

Cross-Validated SNP Density Estimates

Mark Coppejans
Duke University
Durham NC, USA

A. Ronald Gallant ¹
University of North Carolina
Chapel Hill NC, USA

First draft: May 1999
This draft: November 2001

¹Corresponding author: Mark Coppejans, Department of Economics, Duke University, Box 90097, Durham NC 27708-0097 USA; phone 1-919-660-1804; e-mail mtc@econ.duke.edu. This material is based upon work supported by the National Science Foundation under Grant No. 0000176. The most recent version of the paper is available by anonymous ftp as the PostScript file an.ps at ftp.econ.duke.edu in directory pub/arg/papers. We wish to thank the three referees and the associate editor for a prompt and careful reading of the manuscript and for their helpful recommendations.

ABSTRACT

We consider cross-validation strategies for the SNP nonparametric density estimator, which is a truncation (or sieve) estimator based upon a Hermite series expansion with coefficients determined by quasi maximum likelihood. Our main focus is on the use of SNP density estimators as an adjunct to EMM structural estimation. It is known that for this purpose a desirable truncation point occurs at the last point at which the ISE curve of the SNP density estimate declines abruptly. We study the determination of the ISE curve for iid data by means of leave-one-out cross-validation and hold-out-sample cross-validation through an examination of their performance over the Marron-Wand test suite and models related to asset pricing and auction applications. We find that both methods are informative as to the location of abrupt drops, but that neither can reliably determine the minimum of the ISE curve. We validate these findings with a Monte Carlo study. The hold-out-sample method is cheaper to compute because it requires fewer nonlinear optimizations. We consider the asymptotic justification of hold-out-sample cross-validation. For this purpose, we establish rates of convergence of the SNP estimator under the Hellinger norm that are of interest in their own right.

Keywords: Cross-validation, seminonparametric, SNP, efficient method of moments, EMM, Hellinger distance, convergence rates.

JEL Classification: C14, C15.

1 Introduction and Summary

We consider cross-validation strategies for the SNP nonparametric density estimator introduced by Gallant and Nychka (1987), which is a truncation (or sieve) estimator based upon a Hermite series expansion. SNP density estimators were introduced to model the semi-parametric component of nonlinear structural models. Nonlinear models require a numerical optimization procedure to estimate the parameters of the parametric part of the model. A series expansion is the natural choice for modeling the nonparametric part because determining the coefficients of series expansion can be accomplished by just including the coefficients among the parameters to be determined in the numerical optimization. Moreover, if one ignores the sieve origin of the additional parameters and computes tests and confidence intervals as if they were part of the parametric model, inference is often reasonably accurate. Provably so in some instances (Eastwood and Gallant, 1991; Fan, Zhang, and Zhang, 2000).

An open question in these applications is how the truncation point should be determined in a given sample. Early semiparametric applications such as Davidian and Gallant (1992, 1993) and nonparametric applications such as Gallant, Hsieh, and Tauchen (1991), Gallant, Rossi, and Tauchen (1992, 1993) used likelihood based inference. It therefore seemed natural to use standard maximum likelihood truncation rules such as BIC (Schwarz, 1978) to determine the truncation point. This tendency has become habitual and BIC is now the standard choice criterion, regardless of application.

Our interest here is in the choice of the truncation point for SNP estimators used in connection with EMM estimation (Gallant and Tauchen, 1999). In this application it is known that if the truncation point of the SNP estimator is properly selected, then the EMM estimator is fully efficient (Gallant and Long, 1997; Chumacero, 1997; Andersen, Chung, and Sorensen, 1999; and Michaelides and Ng, 2000).

For iid data, more is known. It is known that BIC will most likely truncate too soon (Fenton and Gallant, 1996) and therefore BIC is not likely to suggest the correct truncation point in EMM applications. This same study suggests that more aggressive truncation strategies such as AIC are unlikely to work either. Further, it is known that in typical EMM applications it is only necessary to make sure that the truncation point is large enough to

achieve nearly full efficiency. Once this point is reached, further EMM efficiency gains are marginal, regardless of sample size (Gallant and Tauchen, 1999).

The study by Gallant and Tauchen (1999) also suggests that mean squared error (MSE), which is the Lebesgue integral of the true density times the square of the difference between the SNP fit and the true density, can provide a reasonably reliable estimate of the desired truncation point. Typically in a plot of MSE against truncation point, there are a few abrupt declines that correspond to when the SNP estimator captures the gross features of the true density. The correct truncation point to achieve nearly full efficiency lies just to the right of the last abrupt decline (Gallant and Tauchen, 1999).

In this paper we consider integrated squared error (ISE), which is the Lebesgue integral of the square of the difference between the SNP fit and the true density. A comparison of the figures plotting ISE reported here with the figures of EMM efficiency reported in Gallant and Tauchen (1999) suggests that plots of ISE will serve as well as plots of MSE in determining the correct truncation point for EMM applications.

Because ISE depends on the unknown density, some method for estimating ISE is required if it is to be used in applications. Here we study the estimation of ISE by two types of cross-validation: the more traditional leave-one-out cross-validation method, and an average hold-out-sample method. We find that both methods are informative as to the location of abrupt drops. We find that neither method indicates the minimum of the true ISE curve very reliably, but that hold-out-sample cross-validation seems more reliable than leave-one-out cross-validation. The hold-out-sample method is cheaper to compute, because it requires fewer nonlinear optimizations, and it can be justified theoretically. As an intermediate step we obtain rates of convergence in the Hellinger norm that are of interest in their own right.

In the iid context, a leading application area of simulated method of moment estimators, of which EMM is an example, is to auction data. See, e.g., Laffont, Ossard, and Vuong (1995). We include an application as one of our illustrations. The stochastic volatility model of Tauchen and Pitts (1983) is an example of an iid model for financial markets that is amenable to EMM estimation. It is a reasonable model for a short time series on an individual stock because it can generate fat tails, but it is a poor model for long time series on either an individual stock or an index because a second persistent volatility factor becomes

empirically relevant in longer series (Gallant, Hsu, and Tauchen, 1999; Chernov, Gallant, Ghysels, and Tauchen, 2001). Variants of the stochastic volatility model are also applicable in other cross sectional settings, such as models of outcomes on standardized tests. We include an iid stochastic volatility model as one of our illustrations. Hierarchical models, of which the stochastic volatility model is an example, have wide applicability in the iid setting and are especially amenable to EMM estimation. Some examples are Edwards (1990), Lee and Nelder (1996), and Stasny (1991).

The plan of the remainder of the paper is as follows. In Section 2, we describe the leave-one-out and hold-out-sample cross-validation procedures and examine their behavior over the densities of the Marron-Wand test suite and for two densities representative of asset pricing and auction applications, namely, the second largest order statistic from the log normal distribution and a scale mixture of normals. We also assess performance in a Monte Carlo study of one of the densities from the Marron-Wand test suite, specifically the trimodal density. Section 3 summarizes facts regarding Hermite functions that we use extensively in the sequel. In Section 4 we study the Hellinger distance properties of the SNP estimator using methods devised by Wong and Shen (1995), which we believe to be the sharpest results available to date. In Section 5 we relate Hellinger distance to mean squared error, and derive results for hold-out-sample cross-validation. Section 6 discusses multivariate and time series applications. Section 7 concludes.

2 Test Cases

The SNP estimator is based on the class of densities

$$\mathcal{F}_K = \left\{ f_K : f_K(x, \xi) = \left[\sum_{i=0}^K \xi_i x^i \right]^2 e^{-x^2/2} + \epsilon_o \phi(x), \xi \in \Xi_K \right\}$$

$$\Xi_K = \left\{ \xi : \xi = (\xi_0, \xi_1, \dots, \xi_K), \int f_K(x, \xi) dx = 1 \right\}$$

where ϕ denotes the standard normal density, ϵ_o is a small positive number, and $K = 0, 1, \dots$. Estimation is by quasi maximum likelihood

$$\hat{f}_K = \underset{\substack{f \in \mathcal{F}_K \\ -\infty < \mu < \infty \\ 0 < \sigma < \infty}}{\operatorname{argmax}} \sum_{x_i \in \mathcal{X}} \log \left[\frac{1}{\sigma} f \left(\frac{x_i - \mu}{\sigma} \right) \right]$$

where

$$\mathcal{X} = \{x_1, x_2, \dots, x_n\}$$

is a random sample of size n from the unknown true density f_o . In the later sections we shall suggest specific relationships between K and n of the form $K = K_n$ for some function on the integers K_n . When such a relationship is in force, we will use \mathcal{F}_n and \hat{f}_n to denote \mathcal{F}_K and \hat{f}_K , respectively.

Some structural aspects of the SNP estimator deserve comment. If $K = 0$ then \mathcal{F}_K contains only the normal density which implies that the normal density is the leading term of the SNP expansion. Other choices of the weight function $e^{-x^2/2}$ that multiplies the polynomial term and the density $\phi(x)$ that multiplies ϵ_o are permitted by the theory developed in Gallant and Nychka (1987) and will result in a different leading term. Here, we shall only consider the choices $e^{-x^2/2}$ and $\phi(x)$ because they are particularly convenient and are standard in applications. The term $\epsilon_o\phi(x)$ acts as a lower bound that insures that integrals such as $\int_{-\infty}^{\infty} \log f f_o dx$ exist for all $f \in \mathcal{F}_K$. It also serves to keep terms such as $\log f$ from going out of range during optimizations. In the results reported here, $\epsilon_o = 10^{-5}$. Other nonparametric estimators based on Hermite expansions have been proposed (Devroye and Györfi, 1985). The SNP estimator differs from these others in three respects: it is positive, it is location and scale invariant, and the coefficients are determined by quasi maximum likelihood rather than by method of moments.

The class of densities \mathcal{F} for which the SNP density is dense is described in Gallant and Nychka (1987). Qualitatively this class can be described as follows. On any preselected compact set, the SNP density can approximate any density in the Hellinger norm because the square root of a density is an L_2 function and the Hermite functions are dense in L_2 . Off that compact set, tail behavior must be restricted. To get consistency under either deterministic or random truncation rules, Gallant and Nychka (1987) allow fat t-like tails. To get rate results, as in this article, it is necessary to impose more stringent tail conditions (see our Assumption 1).

The integrated squared error of the estimator is

$$\text{ISE}(\hat{f}_K) = \int \hat{f}_K^2 dx - 2 \int \hat{f}_K f_o dx + \int f_o^2 dx$$

$$= M_{(1)} - 2M_{(2)} + M_{(3)}.$$

The first term on the right can be computed directly and the third term is constant and therefore not needed to determine the shape of a plot of $\text{ISE}(\hat{f}_K)$ against K . This leaves only

$$M_{(2)} = \int \hat{f}_K f_o dx$$

to be determined.

To this end, for a given proportion α that can be expressed as $\alpha = 1/J$ for some positive integer J , partition the sample into J groups, denoted as $\mathcal{X}_{j,\alpha}$ for $j = 1, \dots, J$, making the sizes of these groups as equal as possible. Let $[\alpha n]$ denote the minimum group size. For convenience, we require that $n \geq J$ so that $[\alpha n] \geq 1$. We also require that the allocation scheme is such that the groups $\mathcal{X}_{j,\alpha}$, $j = 1, \dots, J$, are independent. Recall that \mathcal{X} is presumed to be a random sample and is therefore neither sorted nor structured in any other way so that allocation to the groups in contiguous blocks, in recursive serial order, or according to any deterministic function mapping $\{i : i = 1, \dots, n\}$ to $\{j : j = 1, \dots, J\}$ will accomplish this. Let $\hat{f}_{j,K}$ denote the estimate obtained from the sample points that remain after deletion of the j th group. That is,

$$\hat{f}_{j,K} = \underset{\substack{f \in \mathcal{F}_K \\ -\infty < \mu < \infty \\ 0 < \sigma < \infty}}{\text{argmax}} \sum_{x_i \in \tilde{\mathcal{X}}_{j,\alpha}} \log \left[\frac{1}{\sigma} f \left(\frac{x_i - \mu}{\sigma} \right) \right],$$

where $\tilde{\mathcal{X}}_{j,\alpha} = \mathcal{X} \sim \mathcal{X}_{j,\alpha}$. An estimate of $M_{(2)}$ is

$$\hat{M}_{(2)} = \frac{1}{J} \sum_{j=1}^J \left[\frac{1}{[\alpha n]} \sum_{x_i \in \mathcal{X}_{j,\alpha}} \hat{f}_{j,K}(x_i) \right] \doteq \frac{1}{n} \sum_{j=1}^J \sum_{x_i \in \mathcal{X}_{j,\alpha}} \hat{f}_{j,K}(x_i).$$

If $\alpha = 1/n$, then

$$\text{CVL} = M_{(1)} - 2\hat{M}_{(2)}$$

is the traditional leave-one-out cross-validation formula, which requires n nonlinear optimizations to compute. See Marron (1987) and Li (1987) for a discussions of cross-validation strategies and their relation to other bandwidth strategies for classes of density estimators that include kernel estimators, histogram estimators, and orthogonal series estimators (with

coefficients determined differently than here, as mentioned above). See also Sain, Baggerly, and Scott (1994) and the references therein.

The hold-out-sample cross-validation formula is obtained by putting α to some fixed proportion such as $\alpha = 0.1$, which is the value used in the figures that follow. This α does not change as n increases, which is the basis for a distinction between hold-out-sample and leave-one-out cross-validation. Rather than the formula above, we find it more convenient to use

$$\text{CVH} = \hat{M}_{(1)} - 2\hat{M}_{(2)}$$

for hold-out-sample cross-validation, where

$$\hat{M}_{(1)} = \frac{1}{J} \sum_{j=1}^J \int \hat{f}_{j,K}^2(x) dx.$$

Although, we do this is for theoretical convenience, this formula also reduces the computational burden slightly by eliminating one nonlinear optimization.

In the work reported in this section, we know f_o and therefore can compute both $\text{ISE}(\hat{f}_K)$ and $M_{(3)}$. To facilitate comparisons, we actually compute and plot $\text{CVL} + M_{(3)}$ and $\text{CVH} + M_{(3)}$ against K . Of course this cannot be done in applications, nor is it necessary because $M_{(3)}$ does not depend on K .

Here we report computations for single realizations from five densities: a stochastic volatility model, the second largest order statistic from the log normal density, and three densities from the Marron-Wand (1992) test suite, namely, the Gaussian, the trimodal, and the smooth comb. These five densities are plotted in Figure 1. We check our conclusions by considering 1024 Monte-Carlo repetitions of the analysis for the trimodal.

Computations for the remainder of the Marron-Wand test suite are available by anonymous ftp from host [ftp.econ.duke.edu](ftp://ftp.econ.duke.edu) in directory `pub/arg/papers` as file `mwsuite.ps` or by clicking on “browse ftp site” at www.unc.edu/~arg.

Figure 1 about here

The stochastic volatility model is one of the standard models for statistical analysis of financial market data. A recent summary of the literature is Ghysels, Harvey, and Renault

(1995). Although specialized estimators for various specifications of the stochastic volatility model have been proposed, the EMM estimator is often used instead because it permits one to change specifications with little effort and provides informative diagnostic tests of tentative specifications. A recent example is Liu (2000). As discussed above, EMM is usually implemented with an SNP score which requires determination of K . Also, as discussed above, the computations reported in Gallant and Tauchen (1999) suggest (i) that a typical MSE curve in applications will drop sharply and then flatten, (ii) that a reasonable choice of K is the last point at which MSE takes a sharp drop, and (iii) that finding the K that produces the best MSE will have little payoff, even if successful. A comparison of the figures reported here (and the figures reported in `mwsuite.ps` available by ftp as described above) with the figures of EMM efficiency reported in Gallant and Tauchen (1999) suggests that ISE can be substituted for MSE in determining K .

A standard stochastic volatility model, following Tauchen and Pitts (1983), is a lognormal scale mixture of normals. It is an acceptable model for a short series on an individual stock but must be augmented by a serially correlated persistent volatility factor in longer series (Chernov, Gallant, Ghysels, and Tauchen, 2001). This augmentation would destroy its iid structure. Its density is

$$p(y|\rho) = \int_{-\infty}^{\infty} n(y|\rho_1, e^{2u}) n(u|\rho_2, \rho_3^2) du$$

where $n(u|\alpha, \beta^2)$ denotes the normal density with mean α and variance β^2 . The mean μ , variance σ^2 , and raw kurtosis $\kappa = \frac{\varepsilon_p(y-\mu)^4}{(\sigma^2)^2}$ of $p(y|\rho)$ determine ρ as follows

$$\begin{aligned} \rho_1 &= \mu \\ \rho_2 &= \log \sigma - \rho_3^2 \\ \rho_3 &= [\log(\kappa/3)/4]^{1/2}. \end{aligned}$$

Our choice is $(\mu, \sigma^2, \kappa) = (0, 1/4, 8)$; the considerations in Section 4 of Gallant and Tauchen (1999) suggest that this choice is representative and that results are likely to be robust to changes in these choices. The density is plotted in the first panel of Figure 1.

For this choice, ISE, CVL, and CVH are plotted against K in Figure 2 for sample sizes $n = 400, 900, 1600, 2500 = 20^2, 30^2, 40^2, 50^2$. The plots of either CVL or CVH appear to be

adequate for a determination of the point at which ISE drops sharply and begins to flatten. This finding holds for our remaining four cases, as we shall see, and for cases not reported here but available by ftp from the site indicated earlier.

Figure 2 about here

Some interesting subsidiary information is displayed in Figure 2. The K chosen by BIC, by minimum CVL, and by minimum CVH are shown as vertical bars. Interestingly, BIC seems to give a reliable indication of a lower bound for K : one ought not choose K to the left of that indicated by BIC. This finding is robust across cases considered. The minimum values of the CVL and CVH curves do not give a reliable indication of the minimum of the true ISE curve, although CVH seems to be the more reliable of the two. However, as noted earlier, it is more important to locate the points at which the ISE curve drops abruptly than to locate its exact minimum.

An area of current interest in simulation estimation is the analysis of auction data; see for instance, Laffont, Ossard, and Vuong (1995). Under an independent private value assumption, the values assigned to an object by N bidders are a random sample of size N from a common valuation distribution. In either a single object second price auction or a single object English (oral ascending bid) auction, the observed selling price is the second largest order statistic in this sample. In an sealed bid first price auction, and other auction formats, the winning bid is a functional of the order statistics of the valuation distribution. These observations suggest that estimating the density of the second largest order statistic in a sample of size N has relevance to the application of EMM estimators to auction data. Here we consider the case when the common valuation distribution is lognormal.

The second largest order statistic from a sample of size $\rho_1 = N$ from the lognormal distribution has density function

$$p(y|\rho) = \frac{N(N-1)}{y} \left[\Phi\left(\frac{\log y - \rho_2}{\rho_3}\right) \right]^{N-2} \left[1 - \Phi\left(\frac{\log y - \rho_2}{\rho_3}\right) \right] \phi\left(\frac{\log y - \rho_2}{\rho_3}\right)$$

where $y > 0$ and ϕ and Φ denote the standard normal density and distribution functions, respectively.

The lognormal distribution is nearly normal for small values of ρ_3 and and departs from the normal as ρ_3 increases (Johnson and Kotz, 1994, Chapter 14). The choice $\rho_3 = 1$ appears reasonable judging from the tables in Johnson and Kotz (1994, Chapter 14). The considerations in Section 5 of Gallant and Tauchen (1999) suggest that $\rho_1 = N$ and $\rho_2 = -3$ are representative and that results are likely to be robust to changes in these choices. The density is plotted in the second panel of Figure 1.

Figure 3 about here

ISE, CVL, and CVH are plotted in Figure 3. As above, the plots of either CVL or CVH appear to be adequate for a determination of the point at which ISE drops sharply and begins to flatten and BIC gives a reliable indication of a lower bound for K .

Marron and Wand (1992) proposed a test suite as a battery for use in evaluating density estimators. It is designed to evaluate the ability of an estimator to track complex behavior at the center of the distribution. The densities in the suite are mixtures of normals; the mixing proportions $\{\pi_j\}$, the corresponding means $\{\mu_j\}$, and standard deviations $\{\sigma_j\}$ of the constituent normal densities are given in Marron and Wand (1992). Here, we consider three densities from the suite: the trimodal, the Gaussian, and the smooth comb. They are plotted in the panels three through five of Figure 1. The trimodal is a reasonably representative density from the test suite and the Gaussian and smooth comb are at opposite extremes. The Gaussian is actually in \mathcal{F}_K and the smooth comb is a very difficult density to estimate.

Figure 4 about here

For the trimodal density, ISE, CVL, and CVH are plotted in Figure 4. As before, the plots of either CVL or CVH appear to be adequate for a determination of the point at which ISE drops sharply and begins to flatten and BIC gives a reliable indication of a lower bound for K .

Figure 5 about here

Figure 5, which shows overplots of \hat{f}_K and f_o for the case $n = 900$, indicates the source of the sharp declines in ISE, CVL, and CVH seen in Figure 4. As seen from Figure 5, the abrupt declines are the points at which the gross features of f_o are first captured by \hat{f}_K ; specifically, they are the points at which modes are detected. Once the gross features are identified, relatively extreme values of K are required to determine fine detail. The computations reported in Gallant and Tauchen (1999) suggest that there is little payoff in terms of the efficiency of the EMM estimator to capturing fine detail. Determination of the gross features seems to be adequate.

Figure 6 about here

Figure 6 plots ISE, CVL, and CVH for the Gaussian density, which, as remarked above, is actually in \mathcal{F}_K . As one would expect, all methods for determining K eventually detect this fact. Interestingly, minimum CVH does better than minimum CVL.

Figure 7 about here

Lastly, Figure 7 plots ISE, CVL, and CVH for the smooth comb density. This is a density that is very hard to estimate. Correspondingly, the ISE shows less abrupt drops and all three methods for determining K shown in the figure get driven inexorably higher as sample size increases. Nonetheless, the plots CVL and CVH do seem to give a reliable indication of the situation one faces.

In order to confirm these conclusions, we conducted 1024 Monte-Carlo repetitions of the computations for the trimodal density. The results are presented in Tables 1 through 5

Table 1 presents the ISE curve for sample sizes $n=400, 900, 1600,$ and 2500 , averaged over the 1024 repetitions together with the root mean squared error of approximation by CVL

and CVH at $\alpha = 5\%$, 10% , and 20% hold out samples. Formulas describing the computation are in the table legend. The first conclusion from inspecting this table is that CVH is not sensitive to hold out sample size within the range considered and CVH does not differ from CVL in approximation error in any practical sense. The second conclusion from inspecting this table is that the approximation error appears small enough to be able to reliably detect the last abrupt drop in an ISE curve. But once this point is reached, the approximation error is about as large as the curve itself which would imply that neither CVL nor CVH is accurate enough to detect the minimum of the ISE curve.

Tables 2 through 5 examine the ability of CVL, CVH, and BIC to detect the point at which the last abrupt drop in the ISE curve occurs. The tables correspond to sample sizes $n=400$, 900 , 1600 , and 2500 , respectively. The first line of each table gives the frequency with which the abrupt drop \hat{K} assumed one of its possible values. In each column below that value is the frequency with which CVL, CVH_α , for $\alpha = 5\%$, 10% , 20% , and BIC missed \hat{K} by one or more than one. Using the criterion that only misses by two or more are serious, it appears from inspection of tables that CVL and CVH are quite reliable and there is no material difference in reliability across criterion. This is not true for BIC. With BIC there appears to be a serious risk of underestimation in applications.

Table 1 about here

Table 2 about here

Table 3 about here

Table 4 about here

Table 5 about here

The numerical methods employed are as follows. All code is in C++. The major software components are available by anonymous ftp from host ftp.econ.duke.edu in directories pub/arg/npe and pub/arg/libcpp or by clicking on “browse ftp site” at www.unc.edu/~arg. The algorithms used to compute \hat{f}_K and the methods employed to compute and simulate from the Marron-Wand test suite are described in Section 4.3 of Fenton and Gallant (1996); Sections 4 and 5 of Gallant and Tauchen (1999) describe computation of and simulations from the stochastic volatility model and the second largest order statistic from the log normal. Briefly, f_K is represented as a sum of Hermite functions, as described in Section 3, and optimization is carried out by a quasi Newton method with line search using the outer product form of the information matrix to approximate the Hessian; the requisite matrix algebra is by means of a matrix class found in pub/arg/libcpp at ftp.econ.duke.edu. The integrals $\text{ISE}(\hat{f}_K)$ and $M_{(3)}$ are computed as Reimann sums, e.g., $M_{(3)} = \sum f_o^2(x_i)\Delta x_i$ for finely spaced x_i , and $M_{(1)}$ is computed by Gauss-Hermite quadrature.

3 Hermite Expansions

It is convenient to rewrite the SNP density in terms of normalized Hermite polynomials whose definitions and properties are collected together in the following lemma.

LEMMA 1 The Hermite polynomials

$$H_{e_i}(x) = \sum_{m=0}^{\lfloor i/2 \rfloor} (-1)^m \frac{i!}{m!2^m(i-2m)!} x^{i-2m},$$

where $\lfloor i/2 \rfloor$ is the integer part of $i/2$, and their normalized counterpart

$$\bar{H}_{e_i}(x) = (\sqrt{2\pi} i!)^{-1/2} H_{e_i}(x)$$

satisfy the following properties for all real x and all positive integers m and i .

- (a) $\int_{-\infty}^{\infty} H_{e_i}(x)H_{e_m}(x)e^{-x^2/2} dx = \begin{cases} \sqrt{2\pi}i! & i = m \\ 0 & i \neq m \end{cases}$
- (b) $\int_{-\infty}^{\infty} \overline{H}_{e_i}(x)\overline{H}_{e_m}(x)e^{-x^2/2} dx = \begin{cases} 1 & i = m \\ 0 & i \neq m \end{cases}$
- (c) Differential equation: $\frac{d^2}{dx^2}H_{e_i}(x) - x\frac{d}{dx}H_{e_i}(x) + iH_{e_i}(x) = 0$
- (d) Recurrence relation: $H_{e_{i+1}}(x) = xH_{e_i}(x) - iH_{e_{i-1}}(x)$
- (e) Differential relation: $\frac{d}{dx}H_{e_i}(x) = iH_{e_{i-1}}(x)$
- (f) Rodrigues formula: $H_{e_i}(x) = (-1)^i e^{x^2/2} \frac{d^i}{dx^i} e^{-x^2/2}$
- (g) Upper bound: $|\overline{H}_{e_i}(x)e^{-x^2/4}| < \tilde{B} \doteq 0.6862127$.

Proof Abramowitz and Stegun (1972, Chapter 22). □

In this section and hereafter, K will depend explicitly on the sample size n according to $K = K_n$, where K_n is a function defined on the integers that is specified later. We write \mathcal{F}_n for \mathcal{F}_K and f_n for f_K to emphasize that this restriction is in force. Using Lemma 1, the class \mathcal{F}_n can be written in terms of normalized Hermite polynomials as

$$\mathcal{F}_n = \left\{ f_n : f_n(x, \theta) = \left[\sum_{i=0}^{K_n} \theta_i \overline{H}_{e_i}(x) \right]^2 e^{-x^2/2} + \epsilon_o \phi(x), \theta \in \Theta_n \right\}$$

$$\Theta_n = \left\{ \theta : \theta = (\theta_0, \theta_1, \dots, \theta_{K_n}), \sum_{i=0}^{K_n} \theta_i^2 + \epsilon_o = 1 \right\}.$$

The family of orthonormal polynomials $\{\overline{H}_{e_m}(x)\}_{m=0}^{\infty}$ is complete for $L_2(e^{-x^2/2})$ where $L_2(e^{-x^2/2})$ denotes those functions g with

$$\|g\|_H = \left(\int_{-\infty}^{\infty} g^2(x)e^{-x^2/2} dx \right)^{1/2} < \infty$$

(Sansone, 1991, p. 351). Thus every g in $L_2(e^{-x^2/2})$ has the expansion

$$g(x) = \sum_{i=0}^{\infty} \theta_i \overline{H}_{e_i}(x)$$

where $\theta_i = \int_{-\infty}^{\infty} g(x)\overline{H}_{e_i}(x)e^{-x^2/2} dx$ and equality is with respect to $\|\cdot\|_H$. Moreover $\int_{-\infty}^{\infty} g^2(x)e^{-x^2/2} dx = \sum_{i=0}^{\infty} \theta_i^2$. Suppose f is a density that can be written as

$$f(x) = g^2(x)e^{-x^2/2} + \epsilon_o \phi(x)$$

for some g . Because $\int f dx = \int \phi dx = 1$, it follows that g is in $L_2(e^{-x^2/2})$ and that g has a Hermite expansion whose coefficients satisfy

$$\sum_{i=0}^{\infty} \theta_i^2 + \epsilon_o = 1.$$

Consequently, the class of functions that can be attained as limits of functions from \mathcal{F}_n is

$$\mathcal{F}_\infty = \left\{ f : f = g^2 e^{-x^2/2} + \epsilon_o \phi, g \in L_2(e^{-x^2/2}), \int f dx = 1 \right\} = \overline{\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n \right)}$$

where the overbar indicates closure with respect to the norm $\|f\|_{\mathcal{F}} = \|g\|_H$ defined for functions of the form $f = g^2 e^{-x^2/2} + \epsilon_o \phi$ with $g \in L_2(e^{-x^2/2})$.

The following lemma states the rate at which the Hermite coefficients decrease.

LEMMA 2 If g is k -times differentiable with $g^{(j)} \in L_2(e^{-x^2/2})$ for $j = 0, 1, \dots, k$ then the Hermite coefficients of g satisfy

$$\sum_{i=K}^{\infty} \theta_i^2 = o(K^{-k}).$$

Proof If g is k -times differentiable with $g^{(k)} \in L_2(e^{-x^2/2})$ then $g^{(k)}$ has a Hermite expansion with coefficients

$$\theta_i^{(k)} = \int_{-\infty}^{\infty} g^{(k)}(x) \overline{H}_{e_i}(x) e^{-x^2/2} dx = (\sqrt{2\pi} i!)^{-1/2} \int_{-\infty}^{\infty} g^{(k)}(x) H_{e_i}(x) e^{-x^2/2} dx.$$

Using Parts (d) and (e) of Lemma 1 to integrate by parts we have the recursion

$$\int_{-\infty}^{\infty} g^{(k)}(x) H_{e_i}(x) e^{-x^2/2} dx = \int_{-\infty}^{\infty} g^{(k-1)}(x) H_{e_{i+1}}(x) e^{-x^2/2} dx$$

which can be iterated to yield

$$\theta_i^{(k)} = [(i+k) \cdots (i+1)]^{1/2} \int_{-\infty}^{\infty} g(x) \overline{H}_{e_{i+k}}(x) e^{-x^2/2} dx = [(i+k) \cdots (i+1)]^{1/2} \theta_{i+k}.$$

It follows that

$$0 \leq \sum_{i=0}^{\infty} \theta_{i+k}^2 < \sum_{i=0}^{\infty} i \theta_{i+k}^2 < \sum_{i=0}^{\infty} i^2 \theta_{i+k}^2 < \cdots < \sum_{i=0}^{\infty} i^k \theta_{i+k}^2 < \sum_{i=0}^{\infty} [\theta_i^{(k)}]^2 < \infty$$

which implies $\sum_{i=0}^{\infty} (i+k)^k \theta_{i+k}^2 < \infty$ because $(i+k)^k$ is a polynomial in i of degree k . From

$$0 \leq \lim_{K \rightarrow \infty} \sum_{i=K}^{\infty} K^k \theta_i^2 \leq \lim_{K \rightarrow \infty} \sum_{i=K}^{\infty} i^k \theta_i^2 = 0$$

we can conclude that

$$\sum_{i=K}^{\infty} \theta_i^2 = o(K^{-k}).$$

□

4 A Rate on Hellinger Distance

In this section we apply a result of Wong and Shen (1995) to obtain a rate of convergence on the Hellinger distance between the true density and its estimate. This rate is used in the next section to establish the asymptotic validity of cross-validation formulae.

Write the true density

$$f_o(x) = [g_o(x)]^2 e^{-x^2/2} + \epsilon_o \phi(x)$$

as

$$f_\infty(x, \theta_o) = \left[\sum_{i=0}^{\infty} \theta_{oi} \bar{H}_{e_i}(x) \right]^2 e^{-x^2/2} + \epsilon_o \phi(x)$$

where $\theta_o = (\theta_{o0}, \theta_{o1}, \theta_{o2}, \dots)$. When θ_o is truncated for use in the formula

$$f_n(x, \theta) = \left[\sum_{i=0}^{K_n} \theta_i \bar{H}_{e_i}(x) \right]^2 e^{-x^2/2} + \epsilon_o \phi(x),$$

the renormalization

$$\theta_n = \frac{\sqrt{(1 - \epsilon_o)}}{\sqrt{\sum_{i=0}^{K_n} (\theta_{oi})^2}} (\theta_{o0}, \theta_{o1}, \dots, \theta_{oK_n})$$

is required to get θ_n in Θ_n . Thus, there is a distinction between $f_n(x, \theta_o)$ and $f_n(x, \theta_n)$.

To apply the Wong and Shen result, we need to compute the bracketing Hellinger metric entropy of the estimator and the chi-squared truncation error, for which we require the following assumption.

ASSUMPTION 1 Let f_o indicate the true density, the density from which the observed data $\{x_t\}$ is a random sample. We shall require f_o to be in \mathcal{F}_∞ , to at be least once continuously differentiable over $(-\infty, \infty)$, and to have a largest mode. Writing $f_K(x) = g_K^2(x)e^{-x^2/2} + \epsilon_o \phi(x)$ for the truncation of $f_o(x) = g_o^2(x)e^{-x^2/2} + \epsilon_o \phi(x)$, we require that

$$\lim_{K \rightarrow \infty} \sup_{-\infty < x < \infty} \frac{g_o^2(x) + g_K^2(x)}{g_K^2(x) + \epsilon_o/\sqrt{2\pi}} \leq C$$

for some $C < \infty$.

The SNP estimator is location and scale invariant. We will exploit this fact in the theoretical development by assuming that the mode given by Assumption 1 occurs at $x = 0$ with $f(0) = 1$ and considering the simplified estimator

$$\hat{f}_n = f_n(x, \hat{\theta}_n)$$

where

$$\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta_n} \sum_{t=1}^n \log [f_n(x_t, \theta)],$$

K_n increases with n at a specified rate, and $\{x_t\}$ is a random sample from the density $f_o(x)$.

The condition $\lim_{K \rightarrow \infty} \sup_{-\infty < x < \infty} [g_o^2(x) + g_K^2(x)] / [g_K^2(x) + \epsilon_o / \sqrt{2\pi}] \leq C$ restricts the tail behavior of f_o . If f_o has normal tails beyond some points well to the left and right of the 1% and 99% quantiles of f_o , the condition is satisfied. If g_o is any polynomial of finite degree plus a bounded function, the condition is satisfied. Restrictions on tail behavior are common in nonparametric density estimation, and this condition seems relatively mild as such restrictions go. Note that Assumption 1 is a restriction on f_o only, we have not restricted \mathcal{F}_n .

Upon applying the Wong and Shen result, we obtain the following rates.

PROPOSITION 1 Let the true density have the form

$$f_o(x) = [g_o(x)]^2 e^{-x^2/2} + \epsilon_o \phi(x),$$

let $\delta \in (0, 1)$ be given, and impose Assumption 1. If g_o is k -times differentiable with $g^{(j)} \in L_2(e^{-x^2/2})$ for $j = 0, 1, \dots, k$, then the Hermite coefficients of g_o satisfy $\sum_{i=K}^{\infty} \theta_i^2 = o(K^{-k})$ and the SNP estimate truncated at

$$K_n = n^{\frac{1-\delta}{k+1}}$$

converges to f_o in the Hellinger metric at the rate

$$\|\hat{f}_n^{1/2} - f_o^{1/2}\| = n^{-\frac{k(1-\delta)}{2k+2}}$$

almost surely. Under the stronger smoothness condition that $\sum_{i=K}^{\infty} \theta_i^2 \approx K^{-k-\xi}$ for some small $\xi > 0$, δ may be set to zero.

Proof We must first compute the bracketing Hellinger metric entropy. The Hellinger metric $d(\cdot, \cdot)$ is

$$d(f_1, f_2) = \left\{ \int [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 dx \right\}^{1/2} = \|f_1^{1/2} - f_2^{1/2}\|.$$

A Hellinger u -bracketing of \mathcal{F}_n is a set of N pairs of functions $\{(f_j^L, f_j^U) : j = 1, 2, \dots, N\}$ such that (i) $d(f_j^L, f_j^U) \leq u$ for $j = 1, 2, \dots, N$ and (ii) for any $h \in \mathcal{F}_n$, there is a j such

that $f_j^L \leq h \leq f_j^U$. Let N^* denote the the smallest such N . The bracketing Hellinger metric entropy of \mathcal{F}_n is the function $H(\cdot, \mathcal{F}_n)$ defined by $H(u, \mathcal{F}_n) = \log N^*$. To compute it, put

$$f_j = g_j^2 e^{-x^2/2} + \epsilon_o \phi$$

$$f_j^L(x) = \inf_{g \in B(j,r)} g^2(x) e^{-x^2/2} + \epsilon_o \phi(x)$$

$$f_j^U(x) = \sup_{g \in B(j,r)} g^2(x) e^{-x^2/2} + \epsilon_o \phi(x)$$

where

$$B(j, r) = \left\{ g = \sum_{i=0}^{K_n} \theta_i \bar{H}_{e_i} : \theta \in \Theta_n, \int (g - g_j)^2 e^{-x^2/2} dx \leq r^2 \right\}$$

with g_j , to be determined later, of the form

$$g_j = \sum_{i=0}^{K_n} \theta_{j,i} \bar{H}_{e_i} \quad \theta_j \in \Theta_n.$$

Note that $g \in B(j, r)$ implies $|\theta_i - \theta_{j,i}| < r$. Also $B(j, r)$ compact implies the infimum and supremum above are attained by some $g \in B(j, r)$ for each x . For $f = g^2 e^{-x^2/2} + \epsilon_o \phi$,

$$\begin{aligned} d^2(f_j^L, f_j^U) &= \int (\sqrt{f_j^U} - \sqrt{f_j^L})^2 dx \\ &\leq \int (\sqrt{f_j^L} - \sqrt{f_j})^2 dx + \int (\sqrt{f_j^U} - \sqrt{f_j})^2 dx \\ &\leq 2 \int \sup_{g \in B(j,r)} (\sqrt{f} - \sqrt{f_j})^2 dx \\ &= 2 \int \sup_{g \in B(j,r)} (\sqrt{f} - \sqrt{f_j})^2 \frac{[\sqrt{f} + \sqrt{f_j}]^2}{[\sqrt{f} + \sqrt{f_j}]^2} dx \\ &= 2 \int \sup_{g \in B(j,r)} \frac{(f - f_j)^2}{[\sqrt{f} + \sqrt{f_j}]^2} dx \\ &= 2 \int \sup_{g \in B(j,r)} \frac{(|g - g_j| e^{-x^2/4})^2 (|g + g_j| e^{-x^2/4})^2}{(\sqrt{f} + \sqrt{f_j})^2} dx \\ &\leq 2 \int \sup_{g \in B(j,r)} (g - g_j)^2 e^{-x^2/2} \sup_{g \in B(j,r)} \frac{(g + g_j)^2 e^{-x^2/2}}{f + f_j} dx \\ &\leq 4 \int \sup_{g \in B(j,r)} (g - g_j)^2 e^{-x^2/2} \sup_{g \in B(j,r)} \frac{g^2 + g_j^2}{g^2 + g_j^2 + 2\epsilon_o/\sqrt{2\pi}} dx \\ &\leq 4 \int \sup_{g \in B(j,r)} (g - g_j)^2 e^{-x^2/2} dx \\ &= 4 \int \sup_{g \in B(j,r)} \sum_{i=0}^{K_n} \sum_{k=0}^{K_n} [\theta_i - \theta_{j,i}] [\theta_k - \theta_{j,k}] \bar{H}_{e_i}(x) \bar{H}_{e_k}(x) e^{-x^2/2} dx \end{aligned}$$

$$\begin{aligned}
&\leq 4 \int r^2 \sum_{i=0}^{K_n} \sum_{k=0}^{K_n} |\overline{H}_{e_i}(x) \overline{H}_{e_k}(x)| e^{-x^2/2} dx \\
&\leq 4r^2 \sum_{i=0}^{K_n} \sum_{k=0}^{K_n} \left[\int \overline{H}_{e_i}^2(x) e^{-x^2/2} dx \int \overline{H}_{e_k}^2(x) e^{-x^2/2} dx \right]^{1/2} \\
&= 4r^2 (K_n + 1)^2.
\end{aligned}$$

Cover Θ_n with open balls of radius $r = u \left[\sqrt{2}(K_n + 1) \right]^{-1}$ and let θ_j for $j = 1, \dots, N$ be the centers of these balls. Every $f \in \mathcal{F}_n$ must be in one of these balls whence $f_j^L \leq f \leq f_j^U$ for some j , and, by the calculation above, $d(f_j^L, f_j^U) \leq u$ for each j . By Gallant and Souza (1991, Lemma 1), the number of open balls of radius r required to cover Θ_n is bounded by $2(K_n + 1)(2/r + 1)^{K_n}$. Therefore, $N^* \leq 2(K_n + 1)(4K_n/u + 1)^{K_n}$ and

$$H(u, \mathcal{F}_n) \leq \log(2K_n + 2) + K_n \log(4K_n/u + 1).$$

We now have a bound on the bracketing Hellinger metric entropy in hand.

Before considering the chi-squared truncation error, we attend to a detail, which is that Wong and Shen require the computation of ϵ that satisfies

$$\int_{\epsilon^2/2^8}^{\sqrt{2}\epsilon} H^{1/2}(u, \mathcal{F}_n) du \leq c_4 n^{1/2} \epsilon^2$$

for a given c_4 . Because both the square root and log functions are concave, Jensen's inequality implies, for $a = \sqrt{2}\epsilon - \epsilon^2/2^8$, that

$$\begin{aligned}
\int_{\epsilon^2/2^8}^{\sqrt{2}\epsilon} H^{1/2}(u, \mathcal{F}_n) du &\leq a \left[\log(2K_n + 2) + K_n \log \left(a^{-1} \int_{\epsilon^2/2^8}^{\sqrt{2}\epsilon} \frac{4K_n}{u} du + 1 \right) \right]^{1/2} \\
&= a \left[\log(2K_n + 2) + K_n \log \left(4K_n a^{-1} \log \frac{\sqrt{2}\epsilon}{\epsilon^2/2^8} + 1 \right) \right]^{1/2}.
\end{aligned}$$

Suppose that $K_n = n^\alpha$ and $\epsilon = n^{-\beta}$ for some $\alpha, \beta > 0$. Then from the expression above it follows that given any $\delta > 0$ there are bounds B_i such that

$$\begin{aligned}
\int_{\epsilon^2/2^8}^{\sqrt{2}\epsilon} H^{1/2}(u, \mathcal{F}_n) du &\leq B_1 n^{-\beta} \left[\alpha \log n + n^\alpha \log \left(n^{\alpha+\beta} \log n^{-\beta} \right) \right]^{1/2} \\
&\leq B_2 n^{-\beta} \left(n^\alpha \log n^{\alpha+\beta} \right)^{1/2} \\
&\leq B_3 n^{-\beta + \alpha/2 + \delta/2}
\end{aligned}$$

Because δ is arbitrary, we may assume that $B_3 \leq c_4$. If we put $\beta = (1 - \alpha - \delta)/2$ so that

$$\epsilon = n^{-\beta} = n^{(\alpha + \delta - 1)/2},$$

then this inequality implies

$$\int_{\epsilon^2/2^8}^{\sqrt{2}\epsilon} H^{1/2}(u, \mathcal{F}_n) du \leq c_4 n^{1/2} \epsilon^2$$

as required.

Having dealt with this detail, we now compute the rate of decrease of the chi-squared truncation error defined as

$$\delta_n(1) = \inf_{f \in \mathcal{F}_n} \rho_1(f_o, f),$$

where $\rho_1(f_o, f) = \int (f_o - f)^2 / f dx$. Using Assumption 1 and writing $f_n = f(\cdot, \theta_n)$ for the truncation of f_o normalized to integrate to one, we have for some $B > 0$ that

$$\begin{aligned} \delta_n(1) &\leq \int \frac{(f_o - f_n)^2}{f_n} dx \\ &= \int \frac{(g_o^2 - g_n^2)^2 e^{-x^2/2}}{g_n^2 + \epsilon_o / \sqrt{2\pi}} dx \\ &= \int \frac{(g_o - g_n)^2 (g_o + g_n)^2 e^{-x^2/2}}{g_n^2 + \epsilon_o / \sqrt{2\pi}} dx \\ &\leq \int \frac{2(g_o^2 + g_n^2)}{g_n^2 + \epsilon_o / \sqrt{2\pi}} (g_o - g_n)^2 e^{-x^2/2} dx \\ &\leq B \int (g_o - g_n)^2 e^{-x^2/2} dx \quad \text{for large } K \\ &= B \int \left[\sum_{i=0}^{\infty} \theta_{oi} \bar{H}_{e_i} - \sum_{i=0}^{K_n} \theta_i \bar{H}_{e_i} \right]^2 e^{-x^2/2} dx \\ &= B \int \left[\sum_{i=K_n+1}^{\infty} \theta_{oi} \bar{H}_{e_i} + \sum_{i=0}^{K_n} (\theta_i - \theta_{oi}) \bar{H}_{e_i} \right]^2 e^{-x^2/2} dx \\ &\leq 2B \int \left\{ \left[\sum_{i=K_n+1}^{\infty} \theta_{oi} \bar{H}_{e_i} \right]^2 + \left[\sum_{i=0}^{K_n} (\theta_i - \theta_{oi}) \bar{H}_{e_i} \right]^2 \right\} e^{-x^2/2} dx \\ &= 2B \sum_{i=K_n+1}^{\infty} (\theta_{oi})^2 + 2B \sum_{i=0}^{K_n} (\theta_i - \theta_{oi})^2 \\ &= 2B \sum_{i=K_n+1}^{\infty} (\theta_{oi})^2 + 2B \frac{[\sqrt{\sum_{i=1}^{\infty} (\theta_{oi})^2} - \sqrt{\sum_{i=1}^{K_n} (\theta_{oi})^2}]^2}{\sum_{i=0}^{K_n} (\theta_{oi})^2} \sum_{i=0}^{K_n} (\theta_{oi})^2 \\ &= 2B \sum_{i=K_n+1}^{\infty} (\theta_{oi})^2 + 2B \left[\sum_{i=1}^{\infty} (\theta_{oi})^2 + \sum_{i=1}^{K_n} (\theta_{oi})^2 - 2\sqrt{\sum_{i=1}^{\infty} (\theta_{oi})^2} \sqrt{\sum_{i=1}^{K_n} (\theta_{oi})^2} \right] \\ &\leq 2B \sum_{i=K_n+1}^{\infty} (\theta_{oi})^2 + 2B \left[\sum_{i=1}^{\infty} (\theta_{oi})^2 - \sum_{i=1}^{K_n} (\theta_{oi})^2 \right] \end{aligned}$$

$$\begin{aligned}
&= 4B \sum_{i=K_n+1}^{\infty} (\theta_{oi})^2 \\
&= o(K_n^{-k}).
\end{aligned}$$

Therefore, the rate of decrease of the chi-squared truncation error is $\delta_n(1) = o(K_n^{-k})$.

We are now in a position to apply the Wong and Shen result. For every $\theta \in \Theta_n$ and $c_1 > 0$,

$$\left\{ \|f_n^{1/2}(\cdot, \hat{\theta}_n) - f_o^{1/2}\| \geq \epsilon \right\} \subset A(n, \theta)$$

where

$$A(n, \theta) = \left\{ \sup_{\|f^{1/2} - f_o^{1/2}\| \geq \epsilon, f \in \mathcal{F}_n} \prod_{i=1}^n \frac{f(x_i)}{f_n(x_i, \theta)} \geq \exp\left(-\frac{1}{2}c_1 n \epsilon^2\right) \right\}.$$

Using Theorem 3(i) of Wong and Shen, we have

$$P \left\{ \|\hat{f}_n^{1/2} - f_o^{1/2}\| \geq \epsilon \right\} \leq \inf_{\theta \in \Theta_n} P[A(n, \theta)] \leq 5 \exp(-c_2 n \epsilon^2) + \exp[-nc_1 \epsilon^2 / 2 + n \delta_n(1)].$$

Substituting $\epsilon = n^{(\alpha+\delta-1)/2}$ and $\delta_n(1) = o(n^{-\alpha k})$ from above we have

$$P \left\{ \|\hat{f}_n^{1/2} - f_o^{1/2}\| \geq n^{(\alpha+\delta-1)/2} \right\} \leq 5 \exp(-c_2 n^{\alpha+\delta}) + \exp[-c_1 n^{\alpha+\delta} / 2 + o(n^{-\alpha k+1})].$$

The best rate is achieved when $n^{\alpha+\delta} \approx n^{-\alpha k+1}$; that is, when $\alpha = \frac{1-\delta}{k+1}$. This gives,

$$\begin{aligned}
&P \left\{ \|\hat{f}_n^{1/2} - f_o^{1/2}\| \geq n^{-k(1-\delta)/(2k+2)} \right\} \\
&\leq 5 \exp(-c_2 n^{(1+\delta k)/(k+1)}) + \exp \left\{ -n^{(1+\delta k)/(k+1)} [(c_1/2) - o(1)] \right\}.
\end{aligned}$$

Application of the Borel-Cantelli lemma yields

$$\|\hat{f}_n^{1/2} - f_o^{1/2}\| = n^{-k(1-\delta)/(2k+2)}$$

almost surely where $\delta > 0$ is arbitrary.

If we were to further restrict f_o so that $\sum_{i=K}^{\infty} \theta_i^2 \approx K^{-k-\xi}$ for some small $\xi > 0$, rather than using $\sum_{i=K}^{\infty} \theta_i^2 = o(K^{-k})$ given by Lemma 2, then we would have $\delta_n(1) \approx K^{-k-\xi}$ and could set $\delta = 0$ in the expression for $\|\hat{f}_n^{1/2} - f_o^{1/2}\|$ above. \square

5 Cross-Validation

As seen earlier, for the applications we have in mind, we require an estimate of the integrated squared error curve that is uniform in the truncation point of the SNP estimator. In this section, we shall derive this result by pursuing the more traditional goal of trying to find K_n that minimizes integrated squared error and obtain a uniform estimate of the integrated squared error curve as a by-product. As sample size n increases, we shall presume that an observation in $\mathcal{X}_{j,\alpha}$ remains there. This amounts to regarding the $\mathcal{X}_{j,\alpha}$ as bins to which observations are added and has the effect of guaranteeing that $\mathcal{X}_{j,\alpha}$ on n observations and $\tilde{\mathcal{X}}_{j,\alpha}$ on n' observations are independent collections even when $n \neq n'$.

We require a sequence K_n that drives the integrated squared error to zero which does not depend on the hold-out sample $\mathcal{X}_{j,\alpha}$. We shall establish its existence by relating the L_2 norm

$$\|f_1 - f_2\| = \left[\int (f_1 - f_2)^2 dx \right]^{1/2}$$

to Hellinger distance

$$d(f_1, f_2) = \left[\int (f_1^{1/2} - f_2^{1/2})^2 dx \right]^{1/2}$$

and applying the results of the previous section. Assumption 1 implies $f_o(x) \leq 1$ whereas an estimate \hat{f}_n may not satisfy $\hat{f}_n \leq 1$. If we replace \hat{f}_n by $\tilde{f}_n = \min\{\hat{f}_n, 1\}$, we improve the accuracy of an approximation because $\|f_o - \tilde{f}_n\| \leq \|f_o - \hat{f}_n\|$ and $d(f_o, \tilde{f}_n) \leq d(f_o, \hat{f}_n)$. It may be that $\int \tilde{f}_n dx < 1$ due to this truncation, but it is of no concern in this section. Throughout this section we shall merely presume that truncation has been applied as necessary to guarantee that densities are bounded by one without using a special notation. In consequence, the only density that we may presume integrates to one is the true density f_o .

The next Lemma shows that the Hellinger distance is bounded from below by the L_2 norm times a constant.

LEMMA 3 Under Assumption 1

$$\int [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 dx \geq \frac{1}{4} \int [f_1(x) - f_2(x)]^2 dx.$$

Proof Note that

$$\begin{aligned}
\int [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 dx &= \int [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 \frac{[f_1^{1/2}(x) + f_2^{1/2}(x)]^2}{[f_1^{1/2}(x) + f_2^{1/2}(x)]^2} dx \\
&= \int \frac{[f_1(x) - f_2(x)]^2}{[f_1^{1/2}(x) + f_2^{1/2}(x)]^2} dx \\
&\geq \frac{1}{4} \int [f_1(x) - f_2(x)]^2 dx.
\end{aligned}$$

The last line uses the fact that $f_1, f_2 \leq 1$. □

Replacing f_1 with \hat{f}_n and f_2 with f_o , we see this result establishes the existence of a sequence of a functions \hat{f}_n for which $L_2 \rightarrow 0$ a.s. because the results of the previous section establish the existence of such a sequence for Hellinger distance.

We also need to establish a similar type of inequality in the other direction. Here and throughout the remainder of this section, C denotes a generic upper bound.

LEMMA 4 Given the conditions in Lemma 3 and some M , $0 < M < \infty$,

$$\int_{-M}^M [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 dx \leq C \int [f_1(x) - f_2(x)]^2 dx.$$

Proof Similar to the proof of Lemma 3, we have

$$\begin{aligned}
\int_{-M}^M [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 dx &= \int_{-M}^M \frac{[f_1(x) - f_2(x)]^2}{[f_1^{1/2}(x) + f_2^{1/2}(x)]^2} dx \\
&\leq \frac{1}{4\epsilon_o\phi(M)} \int_{-M}^M [f_1(x) - f_2(x)]^2 dx.
\end{aligned}$$

□

This result guarantees that as $L_2^2 \rightarrow 0$, $\int_{-M}^M [f_1^{1/2}(x) - f_2^{1/2}(x)]^2 \rightarrow 0$ at the same rate for $M < \infty$ arbitrarily large. This latter distance is just the Hellinger distance defined on a strict subset of the underlying support. A result which covers the entire support as $n \rightarrow \infty$ is given in the lemma below.

LEMMA 5 Suppose that for some sequence of densities, $\{f_{1,n}, f_{2,n}\}$,

$$\int [f_{1,n}(x) - f_{2,n}(x)]^2 \leq Cn^{-\tau}, \quad \tau > 0.$$

Set $M_n = (2 \log(n^{\tau_M}))^{1/2}$, $0 < \tau_M < \tau$. Then under the conditions in Lemma 3,

$$\int_{-M_n}^{M_n} [f_{1,n}^{1/2}(x) - f_{2,n}^{1/2}(x)]^2 dx \leq Cn^{-\tau+\tau_M}.$$

Proof The proof follows from Lemma 4 since $\phi(M_n) = n^{-\tau_M} / \sqrt{2\pi}$. \square

The goal now is to construct a data driven procedure that selects K . Proposition 1 implies that the number of terms should not grow faster than $\mathcal{O}(n^{1/2})$. Therefore, we choose some positive constant C^* and then search for K over the set

$$\mathcal{K}_n = \{1, 2, \dots, \lfloor C^* n^{1/2} \rfloor\},$$

where $\lfloor a \rfloor$ is the integer part of a .

We will next establish uniform convergence over \mathcal{K}_n . To do this, set

$$\begin{aligned} \hat{Q}_{j,n}(K) &= \int \hat{f}_{j,K}^2(x) dx - \frac{2}{\lfloor \alpha n \rfloor} \sum_{x_i \in \mathcal{X}_{j,\alpha}} \hat{f}_{j,K}(x_i) + \int f_o^2(x) dx, \\ Q_{j,n}(K) &= \int \hat{f}_{j,K}^2(x) dx - 2 \int \hat{f}_{j,K}(x) f_o(x) dx + \int f_o^2(x) dx, \end{aligned}$$

and

$$\begin{aligned} \hat{S}_{j,n}(K) &= \hat{Q}_{j,n}(K) + \frac{2}{\lfloor \alpha n \rfloor} \sum_{x_i \in \mathcal{X}_{j,\alpha}} f_o(x_i), \\ S_{j,n}(K) &= Q_{j,n}(K) + 2 \int f_o^2(x) dx. \end{aligned}$$

In Section 2, choosing \hat{K}_n to minimize $\text{CVH} = \hat{M}_{(1)} - 2\hat{M}_{(2)}$ was proposed as a possible truncation point of the SNP estimator. Minimizing CVH is equivalent to minimizing either $(1/J) \sum_{j=1}^J \hat{Q}_{j,n}(K)$ or $(1/J) \sum_{j=1}^J \hat{S}_{j,n}(K)$ because these two functions only differ from CVH by additive terms that do not depend on K . The population analog of this choice — in the sense of replacing $\hat{M}_{(2)}$ by $M_{(2)}$ — is

$$K_n = \operatorname{argmin}_{K \in \mathcal{K}_n} \frac{1}{J} \sum_{j=1}^J Q_{j,n}(K) = \operatorname{argmin}_{K \in \mathcal{K}_n} \frac{1}{J} \sum_{j=1}^J S_{j,n}(K).$$

The next Assumption is used to help show that $\sup_{K \in \mathcal{K}_n} |\hat{Q}_{j,n}(K) - Q_{j,n}(K)|$ tends to zero at a faster rate than $Q_{j,n}(K_n)$. The rate is needed to establish uniform convergence.

ASSUMPTION 2 For each j ,

$$P \left(\bigcap_{\zeta > 0} \left\{ \liminf_{n \rightarrow \infty} \frac{Q_{j,n}(K_n)}{n^{-1+\zeta}} = 0 \right\} \right) = 0.$$

This assumption rules out the case when $Q_{j,n}(K_n)$ decreases to zero at a rate n^{-1} . Stone (1982) has shown that in the typical nonparametric setting, $Q_{j,n}(K_n)$ must converge to zero slower than n^{-1} . Thus, we are effectively ruling out the parametric case where the underlying density is actually a finite sum of Hermite polynomials. For the anticipated applications described in Section 1, the parametric case is implausible and, therefore, uninteresting.

We are now ready to state and prove the uniform convergence result.

LEMMA 6 Under Assumptions 1 and 2

$$\sup_{K \in \mathcal{K}_n} \left| \frac{\hat{S}_{j,n}(K) - S_{j,n}(K)}{Q_{j,n}(K)} \right| = o_s(1), \quad j = 1, \dots, J.$$

Proof Put

$$B_{j,\zeta} = \left\{ \liminf_{n \rightarrow \infty} \frac{Q_{j,n}(K_n)}{n^{-1+\zeta}} = 0 \right\}.$$

$B_{j,\zeta}$ is independent of $\mathcal{X}_{j,\alpha}$ for every n ; on its complement $\tilde{B}_{j,\zeta}$ there is a constant C such that $Q_{j,n}(K_n) > Cn^{-1+\zeta}$ for large n . Fix $K \in \mathcal{K}_n$ and $j \in \{1, \dots, J\}$ and define

$$\begin{aligned} U_{j,n}(K) &= \hat{S}_{j,n}(K) - S_{j,n}(K) \\ &= \frac{2}{[\alpha n]} \sum_{x_i \in \mathcal{X}_{j,\alpha}} [f_o(x_i) - \hat{f}_{j,K}(x_i)] - 2 \int [f_o(x) - \hat{f}_{j,K}(x)] f_o(x) dx. \end{aligned}$$

Note, $\mathcal{E}[U_{j,n}(K) | \tilde{B}_{j,\zeta}] = 0$, $|U_{j,n}(K)|$ is bounded, and $V_{j,n}(K) = [\alpha n] \text{Var}[U_{j,n}(K) | \tilde{B}_{j,\zeta}] \leq CQ_{j,n}(K)$ for some C . Then for large n and some C , we have by Bernstein's inequality (Pollard, 1984, p. 193) that for small $\eta > 0$,

$$\begin{aligned} P \left[|U_{j,n}(K)| \geq \eta Q_{j,n}(K) \mid \tilde{B}_{j,\zeta} \right] &\leq 2 \exp \left[-\frac{\frac{1}{2}[\alpha n] \eta^2 Q_{j,n}^2(K)}{V_{j,n}(K) + C\eta Q_{j,n}(K)} \right] \\ &\leq 2 \exp \left[-Cn\eta^2 Q_{j,n}(K) \right] \\ &\leq 2 \exp \left[-Cn^\zeta \eta^2 \right]. \end{aligned}$$

Now we will no longer hold K fixed. For an overall probabilistic bound, observe that

$$P \left[\sup_{K \in \mathcal{K}_n} |U_{j,n}(K)/Q_{j,n}(K)| \geq \eta \mid \tilde{B}_{j,\zeta} \right] \leq (\#\mathcal{K}_n) \sup_{K \in \mathcal{K}_n} P \left[|U_{j,n}(K)/Q_{j,n}(K)| \geq \eta \mid \tilde{B}_{j,\zeta} \right],$$

where $\#\mathcal{K}_n$ is the cardinality of \mathcal{K}_n , which is $C^*n^{1/2}$. This implies a bound of

$$2C^*n^{1/2} \exp(-Cn^\zeta \eta^2) \leq \exp(-Cn^{\zeta/2} \eta^2)$$

for large n . By the Borel-Cantelli Lemma

$$P \left[\limsup_{n \rightarrow \infty} \sup_{K \in \mathcal{K}_n} \left| \frac{\hat{S}_{j,n}(K) - S_{j,n}(K)}{Q_{j,n}(K)} \right| \neq 0 \mid \tilde{B}_{j,\zeta} \right] = 0.$$

Letting Q denote the positive rational numbers, Assumption 2 implies $P \left(\bigcup_{\zeta \in Q} \tilde{B}_{j,\zeta} \right) = 1$, whence

$$\begin{aligned} & P \left[\limsup_{n \rightarrow \infty} \sup_{K \in \mathcal{K}_n} \left| \frac{\hat{S}_{j,n}(K) - S_{j,n}(K)}{Q_{j,n}(K)} \right| \neq 0 \right] \\ & \leq \sum_{\zeta \in Q} P \left(\tilde{B}_{j,\zeta} \right) P \left[\limsup_{n \rightarrow \infty} \sup_{K \in \mathcal{K}_n} \left| \frac{\hat{S}_{j,n}(K) - S_{j,n}(K)}{Q_{j,n}(K)} \right| \neq 0 \mid \tilde{B}_{j,\zeta} \right] = 0. \end{aligned}$$

□

THEOREM 1 Under Assumptions 1 through 2

$$\sup_{K \in \mathcal{K}_n} \left| \frac{\sum_{j=1}^J \hat{S}_{j,n}(K) - \sum_{j=1}^J S_{j,n}(K)}{\sum_{j=1}^J Q_{j,n}(K)} \right| = o_s(1).$$

Proof Because

$$\frac{Q_{j,n}(K)}{\sum_{j=1}^J Q_{j,n}(K)} < 1,$$

we have

$$\begin{aligned} \left| \frac{\sum_{j=1}^J \hat{S}_{j,n}(K) - \sum_{j=1}^J S_{j,n}(K)}{\sum_{j=1}^J Q_{j,n}(K)} \right| & \leq \sum_{j=1}^J \frac{|\hat{S}_{j,n}(K) - S_{j,n}(K)|}{\sum_{j=1}^J Q_{j,n}(K)} \\ & = \sum_{j=1}^J \left[\frac{Q_{j,n}(K)}{\sum_{j=1}^J Q_{j,n}(K)} \right] \left[\frac{|\hat{S}_{j,n}(K) - S_{j,n}(K)|}{Q_{j,n}(K)} \right] \\ & \leq \sum_{j=1}^J \frac{|\hat{S}_{j,n}(K) - S_{j,n}(K)|}{Q_{j,n}(K)}. \end{aligned}$$

□

We can now show that \hat{K}_n is consistent.

THEOREM 2 Under Assumptions 1 through 2

$$\frac{\sum_{j=1}^J Q_{j,n}(\hat{K}_n)}{\sum_{j=1}^J Q_{j,n}(K_n)} = 1 \quad \text{a.s.}$$

Proof Applying Theorem 1 we have

$$\begin{aligned}
0 &\leq \sum_{j=1}^J Q_{j,n}(\hat{K}_n) - \sum_{j=1}^J Q_{j,n}(K_n) \\
&= \sum_{j=1}^J S_{j,n}(\hat{K}_n) - \sum_{j=1}^J S_{j,n}(K_n) \\
&= \sum_{j=1}^J \hat{S}_{j,n}(\hat{K}_n) - \sum_{j=1}^J S_{j,n}(K_n) + o_s \left[\sum_{j=1}^J Q_{j,n}(\hat{K}_n) \right] \\
&\leq \sum_{j=1}^J \hat{S}_{j,n}(K_n) - \sum_{j=1}^J S_{j,n}(K_n) + o_s \left[\sum_{j=1}^J Q_{j,n}(\hat{K}_n) \right] \\
&= \sum_{j=1}^J S_{j,n}(K_n) - \sum_{j=1}^J S_{j,n}(K_n) + o_s \left[\sum_{j=1}^J Q_{j,n}(\hat{K}_n) \right] + o_s \left[\sum_{j=1}^J Q_{j,n}(K_n) \right] \\
&= o_s \left[\sum_{j=1}^J Q_{j,n}(\hat{K}_n) \right] + o_s \left[\sum_{j=1}^J Q_{j,n}(K_n) \right] \\
&\leq 2 o_s \left[\sum_{j=1}^J Q_{j,n}(\hat{K}_n) \right] \\
&= Z_n.
\end{aligned}$$

Then

$$\begin{aligned}
1 &\leq \frac{\sum_{j=1}^J Q_{j,n}(\hat{K}_n)}{\sum_{j=1}^J Q_{j,n}(K_n)} \\
&= \frac{\sum_{j=1}^J [Q_{j,n}(\hat{K}_n) - Q_{j,n}(K_n) + Q_{j,n}(K_n)]}{\sum_{j=1}^J [Q_{j,n}(K_n) - Q_{j,n}(\hat{K}_n) + Q_{j,n}(\hat{K}_n)]} \\
&= \frac{\sum_{j=1}^J Q_{j,n}(K_n) + Z_n}{\sum_{j=1}^J Q_{j,n}(\hat{K}_n) - Z_n} \\
&\leq \frac{\sum_{j=1}^J Q_{j,n}(\hat{K}_n) + Z_n}{\sum_{j=1}^J Q_{j,n}(\hat{K}_n) - Z_n} \\
&= \frac{[1 + 2o_s(1)] \sum_{j=1}^J Q_{j,n}(\hat{K}_n)}{[1 - 2o_s(1)] \sum_{j=1}^J Q_{j,n}(\hat{K}_n)} \\
&\rightarrow 1 \quad \text{a.s.}
\end{aligned}$$

□

6 Extensions

Computationally, the extension of cross-validation to the multivariate case for iid data is straightforward. Formulas that define the SNP estimator $\hat{f}_K(x)$ for $x \in \mathfrak{R}^d$ are in Gallant and Nychka (1987). A random sample $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ with $x_i \in \mathfrak{R}^d$ can be divided into groups $\mathcal{X}_{j,\alpha}$ in the same manner as in Section 2 because the allocation schemes there only involve the index i of $x_i \in \mathcal{X}$. The formulas given in Section 2 can be applied to the multivariate form of $\hat{f}_K(x)$ without alteration, thus defining $\hat{f}_{j,K}(x)$, CVL, CVH, etc. The most onerous tasks in an extension of the proofs to the multivariate case are to obtain the results of Section 3 and to find a tractable notation for describing them. The remainder of the arguments would be straightforward. Likely to be of more practical value than the theory would be a careful empirical study of some relevant multivariate test cases in the style of Section 2. Sain, Baggerly, and Scott (1994) find that the performance of cross-validation actually improves in the multivariate case for kernel density estimates; perhaps a careful study of test cases would reveal that this is the case for SNP density estimates as well.

For the extension to the time series case, let

$$\mathcal{Y} = \{y_1, y_2, \dots, y_n\}$$

denote observations on a multivariate time series that is Markovian in the lags $x_{t-1} = (y_{t-L}, \dots, y_{t-1})$ with conditional density $f(y|x)$ and marginal density $f(x)$. Denote the SNP conditional density by $f_K(y|x, \theta)$ whose functional form and parameter space Θ are described in Gallant and Tauchen (1989); see Gallant and Tauchen (2001) for refinements. The SNP estimate computed from the entire sample is

$$\hat{f}_K = \operatorname{argmax}_{\theta \in \Theta} \sum_{y_{t-L}, y_t \in \mathcal{Y}} \log f(y_t | x_{t-1}, \theta).$$

Define

$$\begin{aligned} \text{ISE}(\hat{f}_K) &= \int \left\{ \int [\hat{f}_K(y|x) - f(y|x)]^2 dy \right\} f(x) dx \\ &= \iint \hat{f}_K^2(y|x) dy f(x) dx - 2 \iint \hat{f}_K(y|x) f(y|x) dy f(x) dx + \iint f^2(y|x) dy f(x) dx \\ &= M_{(1)} - 2M_{(2)} + M_{(3)}. \end{aligned}$$

Let \mathcal{Y} be partitioned into contiguous blocks $\mathcal{Y}_{j,\alpha}$ of approximate size $[\alpha n]$ as follows: For a given proportion α that can be expressed as $\alpha = 1/J$, where $J \leq n$ and $[\alpha n]$ is the integer part of αn , set

$$\begin{aligned}\mathcal{Y}_{j,\alpha} &= \{y_{(j-1)[\alpha n]+1}, \dots, y_{j[\alpha n]}\} & j = 1, \dots, J-1 \\ \mathcal{Y}_{J,\alpha} &= \{y_{(J-1)[\alpha n]+1}, \dots, y_n\}.\end{aligned}$$

The SNP estimate computed from the data with the j th block deleted is

$$\hat{f}_{j,K} = \operatorname{argmax}_{\theta \in \Theta} \sum_{y_{t-L}, y_t \in \tilde{\mathcal{Y}}_{j,\alpha}, y_{t-L}, y_t \notin \mathcal{Y}_{j,\alpha}} \log f(y_t | x_{t-1}, \theta)$$

where $\tilde{\mathcal{Y}}_{j,\alpha} = \mathcal{Y} \setminus \mathcal{Y}_{j,\alpha}$. An estimate of ISE is

$$\text{CVH} = \frac{1}{J} \sum_{j=1}^J \sum_{y_{t-L}, y_t \in \tilde{\mathcal{Y}}_{j,\alpha}, y_{t-L}, y_t \notin \mathcal{Y}_{j,\alpha}} \left[\int \hat{f}_{j,K}^2(y | x_{t-1}) dy - 2\hat{f}_{j,K}^2(y_t | x_{t-1}) \right].$$

Developing a proof strategy justifying this formula would entail the difficulties in obtaining the results of Section 3 discussed above. With such results in hand, one would hope to be able to get rate results for quasi maximum likelihood estimates similar to those in Section 4 under mixing conditions by adapting the methods of proof in Wong and Shen (1995), Chen and Shen (1998), and Chen and Carrasco (2001). The cross-validation formula above could then be justified by applying an exponential inequality for dependent processes as in Section 5. But as remarked above, of more practical value than the theory would most likely be a careful empirical study of some relevant test cases in the style of Section 2.

7 Conclusion

Through the use of examples covering a wide range of cases, we exhibited the phenomenon that a plot of the integrated squared error (ISE) of the seminonparametric (SNP) density estimator against the truncation point K of the SNP estimate tends to rapidly decrease until a certain point and then abruptly level off. This phenomenon has been noted previously in the literature where it has also been noted that the point where the abrupt change occurs is where K should be chosen when SNP is used in connection with the efficient method of moments (EMM) simulation estimator. Given this, an ideal methodology for selecting the

K of an SNP estimate intended for use in EMM applications would be one that accurately predicts this change point and that can be computed at reasonable cost. Two of the more common bandwidth procedures in the literature are the Schwarz criterion (BIC) and leave-one-out cross-validation (CVL). But BIC often selects a K which is too small and CVL is relatively expensive to compute because it requires as many nonlinear optimizations as the sample size n . Here, we proposed an average hold-out-sample cross-validation criterion (CVH) that is much cheaper to compute; for instance, CVH computed from 10% hold-out-samples requires 10 nonlinear optimizations on samples of size $.9n$. In a Monte Carlo study, results were not sensitive to the size of the hold-out-sample. In the same Monte Carlo study, and in numerous examples, CVH did as well at finding K as CVL. A theoretical justification for CVH was provided.

8 References

- Abramowitz, M., and I. A. Stegun, I. A., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York.
- Andersen, T. G., Hyung-Jin Chung, H.-J. and Sorensen, B. E., 1999. Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics* 91, 61–87.
- Chen, X., and Carrasco, M., 2001. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, forthcoming.
- Chen, X., and Shen, X., 1998. Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.
- Chernov, M., Gallant, A. R., Ghysels, E., Tauchen, G., 2001. Alternative models for stock price dynamics. Manuscript, Business School, Columbia University, New York.
- Chumacero, R., 1997. Finite sample properties of the efficient method of moments. *Studies in Nonlinear Dynamics and Econometrics* 2, 35–51.

- Davidian, M., and Gallant, A. R., 1992. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine, “ *Journal of Pharmacokinetics and Biopharmaceutics* 20, 531–558.
- Davidian, M., and Gallant, A.R., 1993. The nonlinear mixed effects model with a smooth random effects density, “ *Biometrika* 80, 475–488.
- Devroye, L., and Györfi, L., 1985. *Nonparametric Density Estimation, the L_1 View*. Wiley, New York.
- Eastwood, B. J., and Gallant, A. R., 1991. Adaptive rules for seminonparametric estimators that achieve asymptotic normality. *Econometric Theory* 7, 307–340.
- Edwards, D., 1990. Hierarchical interaction models. *Journal of the Royal Statistical Society, Series B* 52, 3–20.
- Fan, J., Zhang, C., and Zhang, J., 2000. Sieve likelihood ratio statistics and wilks phenomenon. Manuscript, Department of Statistics, University of North Carolina, Chapel Hill NC.
- Fenton, V. M., and Gallant, A. R., 1996. Qualitative and asymptotic performance of SNP density estimators. *Journal of Econometrics* 74, 77–118.
- Gallant, A. R., Hsieh, D. A., and Tauchen, G. E., 1991. On fitting a recalcitrant series: the pound/dollar exchange rate, 1974–83. in Barnett, W. A., Powell, J., and Tauchen, G. E., eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Chapter 8, 199–240, Cambridge University Press, Cambridge.
- Gallant, A. R., Hsu, C-T., and Tauchen, G., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *The Review of Economics and Statistics* 81(4), 617–631.
- Gallant, A. R., and Long, J. R., 1997. Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* 84, 125–141.

- Gallant, A. R., and Nychka, D. W., 1987. Semiparametric maximum likelihood estimation. *Econometrica* 55, 363–390.
- Gallant, A. R., Rossi, P. E., and Tauchen, G., 1992. Stock prices and volume. *The Review of Financial Studies* 5, 199–242.
- Gallant, A. R., Rossi, P. E., and Tauchen, G., 1993. Nonlinear dynamic structures. *Econometrica* 61, 871–907.
- Gallant, A. R., and Souza, G., 1991. On the asymptotic normality of Fourier flexible form estimates. *Journal of Econometrics* 50, 329–353.
- Gallant, A. R., and Tauchen, G., 1989. Semiparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* 57, 1091–1120.
- Gallant, A. R., and Tauchen, G., 1999. The relative efficiency of method of moments estimators. *Journal of Econometrics* 92, 149–172.
- Gallant, A. R., and Tauchen, G., 2001. Efficient method of moments. Manuscript, Department of Economics, Duke University, Durham NC.
- Ghysels, E., Harvey, A., and Renault, E., 1995. Stochastic volatility. in ed. Maddala, G. S., *Handbook of Statistics, Vol. 14, Statistical Methods in Finance*, North Holland, Amsterdam.
- Johnson, N. L., and Kotz, S., 1994. *Continuous Univariate Distributions-1*, 2nd ed., New York, Wiley.
- Laffont, J.-J., Ossard, H., and Vuong, Q., 1995. The econometrics of first price auctions. *Econometrica* 63, 953–980.
- Lee, Y., and Nelder, J. A., 1996. Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B* 58, 619–678.
- Li, K.-C., 1987. Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics* 15, 958–975.

- Liu, M., 2000. Modeling long memory in stock market volatility. *Journal of Econometrics* 99, 139–171.
- Marron, J. S., 1987. A comparison of cross-validation techniques in density estimation. *The Annals of Statistics* 15, 152–162.
- Marron, J. S., and Wand, M. P., 1992. Exact mean integrated squared error. *The Annals of Statistics* 20, 712–736.
- Michaelides, A., and Ng, S., 2000. Estimating the rational expectations model of speculative storage: a Monte Carlo comparison of three simulation estimators. *Journal of Econometrics* 51, 231–266.
- Pollard, D., 1984. *Convergence of Stochastic Processes*. Springer-Verlag, New York.
- Sain, S. R., Baggerly, K. A., and Scott, D. W., 1994. Cross-validation of multivariate densities. *Journal of the American Statistical Association* 89, 807–817.
- Sansone, G., 1991. *Orthogonal Functions*. Dover, New York.
- Schwarz, G., 1978. Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
- Stasny, E. A., 1991. Hierarchical models for the probabilities of a survey classification and nonresponse: an example from the National Crime Survey. *Journal of the American Statistical Association* 86, 296–303.
- Stone, C. J., 1982. Optimal global rates of convergence of nonparametric regression. *Annals of Statistics* 10, 1040–1053.
- Tauchen, G., and Pitts, M., 1983. The price variability-volume relationship on speculative markets. *Econometrica* 51, 485–505.
- Wong, W. H., and Shen, X., 1995. Probability inequalities for likelihood ratios and convergence rates of sieve mles. *The Annals of Statistics* 23, 339–362.

Tables and Figures

Mean	K														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$n = 0400$															
ISE	.031	.030	.024	.023	.013	.013	.011	.009	.005	.004	.004	.005	.005	.005	.005
Err CVH _{.2}	.008	.008	.008	.007	.007	.007	.008	.007	.007	.007	.007	.008	.008	.008	.008
Err CVH _{.1}	.008	.008	.007	.007	.007	.007	.008	.007	.007	.007	.007	.008	.008	.008	.008
Err CVH _{.05}	.008	.008	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007	.008	.008	.008
Err CVL	.008	.008	.007	.007	.007	.007	.007	.007	.007	.007	.007	.008	.008	.008	.008
$n = 0900$															
ISE	.030	.030	.026	.025	.013	.013	.011	.010	.003	.003	.003	.003	.003	.003	.003
Err CVH _{.2}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.004	.004	.004	.005	.005	.005
Err CVH _{.1}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.004	.004	.004	.005	.005	.005
Err CVH _{.05}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.004	.004	.005	.005	.005	.005
Err CVL	.005	.005	.005	.005	.005	.005	.005	.005	.005	.004	.004	.005	.005	.005	.005
$n = 1600$															
ISE	.030	.029	.027	.026	.014	.014	.012	.011	.003	.002	.002	.002	.002	.002	.002
Err CVH _{.2}	.004	.004	.004	.004	.003	.003	.004	.003	.004	.003	.003	.003	.003	.003	.003
Err CVH _{.1}	.004	.004	.004	.004	.003	.003	.004	.003	.004	.003	.003	.003	.003	.003	.003
Err CVH _{.05}	.004	.004	.004	.004	.003	.003	.004	.003	.004	.003	.003	.003	.003	.003	.003
Err CVL	.004	.004	.004	.004	.003	.004	.003	.003	.003	.003	.003	.003	.003	.003	.003
$n = 2500$															
ISE	.030	.029	.028	.027	.016	.016	.014	.013	.003	.002	.002	.002	.002	.002	.001
Err CVH _{.2}	.003	.003	.003	.003	.003	.003	.004	.003	.004	.002	.002	.003	.003	.003	.003
Err CVH _{.1}	.003	.003	.003	.003	.003	.003	.004	.003	.004	.002	.002	.002	.002	.003	.003
Err CVH _{.05}	.003	.003	.003	.003	.003	.003	.004	.003	.004	.002	.002	.002	.002	.003	.003
Err CVL	.003	.003	.003	.003	.003	.003	.003	.003	.003	.002	.002	.003	.003	.003	.003

Table 1. MISE of SNP Estimates and CV Approximation Error for the Trimodal Density. Shown as Mean ISE is $\mathcal{E} \int (\hat{f}_K - f_o)^2 dx$ and as Mean Err CVL is $\sqrt{\mathcal{E} \left[(\hat{f}_K - f_o)^2 - \text{CVL} + M_{(3)} \right]^2}$ where expectation is computed by averaging over 1024 Monte Carlo repetitions. Similarly for Err CVH.

Mean	<i>K</i>														
	16	17	18	19	20	25	30	35	40	45	50	55	60	65	70
<i>n</i> = 0400															
ISE	.005	.006	.006	.006	.006	.007	.008	.010	.011	.013	.015	.017	.018	.020	.022
Err CVH _{.2}	.008	.008	.008	.008	.008	.009	.009	.010	.010	.011	.011	.011	.012	.012	.013
Err CVH _{.1}	.008	.008	.008	.008	.008	.009	.009	.010	.010	.010	.011	.011	.011	.012	.012
Err CVH _{.05}	.008	.008	.008	.008	.008	.009	.009	.009	.010	.010	.011	.011	.011	.011	.012
Err CVL	.008	.008	.008	.008	.008	.009	.009	.010	.010	.011	.011	.012	.012	.013	.013
<i>n</i> = 0900															
ISE	.002	.003	.003	.003	.003	.003	.003	.004	.004	.005	.005	.006	.006	.007	.007
Err CVH _{.2}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.006	.006	.006	.006
Err CVH _{.1}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.006	.006	.006	.006
Err CVH _{.05}	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.006	.006	.006	.006
Err CVL	.005	.005	.005	.005	.005	.005	.005	.005	.005	.005	.006	.006	.006	.006	.006
<i>n</i> = 1600															
ISE	.002	.002	.002	.002	.002	.002	.002	.002	.002	.002	.003	.003	.003	.003	.003
Err CVH _{.2}	.003	.003	.003	.003	.003	.003	.003	.004	.004	.004	.004	.004	.004	.004	.004
Err CVH _{.1}	.003	.003	.003	.003	.003	.003	.003	.004	.004	.004	.004	.004	.004	.004	.004
Err CVH _{.05}	.003	.003	.003	.003	.003	.003	.003	.003	.004	.004	.004	.004	.004	.004	.004
Err CVL	.003	.003	.003	.003	.003	.003	.003	.003	.004	.004	.004	.004	.004	.004	.004
<i>n</i> = 2500															
ISE	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.002	.002	.002
Err CVH _{.2}	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
Err CVH _{.1}	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
Err CVH _{.05}	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
Err CVL	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003

Table 1 Continued.

Freq.	K														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
\hat{K}	0	0	.212	.019	.395	.028	.060	.079	.191	.014	.001	0	0	0	0
CVH ₂ +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₂ +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₂ -1	0	0	0	0	0	0	0	.014	.004	.008	0	0	0	0	0
CVH ₂ -2	0	0	0	0	0	0	0	0	0	0	.001	0	0	0	0
CVH ₁ +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₁ +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₁ -1	0	0	0	0	0	0	0	.014	.004	.007	0	0	0	0	0
CVH ₁ -2	0	0	0	0	0	0	0	0	0	0	.001	0	0	0	0
CVH ₀₅ +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₀₅ +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH ₀₅ -1	0	0	0	0	0	0	0	.014	.004	.007	0	0	0	0	0
CVH ₀₅ -2	0	0	0	0	0	0	0	0	0	0	.001	0	0	0	0
CVL +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -2	0	0	0	0	0	0	0	0	.003	0	.001	0	0	0	0
BIC +2	0	0	.001	0	0	0	0	0	0	0	0	0	0	0	0
BIC +1	0	0	.128	0	0	0	0	0	0	0	0	0	0	0	0
BIC -1	0	0	.083	.011	0	0	0	0	0	0	0	0	0	0	0
BIC -2	0	0	0	.008	.395	.028	.060	.079	.191	.014	.001	0	0	0	0

Table 2. Truncation Point Frequencies for the Trimodal, $n = 0400$.

Shown as Freq. \hat{K} is the frequency with which K was immediatly to the right of the last abrupt drop of ISE in the 1024 Monte Carlo repetitions described in the legend to Table 1. This is the value of K that achieves good EMM efficiency. Shown as Freq. CVL +2 is the frequency with which CVL overestimated the correct K by two or more; shown as CVL +1 is the frequency with which CVL overestimated the correct K by exactly one. Similarly for CVH and BIC. Only underestimates reduce EMM efficiency.

Freq.	K														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
\hat{K}	0	0	.118	.009	.477	.003	.068	.040	.263	.021	0	0	0	0	0
CVH. ₂ +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. ₂ +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. ₂ -1	0	0	0	0	0	0	.001	.011	.003	.017	0	0	0	0	0
CVH. ₂ -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. ₁ +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. ₁ +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. ₁ -1	0	0	0	0	0	0	.001	.011	.003	.017	0	0	0	0	0
CVH. ₁ -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. _{.05} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. _{.05} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH. _{.05} -1	0	0	0	0	0	0	.001	.010	.003	.016	0	0	0	0	0
CVH. _{.05} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIC +2	0	0	.009	0	0	0	0	0	0	0	0	0	0	0	0
BIC +1	0	0	.109	.004	0	0	0	0	0	0	0	0	0	0	0
BIC -1	0	0	0	.004	.477	0	0	0	0	0	0	0	0	0	0
BIC -2	0	0	0	.001	.001	.003	.068	.040	.263	.021	0	0	0	0	0

Table 3. Truncation Point Frequencies for the Trimodal, $n = 0900$. Shown as Freq. \hat{K} is the frequency with which K was immediately to the right of the last abrupt drop of ISE in the 1024 Monte Carlo repetitions described in the legend to Table 1. This is the value of K that achieves good EMM efficiency. Shown as Freq. CVL +2 is the frequency with which CVL overestimated the correct K by two or more; shown as CVL +1 is the frequency with which CVL overestimated the correct K by exactly one. Similarly for CVH and BIC. Only underestimates reduce EMM efficiency.

Freq.	K														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
\hat{K}	0	0	.065	.013	.477	.008	.070	.033	.313	.019	0	0	0	0	0
CVH _{.2} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.2} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.2} -1	0	0	0	0	0	0	0	.014	.002	.018	0	0	0	0	0
CVH _{.2} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} -1	0	0	0	0	0	0	0	.014	.002	.018	0	0	0	0	0
CVH _{.1} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} -1	0	0	0	0	0	0	0	.014	.002	.018	0	0	0	0	0
CVH _{.05} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIC +2	0	0	.008	0	0	0	0	0	0	0	0	0	0	0	0
BIC +1	0	0	.058	.006	0	0	0	0	0	0	0	0	0	0	0
BIC -1	0	0	0	.007	.477	0	0	0	0	0	0	0	0	0	0
BIC -2	0	0	0	0	0	.008	.070	.033	.313	.019	0	0	0	0	0

Table 4. Truncation Point Frequencies for the Trimodal, $n = 1600$. Shown as Freq. \hat{K} is the frequency with which K was immediately to the right of the last abrupt drop of ISE in the 1024 Monte Carlo repetitions described in the legend to Table 1. This is the value of K that achieves good EMM efficiency. Shown as Freq. CVL +2 is the frequency with which CVL overestimated the correct K by two or more; shown as CVL +1 is the frequency with which CVL overestimated the correct K by exactly one. Similarly for CVH and BIC. Only underestimates reduce EMM efficiency. A blurb goes here.

Freq.	K														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
\hat{K}	0	0	.037	.013	.440	.004	.075	.031	.385	.015	0	0	0	0	0
CVH _{.2} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.2} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.2} -1	0	0	0	0	0	0	0	.014	.004	.014	0	0	0	0	0
CVH _{.2} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.1} -1	0	0	0	0	0	0	0	.014	.004	.014	0	0	0	0	0
CVH _{.1} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVH _{.05} -1	0	0	0	0	0	0	0	.014	.004	.014	0	0	0	0	0
CVH _{.05} -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL +1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CVL -2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIC +2	0	0	.004	0	0	0	0	0	0	0	0	0	0	0	0
BIC +1	0	0	.033	.009	.001	0	0	0	0	0	0	0	0	0	0
BIC -1	0	0	0	.004	.439	.002	0	0	0	0	0	0	0	0	0
BIC -2	0	0	0	0	0	.002	.075	.031	.385	.015	0	0	0	0	0

Table 5. Truncation Point Frequencies for the Trimodal, $n = 2500$. Shown as Freq. \hat{K} is the frequency with which K was immediately to the right of the last abrupt drop of ISE in the 1024 Monte Carlo repetitions described in the legend to Table 1. This is the value of K that achieves good EMM efficiency. Shown as Freq. CVL +2 is the frequency with which CVL overestimated the correct K by two or more; shown as CVL +1 is the frequency with which CVL overestimated the correct K by exactly one. Similarly for CVH and BIC. Only underestimates reduce EMM efficiency. A blurb goes here.

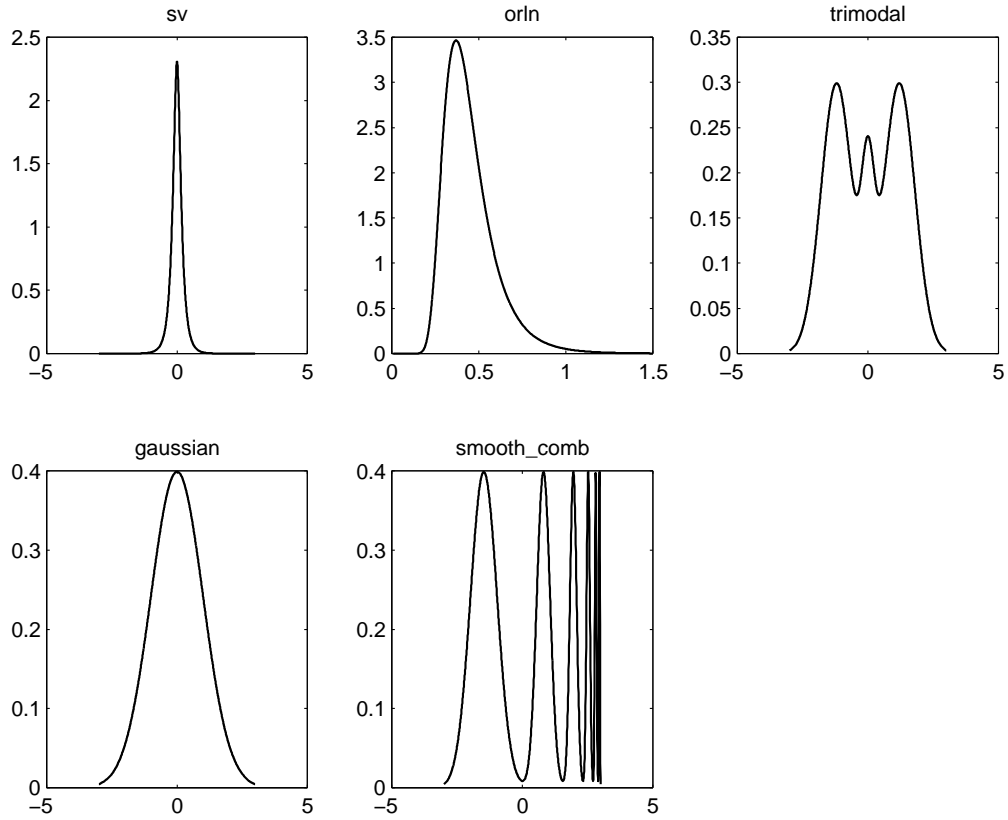


Figure 1. Densities considered. The plot labeled sv is the density of a scale mixture of normals with parameters chosen such that the density has mean 0, variance $1/4$, and raw kurtosis 8; orln is the density of the second largest order statistic in a sample of size 100 from the log normal with location parameter -3 and scale parameter 1. The densities trimodal, gaussian, and smooth_comb are densities from the Marron-Wand test suite.

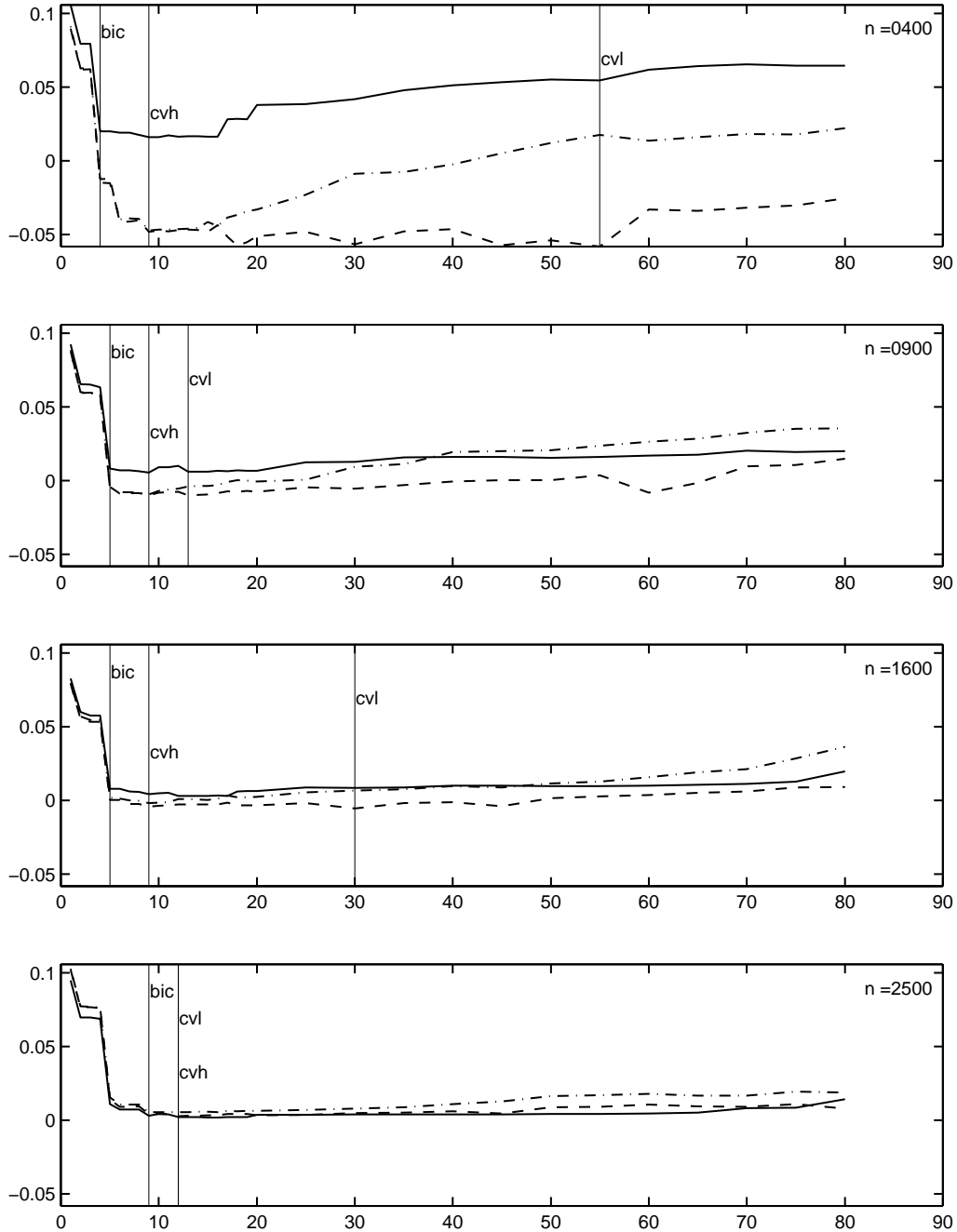


Figure 2. Scale Mixture of Normals. Plotted is the integrated squared error (ISE) and its cross validated estimate (CV) for a realization of size n , as shown in each plot, from the density $p(y|\rho) = \int_{-\infty}^{\infty} n(y|\rho_1, e^{2u}) n(u|\rho_2, \rho_3^2) du$ with ρ chosen so that the density has mean 0, variance $1/4$, and raw kurtosis 8. Solid line is ISE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K , as marked.

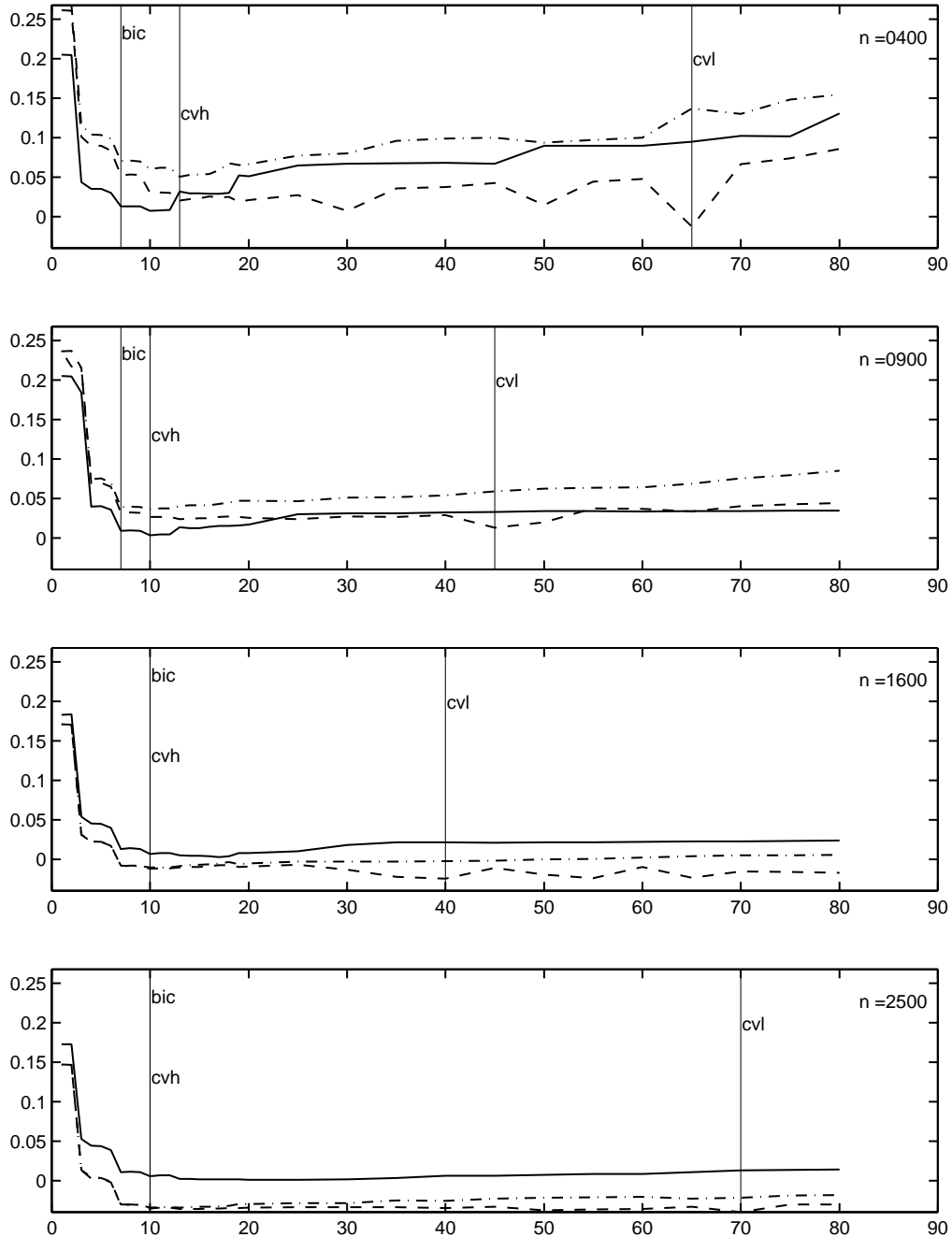


Figure 3. Second Largest Order Statistic of the Lognormal. Plotted is the integrated squared error (ISE) and its cross validated estimate (CV) for a realization of size n , as shown in each plot, from the density $p(y|\rho) = \frac{N(N-1)}{y} \left[\Phi\left(\frac{\log y - \rho_2}{\rho_3}\right) \right]^{N-2} \left[1 - \Phi\left(\frac{\log y - \rho_2}{\rho_3}\right) \right] \phi\left(\frac{\log y - \rho_2}{\rho_3}\right)$ where $y > 0$, ϕ and Φ denote the standard normal density and distribution functions, respectively, and $(N, \rho_2, \rho_3) = (100, -3, 1)$. Solid line is ISE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K , as marked.

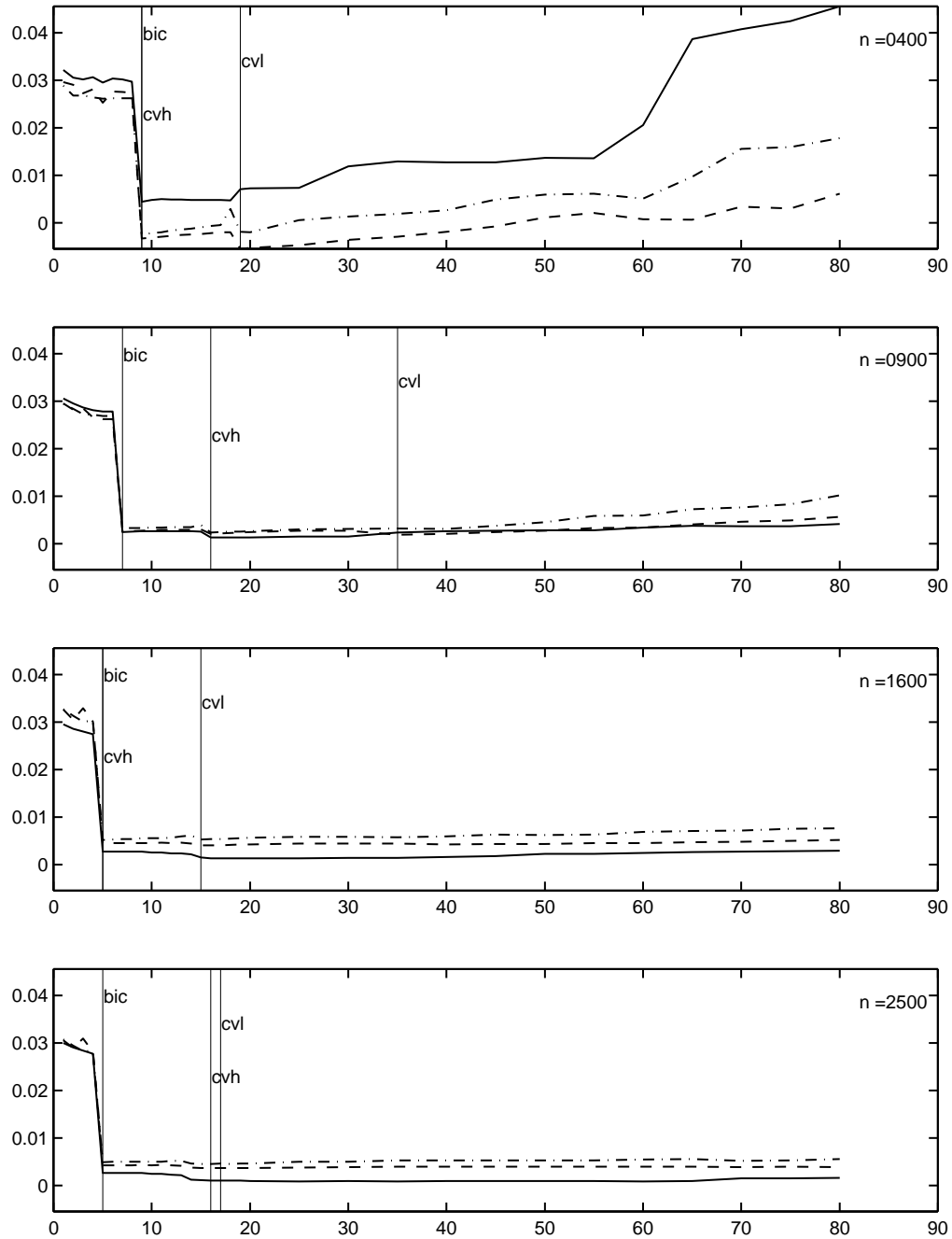


Figure 4. Trimodal. Plotted is the integrated squared error (ISE) and its cross validated estimate (CV) for a realization of size n , as shown in each plot, from the trimodal density of the Marron-Wand test suite. Solid line is ISE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K , as marked.

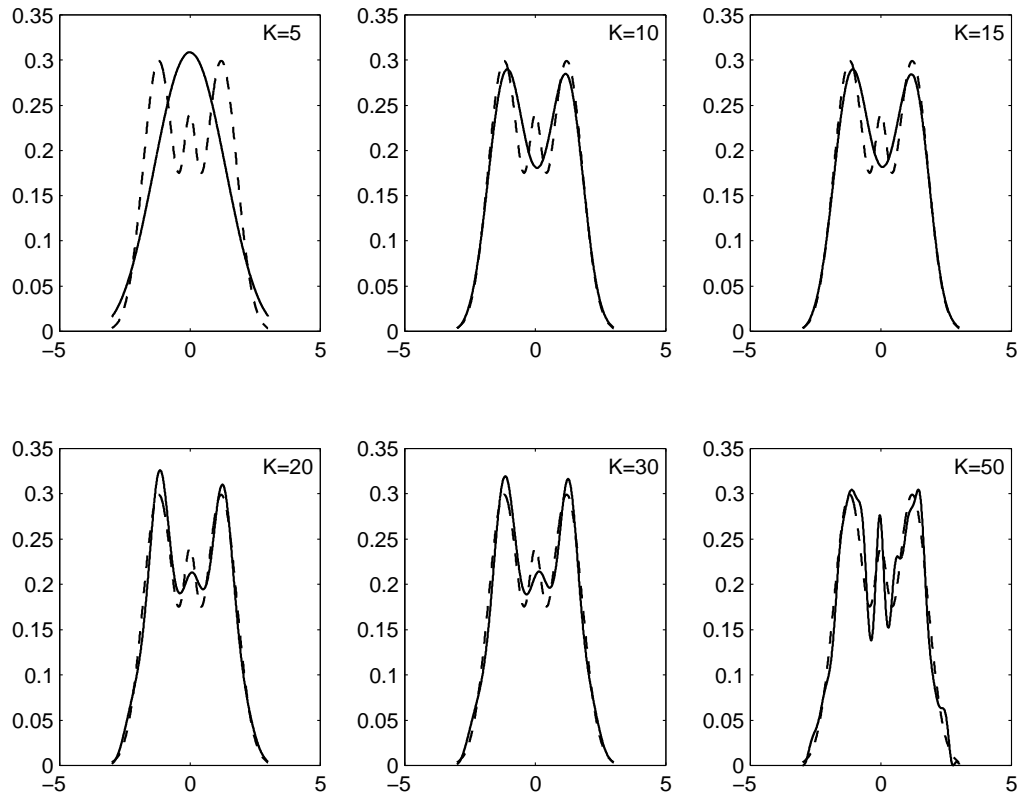


Figure 5. Trimodal. Plotted are SNP density estimates from a realization of size 900 and values of K as shown in each plot, from the trimodal density of the Marron-Wand test suite. Solid line is the estimate, dashed line is true density.

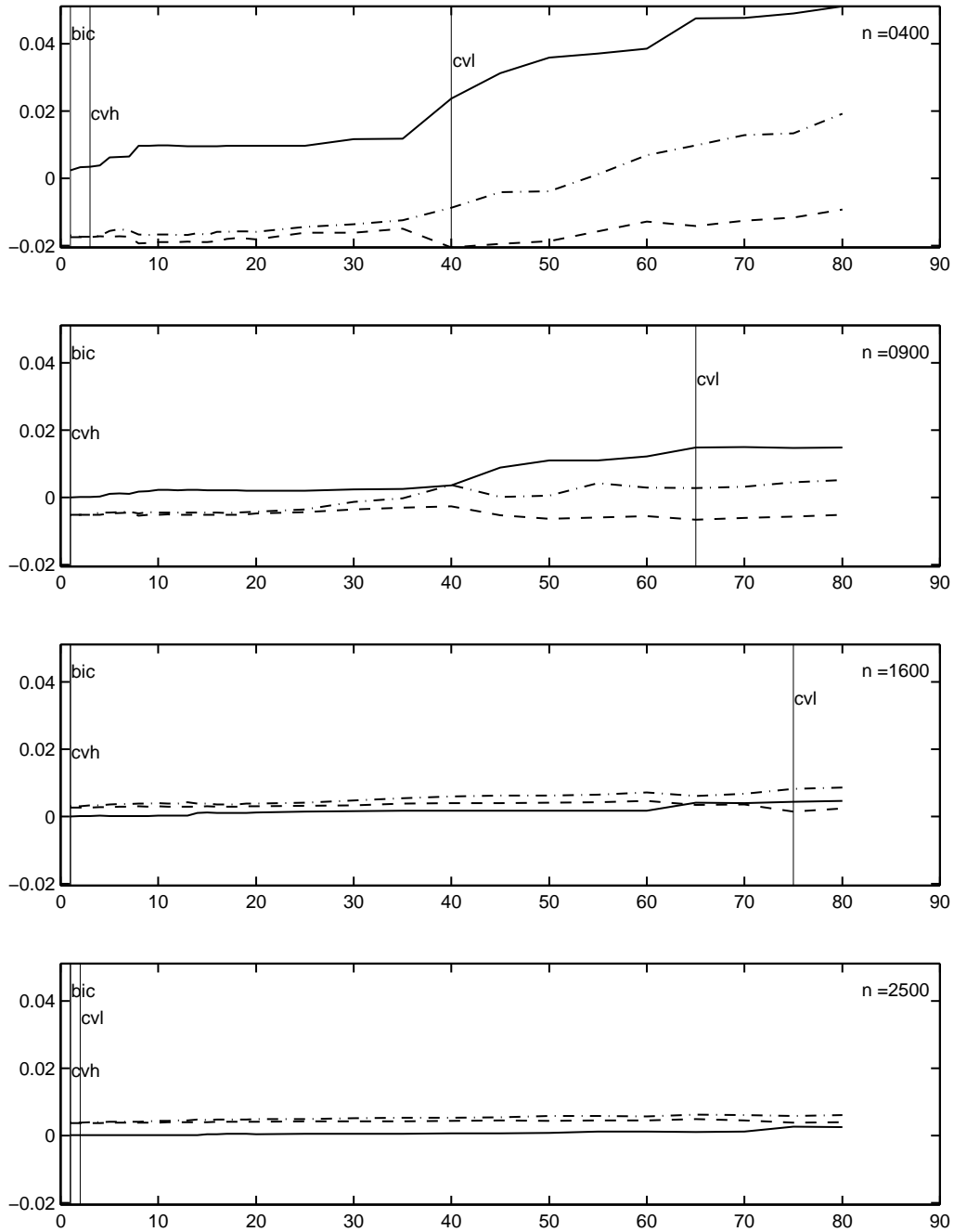


Figure 6. Gaussian. Plotted is the integrated squared error (ISE) and its cross validated estimate (CV) for a realization of size n , as shown in each plot, from the gaussian density of the Marron-Wand test suite. Solid line is ISE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K , as marked.

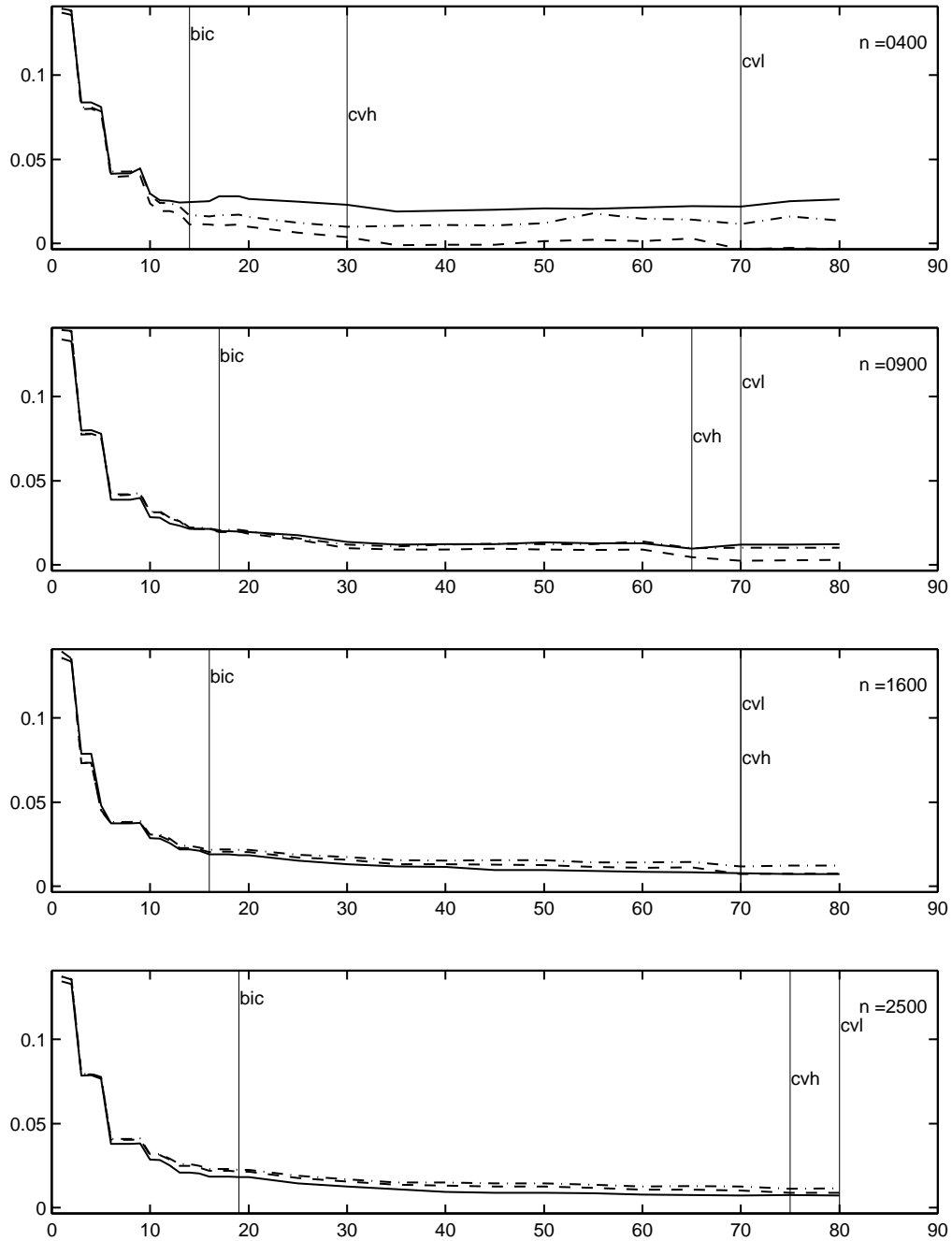


Figure 7. Smooth Comb. Plotted is the integrated squared error (ISE) and its cross validated estimate (CV) for a realization of size n , as shown in each plot, from the smooth comb density of the Marron-Wand test suite. Solid line is ISE, dashed line is its leave-one-out CV estimate (CVL), and dashed and dotted line is the average of ten, 10% hold-out-sample CV estimates (CVH). Vertical lines indicate BIC, CVL, and CVH choices of K , as marked.