). ■

REMARK. If we have L_2-consistency for all stationary noises with absolutely continuous spectral measure having as density a trigonometric polynomial, then by taking $\{Z_n\}$ centered i.i.d. as before, we see that (1.3) must hold. Thus, (1.3) is in a sense a *minimal* assumption on the regressors.

As mentioned in the introduction, (1.3) is *not* sufficient for L_2-consistency for all stationary noises with absolutely continuous spectral measure. In fact, examples are easy to construct (though this construction does not yield the full result of Tempelman (1973), Theorem 5.2).

EXAMPLE 1. *Regressors satisfying* (1.3) *and stationary noises with absolutely continuous spectral measure*, *with no L_2-consistency of the LSE.*

We first note that given $0 \le f \in L_1([-\pi,\pi), dt)$ there is a strict sense stationary centered sequence $\{Z_n\}$ with second moment such that its spectral measure is absolutely continuous with density f; the existence of a wide sense sequence follows from applying Theorem II.3.1 of Doob (1953) to the function $r(k,n) := \int_{-\pi}^{\pi} e^{i(k-n)t} f(t)\,dt$ (with $\mu(n) \equiv 0$ in Doob's notation), and the existence of a strict sense (Gaussian) sequence follows from the discussion on p. 95 of Doob (1953).

Now let f be a positive integrable function on $[-\pi,\pi)$ which is not in $L_2([-\pi,\pi), dt)$, and let $\{Z_n\}$ be stationary having absolutely continuous spectral measure μ with density f. Then $\sum_{n=0}^{\infty} |\widehat{\mu}(n)|^2 = \infty$, and we put $a_n = \widehat{\mu}(n)$. Now (1.3), which for $r = 1$ is just $\sum_n |a_n|^2 = \infty$, is clearly satisfied. For $r = 1$, (2.3) becomes $\widehat{\beta}_m - \beta = (\sum_{k=1}^{m} |a_k|^2)^{-1} \sum_{n=1}^{m} \overline{a}_n Z_n$. But with our choice of $\{a_n\}$, we have $\langle \widehat{\beta}_m - \beta, Z_0\rangle = \mathbf{E}((\widehat{\beta}_m - \beta)\overline{Z}_0) = 1$ for every m, so there is no consistency. Note that $\widehat{\mu}(n) \to 0$ by the Riemann–Lebesgue lemma, so $m^{-1}\sum_{n=1}^{m} |a_n|^2 \to 0$, and if $f \in L_p([-\pi,\pi), dt)$ for some $1 < p < 2$, then $\sum_{n=1}^{\infty} |a_n|^q < \infty$ for $q = p/(p-1)$, by the Hausdorff–Young Theorem (e.g., Zygmund (1968), Vol. II, p. 101).

Below we will obtain a condition that should be added to (1.3) in order to obtain consistency for all noises with absolutely continuous spectral measure; this requires some additional notations. First, we set $\mathbf{a}_n := (a_n^{(1)}, \dots, a_n^{(r)}) \in \mathbb{C}^r$, so (2.1) becomes $Y_n = \langle \vec{\beta}, \overline{\mathbf{a}}_n\rangle + Z_n$.

In the r-dimensional complex Hilbert space $\mathbb{C}^r$ we denote by $\mathbf{e}_k$ the kth unit vector, and by $\|\cdot\|$ the usual Euclidean norm (it will be clear from the context whether we deal with r-dimensional vectors or with random variables, for which the norm is the L_2-norm). With the usual inner product in $\mathbb{C}^r$ we have $\|\mathbf{b}\|^2 = \sum_{j=1}^r |\langle \mathbf{b}, \mathbf{e}_j\rangle|^2$ for $\mathbf{b} = (b_1, \dots, b_r)$.

For an $r \times r$ matrix B, we define $\|B\| := \max\{\|B(b_1, \dots, b_r)^T\| : \|(b_1, \dots, b_r)\| \le 1\}$ (i.e., the norm of B as an operator on $\mathbb{C}^r$), so obviously $\|B\mathbf{b}^T\| \le \|B\| \cdot \|\mathbf{b}\|$ for $\mathbf{b} \in \mathbb{C}^r$, and $\|B_1 B_2\| \le \|B_1\|\,\|B_2\|$ for any $r \times r$ matrices.

If a symmetric matrix B is non-negative definite, then $\|B\|$ equals the maximal eigenvalue of B, denoted by $\lambda_{\max}(B)$. If in addition B is invertible, then its minimal eigenvalue $\lambda_{\min}$ is positive, and $\|B^{-1}\| = \lambda_{\max}(B^{-1}) = 1/\lambda_{\min}(B)$. It is well known that the trace of a symmetric matrix B (the sum of its diagonal elements) equals the sum of the eigenvalues with their multiplicity (i.e., the trace of the diagonalization of B).

Lemma 2.2. $\operatorname{trace}(C_m) = \sum_{n=1}^m \|\overline{\mathbf{a}}_n C_m\|^2 := \sum_{n=1}^m \sum_{j=1}^r |\langle \overline{\mathbf{a}}_n C_m, e_j\rangle|^2$.

Proof. We compute the trace from $C_m = C_m A_m C_m$:

$$[C_m A_m]_{k,j} = \sum_{i=1}^r c_m^{(k,i)} \sum_{n=1}^m a_n^{(i)} \overline{a_n^{(j)}} = \sum_{n=1}^m (C_m \mathbf{a}_n^T)_k \overline{a_n^{(j)}},$$

and we then have

$$\begin{aligned}[C_m A_m C_m]_{k,l} &= \sum_{j=1}^r \Big[\sum_{n=1}^m (C_m \mathbf{a}_n^T)_k \overline{a_n^{(j)}}\Big] c_m^{(j,l)} \\ &= \sum_{n=1}^m (C_m \mathbf{a}_n^T)_k \sum_{j=1}^r \overline{a_n^{(j)}} c_m^{(j,l)} = \sum_{n=1}^m (C_m \mathbf{a}_n^T)_k (\overline{\mathbf{a}}_n C_m)_l.\end{aligned}$$

Hence $\overline{(C_m \mathbf{a}_n^T)^T} = \overline{\mathbf{a}}_n \overline{C_m^T} = \overline{\mathbf{a}}_n C_m$ yields

$$\operatorname{trace}(C_m) = \operatorname{trace}(C_m A_m C_m) = \sum_{k=1}^r \sum_{n=1}^m (C_m \mathbf{a}_n^T)_k (\overline{\mathbf{a}}_n C_m)_k = \sum_{n=1}^m \|\overline{\mathbf{a}}_n C_m\|^2.$$

Proposition 2.3. *Each of the conditions* (1.2) *and* (1.3) *is equivalent to*

$$\lim_{m\to\infty} \sum_{n=1}^m \sum_{j=1}^r |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_j\rangle|^2 = \lim_{m\to\infty} \sum_{n=1}^m \|\overline{\mathbf{a}}_n A_m^{-1}\|^2 = 0. \tag{2.5}$$

Proof. Since

$$\begin{aligned}\sum_{n=1}^m \Big|\sum_{j=1}^r \alpha_j a_n^{(j)}\Big|^2 &= \sum_{n=1}^m \sum_{j=1}^r \sum_{i=1}^r \alpha_i \overline{\alpha}_j a_n^{(i)} \overline{a_n^{(j)}} = \sum_{j=1}^r \sum_{i=1}^r \alpha_i \overline{\alpha}_j A_m(i,j) \\ &= \langle A_m \vec{\alpha}, \vec{\alpha}\rangle\end{aligned}$$

converges monotonely, the compactness of the unit ball of $\mathbb{C}^r$ and Dini's theorem imply that (1.2) is equivalent to (1.3a).

Since the symmetric matrix A_m is positive definite (for large enough m), so is $C_m = A_m^{-1}$, and all its eigenvalues are positive. Hence

$$\|C_m\| = \lambda_{\max}(C_m) \le \operatorname{trace}(C_m) \le r\|C_m\|,$$

and the previous lemma yields the equivalence of (1.3) and (2.5). ■

THEOREM 2.4. *The following are equivalent for a system of regressors* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *with* A_{m_0} *invertible*:

(i) *For every wide sense stationary noise sequence* $\{Z_n\}$ *with absolutely continuous spectral measure, the LSE sequences are* L_2*-consistent.*

(ii) *The system* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *satisfies* (1.3) *and*

$$\sup_{-\pi\le t<\pi}\ \sup_{m\ge m_0} \Big\| \Big(\sum_{n=1}^{m} \overline{\mathbf{a}}_n e^{int}\Big) A_m^{-1} \Big\| < \infty. \tag{2.6}$$

(iii) *The system* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *satisfies* (2.6) *and*

$$\Big\| \Big(\sum_{n=1}^{m} \overline{\mathbf{a}}_n e^{int}\Big) A_m^{-1} \Big\| \to 0 \quad \text{in } dt\text{-measure on } [-\pi,\pi). \tag{2.7}$$

(iv) *For every strict sense stationary centered noise sequence* $\{Z_n\}$ *with finite second moment and absolutely continuous spectral measure, the LSE sequences are* L_2*-consistent.*

Proof. (ii)⇒(iii). We show that (1.3) implies (2.7). Since $\overline{\mathbf{a}}_n C_m$ is a (row) vector in $\mathbb{C}^r$, we have

$$\sum_{n=1}^{m} \|\overline{\mathbf{a}}_n A_m^{-1}\|^2 = \sum_{k=1}^{r}\sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^2,$$

which converges to 0 by (2.5) and the previous proposition. On $[-\pi,\pi)$ we define $\phi_{m,k}(t) := \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle e^{int}$. Then its $L_2([-\pi,\pi), dt)$-norm is $\|\phi_{m,k}\|^2 = \sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^2$, and it tends to 0 as $m\to\infty$. Denote the function on the left hand side of (2.7) by $\psi_m(t)$. Then $[\psi_m(t)]^2 = \sum_{k=1}^{r} |\phi_{m,k}(t)|^2$, so $\int_{-\pi}^{\pi} |\psi_m(t)|^2\,dt = \sum_{k=1}^{r} \|\phi_{m,k}\|^2 \to 0$, which implies (2.7).

(iii)⇒(i). Let $\{Z_n\}$ be wide sense stationary with absolutely continuous spectral measure μ. Then (2.7) implies that $\psi_m(t)\to 0$ in μ-measure and therefore for every k also $\phi_{m,k}(t)\to 0$ in μ-measure, as $m\to\infty$. By (2.3),

$$\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 = \Big\| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle Z_n \Big\|^2, \tag{2.8}$$

so by (2.4),

$$\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 = \int_{-\pi}^{\pi} \Big| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle e^{int} \Big|^2 \, d\mu(t) = \int_{-\pi}^{\pi} |\phi_{m,k}(t)|^2 \, d\mu(t). \tag{2.9}$$

The integrand converges to 0 in measure, and is uniformly bounded, by (2.6), so Lebesgue's bounded convergence theorem yields $\lim_{m\to\infty} \|\widehat{\beta}_m^{(k)} - \beta^{(k)}\| = 0$.

Clearly (i)⇒(iv).

(iv)⇒(ii). Theorem 2.1 yields the necessity of (1.3). Any non-negative $f \in L_1([-\pi,\pi), dt)$ can be the density of the spectral measure of some centered strict sense stationary sequence with second moment (see Example 1).

By (2.9), the L_2-consistency implies that $\lim_m \int_{-\pi}^{\pi} |\phi_{m,k}(t)|^2 f(t)\, dt = 0$ for every $f \in L_1([-\pi,\pi))$, so the functionals $\{|\phi_{m,k}|^2\}$ are norm bounded in L_∞. Since the $\phi_{m,k}$ are continuous, $\sup_{m\ge m_0} \sup_t |\phi_{m,k}(t)| < \infty$ for $1 \le k \le r$. The equality $|\psi_m(t)|^2 = \sum_{k=1}^{r} |\phi_{m,k}(t)|^2$ yields $\sup_{m\ge m_0} \sup_t |\psi_m(t)|^2 < \infty$, which is (2.6). ■

Remark. Tempelman's condition (1.5) implies (2.6), since

$$\begin{aligned}\Big\| \Big(\sum_{n=1}^{m} \overline{\mathbf{a}}_n e^{int} \Big) A_m^{-1} \Big\|^2 &= \sum_{k=1}^{r} \Big| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle e^{int} \Big|^2 \\ &\le \sum_{k=1}^{r} \Big(\sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle| \Big)^2 \\ &\le \sum_{k=1}^{r} \Big(\sum_{n=1}^{m} \sum_{j=1}^{r} |a_n^{(j)}|\, |c_m^{(jk)}| \Big)^2.\end{aligned}$$

Thus, the result of Tempelman (1973) follows from Theorem 2.4. We will later show (see Remarks following Theorem 2.6) that (1.5) is not necessary for (2.6).

Example 2. *Regressor systems satisfying* (1.3) *but not* (2.6).

Let $r = 1$. If $a_n \ge 0$ with $a_n \to 0$ and $\sum_{n=1}^{\infty} a_n^2 = \infty$, then (1.3) holds. For $\varepsilon > 0$ there is N such that $a_n < \varepsilon$ for $n > N$, so for $m > N$ we have

$$\frac{\sum_{n=1}^{m} a_n^2}{\sum_{n=1}^{m} a_n} \le \frac{\sum_{n=1}^{N} a_n^2 + \varepsilon \sum_{n=N+1}^{m} a_n}{\sum_{n=1}^{m} a_n} \le \varepsilon + \frac{\sum_{n=1}^{N} a_n^2}{\sum_{n=1}^{m} a_n},$$

which yields $\limsup_m \sum_{n=1}^{m} a_n^2 / \sum_{n=1}^{m} a_n \le \varepsilon$, since $(\sum_{n=1}^{m} a_n)^2 \ge \sum_{n=1}^{m} a_n^2 \to \infty$. Hence (2.6) fails.

Remark. Theorem 2.4 implies that the sequences in Example 1, which satisfy (1.3), do not satisfy (2.6). Example 2 shows easier constructions, but except for unboundedness, we have no additional information on the density of the absolutely continuous noise for which the LSE is not L_2-consistent.

THEOREM 2.5. *For a fixed* $1 < p < \infty$ *with dual index* $q = p/(p-1)$, *the following are equivalent for a system of regressors* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *with* A_{m_0} *invertible*:

(i) *For every wide sense stationary noise sequence* $\{Z_n\}$ *with absolutely continuous spectral measure having a density in* $L_p([-\pi,\pi),dt)$, *the LSE sequences are* L_2*-consistent.*
(ii) *The system* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *satisfies* (1.3) *and*

$$\sup_{m\ge m_0}\Big\|\Big(\sum_{n=1}^{m}\overline{\mathbf{a}}_n e^{int}\Big)A_m^{-1}\Big\|_{L_{2q}(dt)} < \infty.$$

Proof. (i)⇒(ii) is proved like (iv)⇒(ii) of the previous theorem, except that now for each $1\le k\le r$ the sequence of functionals $\{|\phi_{m,k}|^2\}$ is bounded in $L_q(dt)$-norm, and (ii) follows from

$$\sup_{m\ge m_0}\|\,|\psi_m|^2\|_q \le \sum_{k=1}^{r}\sup_{m\ge m_0}\|\,|\phi_{m,k}|^2\|_q < \infty.$$

(ii)⇒(i). For a wide sense stationary noise sequence with absolutely continuous spectral measure μ with density f, for $1\le k\le r$, (2.9) yields

$$\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 = \int_{-\pi}^{\pi}|\phi_{m,k}(t)|^2 f(t)\,dt,$$

which, by Theorem 2.1, converges to 0 when f is essentially bounded. Since the assumption of the theorem implies that the sequence of functionals $\{|\phi_{m,k}|^2\}_m$ is bounded in $L_q(dt)$-norm, the above convergence to 0 holds for every $f\in L_p([-\pi,\pi),dt)$. ■

DEFINITION. We call a wide sense stationary sequence $\{Z_n\}$ *mixing* if $\lim_{n\to\infty}\mathbf{E}(Z_n\overline{Z}_0)=0$. For centered sequences, this means that the correlation function tends to 0. By Foguel (1963), this is equivalent to $Z_n = U^nZ_0 \to 0$ weakly in L_2, where U is the isometry corresponding to the shift. In terms of the spectral measure μ of $\{Z_n\}$, we have $\widehat{\mu}(n)=\int_{-\pi}^{\pi}e^{int}\,d\mu = \mathbf{E}(Z_n\overline{Z}_0)\to 0$. A measure μ on $[-\pi,\pi)$ with $\lim_{|n|\to\infty}\widehat{\mu}(n)=0$ will be called *mixing.*

In particular, every wide sense stationary sequence with absolutely continuous spectral measure is mixing, by the Riemann–Lebesgue Lemma. Wiener's Lemma (Zygmund (1968), p. 108, Krengel (1985), p. 96) shows that a mixing stationary sequence has a continuous (atomless) spectral measure. When τ is a strongly mixing probability preserving transformation, the stationary sequence $\{f\circ\tau^n\}$ defined by $f\in L_2$ with zero expectation is mixing.

THEOREM 2.6. *Let* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *be a system of regressors with* A_{m_0} *invertible, satisfying condition* (1.3) *and*

$$\sup_{m\ge m_0}\sum_{n=1}^{m}\sum_{k=1}^{r}|\langle \overline{\mathbf{a}}_n C_m,\mathbf{e}_k\rangle| = \sup_{m\ge m_0}\sum_{n=1}^{m}\sum_{k=1}^{r}\Big|\sum_{j=1}^{r}\overline{a_n^{(j)}}c_m^{(j,k)}\Big| < \infty. \tag{2.10}$$

Then for every mixing wide sense stationary noise sequence $\{Z_n\}$ *the LSE sequences are* L_2*-consistent.*

Proof. For $m\ge m_0$ and $1\le k\le r$, define $\alpha_{m,n}^{(k)} := \sum_{j=1}^{r} c_m^{(j,k)}\overline{a_n^{(j)}}$ for $1\le n\le m$, and $\alpha_{m,n}^{(k)} := 0$ for $n>m$. Then (2.3) becomes

$$\widehat{\beta}_m^{(k)} - \beta^{(k)} = \sum_{n=1}^{\infty}\alpha_{m,n}^{(k)} Z_n. \tag{2.11}$$

We want to obtain the consistency when $\{Z_n\}$ is mixing from the generalized Blum–Hanson theorem (Theorem 8.1.3 in Krengel (1985)). Since

$$\sum_{n=1}^{\infty}|\alpha_{m,n}^{(k)}| = \sum_{n=1}^{m}\Big|\sum_{j=1}^{r} c_m^{(j,k)}\overline{a_n^{(j)}}\Big|$$

we have $\sup_{m\ge m_0}\sum_{n=1}^{\infty}|\alpha_{m,n}^{(k)}|<\infty$ for each $k\le r$, by (2.10).

By the definition, $\alpha_{m,n}^{(k)} = \langle \overline{\mathbf{a}}_n C_m,\mathbf{e}_k\rangle$ for $n\le m$. Hence by the Cauchy–Schwarz inequality and Lemma 2.2 we obtain

$$\begin{aligned}\sup_{n\ge 1}|\alpha_{m,n}^{(k)}| &= \sup_{1\le n\le m}|\alpha_{m,n}^{(k)}| \le \sup_{1\le n\le m}\|\overline{\mathbf{a}}_n C_m\|\\ &\le \sqrt{\sum_{n=1}^{m}\|\overline{\mathbf{a}}_n C_m\|^2} = \sqrt{\operatorname{trace}(C_m)} \le \sqrt{r\|C_m\|}.\end{aligned}$$

Thus (1.3) implies that $\sup_{n\ge 1}|\alpha_{m,n}^{(k)}|\to 0$ as $m\to\infty$ for each fixed $k\le r$. Whenever $\{Z_n\}$ is mixing, the proof of Theorem 8.1.3 in Krengel (1985) now yields the norm convergence to 0, as $m\to\infty$, of the right hand side of (2.11). ∎

REMARKS. 1. Since for each k we have

$$\sum_{n=1}^{m}\Big|\sum_{j=1}^{r} c_m^{(j,k)}\overline{a_n^{(j)}}\Big| \le \sum_{n=1}^{m}\sum_{j=1}^{r}|c_m^{(j,k)}\overline{a_n^{(j)}}| = \sum_{j=1}^{r}|c_m^{(j,k)}|\sum_{n=1}^{m}|\overline{a_n^{(j)}}|,$$

(2.10) is implied by (1.5). Thus the theorem improves the result of Tempelman (1973), who proved the L_2-consistency, for systems satisfying (1.3) and (1.5), only for noise sequences with absolutely continuous spectral measure.

2. For $r=1$, (2.10) is equivalent to (1.5). In general, for $r>1$, (2.10) is weaker than (1.5), as shown by the example for $r=2$ with $a_n^{(1)}=1$ for

$n \geq 1$, $a_1^{(2)} = 0$ and $a_n^{(2)} = 1$ for $n \geq 2$. The rows of A_m^{-1} are $(1, -1)$ and $(-1, m/(m-1))$. Then (1.5) fails since $|c_m^{(1,1)}| \sum_{n=1}^m |a_n^{(1)}| = m$, and (2.10) holds since $\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_1 \rangle = 0$ and $\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_2 \rangle = 1/(m-1)$ for $n \geq 2$. In this example, the consistency of the LSE depends on the first observation; without it, we can only estimate $\beta^{(1)} + \beta^{(2)}$.

3. (2.10) implies (2.6), by Theorems 2.6 and 2.4; however, this implication follows directly, since for each k we obtain

$$\sup_{-\pi \leq t < \pi} \sup_{m \geq m_0} \Big| \Big\langle \Big(\sum_{n=1}^m \overline{\mathbf{a}}_n e^{int} \Big) C_m, \mathbf{e}_k \Big\rangle \Big| = \sup_{-\pi \leq t < \pi} \sup_{m \geq m_0} \Big| \sum_{n=1}^m \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle e^{int} \Big|$$
$$\leq \sup_{m \geq m_0} \sum_{n=1}^m |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle|,$$

which is finite by (2.10); summing over k yields (2.6).

4. (2.10) (or even (1.5)) does not imply (1.3): for $r = 1$, take $\{a_n\}$ with $\sum_{n=1}^\infty |a_n| < \infty$. Then also $\sum_{n=1}^\infty |a_n|^2 < \infty$, so (1.5) is obviously satisfied, and (1.3) is not. However, (2.10) always implies $\sup_{m \geq m_0} \|A_m^{-1}\| < \infty$, since Lemma 2.2 and the proof of Proposition 2.3 yield

$$\|A_m^{-1}\| \leq \sum_{n=1}^m \|\overline{\mathbf{a}}_n C_m\|^2 = \sum_{k=1}^r \Big(\sum_{n=1}^m |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle|^2 \Big)$$
$$\leq \sum_{k=1}^r \Big(\sum_{n=1}^m |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle| \Big)^2.$$

THEOREM 2.7. *Let $\{\mathbf{a}^{(j)}\}_{1 \leq j \leq r}$ be a system of regressors satisfying*

$$\liminf_{m \to \infty} \frac{\lambda_{\min}(A_m)}{m} > 0 \quad (i.e., \ \limsup_{m \to \infty} m \|A_m^{-1}\| < \infty). \tag{2.12}$$

Then for every wide sense stationary noise sequence $\{Z_n\}$ with continuous (i.e., atomless) spectral measure the LSE sequences are L_2-consistent.

Proof. By (2.8) we have

$$\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 = \Big\| \sum_{n=1}^m \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle Z_n \Big\|^2$$
$$= \sum_{n=1}^m \sum_{l=1}^m \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k \rangle \overline{\langle \overline{\mathbf{a}}_l C_m, \mathbf{e}_k \rangle} \mathbf{E}(Z_n \overline{Z}_l).$$

Applying the Cauchy–Schwarz inequality (to the double sum) and using Lemma 2.2, we obtain

$$
\begin{aligned}
&\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 \\
&\quad \le \Big(\sum_{n,l=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^2 |\langle \overline{\mathbf{a}}_l C_m, \mathbf{e}_k\rangle|^2\Big)^{1/2} \Big(\sum_{n,l=1}^{m} |\widehat{\mu}(n-l)|^2\Big)^{1/2} \\
&\quad = \Big(\sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^2\Big) \Big(\sum_{n,l=1}^{m} |\widehat{\mu}(n-l)|^2\Big)^{1/2} \\
&\quad \le \Big(\sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\|^2\Big) \Big(\sum_{n,l=1}^{m} |\widehat{\mu}(n-l)|^2\Big)^{1/2} \\
&\quad = \operatorname{trace}(C_m) \Big(\sum_{n,l=1}^{m} |\widehat{\mu}(n-l)|^2\Big)^{1/2} \le r\|C_m\| \Big(\sum_{n=-m}^{m} (m-|n|)|\widehat{\mu}(n)|^2\Big)^{1/2} \\
&\quad \le rm\|C_m\| \bigg(\frac{1}{m}\sum_{n=-m}^{m} |\widehat{\mu}(n)|^2\bigg)^{1/2}.
\end{aligned}
$$

By Wiener's lemma $m^{-1}\sum_{n=-m}^{m} |\widehat{\mu}(n)|^2 \to 2\sum_t \mu(\{t\})^2$ as $m \to \infty$. Since μ is continuous, the limit is 0, and together with (2.12) we obtain the L_2-consistency. ∎

REMARKS. 1. The above proof provides also an upper bound for the rate of convergence:

$$
\begin{aligned}
\mathbf{E}(|\widehat{\beta}_m^{(k)} - \beta^{(k)}|^2) &\le 2rm\|C_m\| \bigg[\frac{1}{m}\sum_{n=0}^{m} |\mathbf{E}(Z_n \overline{Z}_0)|^2\bigg]^{1/2} \\
&\le K \bigg[\frac{1}{m}\sum_{n=0}^{m} |\mathbf{E}(Z_n \overline{Z}_0)|^2\bigg]^{1/2}.
\end{aligned}
$$

2. Condition (2.12) is not sufficient for L_2-consistency of the LSE when the spectral measure of $\mathbf{Z}$ has an atom: let $r = 1$ and take $a_n = (-1)^n$ and $Z_n = (-1)^n Z_0$ for every n, with $Z_0 \neq 0$. Note that $\{Z_n\}$ is always ergodic (its spectral measure has no atom at 0).

3. As shown in Appendix A, when the regressors are obtained as realizations of linearly independent strict sense stationary processes, condition (2.12) is satisfied.

COROLLARY 2.8 *Let* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *be a system of regressors such that* $m^{-1}A_m$ *converges to a non-singular matrix* (*i.e., the system is asymptotically non-degenerate*)*. Then for every wide sense stationary noise sequence* $\{Z_n\}$ *with continuous spectral measure the LSE sequences are* L_2*-consistent.*

Proof. (2.12) holds since by assumption mA_m^{-1} converges to an invertible matrix. ∎

APPLICATION. *Amplitude estimation in sums of harmonic signals with known frequencies.*

A harmonic signal with amplitude $V > 0$, frequency ω (real), and phase $\varphi \in [-\pi, \pi)$ is given at time t by $Ve^{i(\omega t+\varphi)}$. Assume that we are sampling the sum of r harmonic signals with known frequencies and unknown phases and amplitudes, so the observed signal at time n is

$$Y_n = \sum_{j=1}^{r} V_j e^{i(\omega_j n+\varphi_j)} + Z_n.$$

We put $\beta^{(j)} := V_j e^{i\varphi_j}$ and $a_n^{(j)} = (e^{i\omega_j})^n$. When the frequencies are all different, it is easy to compute that $m^{-1}A_m$ converges to the $r \times r$ identity matrix I_r; hence Corollary 2.8 can be applied. The estimate for V_j is $\widehat{V}_{j,m} := |\widehat{\beta}_m^{(j)}|$; since $|\,|\beta^{(j)}| - |\widehat{\beta}_m^{(j)}|\,| \le |\beta^{(j)} - \widehat{\beta}_m^{(j)}|$, for every wide sense stationary noise sequence $\{Z_n\}$ with continuous spectral measure we have

$$\|V_j - \widehat{V}_{j,m}\| \le \|\beta^{(j)} - \widehat{\beta}_m^{(j)}\| \xrightarrow[m\to\infty]{} 0.$$

REMARKS. 1. Clearly (2.12) implies (1.3). Using twice the Cauchy–Schwarz inequality and then Lemma 2.2, we have

$$\sum_{n=1}^{m} \Big|\sum_{j=1}^{r} \overline{a_n^{(j)}} c_m^{(j,k)}\Big| = \sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle| \le \sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\|$$
$$\le \sqrt{m}\Big(\sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\|^2\Big)^{1/2} = \sqrt{m}\sqrt{\operatorname{trace}(C_m)} \le \sqrt{r}\sqrt{m\|C_m\|},$$

which shows that (2.12) implies also (2.10), which in turn implies (2.6).

2. The condition of Corollary 2.8 implies (1.5): $|c_m^{(k,j)}| = |\langle \mathbf{e}_j C_m, \mathbf{e}_k\rangle| \le \|C_m\|$, and $\sum_{n=1}^m |a_n^{(j)}|^2 = [A_m]_{j,j} = |\langle \mathbf{e}_j A_m, \mathbf{e}_j\rangle| \le \|A_m\|$. Together with the Cauchy–Schwarz inequality this yields

$$|c_m^{(k,j)}| \sum_{n=1}^{m} |a_n^{(j)}| \le \|C_m\|\sqrt{m}\Big(\sum_{n=1}^{m} |a_n^{(j)}|^2\Big)^{1/2} \le \|mC_m\|(\|m^{-1}A_m\|)^{1/2},$$

which converges by assumption.

3. For $r = 1$, let $a_n = 1$ if n is a square, and $a_n = 0$ otherwise. Then (1.3) and (2.10) hold, but $m^{-1}\sum_{n=1}^m |a_n|^2 \to 0$. Thus, (1.3) and (2.10) together do not imply (2.12).

4. Any weakly mixing invertible probability preserving transformation which is not strongly mixing (see category theorems in Halmos (1956), or the example of Chacon (1969)), and $f \in L_2$ with zero integral, yield examples of strictly stationary centered ergodic noise sequences with continuous (atomless) spectral measure, which are not mixing.

5. The convergence of $\{m^{-1}A_m\}$ to a non-singular matrix easily implies that the regressors satisfy the set of *Grenander's conditions* for asymptotic efficiency when the noise has continuous spectral density (e.g., Grenander and Rosenblatt (1957), §7.2, pp. 233–234); for $r=1$ Grenander's conditions are (1.3) and $\lim_{m\to\infty}|a_m|^2/\sum_{n=1}^m |a_n|^2 = 0$. In Example 1 Grenander's conditions are satisfied, the noise is stationary with absolutely continuous spectral measure, but there is no L_2-consistency of the LSE.

EXAMPLE 3. *The condition*

$$\limsup_{m\to\infty} \frac{\lambda_{\min}(A_m)}{m} > 0 \qquad (\text{i.e., } \liminf_{m\to\infty} m\|A_m^{-1}\| < \infty). \tag{2.13}$$

does not imply L_2-consistency for all absolutely continuous spectral measures.

We give an example for $r=1$. We construct $\{a_n\}$ inductively by blocks. Let $a_1 = a_2 = 1$ and $m_1 = 2$; at the kth step take a block of 1's of length $m_k' - m_{k-1}$ such that the average $(m_k')^{-1}\sum_{n=1}^{m_k'}|a_n|^2$ is more than 0.75, followed by a block of the constant value $1/k$ of length $m_k - m_k'$ such that $m_k^{-1}\sum_{n=1}^{m_k}|a_n|^2 < 2/k^2$. Since all the m_k terms constructed are at least $1/k$, we have $m_k^{-1}\sum_{n=1}^{m_k}|a_n| \ge 1/k$. It is easy to check that (1.5) does not hold; since $a_n > 0$ for every n, this means that also (2.6) fails, though (1.3) holds. By Theorem 2.4, there exist a strict sense stationary centered sequence with second moment and absolutely continuous spectral measure for which the LSE is not L_2-consistent.

REMARKS. 1. The representation $\langle A_m\vec{\alpha},\vec{\alpha}\rangle = \sum_{n=1}^m |\sum_{j=1}^r \alpha_j a_n^{(j)}|^2$ for $\alpha\in\mathbb{C}^r$ (computed in Proposition 2.3) shows that $\lambda_{\min}(A_m)$ (which equals $\inf\{\langle A_m\vec{\alpha},\vec{\alpha}\rangle : \|\alpha\| = 1\}$) is increasing in m. Hence (2.13) implies (1.3), so by Theorem 2.1 the density of the absolutely continuous spectral measure obtained in Example 3 is not essentially bounded.

2. By the previous remark and Theorem 2.6, (2.13) and (1.5) together imply L_2-consistency of the LSE for noises with mixing spectral measure. We saw before that we may have $m^{-1}\sum_{n=1}^m |a_n|^2 \to 0$ for $\{a_n\}$ satisfying (1.3) and (1.5).

EXAMPLE 4. *Conditions* (2.13) *and* (1.5) *together do not imply L_2-consistency for all continuous (atomless) spectral measures.*

We will give an example for $r=1$. Fix integers $q\ge 3$, $l \ge 1$, let $q_1 = q^l$, and inductively put $q_{j+1} = q^{q_j}$. First, we define $a_1 = 1$, and $a_n = 0$ if $n>1$ is not a power of q. For n a power of q, in general $a_n = 1$, except when $n = q_j$, in which case $a_{q_j} = \sqrt{q_j}$. Since $|a_{q_j}|^2 = q_j$, we have $q_j^{-1}\sum_{n=1}^{q_j}|a_n|^2 \ge 1$, so $\limsup_{m\to\infty}\lambda_{\min}(A_m)/m = \limsup_m m^{-1}\sum_{n=1}^m |a_n|^2 > 0$. We now show by induction that $(q_1+\cdots+q_j)/q_j \le q^2/(q^2-2)$. This is clear for $j=1$. Since the function $f(x) = x/q^x$ is decreasing for $x\ge 1$, we have $q_j/q_{j+1} =$

$f(q_j) \le 2/q^2$, so using the induction hypothesis we obtain

$$\frac{q_1 + \cdots + q_{j+1}}{q_{j+1}} = 1 + \frac{q_1 + \cdots + q_j}{q_j} \cdot \frac{q_j}{q_{j+1}} \le 1 + \frac{q^2}{q^2 - 2} \cdot \frac{2}{q^2} = \frac{q^2}{q^2 - 2}.$$

Let $m = q_J - 1$. There are q_{J-1} non-negative powers of q which are less than m, so $\sum_{n=1}^{m} |a_n|^2 = q_{J-1} - J + 1 + \sum_{j=1}^{J-1} q_j$. Arguments like those in Example 1 show that every finite measure μ on $[-\pi, \pi)$ is the spectral measure of some strict sense stationary centered sequence $\{Z_n\}$ with finite second moment. Let μ be the Lebesgue–Cantor measure obtained with constant ratio of dissection $1/q$ with q *odd*, for which all the Fourier coefficients $\widehat{\mu}(q^k)$ have the same value $c \neq 0$ (see Zygmund (1968), p. 196). The measure μ is continuous (but obviously not mixing), and by our construction of $\{a_n\}$ and (2.3) we have, for $m = q_J - 1$,

$$\langle \widehat{\beta}_m - \beta, \overline{Z}_0 \rangle = \frac{\sum_{n=1}^{m} a_n \widehat{\mu}(n)}{\sum_{k=1}^{m} |a_n|^2} = c \cdot \frac{q_{J-1} - J + 1 + \sum_{j=1}^{J-1} \sqrt{q_j}}{q_{J-1} - J + 1 + \sum_{j=1}^{J-1} q_j}.$$

When we divide the numerator and denominator by q_{J-1}, we get more than 1 in the new numerator, and the new denominator is bounded above by $1 + q^2/(q^2 - 2)$; hence $\widehat{\beta}_m - \beta$ does not converge to 0 in L_2.

In addition to (1.3), (2.12) implies

$$\liminf_{m\to\infty} \frac{\|A_m\|}{m} = \liminf_{m\to\infty} \frac{\lambda_{\max}(A_m)}{m} > 0. \tag{2.14}$$

When $r = 1$, (2.14) and (2.12) are equivalent. However, the next proposition shows that in general (1.3) and (2.14), even reinforced by the condition $\limsup_{m\to\infty} \|A_m\|/m < \infty$, do not yield the consistency for all noises with atomless spectral measure, hence do not imply (2.12).

PROPOSITION 2.9. *Let $\{Z_n\}$ be a wide sense stationary sequence with $\mu(\{0\}) = 0$ which is not mixing. Then there exists a system $\{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}\}$ satisfying* (1.3) *and* (1.5), *for which $\lim_{m\to\infty} \|A_m\|/m$ exists and is not zero, such that the LSE sequence $\{\widehat{\beta}_m^{(2)}\}$ for $\{Z_n\}$ is not L_2-consistent.*

Proof. The condition $\mu(\{0\}) = 0$ means that $\|m^{-1} \sum_{n=1}^{m} Z_n\| \to 0$ as $m \to \infty$, by the mean ergodic theorem (consequence of (2.4)). Since $\{Z_n\}$ is not mixing, for some $\varepsilon > 0$ there is a strictly increasing sequence $\{k_j\}$ of integers such that either $\mathrm{Re}(\langle Z_{k_j}, \overline{Z}_0 \rangle) > \varepsilon$ (or the reverse inequality for every k_j), or $\mathrm{Im}(\langle Z_{k_j}, \overline{Z}_0 \rangle) > \varepsilon$ (or the reverse inequality for every k_j), and we may take only a subsequence of $\{k_j\}$ of density 0 (still denoted by $\{k_j\}$). This shows that $\liminf_l \|l^{-1} \sum_{j=1}^{l} Z_{k_j}\| > 0$. Define $a_n^{(1)} = 1$ if $n \notin \{k_j\}$, and otherwise $a_n^{(1)} = 0$, and put $a_n^{(2)} = 1 - a_n^{(1)}$. Then A_m is a diagonal matrix, with $a_m^{(1,1)} = \sum_{n=1}^{m} a_n^{(1)} = |\{n \le m : n \notin \{k_j\}|$ and $a_m^{(2,2)} = \sum_{n=1}^{m} a_n^{(2)} =$

$|\{n \le m : n \in \{k_j\}|$. Hence $a_m^{(1,1)}/m \to 1$ and $a_m^{(2,2)}/m \to 0$, showing $\|A_m\|/m \to 1$. The above limits yield $\lambda_{\min}(A_m) = a_m^{(2,2)}$ for large m, which tends to ∞ since $\{k_j\}$ is infinite, so (1.3) holds. For $m = k_l$ we have $a_m^{(2,2)} = l$, so $\|C_{k_l}\| = l^{-1}$. For $m = k_l$, (2.8) yields

$$\|\widehat{\beta}_m^{(2)} - \beta^{(2)}\| = \Big\| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_2 \rangle Z_n \Big\| = \Big\| \sum_{j=1}^{l} Z_{k_j} \Big\| / l,$$

which does not converge to 0 by construction. It is easy to verify that (1.5) holds. We can see directly that (2.12) does not hold. ∎

Remark. For $\{Z_n\}$ as in the proposition, there exists a sequence $\{a_n\}$ which satisfies (1.3) and (1.5), such that the LSE is not L_2-consistent: Proposition 8.1.2 in Krengel (1985) yields the existence of a strictly increasing sequence of integers $\{k_j\}$ such that $\|l^{-1} \sum_{j=1}^{l} Z_{k_j}\|$ does not converge to 0. Let $a_{k_j} = 1$, and for $n \notin \{k_j\}$ put $a_n = 0$. Then (1.3) obviously holds, and (1.5) holds since $a_n^2 = a_n \ge 0$. But by construction

$$\limsup_{l\to\infty} \left\| \frac{\sum_{n=1}^{k_l} \overline{a}_n Z_n}{\sum_{n=1}^{k_l} |a_n|^2} \right\| = \limsup_{l\to\infty} \left\| \frac{1}{l} \sum_{j=1}^{l} Z_{k_j} \right\| > 0.$$

Choosing $\{Z_n\}$ with continuous spectral measure but not mixing, we see that (1.3) and (1.5) are not sufficient for L_2-consistency for every atomless spectral measure.

For $0 < \alpha < 1$, an intermediate condition between (1.3) and (2.12) is given by

$$\lim_{m\to\infty} m^{\alpha} \|A_m^{-1}\| = 0. \tag{2.15}$$

However, even for $r = 1$, (2.15) does not yield consistency for every absolutely continuous spectral measure: take $0 < \delta < (1-\alpha)/2$, and define $a_n = n^{-\delta}$. Estimating sums by integrals shows that $m^{\alpha}/\sum_{n=1}^{m} a_n^2 = O(m^{\alpha+2\delta-1}) \to 0$ and $\sum_{n=1}^{m} a_n / \sum_{n=1}^{m} a_n^2 = O(m^{\delta}) \to \infty$, so (2.15) holds and (2.6) fails; by Theorem 2.4 there is a noise with absolutely continuous spectral measure for which the LSE is not L_2-consistent.

Proposition 2.10. *Let $1 < p \le 2$. If the system of regressors satisfies*

$$\liminf_{m\to\infty} \frac{\lambda_{\min}(A_m)}{m^{1/p}} > 0 \quad (\textit{i.e.}, \ \limsup_{m\to\infty} m^{1/p} \|A_m^{-1}\| < \infty), \tag{2.16}$$

then for every wide sense stationary sequence having an absolutely continuous spectral measure with density in $L_p([-\pi,\pi), dt)$, the LSE sequences are L_2-consistent.

Proof. Let $q = p/(p-1)$. For a noise with absolutely continuous spectral measure μ we use the formula obtained in the proof of Theorem 2.7, Hölder's inequality for the double sum, and (since $1 \le p \le 2$) the inequality $\|\cdot\|_p \le \|\cdot\|_2$ in probability spaces, to obtain

$$\begin{aligned}
\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 &= \sum_{n=1}^{m}\sum_{l=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle \overline{\langle \overline{\mathbf{a}}_l C_m, \mathbf{e}_k\rangle} \mathbf{E}(Z_n \overline{Z}_l) \\
&\le \Big(\sum_{n,l=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^p |\langle \overline{\mathbf{a}}_l C_m, \mathbf{e}_k\rangle|^p\Big)^{1/p} \Big(\sum_{n,l=1}^{m} |\widehat{\mu}(n-l)|^q\Big)^{1/q} \\
&= \Big(\sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^p\Big)^{2/p} \Big(\sum_{n=-m}^{m} (m-|n|)|\widehat{\mu}(n)|^q\Big)^{1/q} \\
&= m^{2/p}\bigg(\frac{1}{m}\sum_{n=1}^{m} |\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle|^p\bigg)^{2/p} \bigg(m\sum_{n=-m}^{m} \Big(1-\frac{|n|}{m}\Big)|\widehat{\mu}(n)|^q\bigg)^{1/q} \\
&\le m^{2/p+1/q}\bigg(\frac{1}{m}\sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\|^2\bigg)\Big(\sum_{n=-m}^{m} |\widehat{\mu}(n)|^q\Big)^{1/q} \\
&= m^{1/p}\cdot \operatorname{trace}(C_m)\Big(\sum_{n=-m}^{m} |\widehat{\mu}(n)|^q\Big)^{1/q} \\
&\le m^{1/p} r\|C_m\|\Big(\sum_{n=-m}^{m} |\widehat{\mu}(n)|^q\Big)^{1/q}.
\end{aligned}$$

Since μ has density $f \in L_p$ and $1 < p \le 2$, (2.9) and the Hausdorff–Young theorem yield

$$\int_{-\pi}^{\pi} |\phi_{m,k}(t)|^2 f(t)\, dt = \|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|^2 \le r\cdot m^{1/p}\|C_m\|\,\|f\|_p,$$

with $\phi_{m,k}$ defined in the proof of Theorem 2.4. The hypothesis (2.16) yields

$$\sup_{m\ge m_0} \int_{-\pi}^{\pi} |\phi_{m,k}(t)|^2 f(t)\, dt \le r\|f\|_p \sup_{m\ge m_0} m^{1/p}\|C_m\| \le K\|f\|_p$$

for every $0 \le f \in L_p$, which shows that the sequence $\{|\phi_{m,k}|^2\}_{m\ge m_0}$ of functionals is L_q-norm bounded. Clearly (2.16) implies (1.3). Now the conditions of Theorem 2.5(ii) are satisfied, which proves the claimed consistency. ∎

REMARKS. 1. The condition in the previous proposition is not necessary. For $r = 1$, let $a_n = 1$ if n is a cube, and $a_n = 0$ otherwise. Then (1.3) and (1.5) hold, so we have consistency for every mixing noise, but for $1 < p \le 2$ we have $(m^3)^{1/p}/\sum_{n=1}^{m^3} a_n^2 = m^{3/p-1} \to \infty$.

2. Let $r = 1$. For $p = 2$, the sequence $a_n = n^{-1/4}$ satisfies (2.16) but not (2.15). By Example 2, (2.6) and (1.5) fail.

EXAMPLE 5. *Condition* (2.15) *for every* $0 < \alpha < 1$ *does not imply* L_2*-consistency for all absolutely continuous spectral measures.*

Let $r = 1$, and define $a_1 = 2$, $a_n = 1/\log n$ for $n \ge 2$. Since $\{a_n\}$ is decreasing, for any $0 < \alpha < 1$ we have

$$m^{-\alpha} \sum_{n=1}^{m} |a_n|^2 \ge m^{-\alpha} m/(\log m)^2 \to \infty,$$

so (2.15) holds for α. By Example 2, (2.6) fails, so there exists a stationary noise with absolutely continuous spectral measure for which the LSE is not L_2-consistent; its density is not in any L_p with $p > 1$, since by Proposition 2.10, the LSE is L_2-consistent for every stationary noise with density in $\bigcup_{p>1} L_p([-\pi, \pi), dt)$.

DEFINITION. A wide sense stationary sequence $\{Z_n\}$ is called *mean ergodic* if $\lim_{m\to\infty} \|m^{-1} \sum_{n=1}^{m} Z_n\| = 0$; this means that the spectral measure of $\{Z_n\}$ on $[-\pi, \pi)$ has no atom at 0. In particular, every centered *ergodic* stationary sequence is mean ergodic.

THEOREM 2.11. *The following are equivalent for a system of regressors* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *with* A_{m_0} *invertible*:

(i) *For every wide sense stationary mean ergodic noise sequence* $\{Z_n\}$, *the LSE sequences are* L_2*-consistent.*

(ii) *The system* $\{\mathbf{a}^{(j)}\}_{1\le j\le r}$ *satisfies* (2.6) *and*

$$\Big\|\Big(\sum_{n=1}^{m} \overline{\mathbf{a}}_n e^{int}\Big) A_m^{-1}\Big\| \to 0 \quad \text{for every } 0 \ne t \in [-\pi, \pi). \tag{2.17}$$

Proof. The proof of (ii)⇒(i), using the spectral representation (2.4) and the Lebesgue dominated convergence theorem, is similar to that of (iii)⇒(i) in Theorem 2.4.

By Theorem 2.4, (i) implies (2.6). Fix $0 \ne t \in [-\pi, \pi)$. The sequence $Z_n = e^{int} Z_0$, with Z_0 non-zero with finite second moment, is clearly mean ergodic. If we put this sequence in (2.8), the consistency assumption (i) yields $\| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle e^{int}\|^2 \to 0$ as $m \to \infty$ for each $k \le r$, which yields (2.17). ■

REMARKS. 1. (2.17) together with (1.3) does not imply (2.6), even for $r = 1$: Let $a_n = n^{-1/4}$ for $n \ge 1$. Then (1.3) is obviously satisfied. By Example 2, (2.6) fails. By Theorem 2.6 on p. 4 of Zygmund (1968), for $t \ne 0$, $\sum_{n=1}^{m} a_n e^{int}$ converges as $m \to \infty$, so $\sum_{n=1}^{m} a_n e^{int} / \sum_{n=1}^{m} |a_n|^2$ converges to 0 as $m \to \infty$ for $t \ne 0$, which is (2.17).

2. Conditions (2.6) and (2.7) do not imply condition (2.17). One can even make (2.17) fail on a dense countable subset of $[-\pi,\pi)$. The example is for $r=1$. Let $\{c_j\}$ be a sequence of complex numbers with $|c_j|>0$ for every $j\geq 1$ and $\sum_{j=1}^{\infty}|c_j|<\infty$, and take a dense countable set $\{t_j\}\subset[-\pi,\pi)$. Define $a_n=\sum_{j=1}^{\infty}c_je^{int_j}$. Then $\{a_n\}$ is a bounded almost periodic sequence, so $m^{-1}\sum_{n=1}^{m}|a_n|^2\to\sum_{j=1}^{\infty}|c_j|^2>0$. Thus $\sum_{n=1}^{\infty}|a_n|^2=\infty$, so (1.3) holds. Cauchy's inequality in (1.7) shows that (1.5), which implies (2.6), also holds. By Theorems 2.4 and 2.6, (2.7) also holds. However, condition (2.17) fails, since for $t=-t_j$ we have

$$\frac{\sum_{k=1}^{m}a_ke^{-ikt_j}}{\sum_{k=1}^{m}|a_k|^2}\xrightarrow[m\to\infty]{}\frac{c_j}{\sum_{n=1}^{\infty}|c_n|^2}>0.$$

In this example also (2.12), which implies (2.10), holds. However, if for $t_j\neq 0$ we put $Z_n=e^{int_j}Z_0$ with Z_0 centered with second moment finite, then $\{Z_n\}$ is mean ergodic, but the LSE is not L_2-consistent.

3. If all we want is failure of (2.17), it is enough to fix t_0 with $0<|t_0|<\pi$, and let $a_n=e^{int_0}$. Then $|a_n|=1$ for every n, so (1.3) and (1.5) hold, but (2.17) fails at $t=-t_0$.

4. A necessary and sufficient condition for L_2-consistency for every wide sense stationary noise is obtained by removing the restriction $t\neq 0$ from (2.17).

EXAMPLE 6. *Conditions* (2.6) *and* (2.17) *together do not imply* (2.10).

Let $\{\varepsilon_n\}$ be the Rudin–Shapiro ±1 sequence defined in Rudin (1959), for which there exists a constant C such that

$$\Big|\sum_{n=1}^{N}\varepsilon_ne^{int}\Big|\leq C\sqrt{N},\qquad t\in[-\pi,\pi),\ N=1,2,\dots.\tag{2.18}$$

We take $r=1$ (so (1.5) and (2.10) are equivalent), and define $a_n=\varepsilon_nn^{-1/8}$. Summation by parts, monotonicity of $n^{-1/8}$, and (2.18) yield (see Zygmund (1968), pp. 3–4)

$$\Big|\sum_{n=1}^{N}a_ne^{int}\Big|=\Big|\sum_{n=1}^{N}\varepsilon_nn^{-1/8}e^{int}\Big|\leq\max_{1\leq k\leq N}\Big|\sum_{n=1}^{k}\varepsilon_ne^{int}\Big|\leq C\sqrt{N}\tag{2.19}$$

for every $t\in[-\pi,\pi)$ and $N\geq 1$. Since $\sum_{n=1}^{N}|a_n|^2\geq\frac{4}{3}[N^{3/4}-1]$, (2.19) shows that (2.17) holds uniformly in t, and so also (2.6) holds. It is easy to compute that (1.5), which in this case is equivalent to (2.10), fails (see also Example 2).

REMARKS. 1. By Theorem 2.11, the regressor sequence defined in the previous example satisfies the conclusion of Theorems 2.6 and 2.7, though

(2.10), and hence also (2.12), are not satisfied. Thus (2.10) is not necessary for consistency for all mixing noises, and (2.12) is not necessary for consistency for all noises with continuous spectral measure.

2. For $r = 1$, condition (2.15) with $\alpha = 1/p$ implies, by Hölder's inequality, that $\sum_{n=1}^{\infty} |a_n|^{2q} = \infty$ (where $q = p/(p-1)$ is the dual index). Note that when $a_n \geq 0$ for every n, Theorem 2.4 and Example 2 imply that this divergence is necessary for L_2-consistency for every noise with absolutely continuous spectral measure. However, if we take $a_n = n^{-1/4}\varepsilon_n$ in Example 6, similar estimates show that (2.6) holds. Since (1.3) holds, by Theorem 2.4 we have L_2-consistency for every noise with absolutely continuous spectral measure. On the other hand, $\sum_{n=1}^{\infty} |a_n|^{2q} < \infty$ for every $q > 2$.

Proposition 2.12. *If* $\lim_{m\to\infty} m\|A_m^{-1}\| = 0$, *then for every noise sequence* $\{Z_n\}$ *with* $\sup_n \|Z_n\|_p = K < \infty$ *for some fixed* $1 \leq p < \infty$ *the LSE sequences are* L_p*-consistent.*

Proof. Using (2.3), the triangle inequality, the Cauchy–Schwarz inequality, and Lemma 2.2, we obtain

$$\begin{aligned}
&\|\widehat{\beta}_m^{(k)} - \beta^{(k)}\|_p \\
&\quad = \Big\| \sum_{n=1}^{m} \langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle Z_n \Big\|_p \leq \sum_{n=1}^{m} \|\langle \overline{\mathbf{a}}_n C_m, \mathbf{e}_k\rangle Z_n\|_p \leq \sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\| K \\
&\quad \leq K\sqrt{m}\Big(\sum_{n=1}^{m} \|\overline{\mathbf{a}}_n C_m\|^2\Big)^{1/2} = K\sqrt{m\cdot \operatorname{trace}(C_m)} \leq K\sqrt{r}\sqrt{m\|C_m\|} \to 0.
\end{aligned}$$

Corollary 2.13. *If* $\lim_{m\to\infty} m\|A_m^{-1}\| = 0$, *then for every wide sense stationary noise sequence* $\{Z_n\}$, *the LSE sequences are* L_2*-consistent.*

Table 1 below lists the results of our theorems for the case $r = 1$.

Proposition 2.14. *Let* $r = 1$ *and* $a_n \geq 0$ *for every* n, *and assume* $\sum_{n=1}^{\infty} a_n^2 = \infty$. *Then the following are equivalent*:

(i) *The LSE sequence is* L_2*-consistent for every wide sense stationary noise with mixing spectral measure.*

(ii) *The LSE sequence is* L_2*-consistent for every wide sense stationary noise with absolutely continuous spectral measure.*

(iii) $\sup_{m\geq m_0} \sum_{k=1}^{m} a_k / \sum_{k=1}^{m} a_k^2 < \infty$.

Furthermore, the LSE sequence is L_2*-consistent for every wide sense stationary noise* (*with no restriction on the spectral measure*) *if and only if*

$$\lim_{m\to\infty} \frac{\sum_{k=1}^{m} a_k}{\sum_{k=1}^{m} a_k^2} = 0.$$

Proof. By assumption and the triangle inequality, $\sup_t |\sum_{k=1}^m a_k e^{ikt}| = \sum_{k=1}^m a_k$. Now use the appropriate results from Table 1. ∎

Table 1. L_2-consistency of the LSE in the model (2.1) when $r = 1$

Conditions on $\{a_n\}$ with $\lvert a_{m_0}\rvert > 0$	Implication(s)	Consistency for *all* noises with spectral measure...
$\sum_{n=1}^{\infty} \lvert a_n\rvert^2 = \infty$	$\Longleftrightarrow$	absolutely continuous with bounded density
$\liminf_{m\to\infty} \frac{1}{m^{1/p}} \sum_{n=1}^{m} \lvert a_n\rvert^2 > 0$, $1 < p \le 2$	$\Longrightarrow$	absolutely continuous with density in $L_p(dt)$
(i) $\sum_{n=1}^{\infty} \lvert a_n\rvert^2 = \infty$ (ii) $\sup_{m\ge m_0} \left\lVert \frac{\sum_{k=1}^m a_k e^{ikt}}{\sum_{k=1}^m \lvert a_k\rvert^2} \right\rVert_{L_{2p/(p-1)}(dt)} < \infty$	$\Longleftrightarrow$	absolutely continuous with density in $L_p(dt)$ $(1 < p < \infty)$
(i) $\sum_{n=1}^{\infty} \lvert a_n\rvert^2 = \infty$ (ii) $\sup_{m\ge m_0} \sup_{t\in[-\pi,\pi)} \frac{\lvert\sum_{k=1}^m a_k e^{ikt}\rvert}{\sum_{k=1}^m \lvert a_k\rvert^2} < \infty$	$\Longleftrightarrow$	absolutely continuous
(i) $\sum_{n=1}^{\infty} \lvert a_n\rvert^2 = \infty$ (ii) $\sup_{m\ge m_0} \frac{\sum_{k=1}^m \lvert a_k\rvert}{\sum_{k=1}^m \lvert a_k\rvert^2} < \infty$	$\Longrightarrow$	mixing (i.e., $\widehat{\mu}(n)\to 0$)
$\liminf_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} \lvert a_n\rvert^2 > 0$	$\Longrightarrow$	continuous (no atoms)
(i) $\frac{\sum_{k=1}^m a_k e^{ikt}}{\sum_{k=1}^m \lvert a_k\rvert^2} \to 0$ for all $0 \ne t \in [-\pi,\pi)$ (ii) $\sup_{m\ge m_0} \sup_{t\in[-\pi,\pi)} \frac{\lvert\sum_{k=1}^m a_k e^{ikt}\rvert}{\sum_{k=1}^m \lvert a_k\rvert^2} < \infty$	$\Longleftrightarrow$	no atom at 0
(i) $\frac{\sum_{k=1}^m a_k e^{ikt}}{\sum_{k=1}^m \lvert a_k\rvert^2} \to 0$ for all $t \in [-\pi,\pi)$ (ii) $\sup_{m\ge m_0} \sup_{t\in[-\pi,\pi)} \frac{\lvert\sum_{k=1}^m a_k e^{ikt}\rvert}{\sum_{k=1}^m \lvert a_k\rvert^2} < \infty$	$\Longleftrightarrow$	any spectral measure (no restrictions)
$\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} \lvert a_n\rvert^2 = \infty$	$\Longrightarrow$	any spectral measure (no restrictions)

3. L_p-consistency of the LSE when the noise is strict sense stationary. In this section we study L_p-consistency of the LSE of the regression coefficients, in the case of non-random regressors, under the assumption that the noise is *strict sense stationary* with finite pth moment. For $1 \le p < 2$ this means that second moment is not assumed, so the results of §2 cannot be directly applied. For $p > 2$ we assume more than was assumed in §2, and expect stronger convergence results.

Recall that a sequence $\{\widehat{u}_m\}$ of estimators of a random variable u is called (*strongly*) *consistent* if $\{\widehat{u}_m\}_{m\in\mathbb{N}}$ converges to u in probability (resp. almost surely) as $m \to \infty$. If $1 \le p < \infty$ and u is in L_p (i.e., has finite absolute pth moment), the sequence of estimators is called L_p-*consistent* if each $\widehat{u}_m$ is in L_p, and

$$\lim_{m\to\infty} \mathbf{E}(|\widehat{u}_m - u|^p) = 0.$$

Clearly, L_p-consistency implies consistency, and also L_r-consistency for any $1 \le r < p$.

Proposition 2.12 shows that when the regressor system in the model (2.1) satisfies $\lim_{m\to\infty} m\|A_m^{-1}\| = 0$, we have L_p-consistency for any noise sequence with $\sup_n \|Z_n\|_p < \infty$. When we assume only $\limsup_{m\to\infty} m\|A_m^{-1}\| < \infty$, Theorem 2.7 yields the L_2-consistency of the LSE for every wide sense stationary noise sequence $\mathbf{Z}$ with continuous spectral measure. Our first result deals with L_p-consistency for $p > 2$ with the latter assumption on the regressors. We use the notation of §2, in particular $C_m = A_m^{-1}$.

THEOREM 3.1. *Let $\{\mathbf{a}^{(k)}\}_{1\le k\le r}$ be a system of regressors in the model* (2.1) *satisfying*

$$\limsup_{m\to\infty} m\|A_m^{-1}\| < \infty \tag{3.1}$$

and let $\vec{\beta}_m := (\widehat{\beta}_m^{(1)}, \dots, \widehat{\beta}_m^{(r)})$, defined by (2.2), *be the m-th LSE for the regression coefficients $\beta^{(1)}, \dots, \beta^{(r)}$. Then for every strict sense stationary noise sequence $\{Z_n\}$ with $\mathbf{E}(|Z_0|^p) < \infty$ for some $2 < p < \infty$ and with continuous spectral measure, the LSE sequence $\{\vec{\beta}_m\}$ is L_p-consistent.*

Proof. We first prove the theorem when Z_0 has a moment $p + \delta$. Since $p > 2$, the noise is also wide sense stationary, so L_2-consistency follows from Theorem 2.7, which means $\lim_{m\to\infty} \|\Delta_m^{(k)}\|_2 = 0$, where $\Delta_m^{(k)} := \widehat{\beta}_m^{(k)} - \beta^{(k)}$. By strict sense stationarity, $\|Z_n\|_{p+\delta} = \|Z_0\|_{p+\delta}$ for every n. The proof of Proposition 2.12 yields $\|\Delta_m^{(k)}\|_{p+\delta} \le \|Z_0\|_{p+\delta}\sqrt{r}\sqrt{m\|C_m\|}$, so by the assumption on the regressors $\sup_m \|\Delta_m^{(k)}\|_{p+\delta} < \infty$. Put $\gamma := 2\delta/(p + \delta - 2)$. Then $0 < \gamma < 2$, and $p + \delta = 2(p - \gamma)/(2 - \gamma)$. Hölder's inequality (for

$2/\gamma > 1$) yields

$$\begin{aligned}\mathbf{E}(|\Delta_m^{(k)}|^p) &= \mathbf{E}(|\Delta_m^{(k)}|^\gamma \cdot |\Delta_m^{(k)}|^{p-\gamma}) \\ &\le \{\mathbf{E}(|\Delta_m^{(k)}|^2)\}^{\gamma/2} \cdot \{\mathbf{E}(|\Delta_m^{(k)}|^{2(p-\gamma)/(2-\gamma)}\}^{(2-\gamma)/2} \xrightarrow[m\to\infty]{} 0,\end{aligned}$$

since the first factor converges to 0, and the second is $\|\Delta_m^{(k)}\|_{p+\delta}^{(p+\delta)(2-\gamma)/2}$, which is bounded in m.

We now prove the general case. By strict sense stationarity, $Z_n = T^n Z_0$, with T induced by a probability preserving transformation on $(\mathcal{Z}, \nu)$ and $Z_0 \in L_p$. We now define on L_p the operator

$$R_m^{(k)} := \sum_{n=1}^{m} \Big(\sum_{j=1}^{r} \overline{a_n^{(j)}} c_m^{(jk)}\Big) T^n.$$

By (2.3), for every noise sequence $\{T^n f\}$ with $f \in L_p$, the corresponding LSE estimators will be $\beta^{(k)} + R_m^{(k)} f$, and L_p-consistency is the L_p-norm convergence to 0 of $R_m^{(k)} f$ as $m \to \infty$, for $1 \le k \le r$. By the mean ergodic theorem for linear contractions in reflexive Banach spaces, $E(t)f := \lim_{m\to\infty} m^{-1} \sum_{n=1}^m e^{-int} T^n f$ exists in L_p-norm for every $f \in L_p$ and $t \in [-\pi, \pi)$. Since $p > 2$, the convergence holds also in L_2-norm, and $\{T^n f\}$ has continuous spectral measure if and only if $E(t)f = 0$ for every $t \in [-\pi, \pi)$. But $E(t)f = 0$ for every t is equivalent to f being orthogonal (in the complex L_2) to the closed subspace of L_2 generated by all the eigenvectors of T, denoted by $\mathcal{L}$. Note that all the eigenvalues are of unit modulus, since T is an isometry. Let Σ_a be the sub-σ-algebra generated by the eigenvectors of T. Then $\mathcal{L} = L_2(\Sigma_a)$ by Theorem 2.2 of Krengel (1972). Let $\mathcal{E}$ be the conditional expectation with respect to Σ_a. Then $\mathcal{E}$ and $Q := I - \mathcal{E}$ map $L_{p+\delta}$ into itself. Thus, if $f \in L_{p+\delta}$, then $Qf \in L_{p+\delta}$ and has continuous spectral measure. By the first part of the proof of the theorem, we have L_p-consistency of the LSE for the noise $\{T^n Qf\}$, which means $\lim_{m\to\infty} \|R_m^{(k)} Qf\|_p = 0$. By the remarks following Corollary 2.8, condition (3.1), which is the same as (2.12), implies (2.10), which shows that

$$\sup_{m\ge m_0} \|R_m^{(k)}\|_p \le \sup_{m\ge m_0} \sum_{n=1}^{m} \Big|\sum_{j=1}^{r} \overline{a_n^{(j)}} c_m^{(jk)}\Big| < \infty. \tag{3.2}$$

Let $\{f_l\} \subset L_{p+\delta}$ and $f_l \to Z_0$ in L_p-norm; then $\{Qf_l\} \subset L_{p+\delta}$, and $Qf_l \to Z_0$, since Z_0 is orthogonal to $\mathcal{L}$ and Q projects on $\mathcal{L}^\perp$. Hence (3.2) and $\lim_{m\to\infty} \|R_m^{(k)} Qf_l\|_p = 0$ yield $\lim_{m\to\infty} \|R_m^{(k)} Z_0\|_p = 0$, which proves the theorem. ■

Definition. Let $\mathbf{Z} := \{Z_n\}$ be a strictly stationary sequence with $\mathbf{E}(|Z_0|^p) < \infty$ for some (given) $p \ge 1$. By the ergodic theorems, for every

$t \in [-\pi, \pi)$ the limit $E(t)Z_0 := \lim_{m\to\infty} m^{-1} \sum_{n=1}^{m} e^{-int} Z_n$ exists almost surely and in L_p-norm (also for $p = 1$). The set of t for which $E(t)Z_0 \not\equiv 0$ is called the *point spectrum* of $\{Z_n\}$, and is denoted by $\sigma_{\Pi}(\mathbf{Z})$. We say that $\{Z_n\}$ *has continuous spectrum* if $\sigma_{\Pi}(\mathbf{Z}) = \emptyset$, i.e., for every $t \in [-\pi, \pi)$ we have $\lim_{m\to\infty} \|m^{-1} \sum_{n=1}^{m} e^{-int} Z_n\|_p = 0$; when $p \geq 2$ this is equivalent to $\{Z_n\}$ having a continuous spectral measure, but the definition is valid also for $p < 2$.

THEOREM 3.2. *Let $\{\mathbf{a}^{(k)}\}_{1\leq k\leq r}$ be a system of regressors in the model* (2.1) *satisfying* (3.1), *and let $\vec{\beta}_m := (\widehat{\beta}_m^{(1)}, \ldots, \widehat{\beta}_m^{(r)})$, defined by* (2.2), *be the m-th LSE for the regression coefficients $\beta^{(1)}, \ldots, \beta^{(r)}$. Then for every strict sense stationary noise sequence $\{Z_n\}$ with $\mathbf{E}(|Z_0|^p) < \infty$ for some $1 \leq p < \infty$ and with continuous spectrum, the LSE sequence $\{\vec{\beta}_m\}$ is L_p-consistent.*

Proof. For $p \geq 2$ the result is proved in the previous theorem. Fix p with $1 \leq p < 2$. By strict stationarity, $Z_n = T^n Z_0$. When Z_0 has finite second moment, the L_p-norm convergence to 0 of $m^{-1} \sum_{n=1}^{m} e^{-int} Z_n$ (which holds since we assume continuous spectrum) implies weak-L_2 convergence to 0, and hence L_2-norm convergence to 0—see the mean ergodic theorem on p. 72 of Krengel (1985). Thus, if $\{Z_n\}$ has continuous spectrum and Z_0 has finite second moment, then $\{Z_n\}$ has continuous spectral measure, and Theorem 2.7 yields the L_2-consistency, hence L_p-consistency. In the notation of the previous proof, $\|R_m^{(k)} f\|_p \to 0$ if $f \in L_2$ with continuous spectrum. For $f \in L_p$, the sequence $\{T^n f\}$ has continuous spectrum if and only if f is an L_p-*flight vector* for T, i.e., $T^{n_l} f \to 0$ L_p-weakly for some subsequence $\{n_l\}$ (Lin and Jones (1980), Corollary 4; see also Krengel (1985), §2.4). The proof of Proposition 2.6 in Lin *et al.* (1999) shows that the L_2-flight vectors for T are L_p-dense in the space of L_p-flight vectors. The L_p-consistency for the L_2 flight vectors shown above and (3.2) (which holds for any $p \geq 1$) yield the theorem. ∎

REMARKS. 1. Theorems 3.1 and 3.2 clearly apply when the regressors form an asymptotically non-degenerate A-system, i.e., $\lim_{m\to\infty} A_m/m$ exists and is invertible.

2. For $p = 2$ we can take $\{Z_n\}$ only wide sense stationary (with continuous spectral measure), by Theorem 2.7 or Corollary 2.8.

We now study the consistency of the LSE when the spectral measure of $\mathbf{Z}$ is allowed to have atoms. To obtain consistency, additional assumptions on the regressors will be needed. In order to see what are reasonable assumptions, we first look at the special case of asymptotic non-degeneracy when $r = 1$. By (1.6), the LSE error in this case is

$$\Delta_m := \widehat{\beta}_m - \beta = \frac{\frac{1}{m}\sum_{n=1}^m \overline{a}_n Z_n}{\frac{1}{m}\sum_{n=1}^m |a_n|^2}.$$

The non-degeneracy is the convergence of the denominator to a positive limit. When we consider the stationary noise $Z_n = e^{int}Z_0$, the consistency is reduced to convergence to 0 of $m^{-1}\sum_{n=1}^m \overline{a}_n e^{int}$.

Definition. A sequence $\mathbf{a} := \{a_n : n \geq 0\}$ of complex numbers is called a *Hartman sequence* if for every real $t \in [-\pi, \pi)$ the limit $c(t, \mathbf{a}) := \lim_m m^{-1}\sum_{n=1}^m a_n e^{-int}$ exists. The *spectrum* $\sigma(\mathbf{a})$ of a Hartman sequence $\mathbf{a}$ is the set of t with $c(t, \mathbf{a}) \neq 0$.

Kahane (1961) proved that if a given Hartman sequence $\mathbf{a}$ satisfies $\sup_n n^{-1}\sum_{k=1}^n |a_k| < \infty$, then its spectrum is countable (see Rosenblatt (1994) for an elementary proof when $\sup_n n^{-1}\sum_{k=1}^n |a_k|^2 < \infty$).

Thus, in our problem it is natural to assume that $\mathbf{a}$ is a Hartman sequence, and we have the consistency for the noise $\{e^{int}Z_0\}$ if and only if $t \notin \sigma(\mathbf{a})$. For modulated ergodic theorems for Hartman sequences see Appendix B.

Theorem 3.3. *Let $\{\mathbf{a}^{(k)}\}_{1\leq k\leq r}$ be a system of regressors in the model* (2.1) *satisfying* (3.1) *and*

$$\sup_{m\geq 1} \frac{1}{m}\|A_m\| < \infty \tag{3.3}$$

such that each $\mathbf{a}^{(k)}$ is a Hartman sequence, and set $\sigma(\vec{\mathbf{a}}) = \bigcup_{k=1}^r \sigma(\mathbf{a}^{(k)})$. Let the noise $\mathbf{Z}$ be a wide sense stationary random sequence, and let $\vec{\beta}_m := (\widehat{\beta}_m^{(1)}, \ldots, \widehat{\beta}_m^{(r)})$ be the m-th LSE for the regression coefficients, given by (2.2).

(i) *The sequence $\{\vec{\beta}_m\}$ is L_2-consistent if and only if*

$$\sigma_{\Pi}(\mathbf{Z}) \cap \sigma(\vec{\mathbf{a}}) = \emptyset. \tag{3.4}$$

(ii) *When $\mathbf{Z}$ is also strict sense stationary, with $\mathbf{E}(|Z_0|^p) < \infty$ for some $2 \leq p < \infty$, then $\{\vec{\beta}_m\}$ is L_p-consistent if and only if* (3.4) *holds.*

(iii) *When $\mathbf{Z}$ is also strict sense stationary, and each $\mathbf{a}^{(j)}$ is also a good modulating sequence for $\mathbf{Z}$, then $\{\vec{\beta}_m\}$ is strongly consistent if and only if* (3.4) *holds.*

Proof. With the notation $D_m := mC_m = mA_m^{-1}$, we rewrite (2.3) in the form

$$\Delta_m^{(k)} = \widehat{\beta}_m^{(k)} - \beta^{(k)} = \sum_{j=1}^r d_m^{(jk)}\left(\frac{1}{m}\sum_{n=1}^m Z_n \overline{a_n^{(j)}}\right). \tag{3.5}$$

Assumption (3.3) yields $\sup_{m\geq 1} m^{-1}\sum_{n=1}^m |a_n^{(j)}|^2 < \infty$ for each j, hence $\overline{\mathbf{a}^{(j)}} \in W_2$.

Since each $\overline{\mathbf{a}^{(j)}}$ is Hartman by assumption, we apply Theorem B.2 (from Appendix B) to $\mathbf{Z}$, and obtain the L_2-norm convergence

$$\lim_{m\to\infty}\left(\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{a_n^{(j)}}\right) = L(\overline{\mathbf{a}}^{(j)},\mathbf{Z}) = \sum_{t\in[-\pi,\pi)} c(t,\overline{\mathbf{a}}^{(j)})E(-t,\mathbf{Z}). \tag{3.6}$$

By condition (3.1),

$$\sup_{m>m_0}\max_{1\le j,k\le r}|d_m^{(jk)}| = C < \infty. \tag{3.7}$$

(i) Assume that (3.4) holds. Then the limit in (3.6) is 0, since all terms are 0. Substituting in (3.5), for $m > m_0$ we get

$$\|\varDelta_m^{(k)}\|_2 \le C\sum_{j=1}^{r}\left\|\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{a_n^{(j)}}\right\|_2 \xrightarrow[m\to\infty]{} 0. \tag{3.8}$$

Assume now that the LSE is L_2-consistent for the given $\{Z_n\}$. For $m > m_0$, (3.7) yields

$$\left\|d_m^{(jk)}L(\overline{\mathbf{a}}^{(j)},\mathbf{Z}) - d_m^{(jk)}\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{a_n^{(j)}}\right\|_2 \le C\left\|L(\overline{\mathbf{a}}^{(j)},\mathbf{Z}) - \frac{1}{m}\sum_{n=1}^{m} Z_n\overline{a_n^{(j)}}\right\|_2 \xrightarrow[m\to\infty]{} 0$$

for every $1 \le j,k \le r$. If we substitute this in (3.5), orthogonality of $\{E(-t,\mathbf{Z})\}$ and L_2-consistency yield

$$\sum_{-t\in\sigma_\Pi(\mathbf{Z})}\left|\sum_{j=1}^{r} d_m^{(jk)}c(t,\overline{\mathbf{a}}^{(j)})\right|^2 = \left\|\sum_{j=1}^{r} d_m^{(jk)}L(\overline{\mathbf{a}}^{(j)},\mathbf{Z})\right\|_2^2 \to 0.$$

Thus, for fixed t with $E(-t,\mathbf{Z})\ne 0$, the vector $\vec{c}(t) := (c(t,\overline{\mathbf{a}}^{(1)}),\ldots,c(t,\overline{\mathbf{a}}^{(r)}))$ satisfies $\lim_{m\to\infty}\|\vec{c}(t)D_m\| = 0$, so $\|\vec{c}(t)\| \le \|\vec{c}(t)D_m\|\cdot\|m^{-1}A_m\| \to 0$, using (3.3). Hence $c(-t,\mathbf{a}^{(j)}) = 0$ when $E(-t,\mathbf{Z}) \ne 0$, which proves (3.4).

(ii) Since $p \ge 2$, (3.4) yields L_2-consistency, by (i). (3.4) and Theorem B.3 yield L_p-norm convergence to 0 in (3.6). The proof of L_p-consistency is obtained by taking L_p-norms in (3.8).

Since L_p-consistency implies L_2-consistency, (3.4) is necessary by part (i).

(iii) When $\mathbf{a}^{(j)}$ is a good modulating sequence, we have a.s. convergence in (3.6), in addition to L_2-convergence. Thus, if (3.4) holds, then the limit in (3.6) is 0 (since it equals the L_2 limit, shown above to be 0). The strong consistency follows from (3.5) and (3.7), since

$$|\varDelta_m^{(k)}| \le C\sum_{j=1}^{r}\left|\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{a_n^{(j)}}\right| \xrightarrow[m\to\infty]{} 0.$$

Conversely, if we have strong consistency, (3.5) yields

$$\lim_{m\to\infty}\Big(\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{\mathbf{a}}_n\Big)D_m = 0 \qquad \text{a.s.},$$

where $\mathbf{a}_n = (a_n^{(1)}, \ldots, a_n^{(r)})$, as in §2. Hence (3.3) yields

$$\Big\|\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{\mathbf{a}}_n\Big\| \le \Big\|\Big(\frac{1}{m}\sum_{n=1}^{m} Z_n\overline{\mathbf{a}}_n\Big)D_m\Big\|\cdot\Big\|\frac{1}{m}A_m\Big\| \to 0.$$

We conclude that the L_2 limit in (3.6), which exists, is 0. Therefore (3.8) holds and we have L_2-consistency, which implies (3.4), by (i). ∎

Remarks. 1. An asymptotically non-degenerate system satisfies all the assumptions of the theorem (except the additional assumption in (iii)).

2. Theorem 3.3 generalizes a result in Tempelman (1975), where each $\{\mathbf{a}^{(k)}\}$ is 2-Besicovitch.

3. Theorem 3.1 is *not* a special case of Theorem 3.3, since for $\sigma_{\Pi}(\mathbf{Z}) = \emptyset$ Theorem 3.1 yields the consistency without assuming (3.3) and without requiring the sequences $\{\mathbf{a}^{(j)}\}$ to be Hartman.

Examples. 1. Let $\eta = 1$ or -1, both with probability 1/2. Let $r = 1$, and for every n put $a_n \equiv 1$, $Z_n \equiv \eta$; then $\sigma(\mathbf{a}) = \sigma_{\Pi}(\mathbf{Z}) = \{0\}$, and the LSE $\beta_m = \eta$ is not consistent.

2. Let $0 < \omega < \pi$, and $r = 1$. For $a_n = \cos\omega n$ and $Z_n \equiv \eta$, the LSE β_m is consistent. The same is true when $a_n \equiv 1$ and $Z_n = \eta\cos\omega n$ (with the above η).

3. In the case of periodically missing observations, for example, $a_n = 1$ for $n = kl$, $l = 1, 2, \ldots$, and $a_n = 0$ otherwise, let $Z_n = \eta\cos\frac{2\pi}{k}n$; then the LSE is not consistent.

Application. *Amplitude estimation in sums of harmonic signals with known frequencies.*

We saw in §2 that when sampling the sum of r harmonic signals with known frequencies and unknown phases and amplitudes, the regressors $a_n^{(j)} = (e^{i\omega_j})^n$ satisfy $m^{-1}A_m \to I_r$, so (3.1) and (3.3) are satisfied. Each $\mathbf{a}^{(j)}$ is a trigonometric polynomial, hence bounded Besicovitch (see Appendix A), and therefore Hartman and L_1-universally good (Theorem B.4 in Appendix B). Clearly $\sigma(\vec{\mathbf{a}}) = \{\omega_1, \ldots, \omega_r\}$, so for any wide sense stationary noise satisfying (3.4), i.e., with no resonance with the signals, the amplitude estimates satisfy $\lim_{m\to\infty}\|\widehat{V}_{j,m} - V_j\| = 0$ by Theorem 3.3(i); if the noise is also strict sense stationary, Theorem 3.3(iii) yields $\lim_{m\to\infty}\widehat{V}_{j,m} = V_j$ a.e.

4. L_p and strong consistency of the LSE when the regressors are random. In this section we consider the multiple linear regression model (1.1), with observable random regressors $\mathbf{a}^{(j)} + \mathbf{X}^{(j)}$, $1 \le j \le r$, unknown constant regression coefficients $\beta^{(j)}$, $1 \le j \le r$, and observable response $\mathbf{Y}$. The values of $\mathbf{a}^{(j)} + \mathbf{X}^{(j)}$ are observed, but the value of the pure signal $\mathbf{V} := \sum_{j=1}^{r} \beta^{(j)}(\mathbf{a}^{(j)} + \mathbf{X}^{(j)})$ is observed with an additive random noise $\mathbf{Z}$ which is assumed to be *independent* of the regressors, so actually $\mathbf{Y} = \mathbf{V} + \mathbf{Z}$ is observed. We look at the regression model on the product space $(\Omega, P) = (\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$, where the regressors are defined on a probability space $(\mathcal{X}, \mu)$, and $(\mathcal{Z}, \nu)$ is the model for the noise $\mathbf{Z}$. For each realization $\vec{\mathbf{x}} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(r)})$ of the regressor random component we have a random measurement $\mathbf{Y}(\vec{\mathbf{x}}) = \mathbf{v} + \mathbf{Z}$, where $\mathbf{v} = \sum_{j=1}^{r} \beta^{(j)}(\mathbf{a}^{(j)} + \mathbf{x}^{(j)})$, and the regression coefficients LSE $\vec{\beta}_m(\vec{\mathbf{x}}) := (\widehat{\beta}_m^{(1)}(\vec{\mathbf{x}}), \dots, \widehat{\beta}_m^{(r)}(\vec{\mathbf{x}}))$ (which of course depend also on $\mathbf{Z}$) are computed by (2.2), if possible, for the system $\{\mathbf{b}^{(j)}(\vec{\mathbf{x}}) : 1 \le j \le r\}$, where $\mathbf{b}^{(j)}(\vec{\mathbf{x}}) = \mathbf{a}^{(j)} + \mathbf{x}^{(j)}$ for $1 \le j \le r$. We use the notation $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ to show the dependence on the noise $\mathbf{Z}$.

In Theorem 4.1 below we study the consistency of the LSE for *fixed* realizations $\mathbf{a}^{(j)} + \mathbf{x}^{(j)}$ $(1 \le j \le r)$ of the random regressors; we provide conditions for consistency of the LSE $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ as random variables on $(\mathcal{Z}, \nu)$. Such a situation may arise, for example, in the case where the same random "message" is transmitted to various correspondents through the same type of channel; the same realizations $\mathbf{b}^{(j)}(\vec{\mathbf{x}})$ are transmitted repeatedly, and each time the realization of the noise does not depend on the regressors. In the study of strong consistency, another point of view, considered in Corollary 4.2, is also natural—to consider the LSE $\vec{\beta}_m(\vec{\mathbf{X}}, \mathbf{Z})$ over the space $(\Omega, P) = (\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$; this is the situation in which each measurement selects both a random value of $\vec{\mathbf{X}}$ and a random value of $\mathbf{Z}$.

THEOREM 4.1. *Let* $\{\mathbf{a}^{(k)}\}_{1 \le k \le r}$ *be an asymptotically non-degenerate A-system of* L_2*-universally good sequences. Let* $\vec{\mathbf{X}} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(r)})$ *be a* $\mathbb{C}^r$*-valued ergodic strict sense stationary sequence defined on* $(\mathcal{X}, \mu)$, *with* $\mathbf{E}(|X_0^{(j)}|^2) < \infty$, $1 \le j \le r$, *and set* $\sigma_{\Pi}(\vec{\mathbf{X}}) := \bigcup_{j=1}^{r} \sigma_{\Pi}(\mathbf{X}^{(j)})$. *If*

$$\sigma(\vec{\mathbf{a}}) \cap \sigma_{\Pi}(\vec{\mathbf{X}}) = \emptyset \tag{4.1}$$

then there exists a set of realizations $\vec{\mathbf{x}} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(r)})$ *of* $(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(r)})$, *which has* μ*-probability* 1, *with the following properties*:

(i) *For every random noise sequence* $\mathbf{Z}$ *which is defined on* $(\mathcal{Z}, \nu)$ *independently of* $\vec{\mathbf{X}}$, *the regression coefficients LSE* $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ *are well defined* (*for large enough* m).

(ii) *For every wide sense stationary random sequence* $\mathbf{Z}$, *which is defined on* $(\mathcal{Z}, \nu)$ *independently of* $\vec{\mathbf{X}}$ *and satisfies*

(4.2) $$\sigma_{\Pi}(\mathbf{Z}) \cap [\sigma(\vec{\mathbf{a}}) \cup \sigma_{\Pi}(\vec{\mathbf{X}})] = \emptyset$$

the LSE sequence $\{\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})\}$ *is* $L_2(\nu)$*-consistent.*

(iii) *If*, *in addition to the conditions in* (ii), $\mathbf{Z}$ *is also strict sense stationary, then the sequence* $\{\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})\}$ *is strongly consistent on* $(\mathcal{Z}, \nu)$.

(iv) *If in addition to the conditions in* (iii) *we have* $\mathbf{E}(|Z_0|^p) < \infty$ *for a given* $p > 2$, *then the LSE sequence* $\{\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})\}$ *is also* $L_p(\nu)$*-consistent.*

Proof. (i) The assumed convergence of $m^{-1}A_m$, say to G, implies that each $\mathbf{a}^{(j)}$ is in W_2.

By (B.2) we have $\sigma(\mathbf{x}^{(j)}) = \sigma_{\Pi}(\mathbf{X}^{(j)})$ a.s. For a realization $\vec{\mathbf{x}}$, its mth "normalized" Gram matrix is $[f_m^{(ij)}(\vec{\mathbf{x}})]$ with $f_m^{(ij)}(\vec{\mathbf{x}}) = m^{-1}\sum_{n=1}^m x_n^{(i)}\overline{x}_n^{(j)}$. As in Appendix A, we define $\varphi_i = X_0^{(i)}$. By (A.1), $\lim_m f_m^{(ij)}(\vec{\mathbf{x}}) = \int \varphi_i\overline{\varphi}_j\,d\mu$ a.s.

Since each $\mathbf{a}^{(j)}$ is L_2-universally good, we have both $L_2(\mu)$ and a.s. convergence of $\langle \mathbf{a}^{(i)}, \mathbf{X}^{(j)}\rangle_m$ as $m \to \infty$, and the limit is identified by (B.1) with $\mathbf{Z}$ replaced by $\overline{\mathbf{X}}^{(j)}$. Assumption (4.1) implies that the limit is 0 a.s.

We now fix a realization $\vec{\mathbf{x}}$ with all the above properties, and define the sequences $\mathbf{b}^{(j)}(\vec{\mathbf{x}}) = \mathbf{a}^{(j)} + \mathbf{x}^{(j)}$ for $1 \le j \le r$. The mth "normalized" Gram matrix $H_m(\vec{\mathbf{x}}) := [h_m^{(ij)}(\vec{\mathbf{x}})]$ of the regressor system $\{\mathbf{b}^{(j)}(\vec{\mathbf{x}}) : 1 \le j \le r\}$ is given by

(4.3) $$h_m^{(ij)}(\vec{\mathbf{x}}) = \langle \mathbf{a}^{(i)}, \mathbf{a}^{(j)}\rangle_m + \langle \mathbf{a}^{(i)}, \mathbf{x}^{(j)}\rangle_m + \langle \mathbf{x}^{(i)}, \mathbf{a}^{(j)}\rangle_m + f_m^{(ij)}(\vec{\mathbf{x}}).$$

Now the asymptotic non-degeneracy of the system $\{\mathbf{a}^{(k)} : 1 \le k \le r\}$ yields

(4.4) $$\lim_{m\to\infty} h_m^{(ij)}(\vec{\mathbf{x}}) = g^{(ij)} + \int \varphi_i\overline{\varphi}_j\,d\mu =: h^{(ij)},$$

with $G = [g^{(ij)}]$ invertible. The matrix $F := [\int \varphi_i\overline{\varphi}_j\,d\mu]$ is clearly non-negative definite. The limit matrix $H = [h^{(ij)}]$ is positive definite, since $H = G + F \ge G$ and G is positive definite. Hence the system $\{\mathbf{b}^{(j)}(\vec{\mathbf{x}}) : 1 \le j \le r\}$ is asymptotically non-degenerate, so the LSE $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ is well defined for any noise sequence $\mathbf{Z}$ independent of $\mathbf{X}$.

(ii) For a.e. ξ, the sequences $\mathbf{x}^{(j)} := \mathbf{X}^{(j)}(\xi)$ are Hartman by the Wiener–Wintner Theorem, and belong to W_2 (by the pointwise ergodic theorem). By (B.2), we have $\sigma(\mathbf{x}^{(j)}) = \sigma_{\Pi}(\mathbf{X}^{(j)})$ a.s. Hence, for almost every realization $\vec{\mathbf{x}}$, each $\mathbf{b}^{(j)}(\vec{\mathbf{x}})$ is Hartman and in W_1, with

$$\sigma(\mathbf{b}^{(j)}(\vec{\mathbf{x}})) \subset \sigma(\mathbf{a}^{(j)}) \cup \sigma(\mathbf{x}^{(j)}) = \sigma(\mathbf{a}^{(j)}) \cup \sigma_{\Pi}(\mathbf{X}^{(j)}).$$

This and (4.2) show that condition (3.4) is satisfied by the asymtotically non-degenerate system $\{\mathbf{b}^{(j)}(\vec{\mathbf{x}}) : 1 \le j \le r\}$, to which we now apply Theorem 3.3(i).

(iii) By Theorem B.5, for a.e. ξ the sequences $\mathbf{x}^{(j)}$ are in fact L_2-universally good modulating sequences, and the result follows from applying Theorem 3.3(iii).

(iv) follows from Theorem 3.3(ii). ■

REMARKS. 1. For parts (i) and (ii) of Theorem 4.1, the proof needs only that each $\mathbf{a}^{(j)}$ be Hartman and a good modulating sequence for each $\mathbf{X}^{(i)}$, rather than L_2-universally good. With this modified statement, Theorem 3.3(i) becomes a special case of Theorem 4.1(ii) (with $\mathbf{X}^{(j)} = \mathbf{0}$ for each j).

2. Condition (4.1) is satisfied if each $\mathbf{X}^{(j)}$ is centered and the probability preserving transformation θ is weakly mixing (so $\sigma_\Pi(\mathbf{X}^{(j)}) = \emptyset$ for each j). In this case, also the condition $\sigma_\Pi(\mathbf{Z}) \cap \sigma_\Pi(\vec{\mathbf{X}}) = \emptyset$ is satisfied. If also $\sigma_\Pi(\mathbf{Z}) = \emptyset$ (in particular if $\{Z_n\}$ are independent), all the spectral assumptions are satisfied.

COROLLARY 4.2. *Under the assumptions of Theorem* 4.1 *on* $\{\mathbf{a}^{(k)}\}_{1\le k\le r}$ *and on* $\vec{\mathbf{X}}$, *for every strict sense stationary noise* $\mathbf{Z}$ *with finite variance defined on* $(\mathcal{Z}, \nu)$ *satisfying* (4.2), *the LSE sequence* $\{\vec{\beta}_m(\vec{\mathbf{X}}, \mathbf{Z})\}$ *is strongly consistent on* $(\Omega, P) = (\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$.

Proof. Apply Fubini's theorem to part (iii) of Theorem 4.1. ■

REMARK. Corollary 4.2 can be proved in a straightforward way. Unlike the proof of part (iii) of Theorem 4.1, we do not need to appeal to Theorem B.5 when we consider only one fixed noise sequence $\mathbf{Z}$: the individual ergodic theorem, applied to the Cartesian product of the transformations defining $\mathbf{Z}$ and the $\mathbf{X}^{(j)}$, implies that for a.e. realization $\vec{\mathbf{x}}$ the sequences $\mathbf{x}^{(j)}$ are good modulating sequences for $\mathbf{Z}$, which is what is needed for the proof. However, the return times theorem used in the proof of Theorem 4.1 gives more—it yields a "universal" set of realizations $\vec{\mathbf{x}}$ of full μ-probability, which can be chosen *in advance*, and will give strongly consistent LSE for every possible observation $\mathbf{Y}(\vec{\mathbf{x}}, \mathbf{Z})$ (i.e., for *every possible noise* $\mathbf{Z}$ which satisfies the assumptions of the theorem).

THEOREM 4.3. *Let* $\{\mathbf{a}^{(k)}\}_{1\le k\le r}$ *be an A-system of* L_2*-universally good sequences. Let* $\vec{\mathbf{X}} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(r)})$ *be a* $\mathbb{C}^r$*-valued ergodic strict sense stationary sequence defined on* $(\mathcal{X}, \mu)$, *with* $\mathbf{E}(|X_1^{(j)}|^2) < \infty$, $1 \le j \le r$, *such that* $X_1^{(1)}, \ldots, X_1^{(r)}$ *are linearly independent. If* (4.1) *holds*, *then there exists a set of realizations* $\vec{\mathbf{x}} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)})$ *of* $(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(r)})$, *which has probability* 1, *such that statements* (i)—(iv) *of Theorem* 4.1 *are true.*

Proof. The proof proceeds as that of Theorem 4.1, up to (4.3). By assumption, the convergence (4.4) holds, but the limit matrix G is no longer assumed to be invertible. However, as shown in the Proposition of Appendix A, under the linear independence assumption of the theorem the matrix $F = [\int \varphi_i \overline{\varphi}_j \, d\mu]_{1\le i,j\le r}$ is invertible, which yields the required invertibility of the limit matrix $H = G + F \ge F$. The remainder of the proof is the same as before. ■

REMARKS. 1. Theorem 4.3 applies when $\mathbf{a}^{(j)} = \mathbf{0}$ for every j. This corresponds to a model of a random pure signal $\mathbf{V} = \sum_{j=1}^{r} \beta^{(j)} \mathbf{X}^{(j)}$ with stationary ergodic linearly independent regressors, observed with an additive stationary noise $\mathbf{Z}$ independent of the regressors.

COROLLARY 4.4. *Under the assumptions of Theorem* 4.3 *on* $\{\mathbf{a}^{(k)}\}_{1\le k\le r}$ *and on* $\vec{\mathbf{X}}$, *for every strict sense stationary noise* $\mathbf{Z}$ *with finite variance defined on* $(\mathcal{Z}, \nu)$ *satisfying* (4.2), *the LSE sequence* $\{\vec{\beta}_m(\vec{\mathbf{X}}, \mathbf{Z})\}$ *is strongly consistent on* $(\Omega, P) = (\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$.

We now consider the following model:

$$Y_n = \sum_{j=1}^{r} \beta^{(j)} e^{i\omega_j n} X_n^{(j)} + Z_n. \tag{4.5}$$

THEOREM 4.5. *Let* $\vec{\mathbf{X}} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(r)})$ *be a* $\mathbb{C}^r$*-valued ergodic strict sense stationary sequence defined on* $(\mathcal{X}, \mu)$ *by a* weakly mixing *probability preserving transformation* θ, *with* $0 < \mathbf{E}(|X_1^{(j)}|^2) < \infty$ *for* $1 \le j \le r$. *Then there exists a set of realizations* $\vec{\mathbf{x}} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)})$ *of* $(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(r)})$, *which has probability* 1, *such that for any different real numbers* $\omega_1, \ldots, \omega_r \in [-\pi, \pi)$ *and for every random noise sequence* $\mathbf{Z}$, *defined on* $(\mathcal{Z}, \nu)$ *independently of* $\vec{\mathbf{X}}$, *the regression coefficients LSE* $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ *for the model* (4.5) *are well defined* (*for large enough* m); *if* $\{Z_n\}$ *is strict sense stationary with finite variance such that*

$$\sigma_{\Pi}(\mathbf{Z}) \cap \{\omega_1, \ldots, \omega_r\} = \emptyset, \tag{4.6}$$

then the LSE sequence $\{\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})\}$ *is* L_2 *and strongly consistent on* $(\mathcal{Z}, \nu)$.

Proof. For a.e. $\xi \in \mathcal{X}$, all the sequences $\{e^{int} X_1^{(j)}(\theta^{n-1}\xi)\}_{n\ge1}$, for $1 \le j \le r$ and $t \in [-\pi, \pi)$, are L_2-universally good, by Proposition B.6 (in Appendix B). By the Wiener–Wintner theorem, for a.e. $\xi \in \mathcal{X}$ we have

$$\begin{aligned} &\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} X_1^{(j)}(\theta^{n-1}\xi) e^{-int} = 0, \\ &\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} (X_1^{(k)} \overline{X_1^{(j)}})(\theta^{n-1}\xi) e^{-int} = 0 \end{aligned} \tag{4.7}$$

for every $0 \ne t \in [-\pi, \pi)$, since θ is weakly mixing. We fix a ξ with all the above properties, which defines the realization $\vec{\mathbf{x}}$, and put $x_n^{(j)} := X_n^{(j)}(\xi)$.

We now define $a_n^{(j)} := e^{i\omega_j n} x_n^{(j)}$. The (k,j)th term of the Gram matrix A_m is

$$[A_m]_{k,j} = \sum_{n=1}^{m} a_n^{(k)} \overline{a_n^{(j)}} = \sum_{n=1}^{m} e^{i(\omega_k - \omega_j)n} \big(X_1^{(k)} \overline{X_1^{(j)}}\big)(\theta^{n-1}\xi).$$

By (4.7), $\lim_{m\to\infty} m^{-1}[A_m]_{k,j} = 0$ for $k \neq j$. Since $\lim_m m^{-1}[A_m]_{j,j} = \mathbf{E}(|X_1^{(j)}|^2) > 0$, we see that $m^{-1}A_m$ converges to an invertible diagonal matrix. Hence the LSE $\vec{\beta}_m(\vec{\mathbf{x}}, \mathbf{Z})$ is well defined (for large enough m).

The above convergence of $m^{-1}A_m$ implies (3.1) and (3.3). By definition each $\mathbf{a}^{(j)} := \{a_n^{(j)}\}$ is L_2-universally good (hence Hartman). Now we have $\sigma(\mathbf{a}^{(j)}) \subset \{\omega_j\}$, since for $t \neq \omega_j$, (4.7) yields

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} a_n^{(j)} e^{-int} = \lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} e^{i\omega_j n} x_n^{(j)} e^{-int} = 0.$$

Hence (3.4) is satisfied by any strict sense stationary $\mathbf{Z}$ with finite variance satisfying (4.6), and the strong consistency of $\vec{\beta}_m(\mathbf{x}, \mathbf{Z})$, i.e., convergence a.e. on $(\mathcal{Z}, \nu)$ to $(\beta^{(1)}, \ldots, \beta^{(r)})$, follows from Theorem 3.3(iii); the L_2-consistency follows from Theorem 3.3(i). ∎

REMARKS. 1. When $\mathbf{E}(X_1^{(j)}) = 0$, we have $\sigma(\mathbf{a}^{(j)}) = \emptyset$. If this holds for every $1 \le j \le r$, then condition (4.6) is not needed.

2. When we deal with known fixed values $\omega_1, \ldots, \omega_r$, the convergence of $m^{-1}A_m$ to a diagonal matrix requires only that for $k \neq j$, $e^{i(\omega_k - \omega_j)}$ is not an eigenvalue of θ. If θ is not weakly mixing, then $\sigma(\mathbf{a}^{(j)}) = \sigma_{\Pi}(\mathbf{X}^{(j)}) \dotplus \omega_j := \{t + \omega_j \bmod (-\pi, \pi) : t \in \sigma_{\Pi}(\mathbf{X}^{(j)})\}$. To obtain the result of the theorem, (4.6) should be replaced by

$$\sigma_{\Pi}(\mathbf{Z}) \cap \bigcup_{j=1}^{r} [\sigma_{\Pi}(\mathbf{X}^{(j)}) \dotplus \omega_j] = \emptyset.$$

We now apply Theorem 4.5 to amplitude estimation in a sum of harmonic signals with known frequencies and *observable* (in practice, accurately estimable) stationary random phases. The jth signal is generated with known frequency ω_j, unknown constant amplitude V_j, and random phase $\Phi^{(j)}(t) \in [-\pi, \pi)$. We write $\Phi_n^{(j)} := \Phi^{(j)}(n)$, so the observed signal at time n is

$$Y_n = \sum_{j=1}^{r} V_j e^{i(\omega_j n + \Phi_n^{(j)})} + Z_n. \tag{4.8}$$

We assume that the phase vector sequence $\{\vec{\Phi}_n = (\Phi_n^{(1)}, \ldots, \Phi_n^{(r)})\}$ is strict sense stationary.

Proposition 4.6. *In the model* (4.8) *assume*:

(1) *The vector sequence* $\vec{\mathbf{\Phi}} := \{\vec{\Phi}_n\}$ *is strict sense stationary, generated by a weakly mixing probability preserving transformation* θ *on* $(\mathcal{X}, \mu)$.
(2) *The sequence* $\mathbf{Z} = \{Z_n\}$ *is defined on* $(\mathcal{Z}, \nu)$ *independently of* $\vec{\mathbf{\Phi}} = \{\vec{\Phi}_n\}$, *and is strict sense stationary with finite variance.*
(3) $\sigma_\Pi(\mathbf{Z}) \cap \{\omega_1, \ldots, \omega_r\} = \emptyset$.

Then $\vec{\beta}_m(\vec{\mathbf{\Phi}}, \mathbf{Z})$, *the LSE for* $(V_1, \ldots, V_r)$ *with the random regressors* $\widetilde{X}_n^{(j)} := e^{i(\omega_j n + \Phi_n^{(j)})}$ $(1 \le j \le r)$, *is strongly consistent on* $(\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$.

Proof. We define $X_n^{(j)} = e^{i\Phi_n^{(j)}}$ on $(\mathcal{X}, \mu)$, so (4.8) becomes (4.5), and $\vec{\beta}_m(\vec{\mathbf{\Phi}}, \mathbf{Z}) = \vec{\beta}_m(\vec{\mathbf{X}}, \mathbf{Z})$. Theorem 4.5 now applies, and Fubini's theorem yields the strong consistency on $(\mathcal{X} \times \mathcal{Z}, \mu \times \nu)$. ∎

Remarks. 1. Assumption (1) holds when each sequence $\mathbf{\Phi}^{(j)} := \{\Phi_n^{(j)}\}$ is strict sense stationary, generated by a weakly mixing probability preserving transformation θ_j on $(\mathcal{X}_j, \mu_j)$, and the r sequences $\mathbf{\Phi}^{(1)}, \ldots, \mathbf{\Phi}^{(r)}$ are stochastically independent. Put $\mathcal{X} := \prod \mathcal{X}_j$ and $\mu := \mu_1 \times \ldots \times \mu_r$. On $(\mathcal{X}, \mu)$ define $\theta := \theta_1 \times \ldots \times \theta_r$, which is weakly mixing as a product of weakly mixing transformations (e.g., Krengel (1985), p. 98). For $\vec{\xi} := (\xi_1, \ldots, \xi_r) \in \mathcal{X}$ we put $\widetilde{\Phi}_n^{(j)}(\vec{\xi}) = \Phi_n^{(j)}(\xi_j)$, so $\widetilde{\Phi}_n^{(j)} = \widetilde{\Phi}_1^{(j)} \circ \theta^{n-1}$. By independence, $\vec{\widetilde{\mathbf{\Phi}}}$ is a model for $\vec{\mathbf{\Phi}}$, which is used in the proposition.

2. Condition (3) means that there is no resonance between the noise and the signals. If the noise $\{Z_n\}$ has continuous spectral measure, then condition (3) holds.

3. Under the commonly used assumption that $\Phi_1^{(j)} = \varphi_j + \Psi_1^{(j)}$ with $\Psi_1^{(j)}$ uniformly distributed in $[-\pi, \pi)$, (4.8) becomes

$$Y_n = \sum_{j=1}^{r} V_j e^{i\varphi_j} e^{i(\omega_j n + \Psi_n^{(j)})} + Z_n. \tag{4.9}$$

We now estimate $V_j e^{i\varphi_j}$ by the LSE $\widehat{\beta}_m^{(j)}(\vec{\mathbf{\Psi}}, \mathbf{Z})$; since $\mathbf{E}(e^{i\Psi_1^{(j)}}) = 0$, condition (3) is not needed anymore for consistency (see Remark 1 following Theorem 4.5). The amplitude V_j is then estimated by $|\widehat{\beta}_m^{(j)}(\vec{\mathbf{\Psi}}, \mathbf{Z})|$.

Appendix A. Averageable families of sequences. The convergence of $m^{-1}A_m$, discussed in this paper, means that $m^{-1}\sum_{n=1}^m a_n^{(i)}\overline{a}_n^{(j)}$ converges as $m \to \infty$, for each pair $1 \le i, j \le r$. This is the property we look at in this Appendix.

For any two infinite sequences $\mathbf{a}, \mathbf{b}$ of complex numbers and any integer $m > 0$ we define $\langle \mathbf{a}, \mathbf{b} \rangle_m = m^{-1}\sum_{n=1}^m a_n \overline{b}_n$ (which is equivalent to the $\mathbb{C}^m$ inner product on the first m coordinates), with the norm $\|\mathbf{a}\|_m =$

$[\langle \mathbf{a}, \mathbf{a} \rangle_m]^{1/2}$. For a finite system of sequences $\{\mathbf{a}^{(i)}, 1 \le i \le r\}$ let $g_m^{(ij)} = \langle \mathbf{a}^{(i)}, \mathbf{a}^{(j)} \rangle_m$; the $r \times r$ matrix $G_m = [g_m^{(ij)}]$ is the "normalized" Gram matrix of the first m coordinates of the system.

For $1 \le p < \infty$ the set of infinite sequences $\mathbf{a} = \{a_k\}$ satisfying $\sup_m m^{-1} \sum_{k=1}^m |a_k|^p < \infty$ will be denoted by W_p, and for $\mathbf{a} \in W_p$ we define $\|\mathbf{a}\|_{W_p} := (\limsup_n n^{-1} \sum_{k=1}^n |a_k|^p)^{1/p}$. Then $\| \cdot \|_{W_p}$ is a seminorm on W_p; we say that two sequences $\mathbf{a}$ and $\mathbf{b}$ in W_p are *equivalent* if $\|\mathbf{a} - \mathbf{b}\|_{W_p} = 0$. The set of equivalence classes in W_p is a Banach space with the norm $\| \cdot \|_{W_p}$ (for the completeness, see the discussion in Lin *et al.* (1999)). It is easy to check that $W_{p_1} \subset W_{p_2}$ if $p_1 > p_2$; in particular, we have $W_p \subset W_1$ for $p > 1$. In what follows we write $\| \cdot \|$ instead of $\| \cdot \|_{W_2}$.

DEFINITION. A family of sequences $\mathcal{A}$ will be called an *averageable family* (briefly, *A-family*) if for each pair $\mathbf{a}, \mathbf{b} \in \mathcal{A}$ the following inner product is well defined:

$$\langle \mathbf{a}, \mathbf{b} \rangle := \lim_{m \to \infty} \langle \mathbf{a}, \mathbf{b} \rangle_m.$$

It is clear that each A-family is a subset of W_2, and $\|\mathbf{a}\| := \|\mathbf{a}\|_{W_2} = [\langle \mathbf{a}, \mathbf{a} \rangle]^{1/2}$.

If some sequences form an A-family, then the union of their W_2-equivalence classes is also an A-family. It is easy to see that the W_2-closed linear span of an A-family is an A-family, too, and its equivalence classes form a Hilbert space with respect to the inner product $\langle \mathbf{a}, \mathbf{b} \rangle$.

A finite A-family will be called an *A-system*. For an A-system $\{\mathbf{a}^{(i)}, 1 \le i \le r\}$ we can define its *asymptotic Gram matrix* $G = \lim_{m \to \infty} G_m$; it is a non-negative definite matrix.

DEFINITION. An A-system of sequences $\{\mathbf{a}^{(i)}, 1 \le i \le r\}$ is said to be *asymptotically non-degenerate* if its asymptotic Gram matrix is non-singular.

Evidently, asymptotic non-degeneracy means that the equivalence classes of the sequences $\mathbf{a}^{(i)}, 1 \le i \le r$, form a linearly independent system in W_2. In particular, none of these sequences belongs to the null-class.

Below we consider two large A-families of sequences, important in various applications.

1. *Besicovitch almost periodic sequences.* Sequences $\{a_k\}$ with coordinates of the form $a_k = \sum_n \alpha_n e^{ikt_n}$ with $t_n \in [-\pi, \pi)$ (a finite sum), called trigonometric polynomials, are the simplest *almost periodic sequences*. The well known *uniform almost periodic sequences* are uniform limits of trigonometric polynomials. The *p-Besicovitch almost periodic sequences* (where $1 \le p < \infty$) are the limits of trigonometric polynomials in the W_p-seminorm. A bounded 1-Besicovitch almost periodic sequence is p-Besicovitch for any p.

It is clear that any uniform almost periodic sequence is a bounded Besicovitch almost periodic sequence. The subspace B_2 of all 2-Besicovitch almost periodic sequences is an A-family.

2. *Realizations of strict sense stationary random sequences.* Let $\vec{X}_n = (X_n^{(1)}, \dots, X_n^{(r)})$, $n = 0, 1, 2, \dots$, be an ergodic vector-valued strict sense stationary random sequence over a Lebesgue probability space $(\mathcal{X}, \mu)$. We assume that the coordinates of this sequence have finite second moments. Put $\varphi_i(\xi) = X_0^{(i)}(\xi)$ and $\psi_{i,j}(\xi) = \varphi_i(\xi)\overline{\varphi}_j(\xi)$, $1 \le i, j \le r$. It is clear that $\psi_{i,j} \in L_1(\mu)$. The assumptions of ergodicity and strict stationarity of $\vec{\mathbf{X}}$ imply (e.g., Doob (1953)) that there is an ergodic probability preserving θ such that for each j $(1 \le j \le r)$ we have $X_n^{(j)}(\xi) = \varphi_j(\theta^n \xi)$ for $n \ge 1$. Then $X_n^{(i)} \overline{X}_n^{(j)} = \psi_{ij}(\theta^n \xi))$. Therefore Birkhoff's ergodic theorem yields

$$\text{(A.1)} \qquad \langle \mathbf{X}^{(i)}, \mathbf{X}^{(j)} \rangle_m = \frac{1}{m} \sum_{n=1}^{m} \psi_{ij}(\theta^n \xi) \xrightarrow[m\to\infty]{\text{a.s.}} \int \varphi_i \varphi_j \, d\mu = \mathbf{E}(X_0^{(i)} \overline{X}_0^{(j)}) =: f^{(ij)}.$$

Thus for μ-almost all realizations $\vec{\mathbf{x}}$ of the random sequence $\vec{\mathbf{X}}$ we find that $\{\vec{\mathbf{x}}^{(1)}, \dots, \vec{\mathbf{x}}^{(r)}\}$ is an A-system and $G = [f^{(ij)}]$ is its asymptotic Gram matrix. Clearly the "asymptotic non-degeneracy" condition can be easily checked; it means that *the random variables* $X_0^{(i)}, 1 \le i \le r$, *are linearly independent in* $L_2(\mu)$ (see Davis (1975), p. 178).

Proposition. *In the model* (1.1) *let* $a_n^{(j)} = 0$ *for every* $n > 0$ *and* $1 \le j \le r$, *and let the random regressor sequences* $\{X_n^{(1)}\}, \dots, \{X_n^{(r)}\}$ *be ergodic strict sense stationary on a Lebesgue space, with finite second moment. If* $X_0^{(1)}, \dots, X_0^{(r)}$ *are linearly independent in* L_2, *then for almost every realization the regressor system is asymptotically non-degenerate.*

Appendix B. Modulated ergodic theorems for stationary sequences. General modulated ergodic theorems have recently been studied by Lin *et al.* (1999), where earlier references are given. Below we summarize the special cases which we shall use, restated in terms of stationary sequences.

We first note that if $\mathbf{X} := \{X_k : k \ge 0\}$ is a wide sense complex-valued stationary sequence (see Doob (1953), Section X.1), then so also is $\{e^{-ikt} X_k\}$, for every $t \in [-\pi, \pi)$. By the mean ergodic theorem,

$$E(t, \mathbf{X}) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} e^{-ikt} X_k \quad \text{exists in } L_2\text{-norm.}$$

The set of $t \in [-\pi, \pi)$ for which $E(t, \mathbf{X})$ is not zero is called the *point spectrum* of $\mathbf{X}$, and is denoted by $\sigma_\Pi(\mathbf{X})$. When $E(t, \mathbf{X})$ is not 0, it is an eigenvector of the shift isometry associated with $\mathbf{X}$, with corresponding eigenvalue e^{it}—see Doob (1953). The family $\{E(t, \mathbf{X}) : t \in \sigma_\Pi(\mathbf{X})\}$ is therefore orthogonal, and since the shift isometry is defined on a separable space, $\sigma_\Pi(\mathbf{X})$ is always countable; it is precisely the set of atoms of the spectral measure of $\mathbf{X}$.

THEOREM B.1. *Let $\{a_n\}$ be a sequence of complex numbers such that for some $p > 1$, $\sup_n n^{-1} \sum_{k=1}^n |a_k|^p < \infty$. Then for every wide sense stationary sequence $\mathbf{Z} = \{Z_n\}$ with $\sigma_\Pi(\mathbf{Z}) = \emptyset$ (i.e., with continuous spectral measure) we have $\lim_{n\to\infty} \|n^{-1} \sum_{k=1}^n a_k Z_k\| = 0$.*

The proof follows from the second part of the proof of Theorem 4.1 of Çömez *et al.* (1998). The special case $p = 2$, which is the one of interest for us, follows also from carrying out the computation in the proof of Theorem 2.7 for $r = 1$.

Any Besicovitch almost periodic sequence is a Hartman sequence, but there are Hartman sequences which are not Besicovitch (e.g., Lin *et al.* (1999)). It is easy to construct a bounded sequence consisting of 0's and 1's which is not Hartman.

If we want $n^{-1} \sum_{k=1}^n a_k Z_k$ to converge for *every* wide sense stationary sequence $\mathbf{Z}$, then $\{a_n\}$ must be Hartman (since for $t \in [-\pi, \pi)$ the sequence $\{e^{ikt} Z_0\}$ with $\mathbf{E}(Z_0) = 0$ is wide sense stationary). The following two theorems are proved in Lin *et al.* (1999).

THEOREM B.2. *Let $\mathbf{a} = \{a_k\}$ be a Hartman sequence satisfying $\sup_n n^{-1} \sum_{k=1}^n |a_k| < \infty$. Then for every wide sense stationary sequence $\mathbf{Z}$ the limit $L(\mathbf{a}, \mathbf{Z}) := \lim_{n\to\infty} n^{-1} \sum_{k=1}^n a_k Z_k$ exists in L_2-norm, and*

$$L(\mathbf{a}, \mathbf{Z}) = \sum_{t \in [-\pi, \pi)} c(t, \mathbf{a}) E(-t, \mathbf{Z}) \tag{B.1}$$

with L_2-norm convergence (and countably many non-zero terms) on the right-hand side.

THEOREM B.3. *Let $\mathbf{a} = \{a_k\}$ be a Hartman sequence satisfying $\sup_n n^{-1} \sum_{k=1}^n |a_k| < \infty$. Then for every strict sense stationary sequence $\mathbf{Z}$ with finite p-th moment, the limit $L(\mathbf{a}, \mathbf{Z}) = \lim_{n\to\infty} n^{-1} \sum_{k=1}^n a_k Z_k$ exists in L_p-norm. For $p \geq 2$ the limit is given by* (B.1).

DEFINITION. A sequence $\mathbf{a} := \{a_n\}$ of complex numbers is a *good modulating sequence* for a strict sense stationary sequence $\mathbf{Z}$ if the sequence $n^{-1} \sum_{k=1}^n a_k Z_k$ converges almost surely. The sequence $\mathbf{a}$ is called *L_p-universally good* if it is a good modulating sequence for every strict sense stationary sequence $\mathbf{Z}$ with finite pth moment ($1 \leq p < \infty$).

By considering rotations of the unit circle, we see that *any L_p-universally good modulating sequence is a Hartman sequence.* However, a bounded Hartman sequence need not be L_p-universally good for any $1 \le p < \infty$.

THEOREM B.4. (1) *Fix $1 < p < \infty$ with dual index q. Any q-Besicovitch almost periodic sequence is L_p-universally good.*
(2) *Any bounded Besicovitch almost periodic sequence is L_1-universally good.*

The first part of this theorem was proved by Tempelman (1974) and the second one was proved by Ryll-Nardzewski (1975).

The theorem of Wiener and Wintner (1941) shows that almost all realizations of strict sense ergodic stationary sequences with finite first moment are Hartman sequences. The following *Return Times Theorem* of Bourgain *et al.* (1989) makes a considerably stronger statement.

THEOREM B.5. *Fix $1 \le p < \infty$ with dual index q. Let $\mathbf{X} = \{X_k\}$ be an* ergodic *strict sense stationary sequence with finite q-th moment. Then $\mathbf{X}$ is almost surely an L_p-universally good sequence (and is in W_q).*

By the mean ergodic theorem, $E(t, \mathbf{X})$ is well defined for any strict sense stationary sequence $\mathbf{X}$ with finite first moment. When $\mathbf{X}$ is ergodic, the proof of Wiener and Wintner shows that for almost every realization $\mathbf{x} = \{x_k\}$ we have

$$\sigma(\mathbf{x}) = \sigma_{\Pi}(\mathbf{X}), \quad \text{with } c(t, \mathbf{x}) = E(t, \mathbf{X}) \text{ for } t \in \sigma_{\Pi}(\mathbf{X}). \tag{B.2}$$

PROPOSITION B.6. *Fix $1 \le p < \infty$ with dual index q. Let $\mathbf{X} = \{X_k\}$ be an* ergodic *strict sense stationary sequence with finite q-th moment. Then for almost every realization $\mathbf{x} = \{x_k\}$, for every $t \in [-\pi, \pi)$ the sequence $\{e^{ikt} x_k\}$ is an L_p-universally good sequence (and is in W_q).*

The proof combines Theorem B.5 with the following lemma.

LEMMA B.7. *If $\{a_k\}$ is L_p-universally good, then for every $t \in [-\pi, \pi)$ the sequence $\{e^{ikt} a_k\}$ is L_p-universally good.*

Proof. Fix $t \in [-\pi, \pi)$ and $\mathbf{Z}$ strict sense stationary with finite pth moment, generated by the probability preserving τ on $(\mathcal{Z}, \nu)$. On the unit circle Γ with Lebesgue's measure $d\gamma$ let ϱ be the rotation through the angle t (i.e., $\varrho\gamma := e^{it}\gamma$), and take $g(\gamma) = \gamma$. Then $\widetilde{Z}_n := g \circ \varrho^n \otimes Z_n$ is strict sense stationary on $(\Gamma \times \mathcal{Z}, d\gamma \times \nu)$. Hence

$$\frac{1}{n} \sum_{k=1}^{n} e^{ikt} a_k Z_k = \overline{\gamma}\, \frac{1}{n} \sum_{k=1}^{n} a_k \widetilde{Z}_k$$

converges a.e. on $\Gamma \times \mathcal{Z}$, and choosing an appropriate $\gamma \in \Gamma$ yields convergence of $n^{-1} \sum_{k=1}^{n} a_k \widetilde{Z}_k$ for a.e. $\zeta \in \mathcal{Z}$. ∎

Acknowledgements. Parts of this research were carried out while the second author was a Shapiro visitor at the Department of Mathematics of Penn State University, and the third author was a Dozor visitor at Ben-Gurion University. The first author was partially supported by GIF grant I-526-034.

REFERENCES

T. W. Anderson and J. B. Taylor (1976), *Strong consistency of least squares estimates in normal linear regression*, Ann. Statist. 4, 788–790.

D. Berend, M. Lin, J. Rosenblatt and A. Tempelman (2002), *Modulated and subsequential ergodic theorems in Hilbert and Banach spaces*, Ergodic Theory Dynam. Systems 22, 1653–1665.

J. Bourgain, H. Furstenberg, Y. Katznelson and D. Ornstein (1989), *Return times of dynamical systems*, Publ. Math. IHES 69, 47–50.

V. V. Buldygin and V. A. Koval' (2000), *Strong law of large numbers with operator normalizations for martingales and sums of orthogonal random vectors*, Ukrainian Math. J. 52, 1195–1214.

R. V. Chacon (1969), *Weakly mixing transformations which are not strongly mixing*, Proc. Amer. Math. Soc. 22, 559–562.

G.-J. Chen, T. L. Lai and C. Z. Wei (1981), *Convergence systems and strong consistency of least squares estimates in regression models*, J. Multivariate Anal. 11, 319–333.

X. R. Chen and Y. Wu (1993), *On a necessary condition for the consistency of the L_1 estimates in linear regression models*, Comm. Statist. Theory Methods 22, 631–639.

N. Christopeit and K. Helms (1980), *Strong consistency of least squares estimators in linear regression models*, Ann. Statist. 8, 778–788.

D. Çömez, M. Lin and J. Olsen (1998), *Weighted ergodic theorems for mean ergodic L_1-contractions*, Trans. Amer. Math. Soc. 350, 101–117.

P. J. Davis (1975), *Interpolation and Approximation*, Dover, New York.

Y. Derriennic and M. Lin (2001), *Fractional Poisson equations and ergodic theorems for fractional coboundaries*, Israel J. Math. 123, 93–130.

J. L. Doob (1953), *Stochastic Processes*, Wiley, New York.

H. Drygas (1971), *Consistency of the least squares and Gauss–Markov estimators in regression models*, Z. Wahrsch. Verw. Gebiete 17, 309–326.

H. Drygas (1976), *Weak and strong consistency of the least squares estimators in regression models*, ibid. 34, 119–127.

S. Foguel (1963), *Powers of a contraction in Hilbert space*, Pacific J. Math. 13, 551–562.

V. F. Gaposhkin (1979), *Strong consistency of estimates of the trend of a time series*, Mat. Zametki 26, 643–652 (in Russian); English transl.: Math. Notes 26, 812–818.

V. F. Gaposhkin (1985), *A remark on strong consistency of LS estimates under weakly correlated observation errors*, Teor. Veroyatnost. i Primenen. 30, 160–163 (in Russian); English transl.: Theory Probab. Appl. 30, 177–181.

W. H. Green (2003), *Econometric Analysis*, 5th ed., Prentice-Hall, Upper Saddle River, NJ.

U. Grenander and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

P. Halmos (1949), *A non-homogeneous ergodic theorem*, Trans. Amer. Math. Soc. 66, 284–288.

P. Halmos (1956), *Lectures on Ergodic Theory*, Math. Soc. Japan, Tokyo.

E. J. Hannan (1978), *Rates of convergence for time series regression*, Adv. Appl. Probab. 10, 740–743.

M. Jin and X. Chen (1996), *Necessary and sufficient conditions for strong consistency of LS estimates*, Chinese J. Contemp. Math. 17, 97–106.

M. Jin and X. Chen (1999), *Strong consistency of the least squares estimate in multiple regression when the error variance is infinite*, Statist. Sinica 9, 289–296.

J.-P. Kahane (1961), *Sur les coefficients de Fourier–Bohr*, Studia Math. 21, 103–106.

A. Kholevo (1969), *On estimation of regression coefficients*, Teor. Veroyatnost. i Primenen. 14, 78–101 (in Russian); English transl.: Theory Probab. Appl. 14, 79–104.

A. Kholevo (1971), *On asymptotic efficiency of pseudo-best estimators*, Teor. Veroyatnost. i Primenen. 16, 524–534 (in Russian); English transl.: Theory Probab. Appl. 16, 516–527.

U. Krengel (1972), *Weakly wandering vectors and weakly independent partitions*, Trans. Amer. Math. Soc. 164, 199–226.

U. Krengel (1985), *Ergodic Theorems*, de Gruyter, Berlin.

T. L. Lai (1991), *Some almost sure convergence properties of weighted sums of martingale difference sequences*, in: Almost Everywhere Convergence II, Academic Press, Boston, 179–190.

T. L. Lai and H. Robbins (1977), *Strong consistency of least squares estimates in regression models*, Proc. Nat. Acad. Sci. USA 74, 2667–2669.

T. L. Lai, H. Robbins and C. Z. Wei (1978), *Strong consistency of least squares estimates in multiple regression*, ibid. 75, 3034–3036.

T. L. Lai, H. Robbins and C. Z. Wei (1979), *Strong consistency of least squares estimates in multiple regression II*, J. Multivariate Anal. 9, 343–361.

T. L. Lai and C. Z. Wei (1982), *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems*, Ann. Statist. 10, 154–166.

A. Le Breton and M. Musiela (1987), *Strong consistency of least squares estimates in linear regression models driven by semimartingales*, J. Multivariate Anal. 23, 77–92.

M. Lin and L. Jones (1980), *Unimodular eigenvalues and weak mixing*, J. Funct. Anal. 35, 42–48.

M. Lin, J. Olsen and A. Tempelman (1999), *On modulated ergodic theorems for Dunford–Schwartz operators*, Illinois J. Math. 43, 542–567.

A. Novikov (1985), *Consistency of least squares estimates in regression models with martingale errors*, in: Statistics and Control of Stochastic Processes (Moscow, 1984), Transl. Ser. Math. Engrg., Optimization Software, New York, 389–409.

K. Petersen (1983), *Ergodic Theory*, Cambridge Univ. Press, Cambridge.

C. R. Rao and H. Toutenbourg (1995), *Linear Models*, Springer, New York.

J. Rosenblatt (1994), *Norm convergence in ergodic theory and the behavior of Fourier transforms*, Canad. J. Math. 46, 184–199.

Yu. Rozanov (1969), *On a new class of estimates*, in: Multivariate Analysis II (Dayton, OH, 1968), Academic Press, New York, 437–441.

W. Rudin (1959), *Some theorems on Fourier coefficients*, Proc. Amer. Math. Soc. 10, 855–859.

C. Ryll-Nardzewski (1975), *Topics in ergodic theory*, in: Lecture Notes in Math. 472, Springer, 131–156.

A. Savichev and A. Tempelman (1990), *Consistency of linear regression estimates for random functions*, Litovsk. Mat. Sb. 30, 137–141 (in Russian).

A. S. Slabospitskii (1994), *Convergence conditions for the least squares method*, J. Math. Sci. 72, 3076–3079.

V. Solo (1981), *Strong consistency of least squares estimators in regression with correlated disturbances*, Ann. Statist. 9, 689–693.

J. Stapleton (1995), *Linear Statistical Models*, Wiley, New York.

A. Tempelman (1970), *On linear regression estimates*, Soviet Math. Dokl. 11, 452–455.

A. Tempelman (1973), *On linear regression estimates*, in: 2nd Internat. Symp. on Information Theory, B. N. Petrov and T. Csaki (eds.), Akadémiai Kiadó, Budapest, 329–354. Reprinted in *Reproducing Kernel Hilbert Spaces. Applications in Statistical Signal Processing*, H. L. Weinert (ed.), Hutchinson Ross, Strasbourg, 1982, 301–326.

A. Tempelman (1974), *Ergodic theorems for amplitude modulated homogeneous random fields*, Litovsk. Mat. Sb. 14, 221–229 (in Russian); English transl.: Lith. Math. Transl. 14, 698–704, 1975.

A. Tempelman (1975), *Convergence and consistency of regression estimators*, Doctor of Sciences Dissertation, Vilnius (in Russian).

C. Z. Wei (1985), *Asymptotic properties of least-squares estimates in stochastic regression models*, Ann. Statist. 13, 1498–1508.

N. Wiener and A. Wintner (1941), *Harmonic analysis and ergodic theory*, Amer. J. Math. 63, 415–426.

Y. Yajima (1988), *On estimation of a regression model with long-memory stationary errors*, Ann. Statist. 16, 791–807.

A. Zygmund (1968), *Trigonometric Series*, Vols. I–II, corrected 2nd ed., Cambridge Univ. Press, Cambridge.

Department of Mathematics
Ben-Gurion University
Beer-Sheva, Israel
E-mail: guycohen@math.bgu.ac.il
lin@math.bgu.ac.il

Department of Statistics
Penn State University
University Park, PA 16802, U.S.A.
E-mail: arkady@stat.psu.edu

Received 23 October 2003 (4390)