Econometric Theory, **31**, 2015, 337–361.
doi:10.1017/S0266466614000322

THE INTEGRATED MEAN SQUARED ERROR OF SERIES REGRESSION AND A ROSENTHAL HILBERT-SPACE INEQUALITY

Bruce E. Hansen
University of Wisconsin

This paper develops uniform approximations for the integrated mean squared error (IMSE) of nonparametric series regression estimators, including both least-squares and averaging least-squares estimators. To develop these approximations, we also generalize an important probability inequality of Rosenthal (1970, *Israel Journal of Mathematics* 8, 273–303; 1972, *Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 149–167. University of California Press) to the case of Hilbert-space valued random variables.

1. INTRODUCTION

This paper introduces uniform approximations for the integrated mean squared error (IMSE) of nonparametric series regression estimators. Bounds for the IMSE of series regression estimators have been obtained by Newey (1997) but ours are the first uniform approximations. Related papers include Andrews (1991) who studied the asymptotic normality of series estimators, and Newey (1995), de Jong (2002), Chen (2007, Chap. 76), and Song (2008) who studied uniform convergence. The difference is that we are interested in directly characterizing the IMSE and not just a bound on the rate of convergence. The theory rests developing a bound on the expectation of the norm of the inverse of the sample design matrix, and for this we use an argument due to Ing and Wei (2003). Our results apply to a wide variety of series regressions, including polynomial and spline expansions.

We also extend the results to averaging estimators, which are weighted averages of least-squares estimators of individual series estimators. Averaging estimators are strict generalizations of standard estimators and thereby can achieve lower IMSE. See Hansen (2007) and Hansen and Racine (2012). We obtain uniform approximations to the IMSE of averaging estimators. This result is derived in the slightly more restricted setting of nested series estimators.

Research supported by the National Science Foundation. I thank the Co-Editor Yoon-Jae Whang and two referees for helpful comments, catching a mistake, and suggesting an alternative proof of Theorem 1. I also would like to thank Rustam Ibragimov for very helpful suggestions and for providing a translation of Utev (1985). Address correspondence to Bruce Hansen, Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, USA; e-mail: behansen@wisc.edu. www.ssc.wisc.edu/~bhansen.

To develop these approximations, we also introduce a generalization of the classic Rosenthal inequality. The related inequalities of Marcinkiewicz and Zygmund (1937), Rosenthal (1970, 1972), and Burkholder (1973) are foundational for many problems in applied probability and statistics. The classical form of these inequalities is for real-valued random variables. In this paper we generalize the Rosenthal inequality to allow for random variables taking values in a Hilbert space, which includes the case of random matrices. The result is a simple extension of a Banach space inequality obtained by De Acosta (1981) and a Hilbert space inequality due to Utev (1985).

2. A HILBERT-SPACE ROSENTHAL INEQUALITY

The following is a generalization of the one-sided (upper) inequalities of Marcinkiewicz and Zygmund (1937) and Rosenthal (1970, 1972) to the case of Hilbert spaces. For $2 \leq p < 3$ it is based on a Banach space inequality obtained by De Acosta (1981), and for $p \geq 3$ it is derived from a Hilbert space inequality of Utev (1985) following a suggestion of Ibragimov (1997).

THEOREM 1. *For any $p \geq 2$ there is a finite constant A_p such that for any independent centered array of random variables ξ_{ni} taking values in a Hilbert space with norm $\|\cdot\|$ such that $\mathbb{E}\|\xi_{ni}\|^p < \infty$,*

$$\mathbb{E}\left\|\sum_{i=1}^{n}\xi_{ni}\right\|^p \leq A_p\left\{\left(\sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^2\right)^{p/2} + \sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^p\right\}. \tag{1}$$

For $p \geq 2$ we have $A_p = \left(2B_p^{1/p}+1\right)^p$ where B_p is the constant from the martingale generalization of Rosenthal's inequality (Burkholder, 1973, p. 40 or Hall and Heyde, 1980, Thm. 2.12), and for $p \geq 3$ we have $A_p = 2^p C_p$ where C_p is the exact Rosenthal constant in the independent symmetric case from Ibragimov and Sharakhmetov (1997). If $p = 2$, the expression simplifies to

$$\mathbb{E}\left\|\sum_{i=1}^{n}\xi_{ni}\right\|^2 = \sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^2. \tag{2}$$

Inequalities of the form (1) are widely used in probability and statistical theory. They are commonly applied to random variables, and this is sufficient for many purposes (as the bound can be applied separately to each element in the matrix ξ_{ni}). However, for some purposes it is essential for the bound to involve the same norm as used on the left-side.

Remark 1. The constant A_p in (1) depends only on p, not on the underlying probability structure nor the specific norm.

Remark 2. Theorem 1 applies to random vectors and matrices for any Hilbert space norm, for example, the Euclidean norm $\|\boldsymbol{a}\| = \left(\operatorname{tr} \boldsymbol{a}'\boldsymbol{a}\right)^{1/2}$ for vectors and the Frobenius norm $\|\boldsymbol{A}\|_F = \left(\operatorname{tr} \boldsymbol{A}'\boldsymbol{A}\right)^{1/2}$ for matrices.

Remark 3. When ξ_{ni} are identically distributed across i for given n, then we can write (1) as

$$\mathbb{E}\left\|\sum_{i=1}^{n} \xi_{ni}\right\|^{p} \leq A_p \left\{\left(n\mathbb{E}\|\xi_{ni}\|^2\right)^{p/2} + n\mathbb{E}\|\xi_{ni}\|^p\right\}. \tag{3}$$

If the distribution of ξ_{ni} does not depend on n, then the first term on the right-hand-side of (3) is of order $O(n^{p/2})$ which dominates the second term which is of order $O(n)$ (unless $p = 2$ in which case they are the same). However, this relative ranking can change when the distribution of ξ_{ni} changes with n.

Remark 4. Using Loeve's C_r inequality (e.g., Davidson, 1994, p. 140), we can bound (1) by $2A_p n^{p/2-1} \sum_{i=1}^{n} \mathbb{E}\|\xi_{ni}\|^p$, or $2A_p n^{p/2} \mathbb{E}\|\xi_{ni}\|^p$ in the case where ξ_{ni} are identically distributed across i for given n. However, these bounds are often significantly less tight, so are not typically preferred.

Remark 5. Some other moment bounds for matrices, Hilbert-space variables and Banach-valued variables can be found in de la Peña and Giné (1999) and Nze and Doukhan (2004). A review of moment bounds in econometrics can be found in Lai and Wei (1984).

Remark 6. Theorem 1 is restricted to independent variables. Rosenthal-type inequalities for dependent random variables can be found in Utev (1991), de la Peña, Ibragimov, and Sharakhmetov (2003), and Nze and Doukhan (2004). It would be greatly desirable to extend Theorem 1 to allow for dependence. However, our proof builds on results (De Acosta, 1981; Utev, 1985) which are restricted to independent sequences.

Remark 7. The constant A_p may not be the best possible. See Bestsennaya and Utev (1991) and Ibragimov and Sharakhmetov (2002) for sharp bounds on moments of sums and the best constant in Rosenthal's inequality for the case of random variables and p an even integer, and Ibragimov and Ibragimov (2008) for the case of random variables with zero odd moments.

3. MOMENT BOUNDS FOR SERIES REGRESSION

Consider a sample of independently and identically distributed (iid) observations (y_i, z_i), $i = 1, \ldots, n$ where $z_i \in \mathcal{Z}$, a compact subset of $\mathbb{R}^q$. Define the conditional mean $g(z) = \mathbb{E}(y_i \mid z_i = z)$. We examine the estimation of $g(z)$ by series regression. For $m = 1, \ldots, M_n$, let $x_m(z)$ denote a sequence of $K_m \times 1$ vector of functions from a series expansion. For example, a power series sets $x_m(z) = (1, z, z^2, ..., z^m)$ and a quadratic spline sets $x_m(z) = (1, z, z^2, (z-\tau_1)^2 1(z > \tau_1), ..., (z-\tau_m)^2 1(z > \tau_m))$. Set $x_{mi} = x_m(z_i)$, a $K_m \times 1$ regressor vector.

The m'th series estimator of $g(z)$ is

$$\widehat{g}_m(z) = x_m(z)'\widehat{\beta}_m,$$

where

$$\widehat{\beta}_m = \left(\sum_{i=1}^{n} x_{mi}x_{mi}'\right)^{-1} \sum_{i=1}^{n} x_{mi}y_i$$

is the least squares coefficient from a regression of y_i on x_{mi}.

Many of the challenges arising in the theory of series regression stem from the inversion of the sample design matrix

$$\widehat{Q}_m = \frac{1}{n}\sum_{i=1}^{n} x_{mi}x_{mi}'$$

as an estimate of

$$Q_m = \mathbb{E}\left(x_{mi}x_{mi}'\right).$$

In this section we describe some properties of the moments of $\widehat{Q}_m$ and $\widehat{Q}_m^{-1}$.

Define

$$\zeta_m = \sup_{z\in\mathcal{Z}} \left(x_m(z)' Q_m^{-1} x_m(z)\right)^{1/2}, \tag{4}$$

the largest normalized Euclidean length of the regressor vector. Under standard conditions for series regression (including compact support for the regressors), ζ_m will be a bounded function of the dimension K_m. For example, when x_{mi} is a power series, then $\zeta_m^2 = O(K_m^2)$ (see Andrews, 1991), and when x_{mi} is a regression spline, then $\zeta_m^2 = O(K_m)$ (see Newey, 1995). For further discussion see Newey (1997) and Li and Racine (2006).

We will also define the array of constants

$$\Psi_{nm} = \left(\frac{\zeta_m^2 K_m}{n}\right)^{1/2} \tag{5}$$

which will appear frequently in our bounds.

For convenience, we will state our moment bounds under the assumption that $Q_m = I_{K_m}$. Since the estimator $\widehat{g}_m(z)$ is invariant to rotations of the regressor vector x_{mi}, this is without loss of generality for most results of interest.

For any matrix $\boldsymbol{A}$ let $\|\boldsymbol{A}\|_F = \left(\operatorname{tr} \boldsymbol{A}'\boldsymbol{A}\right)^{1/2}$ denote the Frobenius norm. The space of $\ell \times m$ matrices with the Frobenius norm is a Hilbert space, allowing the application of Theorem 1.

LEMMA 1. *If* $Q_m = I_{K_m}$, *for any* $0 < p \le 2$,

$$\mathbb{E}\left\| \widehat{Q}_m - I_{K_m} \right\|_F^p \le \Psi_{nm}^p \tag{6}$$

and for any $p > 2$

$$\mathbb{E}\left\| \widehat{Q}_m - I_{K_m} \right\|_F^p \le A_p \Psi_{nm}^p \left(1 + 2^p \Psi_{nm}^{p-2}\right). \tag{7}$$

As shown by Ing and Wei (2003), the moment bound of Lemma 1 plus the following regularity conditions can be used together to establish moment bounds on the inverse moment matrix.

Assumption 1.

1. For some $\delta > 0$, $\sup_{1 \le m \le M_n} \zeta_m^2 K_m^{1+\delta}/n \longrightarrow 0$.
2. For some $\alpha > 0$, $\eta > 0$, and $\psi < \infty$, for all $\ell' Q_m \ell = 1$ and $0 \le u \le \eta$, $\sup_m \mathbb{P}\left(\left|\ell' x_{mi}\right| \le u\right) \le \psi u^\alpha$.

Assumption 1.1 puts a bound on the number of series terms K_m relative to the sample size. For a polynomial series expansion this requirement is satisfied when $K_m^{3+\delta}/n = o(1)$ and for a spline expansion it is satisfied when $K_m^{2+\delta}/n = o(1)$. It indirectly bounds the number of models M_n.

Assumption 1.2 is an unusual requirement. It specifies that the all linear combinations $\ell' x_{mi}$ have a Lipschitz continuous distribution near the origin. This is used to ensure existence of the expectation of the inverse of the sample design matrix.

Our next set of results are bounds for the spectral norm $\|\boldsymbol{A}\|_S = (\lambda_{\max}(\boldsymbol{A}'\boldsymbol{A}))^{1/2}$ of the inverse design matrix.

LEMMA 2. *Under Assumption 1 and* $Q_m = I_{K_m}$, *for any* $p > 0$ *and* $\eta > 0$ *there is an n sufficiently large such that*

$$\mathbb{E}\left\| \widehat{Q}_m^{-1} \right\|_S^p \le 1 + \eta. \tag{8}$$

LEMMA 3. *Under Assumption 1 and* $Q_m = I_{K_m}$, *for any* $0 < p < 2$ *and* $\eta > 0$ *there is an n sufficiently large such that*

$$\mathbb{E}\left\| \widehat{Q}_m^{-1} - I_{K_m} \right\|_S^p \le (1+\eta)\, \Psi_{nm}^p, \tag{9}$$

and for any $p \ge 2$ *and* $\eta > 0$, *there is an n sufficiently large such that*

$$\mathbb{E}\left\| \widehat{Q}_m^{-1} - I_{K_m} \right\|_S^p \le (1+\eta)\, A_{2p}^{1/2} \Psi_{nm}^p. \tag{10}$$

4. INTEGRATED MEAN SQUARED ERROR

The integrated mean-squared error (IMSE) of the estimator $\widehat{g}_m(z)$ is

$$IMSE_n(m) = \int_{\mathcal{Z}} \mathbb{E}\left(\widehat{g}_m(z) - g(z)\right)^2 f(z)dz, \tag{11}$$

where $f(z)$ is the marginal density of z_i. We are interested in an approximation for $IMSE_n(m)$ which is uniform across the expansions m.

It will be useful to set up some notation. Let

$$\beta_m = \left(\mathbb{E}\left(x_{mi}x_{mi}'\right)\right)^{-1} \mathbb{E}\left(x_{mi}y_i\right)$$

denote the linear projection coefficient, $e_{mi} = y_i - x_{mi}'\beta_m$ be the projection error, and $e_i = y_i - g(z_i)$ be the regression error. Define the approximation error $r_m(z) = g(z) - x_m(z)'\beta_m$ and $r_{mi} = r_m(z_i)$, and observe that $e_{mi} = r_{mi} + e_i$. Since e_{mi} is a projection error, $\mathbb{E}(x_{mi}e_{mi}) = 0$, and since e_i is a regression error, $\mathbb{E}(x_{mi}e_i) = 0$. It follows that $\mathbb{E}(x_{mi}r_{mi}) = 0$. Set $\phi_m^2 = \mathbb{E}\left(r_{mi}^2\right)$ and $Q_m = \mathbb{E}\left(x_{mi}x_{mi}'\right)$. Let $\sigma_i^2 = \mathbb{E}\left(e_i^2 \mid z_i\right)$ and $\sigma_{mi}^2 = \mathbb{E}\left(e_{mi}^2 \mid z_i\right)$, and set $\Omega_m = \mathbb{E}\left(x_{mi}x_{mi}'\sigma_i^2\right)$ and $\Omega_m^* = \mathbb{E}\left(x_{mi}x_{mi}'\sigma_{mi}^2\right)$.

By definition, $g(z) = x_m(z)'\beta_m + r_m(z)$, so

$$\widehat{g}_m(z) - g(z) = x_m(z)'\left(\widehat{\beta}_m - \beta_m\right) - r_m(z). \tag{12}$$

Thus

$$\begin{aligned}
\int_{\mathcal{Z}} \left(\widehat{g}_m(z) - g(z)\right)^2 f(z)dz &= \int_{\mathcal{Z}} r_m(z)^2 f(z)dz \\
&\quad -2\left(\widehat{\beta}_m - \beta_m\right)' \int_{\mathcal{Z}} x_m(z) r_m(z) f(z)dz \\
&\quad + \left(\widehat{\beta}_m - \beta_m\right)' \int_{\mathcal{Z}} x_m(z) x_m(z)' f(z)dz \left(\widehat{\beta}_m - \beta_m\right) \\
&= \mathbb{E}\left(r_{mi}^2\right) - 2\left(\widehat{\beta}_m - \beta_m\right)' \mathbb{E}\left(x_{mi}r_{mi}\right) \\
&\quad + \left(\widehat{\beta}_m - \beta_m\right)' \mathbb{E}\left(x_{mi}x_{mi}'\right)\left(\widehat{\beta}_m - \beta_m\right) \\
&= \phi_m^2 + \left(\widehat{\beta}_m - \beta_m\right)' Q_m \left(\widehat{\beta}_m - \beta_m\right),
\end{aligned}$$

the second equality since integration over the density $f(z)$ is the same as taking expectations, and the third using the fact that $\mathbb{E}(x_{mi}r_{mi}) = 0$. Taking expectations, we have found that

$$IMSE_n(m) = \phi_m^2 + \mathbb{E}\left(\left(\widehat{\beta}_m - \beta_m\right)' Q_m \left(\widehat{\beta}_m - \beta_m\right)\right). \tag{13}$$

The standard asymptotic covariance matrix for $\widehat{\beta}_m$ is $n^{-1}Q_m^{-1}\Omega_m^* Q_m^{-1}$. Substituting this covariance matrix for the expectation in (13) we might expect $IMSE_n(m)$ to be close to

$$I_n^*(m) = \phi_m^2 + \frac{1}{n}\operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right). \tag{14}$$

Furthermore, we might expect $\Omega_m^* \simeq \Omega_m$, so that perhaps $IMSE_n(m)$ might be close to

$$I_n(m) = \phi_m^2 + \frac{1}{n}\operatorname{tr}\left(Q_m^{-1}\Omega_m\right). \tag{15}$$

To show that $IMSE_n^*(m)$ is uniformly close to both $I_n^*(m)$ and $I_n(m)$, we add the following regularity conditions.

Assumption 2.

1. $\left\| Q_m^{-1} \right\|_S \le B < \infty$.
2. $0 < \underline{\sigma}^2 \le \sigma_i^2 \le \overline{\sigma}^2 < \infty$.
3. $g(z)$ is continuously differentiable on $z \in \mathcal{Z}$.

Assumption 2.1 states that the smallest eigenvalue of Q_m is bounded above zero, and thus Q_m is uniformly invertible. This is a standard condition which is satisfied by typical series expansions. For example, Newey (1997) demonstrates that Assumption 2.1 holds when the support $\mathcal{Z}$ of z_i is a Cartesian product of compact connected intervals on which the density $f(z)$ is bounded away from zero.

Assumption 2.2 controls the degree of conditional heteroskedasticity, bounding the conditional variance away from zero and infinity. Assumption 2.3 is a mild smoothness condition.

We can use the the uniform approximations of Lemmas 1–3 to establish the following technical bound.

LEMMA 4. *Under Assumptions 1 and 2, for n sufficiently large, and all $m \le M_n$,*

$$\left| \mathbb{E}\left(\left(\widehat{\beta}_m - \beta_m\right)' Q_m \left(\widehat{\beta}_m - \beta_m\right) \right) - \frac{\operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)}{n} \right| \le A_4 \Psi_{nm} \frac{\operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)}{n}. \tag{16}$$

Using (13) and Lemma 4, we can show that $I_n^*(m)$ and $I_n(m)$ are uniform approximations to $IMSE_n^*(m)$:

THEOREM 2. *Under Assumptions 1 and 2, as $n \to \infty$*

$$\sup_{1 \le m \le M_n} \left| \frac{IMSE_n(m) - I_n^*(m)}{I_n^*(m)} \right| \longrightarrow 0 \tag{17}$$

and

$$\sup_{1 \le m \le M_n} \left| \frac{IMSE_n(m) - I_n(m)}{I_n(m)} \right| \longrightarrow 0. \tag{18}$$

Another way of stating these results is that $IMSE_n(m) = I_n^*(m)(1+o(1)) = I_n(m)(1+o(1))$ uniformly across the series expansions $m \le M_n$.

Interestingly, Theorem 2 does not put a direct bound on the number of models M_n relative to sample size n. The number of models is only indirectly bounded through Assumption 1, which primarily bounds the dimensionality of the models.

The uniform approximation provided in Theorem 1 is an important step for showing that data-dependent choices of model m can be optimal in the sense of minimizing the IMSE. Hansen (2014) shows that under regularity conditions cross-validation selection is asymptotically equivalent to selecting the model which minimizes $I_n(m)$. Combined with Theorem 2, we conclude that cross-validation is asymptotically optimal with respect to minimizing IMSE.

5. AVERAGING REGRESSIONS

Reductions in IMSE can be achieved by averaging across the individual series estimators $\widehat{g}_m(z)$. Let $w = (w_1, ..., w_{M_n})$ be a set of nonnegative weights which sum to one. Define the averaging estimator

$$\widehat{g}(z) = \sum_{m=1}^{M_n} w_m \widehat{g}_m(z).$$

The IMSE of the averaging estimator is

$$\begin{aligned} IMSE_n(w) &= \int_{\mathcal{Z}} \mathbb{E}\left(\widehat{g}(z) - g(z)\right)^2 f(z)dz \\ &= \sum_{m=1}^{M_n} w_m^2 \int_{\mathcal{Z}} \mathbb{E}\left(\widehat{g}_m(z) - g(z)\right)^2 f(z)dz \\ &\quad + 2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m \int_{\mathcal{Z}} \mathbb{E}\left(\left(\widehat{g}_m(z) - g(z)\right)\left(\widehat{g}_\ell(z) - g(z)\right)\right) f(z)dz. \end{aligned}$$

In general, the series regressions need not be nested, but for simplicity we add that assumption as it greatly simplifies our calculations. Recall that $q = \dim(z_i)$.

Assumption 3. For a power series or spline basis sequence $\tau_j(z),\ j = 1, 2, ...,$

$$x_m(z) = (\tau_1(z), \tau_2(z), \tau_3(z), ..., \tau_m(z)).$$

Assumption 3 is satisfied by nested power series. It is also satisfied by sequences of splines when knots are added but not deleted. With this additional assumption, we can show that $IMSE_n(w)$ is uniformly close to

$$I_n^*(w) = \sum_{m=1}^{M_n} w_m^2 \left(\phi_m^2 + \frac{1}{n} \operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)\right)$$
$$+2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m \left(\phi_\ell^2 + \frac{1}{n} \operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)\right).$$

Let $\mathcal{W}_n$ be the M_n-dimensional unit simplex.

THEOREM 3. *Under Assumptions 1–3, as $n \to \infty$*

$$\sup_{w\in\mathcal{W}_n} \left| \frac{IMSE_n(w) - I_n^*(w)}{I_n^*(w)} \right| \longrightarrow 0.$$

Theorem 3 shows that $I_n^*(w)$ is a uniformly good approximation to the IMSE of the averaging estimator, where the uniformity is over all weight vectors.

In analogy to Theorem 2, we might expect that $IMSE_n(w)$ is also uniformly close to

$$I_n(w) = \sum_{m=1}^{M_n} w_m^2 \left(\phi_m^2 + \frac{1}{n} \operatorname{tr}\left(Q_m^{-1}\Omega_m\right)\right)$$
$$+2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m \left(\phi_\ell^2 + \frac{1}{n} \operatorname{tr}\left(Q_m^{-1}\Omega_m\right)\right). \quad \textbf{(19)}$$

This, however, turns out to be harder to establish. To do so, we need a stronger condition.

Assumption 4.

1. $g(z)$ has s continuous derivatives on $z \in \mathcal{Z}$ with $s \geq q/2$ for a spline, and $s \geq q$ for a power series.
2. $\phi_m^2 > 0$ for all m.

Assumption 4.1 is a strengthening of the smoothness condition Assumption 2.3 when $q \geq 3$ for a spline basis or $q \geq 2$ for a power series. Assumption 4.2 specifies that no finite dimensional expansion equals the true regression function, and thus all expansions are approximations. This is standard in the nonparametrics literature. It excludes the possibility that the true regression function is a finite dimensional element of the sieve space (e.g., a linear function).

THEOREM 4. *Under Assumptions 1–4, as $n \to \infty$*

$$\sup_{w\in\mathcal{W}_n} \left| \frac{IMSE_n(w) - I_n(w)}{I_n(w)} \right| \longrightarrow 0.$$

Theorem 4 is relevant for establishing the optimality of cross-validation selection of the averaging weights. Under similar conditions, Hansen (2014) shows that the cross-validation weight selection is asymptotically equivalent to selecting the model which minimizes $I_n(w)$. Theorem 4 thus establishes that cross-validation is an asymptotically optimal with respect to IMSE for selection of the averaging weights.

6. PROOFS

Proof of Theorem 1. Since $\|\cdot\|$ is a Hilbert space norm, we can write $\|x\|^2 = \langle x, x\rangle$ where $\langle x, y\rangle$ is an inner product. Thus

$$\begin{aligned}
\mathbb{E}\left\|\sum_{i=1}^{n}\xi_{ni}\right\|^2 &= \mathbb{E}\left\langle \sum_{i=1}^{n}\xi_{ni}, \sum_{j=1}^{n}\xi_{nj}\right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\langle \xi_{ni}, \xi_{nj}\rangle \\
&= \sum_{i=1}^{n}\mathbb{E}\langle \xi_{ni}, \xi_{ni}\rangle \\
&= \sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^2 .
\end{aligned} \tag{20}$$

The second and third equalities hold by the linearity of the inner product and $\mathbb{E}\xi_{ni} = 0$, and the fourth is the definition $\|x\|^2 = \langle x, x\rangle$. This is (2).

Define $S_n = \sum_{i=1}^{n}\xi_{ni}$. For any $p \geq 2$, De Acosta (1981, Thm. 2.1, part (2)) established the following inequality, valid for Banach-valued random variables ξ_{ni} (which includes Hilbert spaces)

$$\mathbb{E}\left|\|S_n\| - \mathbb{E}\|S_n\|\right|^p \leq 2^p B_p \left\{\left(\sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^2\right)^{p/2} + \sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^p\right\}. \tag{21}$$

By Minkowski's inequality, Liapunov's inequality, (20), and (21),

$$\begin{aligned}
\left(\mathbb{E}\|S_n\|^p\right)^{1/p} &= \left(\mathbb{E}\left|\|S_n\| - \mathbb{E}\|S_n\| + \mathbb{E}\|S_n\|\right|^p\right)^{1/p} \\
&\leq \left(\mathbb{E}\left|\|S_n\| - \mathbb{E}\|S_n\|\right|^p\right)^{1/p} + \mathbb{E}\|S_n\| \\
&\leq \left(\mathbb{E}\left|\|S_n\| - \mathbb{E}\|S_n\|\right|^p\right)^{1/p} + \left(\mathbb{E}\|S_n\|^2\right)^{1/2} \\
&\leq 2B_p^{1/p}\left\{\left(\sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^2\right)^{p/2} + \sum_{i=1}^{n}\mathbb{E}\|\xi_{ni}\|^p\right\}^{1/p}
\end{aligned}$$

$$+\left(\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^2\right)^{1/2}$$

$$\leq\left(2B_p^{1/p}+1\right)\left\{\left(\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^2\right)^{p/2}+\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^p\right\}^{1/p}$$

which establishes (1) with $A_p=\left(2B_p^{1/p}+1\right)^p$.

For $p\geq 3$ we follow a suggestion from Ibragimov (1997, Rem. 1.10) to obtain a tighter bound. First, to induce symmetry, let $\varepsilon_1,\dots,\varepsilon_n$ be independent Rademacher random variables independent of ξ_{ni}. (A Rademacher random variable is a symmetric random variable on the two points $\{-1,1\}$.) Then by the symmetrization inequality (e.g., de la Peña and Giné, 1999, Lem. 1.2.6)

$$\mathbb{E}\left\|\sum_{i=1}^{n}\xi_{ni}\right\|^p\leq 2^p\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i\xi_{ni}\right\|^p. \tag{22}$$

Since the Hilbert-space variables $\varepsilon_i\xi_{ni}$ are independent and symmetric, we can apply Corollary 4 of Utev (1985) (given below) so that the right-hand of (22) is bounded by

$$2^p\mathbb{E}\left|\sum_{i=1}^{n}\|\varepsilon_i\xi_{ni}\|\,u_i\right|^p=2^p\mathbb{E}\left|\sum_{i=1}^{n}\|\xi_{ni}\|\,u_i\right|^p, \tag{23}$$

where $u_1,\dots,u_n$ are independent Rademacher random variables independent of ξ_{ni} and ε_i, and the equality is $\|\varepsilon_i\xi_{ni}\|=\|\xi_{ni}\|$. The right-hand moment only involves the sum of the symmetric real-valued random variables $\|\xi_{ni}\|\,u_i$, and thus we can apply the classic Rosenthal inequality to bound (23) by

$$2^pC_p\left\{\left(\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^2u_i^2\right)^{p/2}+\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^p\,|u_i|^p\right\}$$

$$=2^pC_p\left\{\left(\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^2\right)^{p/2}+\sum_{i=1}^{n}\mathbb{E}\,\|\xi_{ni}\|^p\right\}.$$

See Ibragimov and Sharakhmetov (1997) for the constant C_p. This is (1) for $p\geq 3$ with $A_p=2^pC_p$. ■

For the convenience of English readers I state here Corollary 4 from Utev (1985), kindly translated for me by Rustam Ibragimov.

LEMMA 5. *(Utev, 1985) Let $\psi_1,\dots,\psi_n$ be independent symmetrically distributed random variables taking values in a separable Hilbert space with* *norm* $\|\cdot\|$. *Let* $u_1, ..., u_n$ *be independent real-valued Rademacher random variables independent of* ψ_i. *Then for all* $p \geq 3$

$$\mathbb{E}\left\|\sum_{i=1}^{n} \psi_i\right\|^p \leq \mathbb{E}\left|\sum_{i=1}^{n} \|\psi_i\| u_i\right|^p.$$

Proof of Lemma 1. Let $\|a\| = \left(a'a\right)^{1/2}$ denote the usual Euclidean norm for vectors a. It is useful to observe that

$$\mathbb{E}\|x_{mi}\|^2 = \mathbb{E}\operatorname{tr} x_{mi}x_{mi}' = \operatorname{tr} I_{K_m} = K_m. \tag{24}$$

Also, since $\|x_{mi}\| \leq \sup_{z \in \mathcal{Z}} \|x_m(z)\| = \zeta_m$, then for any $q \geq 2$,

$$\mathbb{E}\|x_{mi}\|^q = \mathbb{E}\|x_{mi}\|^{q-2}\|x_{mi}\|^2 \leq \zeta_m^{q-2}\mathbb{E}\|x_{mi}\|^2 = \zeta_m^{q-2}K_m. \tag{25}$$

Thus by the MSE minimizing property of the mean and (25) with $q = 4$

$$\mathbb{E}\left\|x_{mi}x_{mi}' - I_{K_m}\right\|_F^2 \leq \mathbb{E}\left\|x_{mi}x_{mi}'\right\|_F^2 = \mathbb{E}\|x_{mi}\|^4 \leq \zeta_m^2 K_m. \tag{26}$$

Suppose $p \leq 2$. By Liapunov's inequality, Theorem 1 (2), and (26),

$$\begin{aligned}
\mathbb{E}\left\|\widehat{Q}_m - I_{K_m}\right\|_F^p &\leq \left(\mathbb{E}\left\|\widehat{Q}_m - I_{K_m}\right\|_F^2\right)^{p/2} \\
&= \left(n^{-2}\mathbb{E}\left\|\sum_{i=1}^{n}\left(x_{mi}x_{mi}' - I_{K_m}\right)\right\|_F^2\right)^{p/2} \\
&\leq \left(\frac{\mathbb{E}\left\|x_{mi}x_{mi}'\right\|_F^2}{n}\right)^{p/2} \\
&\leq \left(\frac{\zeta_m^2 K_m}{n}\right)^{p/2} = \Psi_{nm}^p,
\end{aligned}$$

which is (6).

Next, suppose $p > 2$. Using Loeve's C_r inequality, Liapunov's inequality, and (25) with $q = 2p$,

$$\begin{aligned}
\mathbb{E}\left\|x_{mi}x_{mi}' - I_{K_m}\right\|_F^p &\leq 2^{p-1}\left(\mathbb{E}\left\|x_{mi}x_{mi}'\right\|_F^p + \left\|I_{K_m}\right\|_F^p\right) \\
&= 2^p\mathbb{E}\|x_{mi}\|^{2p} \\
&\leq 2^p\zeta_m^{2p-2}K_m.
\end{aligned} \tag{27}$$

Then by Theorem 1, (26), and (27)

$$
\begin{aligned}
\mathbb{E}\left\|\widehat{Q}_m - I_{K_m}\right\|_F^p &= n^{-p}\mathbb{E}\left\|\sum_{i=1}^n \left(x_{mi}x_{mi}' - I_{K_m}\right)\right\|_F^p \\
&\le A_p\left\{\left(\frac{\mathbb{E}\left\|x_{mi}x_{mi}' - I_{K_m}\right\|_F^2}{n}\right)^{p/2} + \frac{\mathbb{E}\left\|x_{mi}x_{mi}' - I_{K_m}\right\|_F^p}{n^{p-1}}\right\} \\
&\le A_p\left\{\left(\frac{\zeta_m^2 K_m}{n}\right)^{p/2} + \frac{2^p\zeta_m^{2p-2}K_m}{n^{p-1}}\right\} \\
&\le A_p\left\{\left(\frac{\zeta_m^2 K_m}{n}\right)^{p/2} + 2^p\left(\frac{\zeta_m^2 K_m}{n}\right)^{(p-1)}\right\} \\
&= A_p\Psi_{nm}^p\left(1+2^p\Psi_{nm}^{p-2}\right),
\end{aligned} \tag{28}
$$

where the third inequality holds since $K_m \le K_m^{p-1}$ as $p \ge 2$, and the final equality uses definition (5). This is (7). ■

Proof of Lemma 2. The argument follows the proof of Theorem 2 of Ing and Wei (2003). While that result was developed for autoregressive regression, the method carries over to the nonparametric setting under Assumption 1.

If $p < 1$, then

$$
\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^p\right)^{1/p} \le \mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S
$$

so without loss of generality we may assume $p \ge 1$. We may also assume that $K_m \ge 1$ as for $K_m = 0$ the result is trivial.

Let $s > 6/\delta$ be an integer and set $J = 2^s p$. The first step is to establish that for some $\overline{B} < \infty$ which depends only on J, α, η, and ψ,

$$
\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^J\right)^{1/J} \le \overline{B}K_m^3. \tag{29}
$$

The proof of (29) is nearly identical to that of Lemma 1 of Ing and Wei (2003). The only difference is the method to bound the left side of their equation (2.13)

$$
\mathbb{P}\left(\bigcap_{i=1}^{tK_m}\left\{\left|\ell_j' x_{mi}\right| \le 3u^{-1/(2q)}\right\}\right) \tag{30}
$$

by the displayed expression above their (2.14). Ing and Wei develop a detailed argument using the autoregressive structure of their problem. Instead, we use the independence of the observations, and then Assumption 1.2, to observe that (30) equals

$$\mathbb{P}\left(\left|\ell_j' x_{mi}\right| \leq 3u^{-1/(2q)}\right)^{tK_m} \leq \psi^{tK_m}\left(3u^{-1/(2q)}\right)^{\alpha tK_m}$$

which is smaller than the bound above Ing–Wei's (2.14). Their argument otherwise goes through with these substitutions and we find (29).

By Minkowski's inequality, for any $1 \leq r \leq J/2$,

$$\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^r\right)^{1/r} \leq \left\|I_{K_m}\right\|_S + \left(\mathbb{E}\left\|\widehat{Q}_n^{-1} - I_{K_m}\right\|_S^r\right)^{1/r}. \tag{31}$$

By the norm inequality and the fact that $\|\boldsymbol{A}\|_S \leq \|\boldsymbol{A}\|_F$

$$\begin{aligned}\left\|\widehat{Q}_m^{-1} - I_{K_m}\right\|_S &= \left\|\left(\widehat{Q}_m - I_{K_m}\right)\widehat{Q}_m^{-1}\right\|_S \\ &\leq \left\|\widehat{Q}_m - I_{K_m}\right\|_S \left\|\widehat{Q}_m^{-1}\right\|_S \\ &\leq \left\|\widehat{Q}_m - I_{K_m}\right\|_F \left\|\widehat{Q}_m^{-1}\right\|_S.\end{aligned} \tag{32}$$

Applying the Cauchy–Schwarz inequality

$$\left(\mathbb{E}\left\|\widehat{Q}_m^{-1} - I_{K_m}\right\|_S^r\right)^{1/r} \leq \left(\mathbb{E}\left\|\widehat{Q}_m - I_{K_m}\right\|_F^{2r}\right)^{1/2r}\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2r}\right)^{1/2r}. \tag{33}$$

Set $\psi = (1+\eta)^{1/p} - 1$ and

$$\varepsilon = \min\left[\frac{\psi}{2+\psi}, \left(\frac{\psi}{2\overline{B}}\right)^{1/s}\right]. \tag{34}$$

Lemma 1 and Assumption 1.1 imply that there is an n sufficiently large such that

$$\left(\mathbb{E}\left\|\widehat{Q}_m - I_{K_m}\right\|_F^{2r}\right)^{1/2r} \leq A_{2r}^{1/2r}\Psi_{nm}\left(1 + 2^{2r}\Psi_{nm}^{2r-2}\right)^{1/2r} \leq \varepsilon K_m^{-\delta/2}. \tag{35}$$

Equations (31)–(35) plus $\left\|I_{K_m}\right\|_S = 1$ establish that

$$\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^r\right)^{1/r} \leq 1 + \varepsilon K_m^{-\delta/2}\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2r}\right)^{1/2r}. \tag{36}$$

Iterating (36) s times, starting with $r = p$,

$$\begin{aligned}
\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^p\right)^{1/p} &\le 1+\varepsilon K_m^{-\delta/2}\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2p}\right)^{1/2p} \\
&\le 1+\varepsilon K_m^{-\delta/2}+\left(\varepsilon K_m^{-\delta/2}\right)^2 \\
&\quad +\cdots+\left(\varepsilon K_m^{-\delta/2}\right)^s\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2^s p}\right)^{1/2^s p} \\
&\le 1+\varepsilon+\varepsilon^2+\cdots+\varepsilon^{s-1}+\left(\varepsilon K_m^{-\delta/2}\right)^s \overline{B}K_m^3 \\
&\le \frac{1}{1-\varepsilon}+\varepsilon^s\overline{B} \\
&\le 1+\psi \\
&= (1+\eta)^{1/p},
\end{aligned}$$

where the third inequality is (29), the fourth inequality uses $s > 6/\delta$, and the final uses (34). This is (8). ■

Proof of Lemma 3. For $0 < p < 2$, applying Holder's inequality to (32), and then (6) and (8), for some $\varepsilon > 0$

$$\begin{aligned}
\mathbb{E}\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^p &\le \left(\mathbb{E}\left\|\widehat{Q}_m-I_{K_m}\right\|_F^2\right)^{p/2}\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2p/(2-p)}\right)^{(2-p)/2} \\
&\le (1+\varepsilon)^{(2-p)/2}\Psi_{nm}^p,
\end{aligned}$$

which is (9) with $\eta = (1+\varepsilon)^{(2-p)/2}-1$.

For $p \ge 2$, Lemma 1 and Assumption 1.1 imply that for any $\eta > 0$ there is an n sufficiently large such that

$$\left(\mathbb{E}\left\|\widehat{Q}_m-I_{K_m}\right\|_F^{2p}\right)^{1/2} \le A_{2p}^{1/2}\Psi_{nm}^p\left(1+2^{2p}\Psi_{nm}^{2p-2}\right)^{1/2} \le (1+\eta)^{1/2}A_{2p}^{1/2}\Psi_{nm}^p. \tag{37}$$

Then by (33), (37), and (8), we obtain for n sufficiently large

$$\mathbb{E}\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^p \le \left(\mathbb{E}\left\|\widehat{Q}_m-I_{K_m}\right\|_F^{2p}\right)^{1/2}\left(\mathbb{E}\left\|\widehat{Q}_m^{-1}\right\|_S^{2p}\right)^{1/2} \le (1+\eta)A_{2p}^{1/2}\Psi_{nm}^p.$$

This is (10). ■

Proof of Lemma 4. Observe that under Assumption 2.1, the IMSE (16) is unaffected if we replace the regressors x_{mi} with $x_{mi}^* = Q_m^{-1/2}x_{mi}$ which has the implication that $\mathbb{E}\left(x_{mi}^*x_{mi}^{*\prime}\right) = I_{K_m}$. For convenience and without loss of generality we shall simply assume that this transformation has been made, or equivalently that $Q_m = I_{K_m}$, and thus we can apply Lemmas 1–3 without modification.

Define $\widehat{S}_m = \frac{1}{n}\sum_{i=1}^n x_{mi}e_{mi}$ so that $\widehat{\beta}_m - \beta_m = \widehat{Q}_m^{-1}\widehat{S}_m$. Since $\widehat{S}_m$ is the average of iid mean zero random vectors, we calculate that

$$\mathbb{E}\left(\widehat{S}_m'\widehat{S}_m\right) = \frac{1}{n}\mathbb{E}\left(x_{mi}'x_{mi}e_{mi}^2\right) = \frac{\operatorname{tr}\Omega_m^*}{n}.$$

Thus

$$\begin{aligned}\mathbb{E}\left(\left(\widehat{\beta}_m-\beta_m\right)'\left(\widehat{\beta}_m-\beta_m\right)\right)-\left(\frac{\operatorname{tr}\Omega_m^*}{n}\right) &= \mathbb{E}\left(\widehat{S}_m'\widehat{Q}_m^{-1}\widehat{Q}_m^{-1}\widehat{S}_m\right)-\left(\frac{\operatorname{tr}\Omega_m^*}{n}\right)\\ &= \mathbb{E}\left(\widehat{S}_m'\left(\widehat{Q}_m^{-1}\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{S}_m\right). \qquad (38)\end{aligned}$$

Now define $\widehat{S}_m^0 = \frac{1}{n}\sum_{i=1}^n x_{mi}e_i$ and $\widehat{\gamma}_m = \frac{1}{n}\sum_{i=1}^n x_{mi}r_{mi}$ so that $\widehat{S}_m = \widehat{S}_m^0 + \widehat{\gamma}_m$. Define $Z = (z_1, ..., z_n)$ and note that $\widehat{\gamma}_m$ is measurable with respect to Z, $\mathbb{E}\left(\widehat{S}_m^0|Z\right) = 0$ and $\mathbb{E}\left(\widehat{S}_m^0\widehat{S}_m^{0\prime}|Z\right) = \frac{1}{n}\widehat{\Omega}_m$ where $\widehat{\Omega}_m = \frac{1}{n}\sum_{i=1}^n x_{mi}x_{mi}'\sigma_i^2$. Using the law of iterated expectations, we find that (38) equals

$$\begin{aligned}&\mathbb{E}\left(\mathbb{E}\left(\widehat{S}_m'\left(\widehat{Q}_m^{-1}\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{S}_m|Z\right)\right)\\
&\quad = \mathbb{E}\left(\widehat{\gamma}_m'\left(\widehat{Q}_m^{-1}\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\gamma}_m\right)+\frac{1}{n}\mathbb{E}\operatorname{tr}\left(\left(\widehat{Q}_m^{-1}\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\Omega}_m\right)\\
&\quad = 2\mathbb{E}\left(\widehat{\gamma}_m'\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\gamma}_m\right) \qquad (39)\\
&\qquad +\mathbb{E}\left(\widehat{\gamma}_m'\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\gamma}_m\right) \qquad (40)\\
&\qquad +\frac{2}{n}\mathbb{E}\operatorname{tr}\left(\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\Omega}_m\right) \qquad (41)\\
&\qquad +\frac{1}{n}\mathbb{E}\operatorname{tr}\left(\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\Omega}_m\right). \qquad (42)\end{aligned}$$

We now bound the four terms (39)–(42).

> From Assumption 2.2 and (24) we deduce

$$\operatorname{tr}\left(\Omega_m^*\right) \geq \mathbb{E}\left(x_{mi}'x_{mi}\sigma_i^2\right) \geq \mathbb{E}\left(x_{mi}'x_{mi}\right)\underline{\sigma}^2 = K_m\underline{\sigma}^2,$$

so that

$$K_m \leq \frac{\operatorname{tr}\left(\Omega_m^*\right)}{\underline{\sigma}^2}. \qquad (43)$$

Furthermore, note that since K_m is a nonnegative integer and $\zeta_m^2 = 0$ if $K_m = 0$ then (43) implies

$$\frac{\zeta_m^2}{n} \leq \frac{\zeta_m^2 K_m^2}{n} \leq \Psi_{nm}^2\frac{\operatorname{tr}\left(\Omega_m^*\right)}{\underline{\sigma}^2}. \qquad (44)$$

We will also make use of the trace inequality, which states that for symmetric $\ell \times \ell$ A and positive semidefinite B,

$$|\mathrm{tr}(AB)| \leq \|A\|_S \,\mathrm{tr}(B). \tag{45}$$

For vectors a, let $\|a\| = \left(a'a\right)^{1/2}$ denote the Euclidean norm. Notice that

$$\mathbb{E}\,\|x_{mi}r_{mi}\|^2 = \mathbb{E}\left(x'_{mi}x_{mi}r^2_{mi}\right) \leq \mathrm{tr}\left(\Omega^*_m\right). \tag{46}$$

Define

$$\overline{r}_m = \sup_{z\in\mathcal{Z}} |r_m(z)| \tag{47}$$

which is bounded under Assumption 2.3. Using (44), (46), and (47),

$$n^{-3}\mathbb{E}\,\|x_{mi}r_{mi}\|^4 \leq \frac{\overline{r}^2_m \zeta^2_m \mathbb{E}\,\|x_{mi}r_{mi}\|^2}{n^3} \leq \frac{\overline{r}^2_m}{\underline{\sigma}^2}\Psi^2_{nm}\left(\frac{\mathrm{tr}\left(\Omega^*_m\right)}{n}\right)^2. \tag{48}$$

Since $\|a\| = \|a\|_F$, we can apply Theorem 1. Using (1) with $p = 4$, (46), and (48),

$$\begin{aligned}
\left(\mathbb{E}\,\|\widehat{\gamma}_m\|^4\right)^{1/2} &\leq A_4^{1/2}\left\{\left(\frac{1}{n}\mathbb{E}\,\|x_{mi}r_{mi}\|^2\right)^2 + n^{-3}\mathbb{E}\,\|x_{mi}r_{mi}\|^4\right\}^{1/2}\\
&\leq A_4^{1/2}\left\{\left(\frac{\mathrm{tr}\left(\Omega^*_m\right)}{n}\right)^2 + \frac{\overline{r}^2_m}{\underline{\sigma}^2}\Psi^2_{nm}\left(\frac{\mathrm{tr}\left(\Omega^*_m\right)}{n}\right)^2\right\}^{1/2}\\
&= A_4^{1/2}\left(1 + \frac{\overline{r}^2_m}{\underline{\sigma}^2}\Psi^2_{nm}\right)^{1/2}\frac{\mathrm{tr}\left(\Omega^*_m\right)}{n}.
\end{aligned} \tag{49}$$

Now, consider (39). Using the trace inequality (45), the Cauchy–Schwarz inequality, Lemma 3, and (49),

$$\begin{aligned}
\mathbb{E}\left|\widehat{\gamma}'_m\left(\widehat{Q}^{-1}_m - I_{K_m}\right)\widehat{\gamma}_m\right| &\leq \mathbb{E}\left(\left\|\widehat{Q}^{-1}_m - I_{K_m}\right\|_S \widehat{\gamma}'_m\widehat{\gamma}_m\right)\\
&\leq \left(\mathbb{E}\left\|\widehat{Q}^{-1}_m - I_{K_m}\right\|^2_S\right)^{1/2}\left(\mathbb{E}\,\|\widehat{\gamma}_m\|^4\right)^{1/2}\\
&\leq (1+\eta)^{1/2} A_4^{3/4}\left(1 + \frac{\overline{r}^2_m}{\underline{\sigma}^2}\Psi^2_{nm}\right)^{1/2}\Psi_{nm}\frac{\mathrm{tr}\left(\Omega^*_m\right)}{n}.
\end{aligned} \tag{50}$$

Similarly, to bound (40),

$$\begin{aligned}
&\mathbb{E}\left|\widehat{\gamma}_m'\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\gamma}_m\right| \\
&\quad\leq \mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^2\widehat{\gamma}_m'\widehat{\gamma}_m\right) \\
&\quad\leq \left(\mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^4\right)\right)^{1/2}\left(\mathbb{E}\left\|\widehat{\gamma}_m\right\|^4\right)^{1/2} \\
&\quad\leq (1+\eta)^{1/2}A_8^{1/2}A_4^{1/2}\left(1+\frac{\overline{r}_m^2}{\underline{\sigma}^2}\Psi_{nm}^2\right)^{1/2}\Psi_{nm}^2\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}.
\end{aligned} \tag{51}$$

For (41), first observe that $\operatorname{tr}(\widehat{\Omega}_m)=\frac{1}{n}\sum_{i=1}^n x_{mi}'x_{mi}\sigma_i^2$ so

$$\mathbb{E}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)=\mathbb{E}\left(x_{mi}'x_{mi}\sigma_i^2\right)\leq\operatorname{tr}\left(\Omega_m^*\right).$$

Using (44),

$$\begin{aligned}
n^{-3}\operatorname{var}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)=n^{-4}\mathbb{E}\left(\|x_{mi}\|^4\sigma_i^4\right)&\leq\frac{\zeta_m^2\overline{\sigma}^2}{n^4}\mathbb{E}\left(\|x_{mi}\|^2\sigma_i^2\right) \\
&\leq\frac{\overline{\sigma}^2}{\underline{\sigma}^2}\Psi_{nm}^2\left(\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}\right)^2,
\end{aligned}$$

and thus

$$\begin{aligned}
\left(n^{-2}\mathbb{E}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)^2\right)^{1/2}&=\left(\frac{1}{n^3}\operatorname{var}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)+\left(\frac{1}{n}\mathbb{E}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)\right)^2\right)^{1/2} \\
&\leq\left(\frac{\overline{\sigma}^2}{\underline{\sigma}^2}\Psi_{nm}^2\left(\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}\right)^2+\left(\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}\right)^2\right)^{1/2} \\
&=\left(1+\frac{\overline{\sigma}^2}{\underline{\sigma}^2}\Psi_{nm}^2\right)^{1/2}\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}.
\end{aligned} \tag{52}$$

By the trace inequality, Lemma 3, and (52),

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\left|\operatorname{tr}\left(\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\Omega}_m\right)\right|&\leq\frac{1}{n}\mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S\operatorname{tr}\left(\widehat{\Omega}_m\right)\right) \\
&\leq\left(\mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^2\right)\right)^{1/2}\left(n^{-1}\mathbb{E}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)^2\right)^{1/2} \\
&\leq(1+\eta)^{1/2}A_4^{1/4}\Psi_{nm}\left(1+\frac{\overline{\sigma}^2}{\underline{\sigma}^2}\Psi_{nm}^2\right)^{1/2}\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n},
\end{aligned} \tag{53}$$

bounding (41).

Similarly, to bound (42),

$$
\begin{aligned}
&\frac{1}{n}\mathbb{E}\left|\operatorname{tr}\left(\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\left(\widehat{Q}_m^{-1}-I_{K_m}\right)\widehat{\Omega}_m\right)\right| \\
&\quad \le \frac{1}{n}\mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^2 \operatorname{tr}\left(\widehat{\Omega}_m\right)\right) \\
&\quad \le \left(\mathbb{E}\left(\left\|\widehat{Q}_m^{-1}-I_{K_m}\right\|_S^4\right)\right)^{1/2}\left(n^{-2}\mathbb{E}\left(\operatorname{tr}(\widehat{\Omega}_m)\right)^2\right)^{1/2} \\
&\quad \le (1+\eta)^{1/2} A_8^{1/4}\Psi_{nm}^2\left(1+\frac{\overline{\sigma}^2}{\underline{\sigma}^2}\Psi_{nm}^2\right)^{1/2}\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}. \qquad \textbf{(54)}
\end{aligned}
$$

Together, (50), (51), (53), and (54), plus $\Psi_{nm}\to 0$ uniformly in $m\le M_n$ under Assumption 1.1, we find that the absolute value of (39)–(42) is bounded by

$$
\begin{aligned}
&\left[\left(2(1+\eta)^{1/2}A_4^{3/4}+(1+\eta)^{1/2}A_4^{1/4}\right)(1+o(1))+o(1)\right]\Psi_{nm}\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} \\
&\quad \le A_4\Psi_{nm}\frac{\operatorname{tr}\left(\Omega_m^*\right)}{n}
\end{aligned}
$$

the inequality using $3\le A_4^{1/4}$, sufficiently small η, and sufficiently large n. This is (16). ■

Proof of Theorem 2. Using (13) and (14) we see that

$$
IMSE_n(m)-I_n^*(m)=\mathbb{E}\left(\left(\widehat{\beta}_m-\beta_m\right)'Q_m\left(\widehat{\beta}_m-\beta_m\right)\right)-\frac{\operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)}{n}.
$$

Using Lemma 4 and $\sup_{1\le m\le M_n}\Psi_{nm}=o(1)$

$$
\left|\frac{IMSE_n(m)-I_n^*(m)}{I_n^*(m)}\right|\le\frac{A_4\Psi_{nm}\dfrac{\operatorname{tr}\left(Q_m^{-1}\Omega_m^*\right)}{n}}{I_n^*(m)}\le o(1),
$$

uniformly in $m\le M_n$, which is (17).

Next, since $x_{mi}'x_{mi}\le\zeta_m^2=o(n)$ and $\mathbb{E}\left(r_{mi}^2\right)=\phi_m^2\le I_n(m)$

$$
\begin{aligned}
\left|\frac{I_n^*(m)-I_n(m)}{I_n(m)}\right| &= \frac{\operatorname{tr}\left(\Omega_m^*\right)-\operatorname{tr}\left(\Omega_m\right)}{nI_n(m)} \\
&= \frac{\mathbb{E}\left(x_{mi}'x_{mi}r_{mi}^2\right)}{nI_n(m)} \\
&\le \frac{\zeta_m^2}{n}=o(1)
\end{aligned}
$$

uniformly in $m\le M_n$. Combined with (17) this shows (18). ■

Proof of Theorem 3. Define $Q_{m\ell} = \mathbb{E}\left(x_{mi}x'_{\ell i}\right)$ and $\gamma_{m\ell} = \mathbb{E}\left(x_{mi}r_{\ell i}\right)$. Using (12),

$$\begin{aligned}
&\int_{\mathcal{Z}} \left(\left(\widehat{g}_m(z) - g(z)\right)\left(\widehat{g}_\ell(z) - g(z)\right)\right) f(z)dz \\
&\quad = \int_{\mathcal{Z}} r_m(z) r_\ell(z) f(z)dz - \left(\widehat{\beta}_m - \beta_m\right)' \int_{\mathcal{Z}} x_m(z) r_\ell(z) f(z)dz \\
&\qquad - \int_{\mathcal{Z}} r_m(z) x_\ell(z)' f(z)dz \left(\widehat{\beta}_\ell - \beta_\ell\right) \\
&\qquad + \left(\widehat{\beta}_m - \beta_m\right)' \int_{\mathcal{Z}} x_m(z) x_\ell(z)' f(z)dz \left(\widehat{\beta}_\ell - \beta_\ell\right) \\
&\quad = \mathbb{E}\left(r_{mi}r_{\ell i}\right) - \gamma'_{m\ell}\left(\widehat{\beta}_m - \beta_m\right) - \gamma'_{\ell m}\left(\widehat{\beta}_\ell - \beta_\ell\right) \\
&\qquad + \left(\widehat{\beta}_m - \beta_m\right)' Q_{m\ell}\left(\widehat{\beta}_\ell - \beta_\ell\right).
\end{aligned} \tag{55}$$

Assumption 3 (nested regressors) implies that for $m \le \ell$, $\gamma_{m\ell} = 0$. Notice also that $r_{mi} = g(z_i) - x'_{mi}\beta_m = r_{\ell i} + x'_{\ell i}\beta_\ell - x'_{mi}\beta_m$ and thus

$$\begin{aligned}
\mathbb{E}\left(r_{mi}r_{\ell i}\right) &= \mathbb{E}\left(\left(r_{\ell i} + \beta'_\ell x_{\ell i} - \beta'_m x_{mi}\right) r_{\ell i}\right) \\
&= \mathbb{E}\left(r_{\ell i}^2\right) + \beta'_\ell \mathbb{E}\left(x_{\ell i}r_{\ell i}\right) - \beta'_m \mathbb{E}\left(x_{mi}r_{\ell i}\right) \\
&= \phi_\ell^2.
\end{aligned} \tag{56}$$

As in the proof of Lemma 4, without loss of generality we assume that $Q_m = I_{K_m}$ for all m. Combined with Assumption 3 this implies $Q_{m\ell} = [I_{K_m}\ 0]$. Thus for $m \le \ell$ (55) equals

$$\phi_\ell^2 - \gamma'_{\ell m}\left(\widehat{\beta}_\ell - \beta_\ell\right) + \left(\widehat{\beta}_m - \beta_m\right)'\left(\widehat{\beta}_m - \beta_m\right).$$

Thus

$$IMSE_n(w) - I_n^*(w)$$

$$= \sum_{m=1}^{M_n} w_m^2 \left(\mathbb{E}\left(\widehat{\beta}_m - \beta_m\right)'\left(\widehat{\beta}_m - \beta_m\right) - \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} \right) \tag{57}$$

$$+2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m \left(\mathbb{E}\left(\widehat{\beta}_m - \beta_m\right)'\left(\widehat{\beta}_m - \beta_m\right) - \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} \right) \tag{58}$$

$$-2\sum_{\ell=1}^{M_n} w_\ell \overline{\gamma}'_\ell \mathbb{E}\left(\widehat{\beta}_\ell - \beta_\ell\right), \tag{59}$$

where

$$\overline{\gamma}_\ell = \sum_{m=1}^{\ell-1} w_m \gamma_{\ell m} = \mathbb{E}\left(x_{\ell i} \sum_{m=1}^{\ell-1} w_m r_{mi} \right). \tag{60}$$

From Lemma 4, the absolute value of the sum of (57) and (58) is smaller than

$$\begin{aligned}
&\sum_{m=1}^{M_n} w_m^2 A_4 \Psi_{nm} \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} + 2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m A_4 \Psi_{nm} \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} \\
&\quad \leq A_4 \max_{m \leq M_n} \Psi_{nm} \left(\sum_{m=1}^{M_n} w_m^2 \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} + 2\sum_{\ell=1}^{M_n}\sum_{m=1}^{\ell-1} w_\ell w_m \frac{\operatorname{tr}\left(\Omega_m^*\right)}{n} \right) \\
&\quad \leq o(1) I_n^*(w) \qquad (61)
\end{aligned}$$

uniformly in $w \in \mathcal{W}_n$, the final inequality since $\max_{m \leq M_n} \Psi_{nm} = o(1)$.

Now consider (59). Notice that

$$\begin{aligned}
\mathbb{E}\left|\sum_{m=1}^{\ell-1} w_m r_{mi}\right|^2 &= \sum_{m=1}^{\ell-1} w_m^2 \mathbb{E} r_{mi}^2 + 2\sum_{j=1}^{\ell-1}\sum_{m=1}^{j-1} w_j w_m \mathbb{E} r_{ji}^2 \\
&\leq \sum_{m=1}^{M_n} w_m^2 \phi_m^2 + 2\sum_{j=1}^{M_n}\sum_{m=1}^{j-1} w_j w_m \phi_j^2 \\
&\leq I_n^*(w). \qquad (62)
\end{aligned}$$

Applying the Cauchy–Schwarz inequality to (60), using (24) and (62), then

$$\left\|\overline{\gamma}_\ell\right\| \leq \left(\mathbb{E}\left\|x_{\ell i}\right\|^2\right)^{1/2} \left(\mathbb{E}\left|\sum_{m=1}^{M_n} w_m r_{mi}\right|^2\right)^{1/2} = K_\ell^{1/2} I_n^*(w)^{1/2}. \qquad (63)$$

As in the proof of Lemma 2 define $\widehat{\gamma}_\ell = n^{-1}\sum_{i=1}^n x_{\ell i} r_{\ell i}$ which is an average of iid mean zero random vectors so

$$\mathbb{E}\left\|\widehat{\gamma}_\ell\right\|^2 = \frac{1}{n}\mathbb{E}\left(x_{\ell i}' x_{\ell i} r_{\ell i}^2\right) \leq \frac{\zeta_\ell^2 \phi_\ell^2}{n}. \qquad (64)$$

By the law of iterated expectations we can deduce that $\mathbb{E}\left(\widehat{\beta}_\ell - \beta_\ell\right) = \mathbb{E}\widehat{Q}_\ell^{-1}\widehat{\gamma}_\ell = \mathbb{E}\left(\widehat{Q}_\ell^{-1} - I_{K_\ell}\right)\widehat{\gamma}_\ell$. Thus using the Cauchy–Schwarz inequality, Lemma 3, and (64),

$$\begin{aligned}
\left\|\mathbb{E}\left(\widehat{\beta}_\ell - \beta_\ell\right)\right\| &\leq \left(\mathbb{E}\left(\left\|\widehat{Q}_\ell^{-1} - I_{K_\ell}\right\|_S^2\right)\right)^{1/2} \left(\mathbb{E}\left\|\widehat{\gamma}_\ell\right\|^2\right)^{1/2} \\
&\leq 2^{1/2} A_4^{1/4} \Psi_{n\ell} \left(\frac{\zeta_\ell^2 \phi_\ell^2}{n}\right)^{1/2}. \qquad (65)
\end{aligned}$$

Applying the Triangle and Cauchy–Schwarz inequalities, (63), and (65),

$$\begin{aligned}
\left|\sum_{\ell=1}^{M_n} w_\ell \overline{\gamma}'_\ell \mathbb{E}\left(\widehat{\beta}_\ell - \beta_\ell\right)\right| &\le \sum_{\ell=1}^{M_n} w_\ell \left\|\overline{\gamma}_\ell\right\| \left\|\mathbb{E}\left(\widehat{\beta}_\ell - \beta_\ell\right)\right\| \\
&\le (1+\eta)^{1/2} A_4^{1/4} I_n^*(w)^{1/2} \sum_{\ell=1}^{M_n} w_\ell \left(\Psi_{n\ell}^2 \phi_\ell^2\right)^{1/2} \\
&\le (1+\eta)^{1/2} A_4^{1/4} I_n^*(w)^{1/2} \left(\sum_{\ell=1}^{M_n} w_\ell \Psi_{n\ell}^2 \phi_\ell^2\right)^{1/2} \\
&\le o(1) I_n^*(w), \qquad \textbf{(66)}
\end{aligned}$$

where the third inequality is Liapunov's, noting that the weights w_ℓ define a probability measure, and the final inequality uses $\max_{m \le M_n} \Psi_{nm} = o(1)$ and $\sum_{\ell=1}^{M_n} w_\ell \phi_\ell^2 \le I_n^*(w)$.

Together, (61) and (66) show that

$$\sup_{w \in \mathcal{W}_n} \left| \frac{IMSE_n(w) - I_n^*(w)}{I_n^*(w)} \right| \le o(1),$$

as stated. ■

Proof of Theorem 4. As discussed in Newey (1997, p. 150), since $g(z)$ has s continuous derivatives under Assumption 4.1, then

$$\phi_m^2 = \inf_\beta \mathbb{E}\left(g(z_i) - \beta' x_m(z_i)\right)^2 \le \inf_\beta \sup_z \left|g(z) - \beta' x_m(z)\right|^2 \le O\left(K_m^{-2s/q}\right).$$

Furthermore, as discussed earlier, $\zeta_m^2 = O(K_m)$ for splines and $\zeta_m^2 = O(K_m^2)$ for power series (see Andrews, 1991; Newey, 1995). It follows that $\zeta_m^2 \phi_m^2 = O(K_m^{1-2s/q})$ for splines and $\zeta_m^2 \phi_m^2 = O(K_m^{2-2s/q})$ for power series. In either case, Assumption 4.1 implies there is some $D < \infty$ such that for all m, $\mathbb{E}\left(x'_{mi} x_{mi} r_{mi}^2\right) \le \zeta_m^2 \phi_m^2 \le D$.

It follows that

$$\begin{aligned}
n\left|I_n^*(w) - I_n(w)\right| = &\sum_{m=1}^{M_n} w_m^2 \mathbb{E}\left(x'_{mi} x_{mi} r_{mi}^2\right) \\
&+ 2\sum_{\ell=1}^{M_n} \sum_{m=1}^{\ell-1} w_\ell w_m \mathbb{E}\left(x'_{mi} x_{mi} r_{mi}^2\right) \le D. \qquad \textbf{(67)}
\end{aligned}$$

Below we show that

$$\inf_w n I_n(w) \to \infty. \qquad \textbf{(68)}$$

Together, (67) and (68) imply that as $n \to \infty$

$$\sup_{w \in \mathcal{W}_n} \left| \frac{I_n^*(w) - I_n(w)}{I_n(w)} \right| \leq \frac{D}{\inf_w n I_n(w)} \longrightarrow 0$$

as stated.

To complete the proof, we now show (68). Let w^n denote the solution to the infimum problem (68), thus $\inf_w I_n(w) = I_n(w^n)$, and rewrite (19) as

$$n I_n(w) = \sum_{m=1}^{M_n} \sum_{\ell=1}^{M_n} w_m w_\ell \left(n\phi_{m \vee \ell}^2 + g_{m \wedge \ell} \right), \tag{69}$$

where $g_m = \operatorname{tr}\left(Q_m^{-1} \Omega_m\right)$.

Notice that $n I_n(w) \geq n\phi_{M_n}^2$, which will diverge if M_n is bounded, since $\phi_m^2 > 0$ for all m by Assumption 4.2. Thus without loss of generality we can assume that $M_n \to \infty$ as $n \to \infty$.

We prove (68) by contradiction. Suppose (68) is false. Then there is a subsequence $\{n\}$ and a constant $B < \infty$ such that for all n along this subsequence,

$$n I_n(w^n) \leq B. \tag{70}$$

For some $0 < \varepsilon < 1$ set M^* so that

$$K_{M^*} \geq \frac{B}{\underline{\sigma}^2 \varepsilon^2}. \tag{71}$$

Assume that n is sufficiently large so that $M_n \geq M^*$.

Note that (70), (69) $\phi_m^2 \geq 0$ and $g_m \geq 0$ imply that

$$\begin{aligned} B &\geq n I_n(w^n) \\ &\geq \sum_{m=M^*}^{M_n} \sum_{\ell=M^*}^{M_n} w_m^n w_\ell^n g_{m \wedge \ell} \\ &\geq \underline{\sigma}^2 \sum_{m=M^*}^{M_n} \sum_{\ell=M^*}^{M_n} w_m^n w_\ell^n K_{m \wedge \ell} \\ &\geq \underline{\sigma}^2 \left(\sum_{m=M^*}^{M_n} w_m^n \right)^2 K_{M^*} \\ &\geq \left(\sum_{m=M^*}^{M_n} w_m^n \right)^2 \frac{B}{\varepsilon^2}, \end{aligned} \tag{72}$$

where the third inequality uses $g_m \geq \underline{\sigma}^2 K_m$ which is implied by Assumption 2.2, the fourth inequality uses $K_{m \wedge \ell} \geq K_{M^*}$ for $m, \ell \geq M^*$, and the fifth inequality is (71). (72) implies that

$$\sum_{m=M^*}^{M_n} w_m^n \le \varepsilon. \tag{73}$$

By a similar argument,

$$\begin{aligned} nI_n(w^n) &\ge \sum_{m=1}^{M^*-1} \sum_{\ell=1}^{M^*-1} w_m^n w_\ell^n n\phi_{m\vee\ell}^2 \\ &\ge \left(\sum_{m=1}^{M^*-1} w_m^n\right)^2 n\phi_{M^*}^2 \\ &\ge (1-\varepsilon)^2 n\phi_{M^*}^2 \\ &\longrightarrow \infty \end{aligned}$$

the third inequality by (73) and the final convergence since $\varepsilon < 1$ and $\phi_{M^*}^2 > 0$ by Assumption 4.2. This contradicts (70), establishing (68) by contradiction and completing the proof. ■

REFERENCES

Andrews, D.W.K. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica* 59, 307–345.

Bestsennaya, E.V. & S.A. Utev (1991) An exact upper bound for the even moment of sums of independent random variables. *Siberian Mathematical Journal* 32, 139–141.

Burkholder, D.L. (1973) Distribution function inequalities for martingales. *The Annals of Probability* 1, 19–42.

Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B. North-Holland.

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

De Acosta, A. (1981) Inequalities for B-valued random vectors with applications to the strong law of large numbers. *The Annals of Probability* 9, 157–161.

De Jong, R.M. (2002) A note on 'Convergence rates and asymptotic normality for series estimators': Uniform convergence rates. *Journal of Econometrics* 111, 1–9.

de la Peña, V.H. & E. Giné (1999) *Decoupling: From Dependence to Independence*. Springer.

de la Peña, V.H., R. Ibragimov, & S. Sharakhmetov (2003) On extremal distributions and sharp L_p-bounds for sums of multilinear forms. *The Annals of Probability* 31, 630–675.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

Hansen, B.E. (2007) Least squares model averaging. *Econometrica* 75, 1175–1189.

Hansen, B.E. (2014) Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. In J. S. Racine, L. Su & A. Ullah (eds.), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press.

Hansen, B.E. & J.S. Racine (2012) Jackknife model averaging. *Journal of Econometrics*, 167, 38–46.

Ibragimov, R. (1997) Estimates for moments of symmetric statistics. Ph.D. dissertation, Institute of Mathematics of Uzbek Academy of Sciences.

Ibragimov, M. & R. Ibragimov (2008) Optimal constants in the Rosenthal inequality for random variables with zero odd moments. *Statistics and Probability Letters* 78, 186–189.

Ibragimov, R. & S. Sharakhmetov (1997) On an exact constant for the Rosenthal inequality. *Theory of Probability and Its Applications* 42, 294–302.

Ibragimov, R. & S. Sharakhmetov (2002) The exact constant in the Rosenthal inequality for random variables with mean zero. *Theory of Probability and Its Applications* 46, 127–131.

Ing, C.-K. & C.-Z. Wei (2003) On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85, 130–155.

Lai, T.L. & C.-Z. Wei (1984) Moment inequalities with applications to regression and time series models. *Inequalities in Statistics and Probability*. IMS Lecture Notes Monograph Series, vol. 5, pp. 165–172. Institute of Mathematical Statistics.

Li, Q. & J.S. Racine (2006) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Marcinkiewicz, J. & A. Zygmund (1937) Sur les foncions independantes. *Fund. Math.* 28, 60–90.

Newey, W.K. (1995) Convergence rates for series estimators. In G.S. Maddalla, P.C.B. Phillips, & T.N. Srinavasan (eds.), *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao*, pp. 254–275. Blackwell.

Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.

Nze, P.A. & P. Doukhan (2004) Weak dependence: Models and applications to econometrics. *Econometric Theory* 20, 995–1045.

Rosenthal, H.P. (1970) On the subspaces of L_p $(p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics* 8, 273–303.

Rosenthal, H.P. (1972) On the span in L^p of sequences of independent random variables. *Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 149–167. University of California Press.

Song, K. (2008) A uniform convergence of series estimators over function spaces. *Econometric Theory* 24, 1463–1499.

Utev, S.A. (1985) Extremal problems in moment inequalities. *Proceedings of the Mathematical Institute of the Siberian Branch of the USSR Academy of Sciences* 5, 56–75.

Utev, S.A. (1991) A method for investigating the sums of weakly dependent random variables. *Siberian Mathematical Journal* 32, 675–690.