The Asymptotic IMSE of Averaging Series Regression

Bruce E. Hansen*
University of Wisconsin†

Februrary 2013
Preliminary. Do not cite.

Abstract

This paper investigates the asymptotic integrated mean squared error (IMSE) of series regression. Least-squares and averaging least-squares estimators are investigated. We characterize the optimal asymptotic IMSE using the optimal sequence of least-squares estimators and averaging weights. We find that under standard assumptions for series regression, the asymptotic IMSE of the optimal averaging estimator is considerably lower than that of the optimal least-squares estimator. The difference depends on the rate of decay of the series approximation error, the smoothness of the approximation error, and whether or not the averaging weights are constrained to be positive. When the approximation error is not smooth, we find significant reduction in asymptotic IMSE by allowing negative weights.

*Research supported by the National Science Foundation.
†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

Series methods are becoming increasing popular in econometrics. A series method approximates an unknown function by a finite order series expansion, estimates this finite order model, and conducts inference acknowledging the approximation. The most well developed theory is for regression estimation, though series methods are widely applied in other estimation contexts as well.

The most common estimation method is to estimate a single finite order model using a conventional estimator. In regression, this means estimating a finite order approximation by least-squares. An alternative suggested by Hansen (2007, 2014a, 2014b) and Hansen and Racine (2012) is to take an average across a set of finite order approximations. The infeasible optimal averaging estimator must have (weakly) lower integrated mean squared error (IMSE) than the infeasible optimal single model estimator, but it is a priori unclear if this difference is negligible asymptotically.

This paper explores this latter question. We derive explicit expressions for the IMSE of series least-squares estimators and averaging least-squares estimators in the regression context. In general, the IMSE depends in a complicated way on the approximation error of the series approximations, making general statements difficult. To give precise answers, we calculate the asymptotic IMSE under specific structures for the approximation error, allowing for power law decay, exponential decay, and for a specific type of non-smooth decay. Under these specifications we are able to calculate closed-form expressions for the IMSE.

We find that for a wide range of specifications the asymptotic IMSE of the averaging estimators is strictly less than the IMSE of the standard series estimators. The magnitude of the difference depends on whether the approximation error is exponential, power law, or non-smooth. Under non-smooth decay, the IMSE of the unconstrained averaging estimator can be arbitrarily smaller than the IMSE of the standard estimator.

The analysis also points to a previously unnoticed feature concerning weight choice. The asymptotic IMSE of the averaging estimator can be reduced by allowing the averaging weights to be negative. Instead the relevant constraint is for the cumulative weights to be non-negative.

The theory of nonparametric series regression was developed by Andrews (1991a) and Newey (1995, 1997). Series estimation for semiparametric models is studied in Chen (2007). The theory of series selection by cross-validation was developed by Li (1987), Andrews (1991b) and Hansen (2014b), and for averaging regression by Hansen and Racine (2012). Nonparametric series methods are reviewed by Li and Racine (2006).

2 Series Regression

Consider a sample of iid observations (y_i, z_i), $i = 1, ..., n$ where $z_i \in \mathcal{Z}$, a compact subset of $\mathbb{R}^q$. Define the conditional mean $g(z) = \mathbb{E}(y_i \mid z_i = z)$, the regression error $e_i = y_i - g(z_i)$, and the conditional variance $\sigma_i^2 = \mathbb{E}\left(e_i^2 \mid z_i\right)$.

We examine the estimation of $g(z)$ by series regression under the assumption that $g(z)$ is continuous. Let $\tau_j(z)$, $j = 1, 2, ...$ be a sequence of basis functions from a polynomial series or

nested spline expansion. For example, a power series sets $\tau_j(z) = z^{j-1}$. Construct the regressors $x_{ji} = \tau_j(z_i)$. For $m = 1, 2, ...$, let $X_m(z) = (\tau_1(z), ..., \tau_m(z))'$ and $X_{mi} = (x_{1i}, ..., x_{mi})'$ be the first m terms listed in an $m \times 1$ vector.

A series regression approximates the conditional mean $g(z)$ by a linear projection of y_i on X_{mi}. We can write this approximating model as

$$y_i = X_{mi}'\beta_m + e_{mi} \tag{1}$$

where the coefficient is defined by linear projection

$$\beta_m = \left(\mathbb{E}\left(X_{mi}X_{mi}'\right)\right)^{-1}\mathbb{E}\left(X_{mi}y_i\right). \tag{2}$$

The m^{th} series approximation to $g(z)$ is the linear function $X_m(z)'\beta_m$. The corresponding approximation error is $r_m(z) = g(z) - X_m(z)'\beta_m$. Set $r_{mi} = r_m(z_i)$ and define its squared error $\phi_m = \mathbb{E}\left(r_{mi}^2\right)$.

The approximation errors $\phi_m = \mathbb{E}\left(r_{mi}^2\right)$ and their differences $\Delta\phi_m = \phi_{m-1} - \phi_m$ will play a major role in our analysis. It is useful to observe that ϕ_m is non-negative, (weakly) monotonically decreasing, and asymptotes to zero as $m \to \infty$. The differences $\Delta\phi_m$ are also non-negative and asymptote to zero, but are not necessarily monotonic.

To gain some intuition consider the case of orthonormal regressors in which case the j^{th} coefficient is $b_j = \mathbb{E}\left(x_{ji}y_i\right)$,

$$\phi_m = \sum_{j=m+1}^{\infty} b_j^2,$$

and $\Delta\phi_m = b_m^2$. It is easy to see in this case that both ϕ_m and $\Delta\phi_m$ decline to zero as $m \to \infty$, but while ϕ_m is monotonic, $\Delta\phi_m = b_m^2$ is not generally monotonic.

Another set of important constants are

$$v_m = \mathrm{tr}\left(Q_m^{-1}\Omega_m\right)$$

and their differences $\Delta v_m = v_m - v_{m-1}$, where $Q_m = \mathbb{E}\left(X_{mi}X_{mi}'\right)$ and $\Omega_m = \mathbb{E}\left(X_{mi}X_{mi}'\sigma_i^2\right)$. To gain some intuition consider the case when the errors are conditionally homoskedastic $\mathbb{E}\left(e_i^2 \mid z_i\right) = \sigma^2$, in which case $v_m = \sigma^2 m$ and $\Delta v_m = \sigma^2$.

We will be considering series approximations of order $m = 0, 1, 2, ..., M_n$ for some $M_n \to \infty$. For $m = 0$ set $X_{mi} = \emptyset$ and $v_0 = 0$.

We now describe our regularity conditions. As is common in series regression, the largest estimated model M_n will be constrained by the rate of growth of the constants

$$\zeta_m = \sup_{z \in \mathcal{Z}}\left(X_m(z)'Q_m^{-1}X_m(z)\right)^{1/2}, \tag{3}$$

the largest normalized Euclidean length of the regressor vector. Under standard conditions for

series regression, ζ_m will be a bounded function of the dimension m. For example, when $\tau_j(z)$ is a power series then $\zeta_m^2 = O(m^2)$ (see Andrews (1991a)), and when $\tau_j(z)$ is a regression spline then $\zeta_m^2 = O(m)$ (see Newey (1995)). For further discussion see Newey (1997) and Qi and Racine (2006).

Assumption 1

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. *For some $\delta > 0$, $\eta > 0$, and $\psi < \infty$, for all $\ell' Q_m \ell = 1$ and $0 \leq u \leq \eta$, $\sup_{m \geq 1} \mathbb{P}\left(|\ell' X_{mi}| \leq u\right) \leq \psi u^{\delta}$.*
3. *For $m \geq 1$, $\left\|Q_m^{-1}\right\|_S \leq B < \infty$.*
4. *$0 < \underline{\sigma}^2 \leq \sigma_i^2 \leq \overline{\sigma}^2 < \infty$.*
5. *For some $\delta > 0$, $\zeta_{M_n}^2 M_n^{1+\delta}/n \longrightarrow 0$.*
6. *$\phi_m^2 > 0$ for all m.*

Assumption 1.1 is a standard smoothness condition.

Assumption 1.2 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.

Assumption 1.3 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 1.3 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 1.4 controls the degree of conditional heteroskedasticity, bounding the conditional variance away from zero and infinity.

Assumption 1.5 puts a bound on the maximal number of series terms M_n relative to the sample size. For a polynomial series expansion this requirement is satisfied when $M_n^{3+\delta}/n = o(1)$ and for a spline expansion it is satisfied when $M_n^{2+\delta}/n = o(1)$. It indirectly bounds the number of possible series approximations $M_n + 1$.

Assumption 1.6 states that all approximating projection models are approximations, so that no approximating model has zero error.

3 Estimators

The standard estimator of (2) is least-squares of y_i on X_{mi}:

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

and the corresponding series estimator of $g(z)$ is

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

For $m = 0$ we define the estimator as $\widehat{g}_0(z) = 0$.

Averaging least-squares estimators are obtained by averaging across the individual least squares estimators. Let $w = (w_0, w_1, ..., w_{M_n})$ be a set of weights which sum to one. An averaging least-squares estimator is defined as

$$\widehat{g}_w(z) = \sum_{m=0}^{M_n} w_m \widehat{g}_m(z).$$

4 Integrated Mean Squared Error

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

$$IMSE_n(m) = \int_{\mathcal{Z}} \mathbb{E}\left(\widehat{g}_m(z) - g(z)\right)^2 f(z) dz \tag{4}$$

where $f(z)$ is the marginal density of z_i. The IMSE of the averaging estimator with weight vector w is

$$IMSE_n(w) = \int_{\mathcal{Z}} \mathbb{E}\left(\widehat{g}_w(z) - g(z)\right)^2 f(z) dz. \tag{5}$$

Hansen (2014a) established the following uniform approximations to $IMSE_n(m)$ and $IMSE_n(w)$.

Theorem 1 *Under Assumption 1, as* $n \to \infty$

$$\sup_{1 \le m \le M_n} \left| \frac{IMSE_n(m) - I_n(m)}{I_n(m)} \right| \longrightarrow 0 \tag{6}$$

and

$$\sup_{w \in \mathcal{W}_n} \left| \frac{IMSE_n(w) - I_n(w)}{I_n(w)} \right| \longrightarrow 0 \tag{7}$$

where $\mathcal{W}_n$ *is the* M_n*-dimensional unit simplex,*

$$I_n(m) = \phi_m + \frac{v_m}{n} \tag{8}$$

and

$$I_n(w) = \sum_{m=0}^{M_n} w_m^2 \left(\phi_m + \frac{v_m}{n}\right) + 2 \sum_{\ell=0}^{M_n} \sum_{m=0}^{\ell-1} w_\ell w_m \left(\phi_\ell + \frac{v_m}{n}\right). \tag{9}$$

Theorem 1 shows that the IMSE of the series estimator $\widehat{g}_m$ and the averaging estimator $\widehat{g}_w$ are asymptotically equivalent to $I_n(m)$ and $I_n(w)$, respectively, uniformly in $m \le M_n$ and $w \in \mathcal{W}_n$. Thus to characterize the asymptotic optimal IMSE of these estimators, it is sufficient to focus on

$I_n(m)$ and $I_n(w)$. We define the asymptotic optimal IMSE of the series estimator as

$$I_n^1 = \inf_m I_n(m),$$

the asymptotic optimal IMSE of the averaging estimator where the weights are constrained to the unit simplex as

$$I_n^2 = \inf_{w \in \mathcal{W}_n} I_n(w),$$

and the asymptotic optimal IMSE of the averaging estimator with unconstrained weights as

$$I_n^3 = \inf_w I_n(w).$$

By construction, $I_n^1 \leq I_n^2 \leq I_n^3$. What is unclear is whether or not (and in which situation) the differences are strict inequalities or meaningfully large. To answer this question, our goal in the subsequent sections is to calculate the asymptotic behavior of I_n^1, I_n^2 and I_n^3.

As will become clear shortly, allowing for unconstrained weights will be of particular interest. A difficulty is that Theorem 1 only established the uniform equivalence of IMSE and $I_n(w)$ for weights in the unit simplex, but this was a technical and not a substantive restriction. For the moment we will investigate $I_n(w)$ as if it is a valid approximation to $IMSE_n(w)$ beyond the unit simplex, and postpone to later the technical justification of this extension.

5 Cumulative Weights

As noted by Hansen (2014b), (9) can be written as

$$I_n(w) = \sum_{m=0}^{M_n} \left(w_m^* \phi_m + w_m^{**} \frac{v_m}{n} \right)$$

where

$$w_m^* = w_m^2 + 2w_m \sum_{\ell=0}^{m-1} w_\ell$$

$$w_m^{**} = w_m^2 + 2w_m \sum_{\ell=m+1}^{M_n} w_\ell.$$

Now define the cumulative weights

$$c_m = \sum_{\ell=0}^{m} w_\ell$$

which satisfy $c_{M_n} = 1$. A simple calculation reveals that

$$c_m^2 - c_{m-1}^2 = w_m^*$$

and

$$(1-c_{m-1})^2 - (1-c_m)^2 = w_m^{**}$$

(using the convention $c_{-1}=0$). Given these relationships, $I_n(w)$ can be equivalently written as a function of $c=(c_0,c_1,...,c_{M_n})$

$$\begin{aligned} I_n(c) &= \sum_{m=0}^{M_n}\left(\left(c_m^2 - c_{m-1}^2\right)\phi_m + \left((1-c_{m-1})^2-(1-c_m)^2\right)\frac{v_m}{n}\right) \\ &= \phi_{M_n} + \sum_{m=0}^{M_n-1}\left(c_m^2\Delta\phi_{m+1} + (1-c_m)^2\frac{\Delta v_{m+1}}{n}\right) \end{aligned} \tag{10}$$

where we have used the fact that $v_0=0$.

Expression (10) is particularly useful for it shows that the asymptotic IMSE is a simple quadratic function of the cumulative weights c_m. It is therefore elementary to calculate the IMSE-minimizing weights.

Specifically, the cumulative weights c_m which minimize $I_n(c)$ are

$$c_m = \frac{\Delta v_{m+1}}{n\Delta\phi_{m+1}+\Delta v_{m+1}} \tag{11}$$

with minimized value

$$I_n^3 = \inf_c I_n(c) = \phi_{M_n} + \sum_{m=1}^{M_n}\left(\frac{\Delta v_m \Delta\phi_m}{n\Delta\phi_m+\Delta v_m}\right). \tag{12}$$

Interestingly and somewhat surprisingly, the optimal cumulative weights c_m satisfy $0\le c_m\le 1$ but are not necessarily monotonic. Thus the corresponding optimal weights $w_m = c_m - c_{m-1}$ satisfy $-1\le w_m\le 1$ and are not necessarily non-negative. This suggests that the restriction of the weights to $0\le w_m\le 1$ may disallow potential reductions in IMSE.

It is instructive to examine the formula for the optimal cumulative weights (11) to understand the benefits of allowing for negative weights w_m. If $\Delta v_m = v$ is a constant (as occurs under conditional homoskedasticity) then the cumulative weights c_m are monotonically increasing in m if the differences $\Delta\phi_m$ are monotonically decreasing (and in this case the optimal weights w_m are non-negative). To understand this situation consider the case of orthonormal regressors so that $\Delta\phi_m = b_m^2$. These are monotonically decreasing when the coefficients b_m^2 are monotonically decreasing, which requires the series approximation to be very smooth.

6 Non-Smooth Approximation Decay

It is difficult to characterize the asymptotic IMSE of the averaging estimator (12) without imposing more structure on the approximation errors ϕ_m and the variance components v_m. It is well known that under our assumptions $\phi_m = O(m^{-\alpha})$ with $\alpha = 2s/q$. This, however, is not

sufficient to calculate the asymptotic limit of (12), mostly because the latter is a function of the changes $\Delta\phi_m$ which are not usefully characterized by an upper bound on the level ϕ_m. To make progress, we need to impose some structure on ϕ_m. We would like this structure to satisfy the known bound $\phi_m = O(m^{-\alpha})$ and yet allow ϕ_m to decline non-smoothly, in particular to allow $\Delta\phi_m$ to be non-monotonic.

A useful class of such approximation errors takes the form

$$\phi_m = \phi\left(\left[\frac{m}{d}\right]\right) \tag{13}$$

where $[x]$ denotes the integer part of x, d is a positive integer, and $\phi(u)$ is a function satisfying $\phi'(u) \leq 0$ and $\phi''(u) \geq 0$. If $d = 1$ then $\phi_m = \phi(m)$ and the changes $\Delta\phi_m$ are monotonic. If $d = 2$, then $\Delta\phi_m = 0$ for odd m, and the even-indexed $\Delta\phi_m$ are monotonic. In general, $\Delta\phi_m \neq 0$ only for indices which are integer multiples of d, e.g. $m = dj$ for some integer j.

We also need to impose structure on the variance component $v_m = \operatorname{tr}\left(Q_m^{-1}\Omega_m\right)$ and its differences $\Delta v_m = v_m - v_{m-1}$. We impose the condition that the changes converge to a constant. This broadens application beyond conditional homoskedasticity, though it is unclear which contexts are excluded.

Assumption 2 *$\Delta v_m \longrightarrow v > 0$ as $m \to \infty$.*

We will consider two specific models for the smooth function $\phi(u)$ in (13), power law decay and exponential decay. We present results each case in the next two sections.

7 Power Law Decay

Assumption 3

1. *$\phi_m = \phi\left(\left[\frac{m}{d}\right]\right)$ where $\phi(u) = Au^{-\alpha}$, for some $A > 0$, $\alpha > 0$ and positive integer d.*
2. *$n/M_n^{\alpha+1} = o(1)$.*

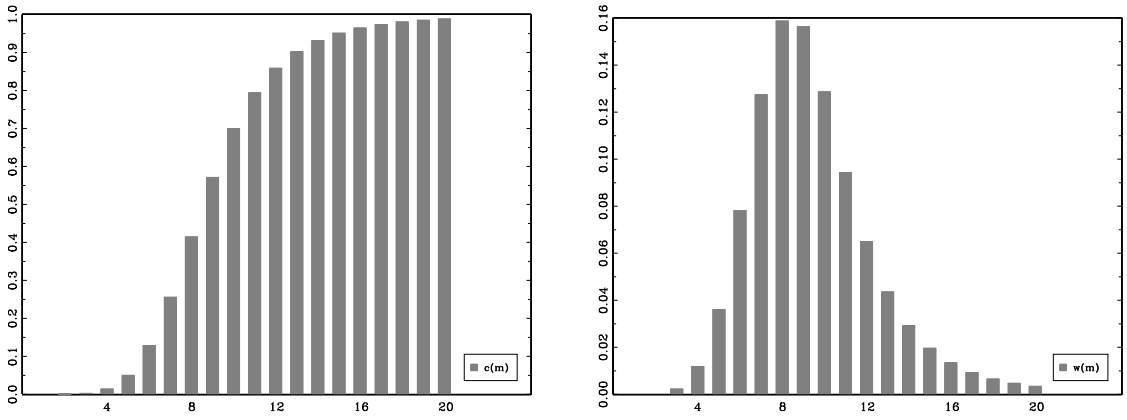
Assumption 3.1 specifies that the approximation errors decay as a power law every d^{th} model m. The coefficient α indexes the speed of the decay, with a small α meaning slow decay and a large α meaning fast decay.

Assumption 3.2 specifies that the largest estimated model M_n diverges to infinity faster than $n^{1/1+\alpha}$. This is required to ensure that the optimal model (which is of order $n^{1/(1+\alpha)}$ under Assumption 3.1) is asymptotically in the choice set, and to ensure that the bias of the largest estimated model is of smaller order than the optimal IMSE.

To understand the implications of Assumption 3, we start by calculating the optimal weights c_m and w_m for leading examples. We set $\alpha = 4$, $n = 100$ and $\mathbb{E}\left(e_i^2 \mid z_i\right) = 1$, and set A so that the optimal series estimator uses $m_n = 8$.

For our first example we set $d = 1$ (so that the decay is smooth). We plot the optimal cumulative weights c_m and weights w_m in Figures 1 and 2. We can see that the optimal weights are spread out on series estimators of diverse order, centered around the optimal single series order $m_n = 8$. The weights are smoothly distributed around this point, reflecting the smooth decay in the coefficients.

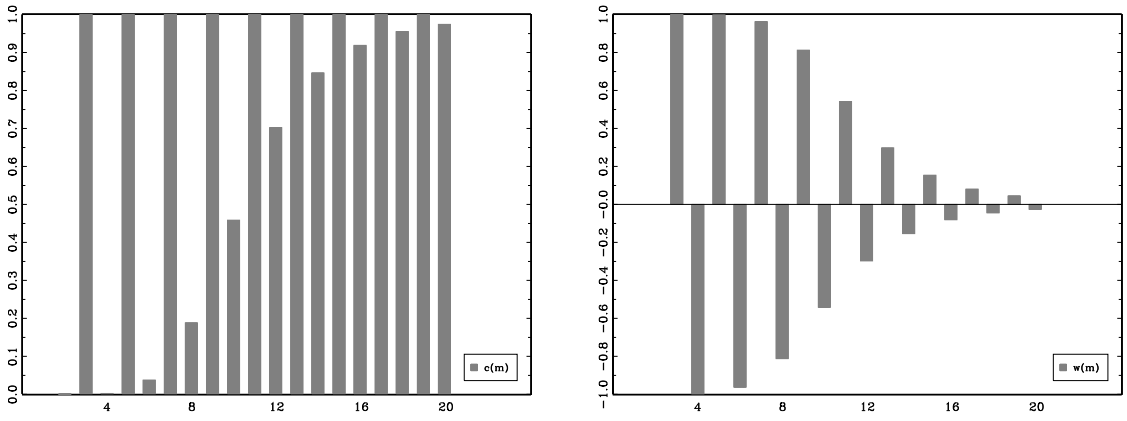
Figure 1: Optimal Weights under Power Law Decay, $d = 1$



(a) c_m (b) w_m

For our second example we set $d = 2$ (so that every second coefficient is zero). We plot the optimal cumulative weights c_m and weights w_m in Figures 3 and 4. The cumulative weight plot shows that the optimal c_m are not smooth in m, but oscillate between 1 and a lower bound which gradually approaches 1. The weight plot shows that the optimal weights w_m oscillate between positive and negative values, converging to zero as m grows. The reason for this behavior is that if for some m, $\Delta\phi_m = 0$ yet $\Delta\phi_{m+1} > 0$ (e.g. if $b_m = 0$ yet $b_{m+1} \neq 0$) then it is optimal to set $c_{m-1} = 1$ and $c_m < 1$, implying $w_m < 0$.

Figure 2: Optimal Weights under Power Law Decay, $d = 2$



(a) c_m (b) w_m

Assumption 3 is useful as we are able to precisely characterize the asymptotic risk of the

estimators under this specification of the approximation errors.

Theorem 2 *Under Assumptions 1-3,*

$$\lim_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^1 = \left(\frac{dv}{\alpha}\right)^{\alpha/(1+\alpha)} A^{1/(1+\alpha)} (1+\alpha), \tag{14}$$

$$\lim_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^2 = (A\alpha)^{1/(1+\alpha)} (dv)^{\alpha/(1+\alpha)} \frac{\Gamma\left(\frac{1}{1+\alpha}\right)\Gamma\left(\frac{\alpha}{1+\alpha}\right)}{1+\alpha}, \tag{15}$$

and

$$\lim_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^3 = (A\alpha)^{1/(1+\alpha)} v^{\alpha/(1+\alpha)} \frac{\Gamma\left(\frac{1}{1+\alpha}\right)\Gamma\left(\frac{\alpha}{1+\alpha}\right)}{1+\alpha}. \tag{16}$$

Theorem 2 shows that under power law decay, the IMSE of the three estimators converge at the common rate $n^{-\alpha/(1+\alpha)}$. For small α (slow decay) this rate can be arbitrarily slow, and for large α (fast decay) this rate can be arbitrarily close to n^{-1}. Theorem 2 also characterizes the normalized asymptotic IMSE of the three estimators, and shows that they have distinct limits. What is of interest to us is the relative performance of the estimators. In the next corollary we present their asymptotic ratios.

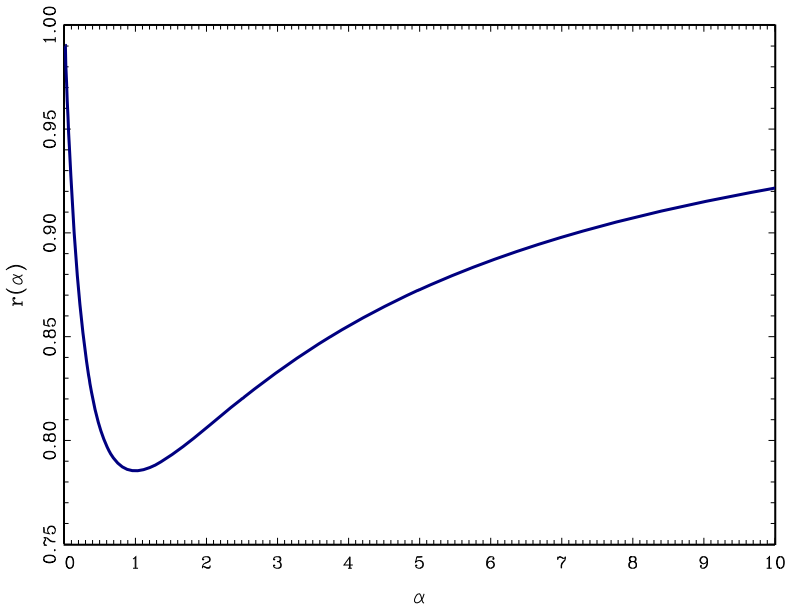


Figure 3: Asymptotic Ratio of IMSE of Constrained Averaging to Series Estimators

Corollary 1 *Under Assumptions 1-3,*

$$\lim_{n\to\infty} \frac{\inf_{w\in\mathcal{W}_n} IMSE_n(w)}{\inf_m IMSE_n(m)} = \lim_{n\to\infty} \frac{I_n^2}{I_n^1} = r(\alpha)$$

$$\lim_{n\to\infty}\frac{\inf_w I_n(w)}{\inf_{w\in\mathcal{W}_n} I_n(w)} = \lim_{n\to\infty}\frac{I_n^3}{I_n^2} = \frac{1}{d^{\alpha/(1+\alpha)}}$$

where

$$r(\alpha) = \frac{\alpha}{(1+\alpha)^2}\Gamma\left(\frac{1}{1+\alpha}\right)\Gamma\left(\frac{\alpha}{1+\alpha}\right).$$

Corollary 1 shows that the asymptotic ratio of the IMSE of the optimal series estimator to that of the constrained averaging estimator equals $r(\alpha)$ which is a function only of the power law decay rate α. It appears to be minimized at $\alpha = 1$ with minimum value $r(1) = \pi/4 \simeq 0.785$. This means that the constrained averaging estimator has up to 21% smaller asymptotic IMSE than the series estimator.

Corollary 1 also compares the constrained and unconstrained averaging estimators. It shows that the unconstrained estimator has smaller asymptotic IMSE when $d > 1$, and the percentage reduction approaches 100% as d gets large. What this shows is that when the series approximation errors ϕ_m decline non-smoothly, the unconstrained averaging estimator can achieve much improved precision relative to the other estimators.

8 Exponential Decay

Assumption 4

1. $\phi_m = \phi\left(\left[\frac{m}{d}\right]\right)$ *where* $\phi(u) = A\exp(-\beta m)$, d *is a positive integer,* $A > 0$ *and* $\beta > 0$.
2. $\ln(n)/M_n = o(1)$.

Assumption 4.1 specifies that the approximation errors decay exponentially, every d^{th} model m. The coefficient β indexes the speed of the decay with larger β indicating faster decay.

Assumption 4.2 specifies that the largest estimated model M_n diverges to infinity faster than $\ln(n)$. This is required to ensure that the optimal model (which is of order $\ln(n)$) is asymptotically in the choice set.

Theorem 3 *Under Assumptions 1, 2, and 4,*

$$\lim_{n\to\infty}\frac{n}{\ln(n)}I_n^1 = \frac{dv}{\beta}, \tag{17}$$

$$\lim_{n\to\infty}\frac{n}{\ln(n)}I_n^2 = \frac{dv}{\beta}, \tag{18}$$

and

$$\lim_{n\to\infty}\frac{n}{\ln(n)}I_n^3 = \frac{v}{\beta}. \tag{19}$$

Theorem 3 shows that under exponential decay, the IMSE of the three estimators converge at the common rate $\ln(n)/n$. Theorem 3 also characterizes the normalized asymptotic IMSE of the estimators, and shows that they take very simple forms.

Corollary 2 *Under Assumptions 1, 2, and 4,*

$$\lim_{n\to\infty} \frac{\inf_{w\in\mathcal{W}_n} IMSE_n(w)}{\inf_m IMSE_n(m)} = 1$$

$$\lim_{n\to\infty} \frac{\inf_w I_n(w)}{\inf_{w\in\mathcal{W}_n} I_n(w)} = \lim_{n\to\infty} \frac{I_n^3}{I_n^2} = \frac{1}{d}.$$

Corollary 2 shows that the asymptotic IMSE of the optimal series and optimal constrained averaging estimators are equivalent under these assumptions, so there is no asymptotic gain from averaging if the weights are constrained to be positive. However, the corollary also shows that the unconstrained optimal averaging estimator has much smaller IMSE when $d > 1$. As for the case of power decay, when the approximation errors decay non-smoothly, the averaging estimator can achieve much smaller IMSE than the other estimators.

9 Simulation Illustration

We investigate the finite sample behavior of feasible selection and averaging series regression estimators in a simple simulation experiment.

Our data generating process is

$$\begin{aligned} y_i &= a\sin(2\pi x_i + c) + e_i \\ x_i &\sim U[0,1] \\ e_i &\sim N(0,1) \end{aligned}$$

The parameter a is selected so that the popular R^2 varies between $\{0.25, 0.50, 0.75\}$. The parameter c is varied between $\{0, 1\}$. The sample size n is varied among $\{50, 75, 100, 200, 400, 600, 1000\}$.

We select this DGP as the sine function is a fairly strong nonlinear shape which is difficult to approximate with low-order polynomials. When $c = 0$ then the function is odd, so the power series expansion is a function only of odd powers of x, which is similar to the case $d = 2$ explored in the previous sections. When $c = 1$ then the power series expansion is a function of all coefficients, which is similar to the case $d = 1$ (though the coefficients do not decay as smoothly as in the simplified models of the previous sections.)

Our estimates are based on simple power series regression. The largest estimated model is $M_n = 4n^{1/5}$. We estimate all models for $m = 0$ to M_n, and construct feasible series estimators by three methods:

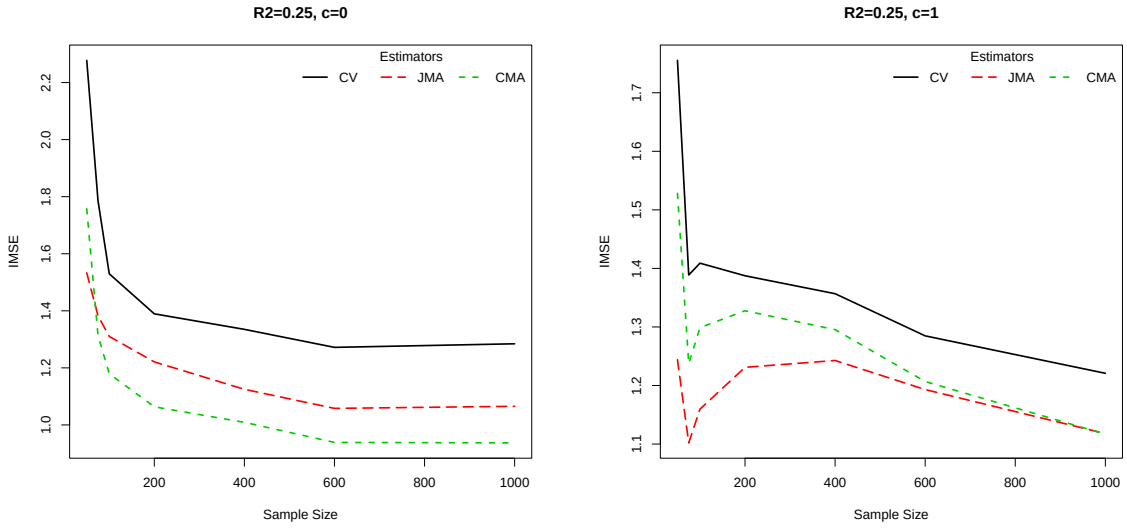
1. Cross-Validation Selection (CV)
2. Cross-Validation Weight Selection with constrained positive weights (JMA)
3. Cross-Validation Weight Selection with cumulative weights constrained to $[0, 1]$ (CMA)

The first method is standard cross-validation selection. The second method is the Jacknife Model Averaging (JMA) method of Hansen and Racine (2012). The third method analogous to JMA, but with different weight contraints. This properties of this latter estimator have not been formally studied (since the previous literature imposes the restriction of positive weights).

For each estimator, we calculate the finite sample integrated mean squared error (IMSE) by numerical integration using a grid on $x \in [0, 1]$ with 100 gridpoints and 10,000 simulation replications. The IMSE is normalized by the IMSE of the individual series estimator with the smallest finite sample IMSE. Thus plots of IMSE are relative to the infeasible best series approximation.

We plot the IMSE of the estimators as a function of sample size n, and show six plots, for varying choices of R^2 and c.

Figure 4: Finite Sample Relative IMSE, $R^2 = 0.25$

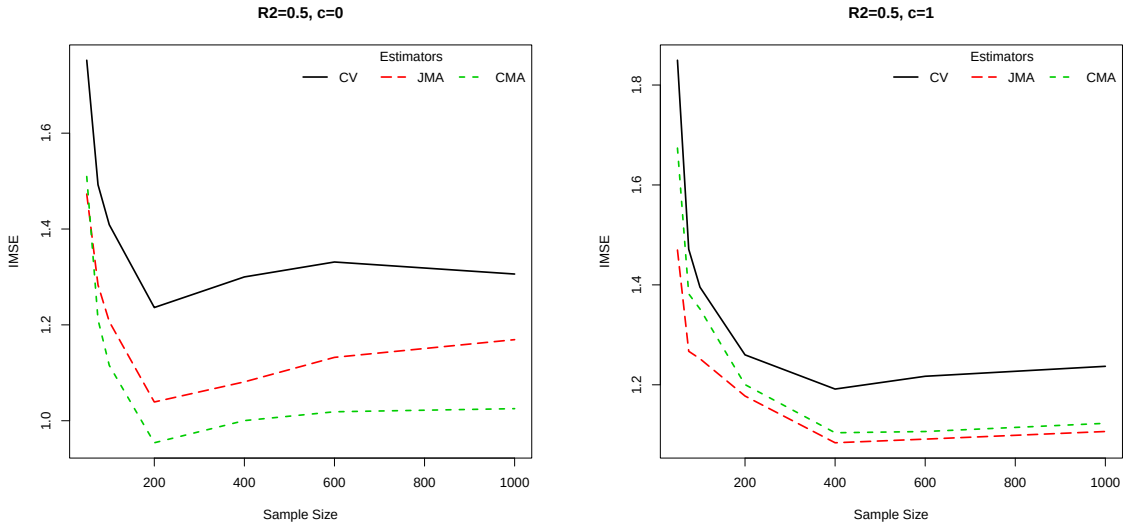


(a) $c = 0$ (b) $c = 1$

Figure 4 shows plots of the relative IMSE for $R^2 = 0.25$, with $c = 0$ on the left and $c = 1$ on the right. In both plots, the averaging methods (JMA and CMA) have uniformly smaller IMSE than CV selection. This shows the benefits of averaging methods. In the left-hand plot ($c = 0$) CMA achieves significantly lower IMSE than JMA, and even has lower IMSE than infeasible optimal selection. The reason why the CMA estimator performs well in this context is because the true regression is an odd function and thus only a function of the odd powers, whcih is exactly the context where we expect the CMA estimator to have lower IMSE. In the right-hand plot the ordering is reversed. The JMA estimator achieves the lower IMSE than the CMA estimator, though for large n the difference becomes negligible. While our theory suggests that CMA should have asymptotically lower IMSE, perhaps the difference shown here reflects extra finite noise induced by allowing negative weights.

Figures 5 and 6 shows similar plots for $R^2 = 0.5$ and $R^2 = 0.75$. They are qualitatively similar to the plots in Figure 4. We can see that the feasible averaging estimators have smaller IMSE than

Figure 5: Finite Sample Relative IMSE, $R^2 = 0.50$



(a) $c = 0$ (b) $c = 1$

the feasible selection estimator. While allowing for negative weights can further decrease the IMSE in some cases, it also increases the IMSE in others. This difference calls for further investigation.

10 Proofs

Proof of Theorem 1 (14): Set $\varepsilon > 0$. Under Assumption 2, there is an $N_\varepsilon < \infty$ such that for all $m \geq N_\varepsilon$, $|\Delta v_m - v| \leq \varepsilon$. Furthermore, under Assumption 1.4,

$$\underline{\sigma}^2 \leq \Delta v_m \leq \overline{\sigma}^2 \tag{20}$$

for all m. Thus

$$v_m = \sum_{j=1}^{m} \Delta v_m \leq mv\left(1+\varepsilon\right) + N_\varepsilon\left(\overline{\sigma}^2 - v\left(1+\varepsilon\right)\right) \tag{21}$$

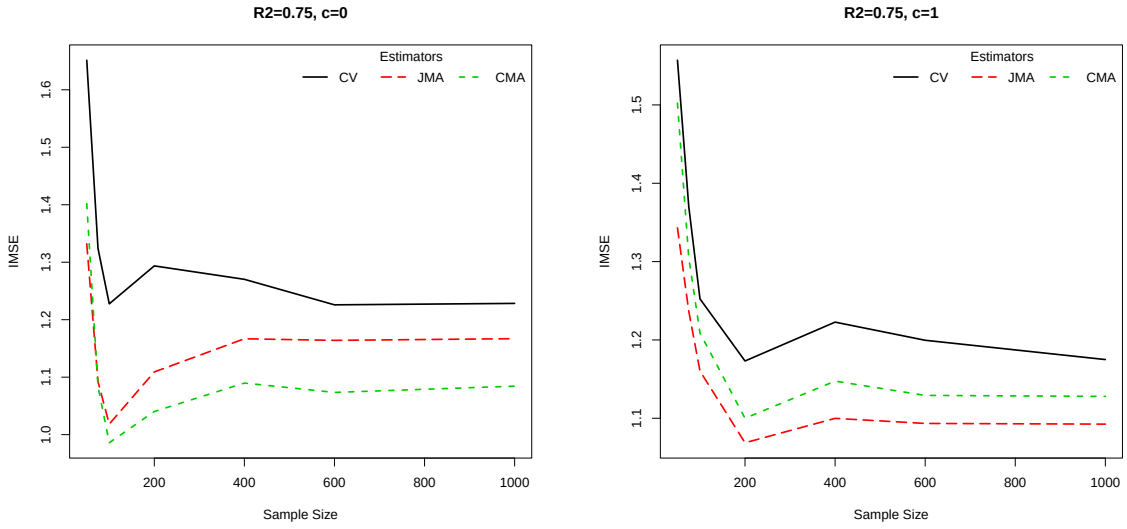
and similarly

$$v_m \geq mv\left(1-\varepsilon\right) + N_\varepsilon\left(\underline{\sigma}^2 - v\left(1-\varepsilon\right)\right). \tag{22}$$

Since $I_n(m) = \phi_m + \dfrac{v_m}{n}$ by (8), and under (13) ϕ_m only decreases for $m = dj$ for integer j, the minimum of $I_n(m)$ will be attained at such a value, and thus we can restrict attention to $m = dj$. Using Assumption 3.1, $I_n(dj) = Aj^{-\alpha} + \dfrac{v_{dj}}{n}$. Combined with (21) we find

$$I_n(dj) \leq Aj^{-\alpha} + j\frac{dv\left(1+\varepsilon\right)}{n} + \frac{N_\varepsilon\left(\overline{\sigma}^2 - v\left(1+\varepsilon\right)\right)}{n}.$$

Figure 6: Finite Sample Relative IMSE, $R^2 = 0.75$



(a) $c = 0$ (b) $c = 1$

The right-hand side is minimized at

$$j = \left(\frac{n\alpha A}{dv\,(1+\varepsilon)} \right)^{1/(1+\alpha)}$$

with minimized value

$$n^{-\alpha/(1+\alpha)} \left(\frac{dv\,(1+\varepsilon)}{\alpha} \right)^{\alpha/(1+\alpha)} A^{1/(1+\alpha)}\,(\alpha+1) + O(n^{-1}).$$

Hence

$$\limsup_{n\to\infty} n^{\alpha/(1+\alpha)} \inf_m I_n(m) \leq \left(\frac{dv\,(1+\varepsilon)}{\alpha} \right)^{\alpha/(1+\alpha)} A^{1/(1+\alpha)}\,(\alpha+1)\,.$$

By a similar calculation using (22),

$$\liminf_{n\to\infty} n^{\alpha/(1+\alpha)} \inf_m I_n(m) \geq \left(\frac{dv\,(1-\varepsilon)}{\alpha} \right)^{\alpha/(1+\alpha)} A^{1/(1+\alpha)}\,(\alpha+1)\,.$$

Since ε is arbitrary, we conclude that

$$\lim_{n\to\infty} n^{\alpha/(1+\alpha)} \inf_m I_n(m) = \left(\frac{dv}{\alpha} \right)^{\alpha/(1+\alpha)} A^{1/(1+\alpha)}\,(1+\alpha)\,.$$

as stated. ■

Proof of Theorem 1 (15): The restriction of the weights to the unit simplex is identical to restricting the cummulative weights to be monotonic: $0 \leq c_0 \leq \cdots \leq c_{M_n} = 1$. Let C_n be the set

of such cumulative weights. Thus

$$I_n^2 = \inf_{w \in \mathcal{W}_n} I_n(w) = \inf_{c \in C_n} I_n(c) \tag{23}$$

where $I_n(c)$ is defined in (10). In general, it is difficult to characterize problems of this form (because of the inequality constraints) but in this case the solution is simple. Since ϕ_m only declines when $m = dj$ for integer j, and v_m is monotonically increasing, the solution to (23) puts all weight on the models $\{0, d, 2d, ..., M_n/d\}$, similar to the optimal series regression problem described in the the proof of (14). Thus we can focus on the simplified criteria

$$\begin{aligned} I_n(c) &= \sum_{j=0}^{M_n/d} \left(\left(c_{jd}^2 - c_{(j-1)d}^2 \right) \phi_{jd} + \left(\left(1 - c_{(j-1)d}\right)^2 - \left(1 - c_{jd}\right)^2 \right) \frac{v_{jd}}{n} \right) \\ &= \phi\left(\frac{M_n}{d}\right) + \sum_{j=0}^{M_n/d-1} \left(c_{jd}^2 \left(\phi(j) - \phi(j+1)\right) + \left(1 - c_{jd}\right)^2 \frac{\left(v_{(j+1)d} - v_{jd}\right)}{n} \right) \end{aligned}$$

where we have used $\phi_{jd} = \phi(j)$ and for notational simplicity we have assumed that M_n is an integer multiple of d.

Minimizing over the c_{jd}, we find

$$I_n^2 = \phi\left(\frac{M_n}{d}\right) + \sum_{j=1}^{M_n/d} \frac{\left(v_{jd} - v_{(j-1)d}\right)\left(\phi(j-1) - \phi(j)\right)}{n\left(\phi(j-1) - \phi(j)\right) + \left(v_{jd} - v_{(j-1)d}\right)}. \tag{24}$$

As in the proof of Theorem 1 (14), set $\varepsilon > 0$ and $N_\varepsilon < \infty$ such that for all $m \geq N_\varepsilon$, $|\Delta v_m - v| \leq \varepsilon$. Then for $j \geq N_\varepsilon/d$, $\left|v_{jd} - v_{(j-1)d} - dv\right| \leq d\varepsilon$. Notice that (24) is increasing in the arguments $\left(v_{jd} - v_{(j-1)d}\right)$. For any upper bound on I_n^2, for $j < N_\varepsilon/d$ we use $v_{jd} - v_{(j-1)d} \leq d\overline{\sigma}^2$ from (20) and the inequality $ba/(na+b) \leq b/n$, and for $j \geq N_\varepsilon/d$ we use $v_{jd} - v_{(j-1)d} \leq dv(1+\varepsilon)$. Thus

$$I_n^2 \leq \phi\left(\frac{M_n}{d}\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + \sum_{j=N_\varepsilon/d}^{M_n/d} \frac{dv(1+\varepsilon)\left(\phi(j-1) - \phi(j)\right)}{n\left(\phi(j-1) - \phi(j)\right) + dv(1+\varepsilon)}. \tag{25}$$

For a lower bound, for $j < N_\varepsilon/d$ we use $v_{jd} - v_{(j-1)d} \geq 0$, and for $j \geq N_\varepsilon/d$ we use $v_{jd} - v_{(j-1)d} \geq dv(1-\varepsilon)$. Thus

$$I_n^2 \geq \phi\left(\frac{M_n}{d}\right) + \sum_{j=N_\varepsilon/d}^{M_n/d} \frac{dv(1-\varepsilon)\left(\phi(j-1) - \phi(j)\right)}{n\left(\phi(j-1) - \phi(j)\right) + dv(1-\varepsilon)}. \tag{26}$$

Since the function $\phi(u)$ is decreasing but convex, by the mean-value theorem

$$-\phi'(j) \leq \phi(j-1) - \phi(j) \leq -\phi'(j-1). \tag{27}$$

Noting that (25) and (26) are both increasing in the arguments $\left(\phi(j-1) - \phi(j)\right)$, we find that I_n^2

is bounded using (27) from above by

$$I_n^2 \leq \phi\left(\frac{M_n}{d}\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + \sum_{j=N_\varepsilon/d}^{M_n/d} \frac{dv(1+\varepsilon)\left(-\phi'(j-1)\right)}{n\left(-\phi'(j-1)\right) + dv(1+\varepsilon)}$$
$$\leq \phi\left(\frac{M_n}{d}\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + \int_0^\infty \frac{dv(1+\varepsilon)\left(-\phi'(x)\right)}{n\left(-\phi'(x)\right) + dv(1+\varepsilon)} dx \tag{28}$$

and from below by

$$I_n^2 \geq \phi\left(\frac{M_n}{d}\right) + \sum_{j=N_\varepsilon/d}^{M_n/d} \frac{dv(1-\varepsilon)\left(-\phi'(j)\right)}{n\left(-\phi'(j)\right) + dv(1-\varepsilon)}$$
$$\geq \phi\left(\frac{M_n}{d}\right) + \int_{N_\varepsilon/d}^{M_n/d} \frac{dv(1-\varepsilon)\left(-\phi'(x)\right)}{n\left(-\phi'(x)\right) + dv(1-\varepsilon)} dx. \tag{29}$$

Now $\phi(u) = Au^{-\alpha}$ implies $-\phi'(x) = A\alpha x^{-\alpha-1}$, so (28) equals

$$I_n^2 \leq Ad^\alpha M_n^{-\alpha} + \frac{N_\varepsilon \overline{\sigma}^2}{n} + A\alpha \int_0^\infty \frac{1}{nA\alpha/dv(1+\varepsilon) + x^{1+\alpha}} dx$$
$$= Ad^\alpha M_n^{-\alpha} + \frac{N_\varepsilon \overline{\sigma}^2}{n} + n^{-\alpha/(1+\alpha)} A\alpha \int_0^\infty \frac{1}{A\alpha/dv(1+\varepsilon) + u^{1+\alpha}} du$$

the equality by the change-of-variables $x = n^{1/(1+\alpha)}u$. Assumption 3.2 implies that $n^{\alpha/(1+\alpha)} M_n^{-\alpha} = o(1)$. Thus

$$\limsup_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^2 \leq A\alpha \int_0^\infty \frac{1}{A\alpha/dv(1+\varepsilon) + u^{1+\alpha}} du.$$

Similarly, (29) equals

$$I_n^2 \geq Ad^\alpha M_n^{-\alpha} + A\alpha \int_{N_\varepsilon/d}^{M_n/d+1} \frac{1}{nA\alpha/dv(1-\varepsilon) + x^{1+\alpha}} dx$$
$$= Ad^\alpha M_n^{-\alpha} + n^{-\alpha/(1+\alpha)} A\alpha \int_{n^{-1/(1+\alpha)}N_\varepsilon/d}^{n^{-1/(1+\alpha)}M_n/d} \frac{1}{A\alpha/dv(1-\varepsilon) + u^{1+\alpha}} du$$

and thus since $M_n n^{-1/(\alpha+1)} \to \infty$ and $n^{-1/(\alpha+1)} \to 0$,

$$\liminf_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^2 \geq A\alpha \int_0^\infty \frac{1}{A\alpha/dv(1-\varepsilon) + u^{1+\alpha}} du.$$

Since ε is arbitrary, we deduce that

$$\lim_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^2 = A\alpha \int_0^\infty \frac{1}{A\alpha/dv + u^{1+\alpha}} du$$
$$= (A\alpha)^{1/(1+\alpha)} (dv)^{\alpha/(1+\alpha)} \frac{1}{1+\alpha} \Gamma\left(\frac{1}{1+\alpha}\right) \Gamma\left(\frac{\alpha}{1+\alpha}\right)$$

as stated, the second equality using the result

$$\int_0^\infty \frac{1}{a+x^b}dx = a^{-(b-1)/b}\frac{1}{b}\Gamma\left(\frac{1}{b}\right)\Gamma\left(\frac{b-1}{b}\right).$$

■

Proof of Theorem 1 (16): From (12) and (13)

$$\begin{aligned} I_n^3 &= \phi_{M_n} + \sum_{m=1}^{M_n}\left(\frac{\Delta v_m \Delta\phi_m}{n\Delta\phi_m + \Delta v_m}\right) \\ &= \phi_{M_n} + \sum_{j=1}^{M_n/d}\left(\frac{\Delta v_{dj}\left(\phi(j-1)-\phi(j)\right)}{n\left(\phi(j-1)-\phi(j)\right)+\Delta v_{dj}}\right). \end{aligned}$$

where the second equality holds under specification (13), which implies that $\Delta\phi_m \neq 0$ only for indices of the form $m = dj$ for integer j, and $\Delta\phi_{dj} = \phi(j-1) - \phi(j)$.

As in the proof of (15), set $\varepsilon > 0$ and $N_\varepsilon < \infty$ such that for all $m \geq N_\varepsilon$, $|\Delta v_m - v| \leq \varepsilon$. For any upper bound on I_n^3, for $j < N_\varepsilon/d$ we use $\Delta v_{dj} \leq \overline{\sigma}^2$ from (20) and the inequality $ba/(na+b) \leq b/n$, and for $j \geq N_\varepsilon/d$ we use $\Delta v_{dj} \leq v(1+\varepsilon)$. Combined with (27) we find

$$\begin{aligned} I_n^3 &\leq \phi\left(\frac{M_n}{d}\right) + \frac{N_\varepsilon\overline{\sigma}^2}{dn} + \sum_{j=N_\varepsilon/d}^{M_n/d}\frac{v(1+\varepsilon)\left(-\phi'(j-1)\right)}{n\left(-\phi'(j-1)\right)+v(1+\varepsilon)} \\ &\leq \phi\left(\frac{M_n}{d}\right) + \frac{N_\varepsilon\overline{\sigma}^2}{dn} + \int_0^\infty \frac{v(1+\varepsilon)\left(-\phi'(x)\right)}{n\left(-\phi'(x)\right)+v(1+\varepsilon)}dx. \end{aligned} \tag{30}$$

Applying $\phi(u) = Au^{-\alpha}$, this equals

$$\begin{aligned} &Ad^\alpha M_n^{-\alpha} + \frac{N_\varepsilon\overline{\sigma}^2}{dn} + \int_0^\infty \frac{A\alpha}{nA\alpha/v(1+\varepsilon)+x^{1+\alpha}}dx \\ &= Ad^\alpha M_n^{-\alpha} + \frac{N_\varepsilon}{dn} + n^{-\alpha/(1+\alpha)}\int_0^\infty \frac{A\alpha}{A\alpha/v(1+\varepsilon)+u^{1+\alpha}}du \end{aligned}$$

using the change of variables $x = n^{1/(1+\alpha)}u$. Assumption 3.2 implies that $n^{\alpha/(1+\alpha)}M_n^{-\alpha} = o(1)$. Thus

$$\limsup_{n\to\infty} n^{\alpha/(1+\alpha)} I_n^3 \leq A\alpha\int_0^\infty \frac{1}{A\alpha/dv(1+\varepsilon)+u^{1+\alpha}}du.$$

Similarly

$$\begin{aligned} I_n^3 &\geq \phi\left(\frac{M_n}{d}\right) + \sum_{j=N_\varepsilon/d}^{M_n/d} \frac{v(1-\varepsilon)\left(-\phi'(j)\right)}{n\left(-\phi'(j)\right) + v(1-\varepsilon)} \\ &\geq \phi\left(\frac{M_n}{d}\right) + \int_{N_\varepsilon/d}^{M_n/d+1} \frac{v(1-\varepsilon)\left(-\phi'(x)\right)}{n\left(-\phi'(x)\right) + v(1-\varepsilon)} dx \\ &= Ad^\alpha M_n^{-\alpha} + \int_{N_\varepsilon/d}^{M_n/d+1} \frac{A\alpha}{nA\alpha/v(1-\varepsilon) + x^{1+\alpha}} dx \\ &= Ad^\alpha M_n^{-\alpha} + n^{-\alpha/(1+\alpha)} \int_{n^{-1/(1+\alpha)}N_\varepsilon/d}^{n^{-1/(1+\alpha)}(M_n/d+1)} \frac{A\alpha}{A\alpha/v(1-\varepsilon) + u^{1+\alpha}} du \end{aligned} \tag{31}$$

and thus

$$\liminf_{n\to\infty} I_n^3 \geq \int_0^\infty \frac{A\alpha}{A\alpha/v(1-\varepsilon) + u^{1+\alpha}} du.$$

Since ε is arbitrary we conclude that

$$\begin{aligned} \lim_{n\to\infty} I_n^3 &= A\alpha \int_0^\infty \frac{1}{A\alpha/v + u^{1+\alpha}} du \\ &= (A\alpha)^{1/(1+\alpha)}\, v^{\alpha/(1+\alpha)} \frac{1}{1+\alpha} \Gamma\left(\frac{1}{1+\alpha}\right) \Gamma\left(\frac{\alpha}{1+\alpha}\right) \end{aligned}$$

as stated. ■

Proof of Theorem 2 (17): As in the proof of Theorem 1 (14), we can restrict attention to $m = dj$ with integer j. Using Assumption 4.1, $I_n(dj) = A\exp(-\beta j) + \frac{v_{dj}}{n}$. Combined with (21) we find

$$I_n(dj) \leq A\exp(-\beta j) + j\frac{dv\,(1+\varepsilon)}{n} + \frac{N_\varepsilon\left(\overline{\sigma}^2 - v\,(1+\varepsilon)\right)}{n}.$$

The right-hand side is minimized at

$$j = \frac{1}{\beta}\ln\left(\frac{nA\beta}{dv\,(1+\varepsilon)}\right)$$

with minimized value

$$\begin{aligned} &\frac{dv\,(1+\varepsilon)}{n\beta} + \frac{\ln(n)}{n}\frac{dv\,(1+\varepsilon)}{\beta} - \frac{1}{\beta}\ln\left(\frac{dv\,(1+\varepsilon)}{A\beta}\right)\frac{dv\,(1+\varepsilon)}{n} + \frac{N_\varepsilon\left(\overline{\sigma}^2 - v\,(1+\varepsilon)\right)}{n} \\ &= \frac{\ln(n)}{n}\frac{dv\,(1+\varepsilon)}{\beta} + O(n^{-1}). \end{aligned}$$

Hence

$$\limsup_{n\to\infty} \frac{n}{\ln(n)} \inf_m I_n(m) \leq \frac{dv\,(1+\varepsilon)}{\beta}.$$

By a similar calculation using (22),

$$\liminf_{n\to\infty} \frac{n}{\ln(n)} \inf_m I_n(m) \geq \frac{dv\,(1-\varepsilon)}{\beta}.$$

Since ε is arbitrary, we conclude that

$$\lim_{n\to\infty} \frac{n}{\ln(n)} \inf_m I_n(m) = \frac{dv}{\beta}.$$

as stated. ■

Proof of Theorem 2 (18): Equations (28) and (29) apply. Assumption 4.1 $\phi(u) = A\exp(-\beta u)$ implies $-\phi'(x) = A\beta\exp(-\beta x)$, so (28) equals

$$\begin{aligned} I_n^2 &\leq A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + A\beta\int_0^\infty \frac{1}{\exp(\beta x) + nA\beta/dv(1+\varepsilon)}dx \\ &= A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + \frac{dv(1+\varepsilon)\ln\left(1+\dfrac{nA\beta}{dv(1+\varepsilon)}\right)}{n\beta} \end{aligned} \tag{32}$$

using the result

$$\int_a^b \frac{1}{\exp(\beta x) + N}dx = \frac{\ln\left(1+N\exp(-\beta a)\right) - \ln\left(1+N\exp(-\beta b)\right)}{N\beta}$$

which implies

$$\int_0^\infty \frac{1}{\exp(\beta x) + N}dx = \frac{\ln(1+N)}{N\beta}.$$

Assumption 4.2 implies the first term in (32) is $o(n^{-1})$. Thus

$$\limsup_{n\to\infty} \frac{n}{\ln(n)} I_n^2 \leq \frac{dv(1+\varepsilon)}{\beta}.$$

(29) equals

$$\begin{aligned} I_n^2 &\geq A\exp\left(-\frac{\beta}{d}M_n\right) + \int_{N_\varepsilon/d}^{M_n/d} \frac{1}{\exp(\beta x) + nA\beta/dv(1-\varepsilon)}dx. \\ &= A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{1}{n}\frac{dv(1-\varepsilon)}{A\beta^2}\ln\left(1+n\frac{A\beta\exp(-\beta N_\varepsilon/d)}{dv(1-\varepsilon)}\right) \\ &- \frac{1}{n}\frac{dv(1-\varepsilon)}{A\beta^2}\ln\left(1+n\frac{A\beta}{dv(1-\varepsilon)}\exp\left(-\frac{\beta}{d}M_n\right)\right) \end{aligned}$$

The first term is $o\left(n^{-1}\right)$ and the third term is $O(n^{-1})$ (since the term in logarithms is $O(1)$ by the same argument). Hence

$$\liminf_{n\to\infty} \frac{n}{\ln(n)} I_n^2 \geq \frac{dv(1-\varepsilon)}{A\beta^2}.$$

Since ε is arbitrary we conclude that

$$\lim_{n\to\infty} \frac{n}{\ln(n)} I_n^2 = \frac{dv}{\beta}.$$

as stated. ■

Proof of Theorem 2 (19): Equations (30) and (31) apply. Assumption 4.1 $\phi(u) = A\exp(-\beta u)$ implies $-\phi'(x) = A\beta\exp(-\beta x)$, so (30) equals

$$\begin{aligned} I_n^3 &\leq A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{N_\varepsilon \overline{\sigma}^2}{dn} + A\beta\int_0^\infty \frac{1}{\exp(\beta x) + nA\beta/v(1+\varepsilon)}dx \\ &= A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{N_\varepsilon \overline{\sigma}^2}{n} + \frac{v(1+\varepsilon)\ln\left(1+\dfrac{nA\beta}{v(1+\varepsilon)}\right)}{n\beta} \end{aligned}$$

and thus

$$\limsup_{n\to\infty} \frac{n}{\ln(n)} I_n^3 \leq \frac{v(1+\varepsilon)}{\beta}.$$

(31) equals

$$\begin{aligned} I_n^3 &\geq A\exp\left(-\frac{\beta}{d}M_n\right) + A\beta\int_{N_\varepsilon/d}^{M_n/d+1} \frac{1}{\exp(\beta x) + nA\beta/v(1-\varepsilon)}dx \\ &= A\exp\left(-\frac{\beta}{d}M_n\right) + \frac{1}{n}\frac{v(1-\varepsilon)}{A\beta^2}\ln\left(1+n\frac{A\beta\exp(-\beta N_\varepsilon/)}{v(1-\varepsilon)}\right) \\ &- \frac{1}{n}\frac{v(1-\varepsilon)}{A\beta^2}\ln\left(1+n\frac{A\beta}{v(1-\varepsilon)}\exp\left(-\frac{\beta}{d}M_n\right)\right). \end{aligned}$$

We find

$$\limsup_{n\to\infty} \frac{n}{\ln(n)} I_n^3 \geq \frac{v(1-\varepsilon)}{\beta}.$$

Since ε is arbitrary we conclude that

$$\lim_{n\to\infty} \frac{n}{\ln(n)} I_n^3 = \frac{v}{\beta}.$$

as stated. ■

References

[1] Andrews, Donald W.K. (1991a): "Asymptotic normality of series estimators for nonparametric and semiparametric models," *Econometrica,* 59, 307-345.

[2] Andrews, Donald W. K. (1991b): "Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors," *Journal of Econometrics*, 47, 359-377.

[3] Chen, Xiaohong (2007): "Large sample sieve estimation of semi-nonparametric models" *Handbook of Econometrics*, Vol. 6B, Chapter 76, eds. James J. Heckman and Edward E. Leamer, North-Holland.

[4] Hansen, Bruce E. (2007): "Least squares model averaging," *Econometrica,* 75, 1175-1189.

[5] Hansen, Bruce E. (2014a): "The integrated mean squared error of series regression and a Rosenthal Hilbert-space inequality," *Econometric Theory*, forthcoming.

[6] Hansen, Bruce E. (2014b): "Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation," *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, forthcoming.

[7] Hansen, Bruce E. and Jeffrey S. Racine (2012): "Jackknife model averaging," *Journal of Econometrics*, 167, 38–46.

[8] Li, Ker-Chau (1987): "Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete Index Set," *Annals of Statistics*, 15, pp. 958-975.

[9] Li, Qi, and Jeffrey S. Racine (2006): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

[10] Newey, Whitney K. (1995): "Convergence rates for series estimators," in Maddalla, G.S., Phillips, P.C.B., Srinavasan, T.N. (eds.) *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao*. Backwell, Cambridge, pp. 254-275.

[11] Newey, Whitney K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147-168.