



SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY



CREATES
Center for Research in Econometric
Analysis of Time Series

CREATES Research Paper 2009-46

Robust Data-Driven Inference for Density-Weighted Average Derivatives

Matias D. Cattaneo, Richard K. Crump and Michael Jansson

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Robust Data-Driven Inference for Density-Weighted Average Derivatives*

MATIAS D. CATTANEO

DEPARTMENT OF ECONOMICS, UNIVERSITY OF MICHIGAN

RICHARD K. CRUMP

FEDERAL RESERVE BANK OF NEW YORK

MICHAEL JANSSON

DEPARTMENT OF ECONOMICS, UC BERKELEY AND *CREATES*

September 28, 2009

ABSTRACT. This paper presents a new data-driven bandwidth selector compatible with the small bandwidth asymptotics developed in Cattaneo, Crump, and Jansson (2009) for density-weighted average derivatives. The new bandwidth selector is of the plug-in variety, and is obtained based on a mean squared error expansion of the estimator of interest. An extensive Monte Carlo experiment shows a remarkable improvement in performance when the bandwidth-dependent robust inference procedure proposed by Cattaneo, Crump, and Jansson (2009) is coupled with this new data-driven bandwidth selector. The resulting robust data-driven confidence intervals compare favorably to the alternative procedures available in the literature.

Keywords: Average derivatives, Bandwidth selection, Robust inference, Small bandwidth asymptotics.

JEL: C12, C14, C21, C24.

*The authors thank Sebastian Calonico and Lutz Kilian, and seminar participants at Michigan for comments. The first author gratefully acknowledges financial support from the National Science Foundation (SES 0921505). The third author gratefully acknowledges financial support from the National Science Foundation (SES 0920953) and the research support of CREATES (funded by the Danish National Research Foundation).

1. INTRODUCTION

Semiparametric models, which include both a finite dimensional parameter of interest and an infinite dimensional nuisance parameter, play a central role in modern statistical and econometric theory, and are potentially of great interest in empirical work. However, the applicability of semiparametric estimators is seriously hampered by the sensitivity of their performance to seemingly ad hoc choices of “smoothing” and “tuning” parameters involved in the estimation procedure. Although classical large sample theory for semiparametric estimators is now well developed, these theoretical results are typically invariant to the particular choice of parameters associated with the nonparametric estimator employed, and usually require strong untestable assumptions (e.g., smoothness of the infinite dimensional nuisance parameter). As a consequence, inference procedures based on these estimators are in general not robust to changes in the choice of tuning and smoothing parameters underlying the nonparametric estimator, and to departures from key unobservable model assumptions. Thus, classical asymptotic results for semiparametric estimators may not always accurately capture their behavior in finite samples, posing considerable restrictions on the overall applicability they may have for empirical work.

This paper proposes a robust data-driven inference procedure for the density-weighted average derivative estimator, an important semiparametric estimator commonly used in empirical work. The main idea is to develop a new data-driven bandwidth selector compatible with the small bandwidth asymptotic theory presented in Cattaneo, Crump, and Jansson (2009). This alternative (first-order) large sample theory encompasses the classical large sample theory available in the literature, and also enjoys several robustness properties. In particular, (i) it provides a valid inference procedure for (small) bandwidth sequences that would render the classical results invalid, (ii) it permits the use of a second-order kernel regardless of the dimension of the regressors and therefore removes strong smoothness assumptions, and (iii) it provides a limiting distribution that is not invariant to the particular choices of smoothing and tuning parameters, without necessarily forcing a slower than root- n rate of convergence (where n is the sample size). The key theoretical insight behind these results is to accommodate bandwidth sequences that break down the asymptotic linearity of the estimator of interest, leading to a more general first-order asymptotic theory that is no longer invariant to the particular choices of parameters underlying the preliminary nonparametric estimator. Consequently, it is expected that a new inference procedure based on this alternative asymptotic theory would (at least partially) “adapt” to the particular choices of these parameters.

The preliminary simulation results in Cattaneo, Crump, and Jansson (2009) show that this alternative asymptotic theory opens the possibility for the construction of a robust inference procedure, providing a range of (small) bandwidths for which the appropriate test statistic enjoys approximately correct size. However, the bandwidth selectors available in the literature turn out to be incompatible with these new results in the sense that they would not deliver a bandwidth choice within the robust range. This paper presents a new data-driven bandwidth selector that

achieves this goal, thereby providing a robust automatic (i.e., fully data-driven) inference procedure for the estimand of interest. These results are corroborated by an extensive Monte Carlo experiment, which shows that the asymptotic theory developed in Cattaneo, Crump, and Jansson (2009) coupled with the data-driven bandwidth selector proposed here lead to remarkable improvements in inference when compared to the alternative procedures available in the literature. In particular, the resulting new data-driven confidence intervals exhibit close-to-correct empirical coverage across all designs considered. Among other advantages, these new data-driven statistical procedures allow for the use of a second-order kernel, which is believed to deliver more stable results in applications (see, e.g., Horowitz and Härdle (1996)), and appear to be considerably more robust to the impact associated with the additional variability introduced by the estimation of the bandwidth selectors.

This paper contributes to the important literature of semiparametric inference for weighted average derivatives. This population parameter of interest was originally introduced by Stoker (1986), and has been intensely studied in the literature since then. Härdle and Stoker (1989) and Härdle, Hart, Marron, and Tsybakov (1992) study general weighted average derivative estimators, although their results are considerably complicated by the fact that their representation requires handling stochastic denominators and appears to be very sensitive to the choice of trimming parameters. Fortunately, the density-weighted average derivative estimator circumvents this problem, while retaining the desirable properties of the general weighted average derivative, and leads to a simple and useful semiparametric estimator. Powell, Stock, and Stoker (1989) study the first-order large sample properties of this estimator and provide sufficient (but not necessary) conditions for root- n consistency and asymptotic normality. Nishiyama and Robinson (2000, 2001, 2005) study its second-order large sample properties by deriving valid Edgeworth expansions for this estimator (see also Robinson (1995)), while Härdle and Tsybakov (1993) and Powell and Stoker (1996) provide second-order mean squared error expansions for this estimator (see also Newey, Hsieh, and Robins (2004)). Both types of higher-order expansions provide simple plug-in bandwidth selectors targeting different properties of this estimator, and are compatible with the classical large sample theory available in the literature. Ichimura and Todd (2007) provide a recent survey, with particular emphasis on implementation, of the results available in the literature. For an interesting empirical example focusing on density-weighted average derivatives, see Deaton and Ng (1998).

The rest of the paper is organized as follows. Section 2 describes the model and reviews the main results available in the literature regarding first-order large sample inference for density-weighted average derivatives. Section 3 presents the higher-order mean squared error expansion and develops the new (infeasible) theoretical bandwidth selector, while Section 4 describes how to construct a feasible (i.e., data-driven) bandwidth selector and establishes its consistency. Section 5 summarizes the results of an extensive Monte Carlo experiment. Section 6 concludes.

2. MODEL AND PREVIOUS RESULTS

Let $z_i = (y_i, x_i')'$, $i = 1, \dots, n$, be a random sample from a vector $z = (y, x')'$, where $y \in \mathbb{R}$ is a dependent variable and $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ is a continuous explanatory variable with a density $f(\cdot)$. The population parameter of interest is the density-weighted average derivative given by

$$\theta = \mathbb{E} \left[f(x) \frac{\partial}{\partial x} g(x) \right],$$

where $g(x) = \mathbb{E}[y|x]$ denotes the population regression function. The following assumption collects typical regularity conditions imposed on this model.

Assumption 1. (a) $\mathbb{E}[y^4] < \infty$.

(b) $\mathbb{E}[\sigma^2(x) f(x)] > 0$ and $\mathbb{V}[\partial e(x)/\partial x - y \partial f(x)/\partial x]$ is positive definite, where $\sigma^2(x) = \mathbb{V}[y|x]$ and $e(x) = f(x) g(x)$.

(c) f is $(Q+1)$ times differentiable, and f and its first $(Q+1)$ derivatives are bounded, for some $Q \geq 2$.

(d) g is twice differentiable, and e and its first two derivatives are bounded.

(e) v is differentiable and

$$\sup_{x \in \mathbb{R}^d} [v(x) f(x) + v(x) \|\partial f(x)/\partial x\| + \|\partial v(x)/\partial x\|] < \infty,$$

where $\|\cdot\|$ is the Euclidean norm and $v(x) = \mathbb{E}[y^2|x]$.

(f) $\lim_{\|x\| \rightarrow \infty} [f(x) + |e(x)|] = 0$.

Assumption 1 and integration by parts lead to $\theta = -2\mathbb{E}[y \partial f(x)/\partial x]$, which in turn motivates the analogue estimator of Powell, Stock, and Stoker (1989) given by

$$\hat{\theta}_n = -2 \frac{1}{n} \sum_{i=1}^n y_i \frac{\partial}{\partial x} \hat{f}_{n,i}(x_i),$$

where $\hat{f}_{n,i}(\cdot)$ is a “leave-one-out” kernel density estimator defined as

$$\hat{f}_{n,i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n \frac{1}{h_n^d} K\left(\frac{x_j - x}{h_n}\right),$$

for some kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ and some positive (bandwidth) sequence h_n . Typical regularity conditions on the kernel-based nonparametric estimator entering this semiparametric estimator $\hat{\theta}_n$ are imposed in the following assumption.

Assumption 2. (a) K is even.

(b) K is differentiable, and K and its first derivative are bounded.

- (c) $\int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' du$ is positive definite, where $\dot{K}(u) = \partial K(u) / \partial u$.
 (d) For some $P \geq 2$,

$$\int_{\mathbb{R}^d} |K(u)| (1 + \|u\|^P) du + \int_{\mathbb{R}^d} \|\dot{K}(u)\| (1 + \|u\|^2) du < \infty,$$

and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = \begin{cases} 1, & \text{if } l_1 + \cdots + l_d = 0, \\ 0, & \text{if } 0 < l_1 + \cdots + l_d < P \end{cases}.$$

Powell, Stock, and Stoker (1989) showed that, under appropriate restrictions on the bandwidth sequence and kernel function, the estimator $\hat{\theta}_n$ is asymptotically linear with influence function given by $L(z) = 2[\partial e(x)/\partial x - y \partial f(x)/\partial x - \theta]$. Thus, the asymptotic variance of this estimator is given by $\Sigma = \mathbb{E}[L(z)L(z)']$. The following result describes the exact conditions and summarizes the main conclusion. (Limits are taken as $n \rightarrow \infty$ unless otherwise noted.)

Result 1. (Powell, Stock, and Stoker (1989)) *If Assumptions 1 and 2 hold, and if $nh_n^{2\min(P,Q)} \rightarrow 0$ and $nh_n^{d+2} \rightarrow \infty$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n L(z_i) + o_p(1) \rightarrow_d \mathcal{N}(0, \Sigma).$$

Result 1 follows from noting that the estimator $\hat{\theta}_n$ admits a n -varying U -statistic representation given by

$$\hat{\theta}_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n U(z_i, z_j; h_n), \quad U(z_i, z_j; h) = -h^{-(d+1)} \dot{K}\left(\frac{x_i - x_j}{h}\right) (y_i - y_j),$$

which leads to the Hoeffding decomposition $\hat{\theta}_n = \theta_n + \bar{L}_n + \bar{W}_n$, where

$$\theta_n = \theta(h_n), \quad \bar{L}_n = \frac{1}{n} \sum_{i=1}^n L(z_i; h_n), \quad \bar{W}_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n W(z_i, z_j; h_n),$$

with

$$\theta(h) = \mathbb{E}[U(z_i, z_j; h)], \quad L(z_i; h) = 2[\mathbb{E}(U(z_i, z_j; h) | z_i) - \theta(h)],$$

$$W(z_i, z_j; h) = U(z_i, z_j; h) - [L(z_i; h) + L(z_j; h)]/2 - \theta(h).$$

This decomposition makes clear the need for the conditions on the bandwidth sequence and the kernel function: (i) condition $nh_n^{2\min(P,Q)} \rightarrow 0$ ensures that the bias of the estimator is asymptotically negligible since $\theta_n - \theta = O(h_n^{\min(P,Q)})$, and (ii) condition $nh_n^{d+2} \rightarrow \infty$ ensures that the “quadratic term” of the Hoeffding decomposition is also asymptotically negligible since

$\bar{W}_n = O_p(n^{-1}h_n^{-(d+2)/2})$. Under the same conditions, Powell, Stock, and Stoker (1989) also develop a simple consistent estimator for Σ , which is given by the analogue estimator

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \hat{L}_{n,i} \hat{L}'_{n,i}, \quad \hat{L}_{n,i} = 2 \left[\frac{1}{n-1} \sum_{j=1, j \neq i}^n U(z_i, z_j; h_n) - \hat{\theta}_n \right].$$

Consequently, under the conditions imposed in Result 1, it is straightforward to form a studentized version of $\hat{\theta}_n$, leading to a simple, asymptotically pivotal test statistic for the testing problem:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0, \quad (1)$$

which is based on $\sqrt{n} \hat{\Sigma}_n^{-1/2} (\hat{\theta}_n - \theta) \rightarrow_d \mathcal{N}(0, I_d)$, with $\hat{\Sigma}_n \rightarrow_p \Sigma$.

As discussed in Newey (1994), asymptotic linearity of a semiparametric estimator has several distinct features that may be considered attractive from a theoretical point of view. In particular, asymptotic linearity is a necessary condition for semiparametric efficiency and leads to a limiting distribution of the statistic of interest that is invariant to the choice of the nonparametric estimator used in the construction of the semiparametric procedure. In other words, regardless of the particular choice of preliminary nonparametric estimator used, the limiting distribution will not depend on the specific nonparametric estimator whenever the semiparametric estimator admits an asymptotic linear representation.

However, achieving an asymptotic linear representation of a semiparametric estimator imposes several strong model assumptions and leads to a large sample theory than may not accurately represent the finite sample behavior of the estimator. In the case of $\hat{\theta}_n$, asymptotic linearity would require $P > 2$ unless $d = 1$, which in turn requires strong smoothness conditions ($Q \geq P$). Consequently, classical asymptotic theory will require the use of a higher-order kernel whenever more than one covariate is included. In addition, classical asymptotic theory (whenever valid) leads to a limiting experiment which is invariant to the particular choices of smoothing (K) and tuning (h_n) parameters involved in the construction of the estimator, and therefore it is unlikely to be able to “adapt” to changes in these parameters. In other words, inference based on classical asymptotic theory is silent with respect to the impact that these parameters may have on the finite sample behavior of $\hat{\theta}_n$.

In an attempt to better characterize the finite sample behavior of $\hat{\theta}_n$, Cattaneo, Crump, and Jansson (2009) show that it is possible to increase the robustness of this estimator by considering a different asymptotic experiment. In particular, instead of forcing asymptotic linearity of the estimator, the authors develop an alternative first-order asymptotic theory that accommodates (but does not require) weaker assumptions than those imposed in the classical first-order asymptotic theory discussed above. The following result collects the main findings.

Result 2. (Cattaneo, Crump, and Jansson (2009)) *If Assumptions 1 and 2 hold, and if $\min(nh_n^{d+2}, 1)nh_n^{2\min(P,Q)} \rightarrow 0$ and $n^2h_n^d \rightarrow \infty$, then*

$$(\mathbb{V}[\hat{\theta}_n])^{-1/2}(\hat{\theta}_n - \theta) \rightarrow_d \mathcal{N}(0, I_d),$$

where

$$\mathbb{V}[\hat{\theta}_n] = \frac{1}{n} [\Sigma + o(1)] + \binom{n}{2}^{-1} h_n^{-(d+2)} [\Delta + o(1)],$$

with $\Delta = 2\mathbb{E}[\sigma^2(x)f(x)] \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' du$. In addition,

$$\frac{1}{n} \hat{\Sigma}_n = \frac{1}{n} \Sigma + 2 \binom{n}{2}^{-1} h_n^{-(d+2)} \Delta + o_p(n^{-1} + n^{-2} h_n^{-(d+2)}).$$

Result 2 shows that the conditions on the bandwidth sequence may be considerably weakened without invalidating the limiting Gaussian distribution. In particular, whenever h_n is chosen so that nh_n^{d+2} is bounded, the limiting distribution will cease to be invariant with respect to the underlying preliminary nonparametric estimator because $\hat{\theta}_n$ is no longer asymptotically linear. (In particular, note that $nh_n^{d+2} \rightarrow \kappa > 0$ retains the root- n consistency of $\hat{\theta}_n$.) In addition, because h_n is allowed to be “smaller” than usual, the bias of the estimator is controlled in a different way, removing the need for higher-order kernels.

Result 2 also shows that the feasible classical testing procedure based on $\sqrt{n} \hat{\Sigma}_n^{-1/2}(\hat{\theta}_n - \theta)$ will be invalid unless $nh_n^{d+2} \rightarrow \infty$, which corresponds to the classical large sample theory case (Result 1). To solve this problem, Cattaneo, Crump, and Jansson (2009) propose two alternative corrections to the standard error matrix $\hat{\Sigma}_n$, leading to two options for “robust” standard errors. To construct the first “robust” standard error formula, the authors introduce a simple consistent estimator for Δ , under the same conditions of Result 2, which is given by the analogue estimator

$$\hat{\Delta}_n = h_n^{d+2} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{W}_{n,ij} \hat{W}_{n,ij}', \quad \hat{W}_{n,ij} = U(z_i, z_j; h_n) - \frac{1}{2} (\hat{L}_{n,i} + \hat{L}_{n,j}) - \hat{\theta}_n.$$

Thus, using this estimator,

$$\hat{V}_{1,n} = \frac{1}{n} \hat{\Sigma}_n - \binom{n}{2}^{-1} h_n^{-(d+2)} \hat{\Delta}_n \tag{2}$$

yields a consistent standard error estimate under small bandwidth asymptotics (i.e., under the weaker conditions imposed in Result 2, which include in particular those imposed in Result 1). To describe the second “robust” standard error formula, let $\hat{\Sigma}_n(H_n)$ be the estimator $\hat{\Sigma}_n$ constructed using a bandwidth sequence H_n (e.g., $\hat{\Sigma}_n = \hat{\Sigma}_n(h_n)$ by definition). Then, under the same conditions of Result 2,

$$\hat{V}_{2,n} = \frac{1}{n} \hat{\Sigma}_n(2^{1/(2+d)} h_n) \tag{3}$$

yields also a consistent standard error estimate under small bandwidth asymptotics.

Consequently, under the conditions imposed in Result 2, it is straightforward to form a studentized version of $\hat{\theta}_n$, leading to two simple, robust test statistics for the testing problem (1), which are based on $\hat{V}_{k,n}^{-1/2}(\hat{\theta}_n - \theta) \rightarrow_d \mathcal{N}(0, I_d)$, with $\hat{V}_{k,n}^{-1}\mathbb{V}[\hat{\theta}_n] \rightarrow_p I_d$, $k = 1, 2$.

Although an interesting theoretical improvement, these results have the obvious drawback of depending on the choice of h_n , which is unrestricted beyond the rate restrictions imposed in Result 2. A preliminary Monte Carlo experiment reported in Cattaneo, Crump, and Jansson (2009) shows that the new, robust standard error formulas have the potential to deliver good finite sample behavior if the initial h_n is chosen to be small enough. As suggested above, this empirical finding may be (partially) justified by the fact that for those “small” bandwidths asymptotic linearity ceases to hold and therefore the limiting distribution is no longer invariant to the choice of the smoothing and tuning parameters.

As mentioned in the introduction, the plug-in rules available in the literature for h_n fail to deliver a choice of h_n that would enjoy the robustness property introduced by the new asymptotic theory described in Result 2. This is not too surprising, since these bandwidth selectors are typically constructed to balance (higher-order) bias and variance in a way that is “appropriate” for the classical large sample theory.

3. MSE EXPANSION AND “OPTIMAL” BANDWIDTH SELECTORS

Higher-order expansions provide a simple and intuitive way of constructing plug-in bandwidth selectors for semiparametric estimators. For the case of the density-weighted average derivative there exist three bandwidth selectors of the plug-in variety. Härdle and Tsybakov (1993) and Powell and Stoker (1996) construct a bandwidth selector based on the minimization of the mean squared error of $\hat{\theta}_n$, while Nishiyama and Robinson (2000, 2005) construct two plug-in bandwidth selectors based on an Edgeworth expansion for the one-sided and two-sided corresponding test statistics. See Ichimura and Todd (2007, Section 6.3) for a general discussion on these results and their implementation.

This paper also considers the mean squared error expansion of $\hat{\theta}_n$ as the starting point for the construction of the plug-in “optimal” bandwidth selector. In order to compute such an expansion it is necessary to strengthen the assumptions concerning the data generating process. The following assumption contains a set of additional mild conditions sufficient to provide a valid higher-order mean squared error expansion of $\hat{\theta}_n$, up to the order needed for this paper.

Assumption 3. (a) g is $(Q + 1)$ times differentiable, and e and its first $(Q + 1)$ derivatives are bounded.

(b) v is three times differentiable, and vf and its first three derivatives are bounded.

(c) $\lim_{\|x\| \rightarrow \infty} [\sigma(x)f(x) + \|\partial\sigma(x)/\partial x\|f(x)] = 0$.

$$(d) \mathbb{E}[\|\partial g(x)/\partial x\|^2 f(x)] < \infty.$$

Assumptions 3(a) and 3(b) are natural and in agreement with those imposed in Powell and Stoker (1996) and Nishiyama and Robinson (2000, 2005), while Assumption 3(c) is slightly stronger than the analogue restriction imposed in those papers. Assumption 3(d) is used to ensure that the higher-order mean squared expansion is valid up to the order needed.

Theorem 1. *If Assumptions 1, 2 and 3 hold, then for $s = \min(P, Q)$ and $\dot{f}(x) = \partial f(x)/\partial x$,*

$$\begin{aligned} \mathbb{E}[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)'] &= \frac{1}{n}\Sigma + \binom{n}{2}^{-1} h_n^{-(d+2)}\Delta + \binom{n}{2}^{-1} h_n^{-d}\mathcal{V} + h_n^{2s}\mathcal{B}\mathcal{B}' \\ &\quad + O(n^{-1}h_n^s) + o(n^{-2}h^{-d} + h_n^{2s}), \end{aligned}$$

where

$$\mathcal{B} = -\frac{2(-1)^s}{s!} \sum_{\substack{0 \leq l_1, \dots, l_d \leq s \\ l_1 + \dots + l_d = s}} \left[\int_{\mathbb{R}^d} u_1^{l_1} \dots u_d^{l_d} K(u) du \right] \mathbb{E} \left[\left(\frac{\partial^{(l_1 + \dots + l_d)}}{\partial x_1^{l_1} \dots \partial x_d^{l_d}} \dot{f}(x) \right) g(x) \right],$$

and

$$\mathcal{V} = \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' \left(u' \mathbb{E} \left[\sigma^2(x) \frac{\partial^2}{\partial x \partial x'} f(x) + \left(\frac{\partial}{\partial x} g(x) \right) \left(\frac{\partial}{\partial x} g(x) \right)' f(x) \right] u \right) du.$$

The result in Theorem 1 is similar to the one obtained by Härdle and Tsybakov (1993) and Powell and Stoker (1996), the key difference being that the additional term of order $O(n^{-2}h_n^{-d})$ is explicitly retained here. (Recall that Result 2 requires $n^2 h_n^d \rightarrow \infty$.)

To motivate the new “optimal” bandwidth selector, recall that the “robust” variance matrix in Result 2 is given by the first two terms of the mean squared error expansion presented in Theorem 1, which suggests considering the next two terms of the expansion to construct an “optimal” bandwidth selector. (Note that, as it is common in the literature, this approach implicitly assumes that both \mathcal{B} and \mathcal{V} are not equal to zero.) Intuitively, balancing these terms corresponds to the case of $nh_n^{d+2} \rightarrow \kappa < \infty$, and therefore pushes the selected bandwidth to the “small bandwidth” region. This approach may be considered “optimal” in a mean square error sense because it makes the leading terms ignored in the general large sample approximation presented in Result 2 as small as possible.

To describe the new bandwidth selector, let $\lambda \in \mathbb{R}^d$ and consider (for simplicity) a bandwidth that minimizes the next two terms of $\mathbb{E}[(\lambda'(\hat{\theta}_n - \theta))^2]$. This “optimal” bandwidth selector is given

by

$$h_{CCJ}^* = \begin{cases} \left(\frac{d(\lambda' \mathcal{V} \lambda)}{s(\lambda' \mathcal{B})^2} \right)^{\frac{1}{2s+d}} n^{-\frac{2}{2s+d}} & \text{if } \lambda' \mathcal{V} \lambda > 0 \\ \left(\frac{2|\lambda' \mathcal{V} \lambda|}{(\lambda' \mathcal{B})^2} \right)^{\frac{1}{2s+d}} n^{-\frac{2}{2s+d}} & \text{if } \lambda' \mathcal{V} \lambda < 0 \end{cases}.$$

This new theoretical bandwidth selector is consistent with the “small bandwidth” asymptotics described in Result 2, since $n^2 (h_{CCJ}^*)^d \rightarrow \infty$. In addition, observe that $n^{-1} h_n^s = o(n^{-2} h_n^{-d})$ whenever $n h_n^{s+d} \rightarrow 0$, which is satisfied when $h_n = h_{CCJ}^*$.

This new bandwidth selector may be compared to the two competing plug-in bandwidth selectors available in the literature, proposed by Powell and Stoker (1996) and Nishiyama and Robinson (2005), and given by

$$h_{PS}^* = \left(\frac{(d+2)(\lambda' \Delta \lambda)}{s(\lambda' \mathcal{B})^2} \right)^{\frac{1}{2s+d+2}} n^{-\frac{2}{2s+d+2}}, \quad \text{and} \quad h_{NR}^* = \left(\frac{2(\lambda' \Delta \lambda)}{(\lambda' \mathcal{B})^2} \right)^{\frac{1}{2s+d+2}} n^{-\frac{2}{2s+d+2}},$$

respectively. Inspection of these bandwidth selectors shows that $h_{CCJ}^* \prec h_{PS}^* \asymp h_{NR}^*$, leading to a bandwidth selection of smaller order.¹

4. DATA-DRIVEN BANDWIDTH SELECTORS

The previous section described a new (infeasible) plug-in bandwidth selector that is compatible with the small bandwidth asymptotic theory proposed by Cattaneo, Crump, and Jansson (2009). In order to implement this selector in practice, as well as its competitors h_{PS}^* and h_{NR}^* , it is necessary to construct consistent estimates for each of the leading constants. These estimates would lead to a data-driven (i.e., automatic) bandwidth selector, denoted \hat{h}_{CCJ} .

A straightforward, somewhat unsatisfactory way of constructing estimates for the leading constants is to provide a “rule-of-thumb” estimator, which is typically motivated by assuming a parametric distribution of the underlying model. However, it is well-known that this kind of rule-of-thumb bandwidth selectors tend to underperform whenever the underlying distributional assumptions are invalid. As an alternative, it is possible to construct a plug-in bandwidth selector, which nonparametrically estimates each quantity \mathcal{B} , Δ and \mathcal{V} using a preliminary bandwidth choice.

To describe the data-driven plug-in bandwidth selectors, let b_n be a preliminary positive bandwidth sequence, which may be different for each estimator. A simple analogue estimator of Δ was introduced in Section 2. In particular, let $\hat{\Delta}_n(b_n)$ be the estimator $\hat{\Delta}_n$ constructed using a bandwidth sequence b_n (e.g., $\hat{\Delta}_n = \hat{\Delta}_n(b_n)$ by definition). Note that this estimator is a n -varying U -statistic as well. Theorem 1 and the calculations provided in Cattaneo, Crump, and Jansson

¹Nishiyama and Robinson (2000) derives a third alternative bandwidth selector which is not explicitly discussed here because this procedure is targeted for one-sided hypothesis testing. Nonetheless, inspection of this alternative bandwidth selection procedure, denoted h_{NR00}^* , shows that $h_{CCJ}^* \prec h_{NR00}^*$ whenever $d+8 > 2s$. Therefore, h_{CCJ}^* is of smaller order unless strong smoothness assumption are imposed in the model and a corresponding higher-order kernel is employed.

(2009) show that, if Assumptions 1, 2 and 3 hold, then

$$\hat{\Delta}_n(b_n) = \Delta + b_n^2 \mathcal{V} + O_p(b_n^3 + n^{-1/2} + n^{-1}b_n^{-d/2}),$$

which gives the consistency of this estimator if $b_n \rightarrow 0$ and $n^2 b_n^d \rightarrow \infty$.

Next, consider the construction of consistent estimators of \mathcal{B} and \mathcal{V} , the two parameters entering the new bandwidth selector h_{CCJ}^* . To this end, let k be a kernel function, which may be different for each estimator, and may be different from K . The following assumption collects a set of sufficient conditions to establish consistency of the plug-in estimators proposed in this paper for \mathcal{B} and \mathcal{V} .

Assumption 4. (a) f , v and e are $(s+1+S)$ times differentiable, and f , vf , e and their first $(s+1+S)$ derivatives are bounded, for some $S \geq 1$.
 (b) k is even.
 (c) k is M times differentiable, and k and its first M derivatives are bounded, for some $M \geq 0$.
 (d) For some $R \geq 2$, $\int_{\mathbb{R}^d} |k(u)| (1 + \|u\|^R) du < \infty$, and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} k(u) du = \begin{cases} 1, & \text{if } l_1 + \cdots + l_d = 0, \\ 0, & \text{if } 0 < l_1 + \cdots + l_d < R \end{cases}.$$

For the bias, a plug-in estimator is given by

$$\hat{\mathcal{B}}_n = -\frac{2(-1)^s}{s!} \sum_{\substack{0 \leq l_1, \dots, l_d \leq s \\ l_1 + \dots + l_d = s}} \left[\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du \right] \hat{\vartheta}_{l_1, \dots, l_d, n},$$

where

$$\hat{\vartheta}_{l_1, \dots, l_d, n} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n b_n^{-(d+1)} \left(\frac{\partial^{(l_1 + \dots + l_d)}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} k \left(\frac{x_i - x_j}{b_n} \right) \right) y_i.$$

The estimator $\hat{\vartheta}_{l_1, \dots, l_d, n}$ is the sample analogue estimator of

$$\vartheta_{l_1, \dots, l_d} = \mathbb{E} \left[\left(\frac{\partial^{(l_1 + \dots + l_d)}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \dot{f}(x) \right) y \right],$$

and is also a n -varying U -statistic estimator employing a leave-one-out kernel-based density estimator.

For \mathcal{V} , an obvious plug-in estimator would be (letting a “hat” denote a sample analogue estimate)

$$\int \dot{K}(u) \dot{K}(u)' (u' \hat{\zeta}_n u) du, \quad \hat{\zeta}_n = \hat{\mathbb{E}} \left[\hat{\sigma}^2(x) \frac{\partial^2}{\partial x \partial x'} \hat{f}(x) + \left(\frac{\partial}{\partial x} \hat{g}(x) \right) \left(\frac{\partial}{\partial x} \hat{g}(x) \right)' \hat{f}(x) \right].$$

However, this estimator has the unappealing property of requiring the estimation of several non-parametric objects, some of which would require handling stochastic denominators. Thus, this direct plug-in approach is likely to be less stable when implemented. Fortunately, it is possible to construct an alternative, indirect estimator much easier to implement in practice. This estimator is intuitively justified as follows: the results presented above show that, under appropriate regularity conditions,

$$b_n^{-2}(\hat{\Delta}_n(b_n) - \Delta) = \mathcal{V} + O_p\left(b_n + n^{-1/2}b_n^{-2} + n^{-1}b_n^{-d/2-2}\right),$$

and therefore an estimator satisfying $\tilde{\Delta}_n = \Delta + o_p(b_n^2) + O_p(n^{-1/2} + n^{-1}b_n^{-d/2})$ would lead to

$$\hat{\mathcal{V}}_n = b_n^{-2}(\hat{\Delta}_n(b_n) - \tilde{\Delta}_n) = \mathcal{V} + o_p(1),$$

if $b_n \rightarrow 0$, $nb_n^4 \rightarrow \infty$ and $n^2b_n^{d+4} \rightarrow \infty$. Under appropriate conditions, an estimator having these properties is given by

$$\tilde{\Delta}_n = \hat{\delta}_n \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' du, \quad \hat{\delta}_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n b_n^{-d} k\left(\frac{x_j - x_i}{b_n}\right) (y_i - y_j)^2.$$

In this case, $\hat{\delta}_n$ is a sample analogue estimator of $\delta = 2\mathbb{E}[\sigma^2(x)f(x)]$, which is also a n -varying U -statistic estimator employing a leave-one-out (higher-order) kernel-based density estimator.

Theorem 2. *If Assumptions 1, 3 and 4 hold, then:*

(i) *For $M \geq s + 1$,*

$$\hat{\vartheta}_{l_1, \dots, l_d, n} = \mathbb{E} \left[\left(\frac{\partial^{(l_1 + \dots + l_d)}}{\partial x_1^{l_1} \dots \partial x_d^{l_d}} \dot{f}(x) \right) y \right] + O_p \left(b_n^{\min(R, S)} + n^{-1/2} + n^{-1}b_n^{-(d+2+2s)/2} \right).$$

(ii) *For $R \geq 3$,*

$$\hat{\delta}_n = 2\mathbb{E}[\sigma^2(x)f(x)] + O_p \left(b_n^{\min(R, s+1+S)} + n^{-1/2} + n^{-1}b_n^{-d/2} \right).$$

This theorem gives simple sufficient conditions to construct a robust data-driven bandwidth selector consistent with the small bandwidth asymptotics derived in Cattaneo, Crump, and Jansson (2009). In particular, define

$$\hat{h}_{CCJ} = \begin{cases} \left(\frac{d(\lambda' \hat{\mathcal{V}}_n \lambda)}{s(\lambda' \hat{\mathcal{B}}_n)^2} \right)^{\frac{1}{2s+d}} n^{-\frac{2}{2s+d}} & \text{if } \lambda' \hat{\mathcal{V}}_n \lambda > 0 \\ \left(\frac{2|\lambda' \hat{\mathcal{V}}_n \lambda|}{(\lambda' \hat{\mathcal{B}}_n)^2} \right)^{\frac{1}{2s+d}} n^{-\frac{2}{2s+d}} & \text{if } \lambda' \hat{\mathcal{V}}_n \lambda < 0 \end{cases}.$$

The following corollary establishes the consistency of the new bandwidth selector \hat{h}_{CCJ} .

Corollary 1. *If Assumptions 1, 2, 3 and 4 hold with $M \geq s + 1$ and $R \geq 3$, and if $b_n \rightarrow 0$ and $n^2 b_n^{\max(8, d+2+2s)} \rightarrow \infty$, then for $\lambda \in \mathbb{R}^d$ such that $\lambda' \mathcal{B} \neq 0$ and $\lambda' \mathcal{V} \lambda \neq 0$,*

$$\frac{\hat{h}_{CCJ}}{h_{CCJ}^*} \rightarrow_p 1.$$

(The analogous result also holds for \hat{h}_{PS} and \hat{h}_{NR} .)

The results presented so far are silent about the selection of the initial bandwidth choice b_n in applications, beyond the rate restrictions imposed by Corollary 1. A simple choice for the preliminary bandwidth b_n may be based on some data-driven bandwidth selector developed for a nonparametric object present in the corresponding target estimands \mathcal{B} , Δ and \mathcal{V} . Typical examples of procedures which may be used include simple rule-of-thumbs, plug-in bandwidth selectors and (smoothed) cross-validation. See, for example, Ichimura and Todd (2007).

As shown in the simulations presented in the next section, it appears that a simple data-driven bandwidth selector from the literature of nonparametric estimation works well for the choice of b_n . Nonetheless, it may be desirable to improve upon this preliminary bandwidth selector in order to obtain better finite sample behavior. Although beyond the scope of this paper, a conceptually feasible (but computationally demanding) idea would be to compute second-order mean squared error expansions for $\hat{\vartheta}_{l_1, \dots, l_d, n}$, $\hat{\Delta}_n$ and $\hat{\delta}_n$. Since these three estimators are n -varying U -statistics, the results from Powell and Stoker (1996) may be applied to obtain a corresponding set of “optimal” bandwidth choices. These procedures will, in turn, also depend on a preliminary bandwidth when implemented empirically, which again would need to be chosen in some way. This idea mimics, in the context of semiparametric estimation, the well-known second-generation direct plug-in bandwidth selector (of level 2) from the literature of nonparametric density estimation. (See, e.g., Wand and Jones (1995) for a detailed discussion.) Although the validity of such bandwidth selectors would require stronger assumptions, by analogy from the nonparametric density estimation literature, they would be expected to improve the finite sample properties of the bandwidth selector for h_n and, in turn, the performance of the semiparametric inference procedure.

5. MONTE CARLO SIMULATIONS

This section reports the main findings from an extensive Monte Carlo experiment conducted to analyze the finite sample properties of the robust data-driven procedure proposed in this paper as well as its relative merits when compared to the other procedures available.

Following the results reported in Cattaneo, Crump, and Jansson (2009), we consider six different models of the (“single index”) form, given by

$$y_i = \tau(y_i^*), \quad y_i^* = x_i' \beta + \varepsilon_i,$$

where $\tau(\cdot)$ is a nondecreasing (link) function and $\varepsilon_i \sim \mathcal{N}(0, 1)$ is independent of the vector of regressors $x_i \in \mathbb{R}^d$. Three different link functions are considered: $\tau(y^*) = y^*$, $\tau(y^*) = \mathbf{1}(y^* > 0)$, and $\tau(y^*) = y^* \mathbf{1}(y^* > 0)$, which correspond to a linear regression, probit, and Tobit model, respectively. ($\mathbf{1}(\cdot)$ represents the indicator function.) The vector of regressors is generated using independent random variables and standardized to have $\mathbb{E}[x_i] = 0$ and $\mathbb{E}[x_i x_i'] = I_d$, with the first component x_{1i} having either a Gaussian distribution or a chi-squared distribution with 4 degrees of freedom (denoted χ_4), while the remaining components have a Gaussian distribution throughout the experiment. All the components of β are set equal to unity, and for simplicity only results for the first component of θ (i.e., θ_1) are reported.

TABLE I: MONTE CARLO MODELS

	$y_i = y_i^*$	$y_i = \mathbf{1}(y_i^* > 0)$	$y_i = y_i^* \mathbf{1}(y_i^* > 0)$
$x_{1i} \sim \mathcal{N}(0, 1)$	Model 1: $\theta_1 = \frac{1}{4\pi}$	Model 3: $\theta_1 = \frac{1}{8\pi^{3/2}}$	Model 5: $\theta_1 = \frac{1}{8\pi}$
$x_{1i} \sim \frac{\chi_4 - 4}{\sqrt{8}}$	Model 2: $\theta_1 = \frac{1}{4\sqrt{2\pi}}$	Model 4: $\theta_1 = 0.02795$	Model 6: $\theta_1 = 0.03906$

Table I summarizes the Monte Carlo models, reports the value of the population parameter of interest, and provides the corresponding label of each model considered. (Whenever unavailable in closed form, the population parameters are computed by a numerical approximation.) The simulation study considers three sample sizes ($n = 100$, $n = 400$ and $n = 700$), two dimensions of the regressors vector ($d = 2$ and $d = 4$), and two kernel orders ($P = 2$ and $P = 4$). The kernel function $K(\cdot)$ is chosen to be a standard Gaussian product kernel when $P = 2$, or a Gaussian density-based multiplicative product kernel when $P = 4$. The preliminary kernel function $k(\cdot)$ is chosen to be a fourth order Gaussian density-based multiplicative product kernel, since $R \geq 3$ is required by Corollary 1. For all possible combinations of the parameters 10,000 replications are carried out.

The simulation experiment considers the three (infeasible) population bandwidth choices described in Section 3, denoted h_{PS}^* , h_{NR}^* and h_{CCJ}^* , and their corresponding data-driven estimates, denoted \hat{h}_{PS} , \hat{h}_{NR} and \hat{h}_{CCJ} . The three estimated bandwidth are obtained using the results described in Section 4 with a common initial bandwidth plug-in estimate used to construct $\hat{\mathcal{B}}_n$, $\hat{\Delta}_n$ and $\hat{\mathcal{V}}_n$. To provide a parsimonious data-driven procedure, the initial bandwidth b_n is constructed as a sample average of a second-generation direct plug-in level-two estimate for the (marginal) density of each dimension of the regressors vector, as described in, for example, Wand and Jones (1995). Confidence intervals for θ_1 are constructed using each of the six bandwidth choices for the classical test statistic of Powell, Stock, and Stoker (1989), denoted PSS, and the two alternative

robust test statistics proposed by Cattaneo, Crump, and Jansson (2009) based on the robust standard errors given by equation (2) and equation (3), and denoted by CCJ1 and CCJ2, respectively. Notice that the classical inference procedure PSS is only theoretically valid when $P = 4$, while the robust procedures CCJ1 and CCJ2 are always valid across all simulation designs.

Figures 1 through 8 plot the empirical coverage for the three competing 95% confidence intervals as a function of the choice of bandwidth for each of the six models. To facilitate the comparison only a restricted range of bandwidths is plotted and two additional horizontal lines at 0.90 and at the nominal coverage rate 0.95 are included for reference. In addition, the three population bandwidth selectors h_{PS}^* , h_{NR}^* and h_{CCJ}^* are plotted as vertical lines. (Note that $h_{PS}^* = h_{NR}^*$ for the case $d = 2$ and $P = 2$.) Each figure depicts the results for a combination of P , d and n , although results for $n = 700$ are not included to conserve space. For example, Figure 2 plots the simulation results using a standard Gaussian product kernel ($P = 2$), $d = 2$ and $n = 400$. In all cases, these figures highlight the potential robustness properties that the test statistics CCJ1 and CCJ2 may have when using the new data-driven plug-in bandwidth selector. In particular, the theoretical bandwidth selector h_{CCJ}^* lays within the robust region for which both CCJ1 and CCJ2 have correct empirical coverage for a range of bandwidths. This, in turn, suggests that (at least) some of the variability introduced by the estimation of this bandwidth selector will not affect the performance of these (robust) test statistics. In contrast, these figures show that this property is unlikely to be enjoyed by the classical procedure, denoted PSS.

To further describe the properties of the new bandwidth selector, Figures 9 through 13 plot corresponding kernel density estimates for the test statistic PSS coupled with either h_{PS}^* and h_{NR}^* , and for the test statistics CCJ1 and CCJ2 coupled with h_{CCJ}^* . To facilitate the comparison the density of the standard normal is also depicted, and to conserve space only the case $n = 400$ is included. These figures show that the Gaussian approximation of the robust test statistics using the new bandwidth selector is considerably better than the corresponding approximation for PSS when constructed using either of the classical bandwidth selectors.

To analyze the performance of the new data-driven bandwidth selector, and the resulting robust data-driven confidence intervals, Tables 1 through 4 present the empirical coverage of each possible confidence interval (PSS, CCJ1 and CCJ2) when using each possible bandwidth selector (the infeasible h_{PS}^* , h_{NR}^* and h_{CCJ}^* , and the feasible \hat{h}_{PS} , \hat{h}_{NR} and \hat{h}_{CCJ}). In general, these tables provide concrete evidence of the superior performance of the robust test statistics when coupled with the new estimated bandwidth \hat{h}_{CCJ} , leading to two robust data-driven confidence intervals. For example, Table 2 reports the case of $P = 2$ and $d = 2$ for all three sample sizes. This table shows that the theoretical bandwidth h_{CCJ}^* and the empirical bandwidth \hat{h}_{CCJ} deliver approximately correct coverage for all models and sample sizes when using either CCJ1 or CCJ2, while this is not the case for the test statistic PSS. However, this case is not theoretically justified for the classical procedure, which may partially explain its poor performance. Nonetheless, for example,

Table 3 shows that even when $P = 4$ and $d = 2$, the classical procedure PSS coupled with either \hat{h}_{PS} or \hat{h}_{NR} is unable to achieve correct coverage. On the other hand, CCJ1 or CCJ2 using \hat{h}_{CCJ} do provide close-to-correct coverage across models and sample sizes. This provides additional evidence of the robustness properties of the new procedures.

In addition, it is interesting to note that the good performance of CCJ1 or CCJ2 using \hat{h}_{CCJ} is maintained even when the dimension grows, which provides empirical evidence of the relatively low sensitivity of the new robust data-driven procedures to the so-called “curse of dimensionality.” This finding may be (heuristically) justified by the fact that under the small bandwidth asymptotics, the limiting distribution is not invariant to the “parameter” d , which in turn may lead to further robustness properties of CCJ1 and CCJ2 in this additional direction.

Finally, as suggested by the good Gaussian approximation reported in Figures 9 through 13 for the new procedures, the main findings summarized in this section continue to hold if other nominal confidence levels are considered. In particular, although not reported to conserve space, the same results are found when 90% or 99% confidence intervals are considered.

6. FINAL REMARKS

This paper introduced a new data-driven plug-in bandwidth selector compatible with the small bandwidth asymptotics developed in Cattaneo, Crump, and Jansson (2009) for density-weighted average derivatives. This new bandwidth selector is of the plug-in variety, and is obtained based on a mean squared error expansion of the estimator of interest. An extensive Monte Carlo experiment showed a remarkable improvement in performance of the resulting new robust data-driven inference procedure. In particular, the new data-driven confidence intervals provide approximately correct coverage in cases where there does not exist valid alternative inference procedures (i.e., using a second-order kernel with at least two regressors), and also compares favorably to the alternative, classical confidence intervals when they are theoretically justified.

7. APPENDIX

Proof of Theorem 1. To save notation, for any function $a : \mathbb{R}^d \rightarrow \mathbb{R}$ let $\dot{a}(x) = \partial a(x) / \partial x$ and $\ddot{a}(x) = \partial^2 a(x) / \partial x \partial x'$. A Hoeffding decomposition of $\hat{\theta}_n$ gives

$$\begin{aligned} \mathbb{E}[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)'] &= \mathbb{V}[\hat{\theta}_n] + (\mathbb{E}[\hat{\theta}_n] - \theta)(\mathbb{E}[\hat{\theta}_n] - \theta)' \\ &= \mathbb{V}[\bar{L}_n] + \mathbb{V}[\bar{W}_n] + h_n^{2s} \mathcal{B} \mathcal{B}' + o(h_n^{2s}), \end{aligned}$$

where the bias expansion follows immediately by a Taylor series expansion.

For $\mathbb{V}[\bar{L}_n]$, using integration by parts,

$$\mathbb{E}[U_n(z_i, z_j) | z_i] = \int_{\mathbb{R}^d} \dot{e}(x_i + u h_n) K(u) du - y_i \int_{\mathbb{R}^d} \dot{f}(x_i + u h_n) K(u) du,$$

and therefore

$$\mathbb{V}[\bar{L}_n] = \frac{4}{n} \mathbb{V}[\mathbb{E}[U_n(z_i, z_j) | z_i] - \theta_n] = \frac{1}{n} \Sigma + O(n^{-1} h_n^s).$$

For $\mathbb{V}[\bar{W}_n]$, by standard calculations,

$$\begin{aligned} \mathbb{V}[\bar{W}_n] &= \binom{n}{2}^{-1} \mathbb{E}[U_n(z_i, z_j) U_n(z_i, z_j)'] + O(n^{-2}) \\ &= \binom{n}{2}^{-1} h_n^{-(d+2)} \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' T(x, u h_n) dx du + O(n^{-2}), \end{aligned}$$

with $T(x, u) = (v(x) + v(x + u) - 2g(x)g(x + u))f(x)f(x + u)$. Then, using a Taylor series expansion, $T(x, u h_n) = T_1(x) + T_2(x)' u h_n + u' T_3(x) u h_n^2 + o(h_n^2)$, where

$$T_1(x) = 2\sigma^2(x) f(x)^2, \quad T_2(x) = 2\sigma^2(x) f(x) \dot{f}(x) + f(x)^2 \dot{\sigma}^2(x),$$

$$T_3(x) = \sigma^2(x) f(x) \ddot{f}(x) + f(x) \dot{\sigma}^2(x) \dot{f}(x) + \left(\frac{1}{2} \ddot{v}(x) - g(x) \ddot{g}(x) \right) f(x)^2.$$

Clearly,

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' T_1(x) dx du = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' 2\sigma^2(x) f(x)^2 dx du = \Delta,$$

and, using integration by parts,

$$\begin{aligned} &h_n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' (T_2(x)' u) dx du \\ &= h_n \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' \left[\left(\int_{\mathbb{R}^d} [\sigma^2(x) 2f(x) \dot{f}(x) + f(x)^2 \dot{\sigma}^2(x)] dx \right)' u \right] du = 0. \end{aligned}$$

Finally, using integration by parts and the fact that $\ddot{\sigma}^2(x) = \ddot{v}(x) - 2\dot{g}(x)\dot{g}(x)' - 2g(x)\ddot{g}(x)$,

$$\begin{aligned} & h_n^2 \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' (u' T_3(x) u) dx du \\ &= h_n^2 \int_{\mathbb{R}^d} \dot{K}(u) \dot{K}(u)' \left[u' \left(\int_{\mathbb{R}^d} \sigma^2(x) \ddot{f}(x) f(x) dx + \int_{\mathbb{R}^d} \dot{g}(x) \dot{g}(x)' f(x)^2 dx \right) u \right] du. \end{aligned}$$

Therefore,

$$\mathbb{V}[\bar{W}_n] = \binom{n}{2}^{-1} h_n^{-(d+2)} \Delta + \binom{n}{2}^{-1} h_n^{-d} \mathcal{V} + o(n^{-2} h_n^{-d}),$$

which establishes the result. \blacksquare

Proof of Theorem 2. For part (i), note that $\hat{\vartheta}_{l_1, \dots, l_d, n}$ may be written as a n -varying U -statistic (assuming without loss of generality that s is even), given by

$$\hat{\vartheta}_{l_1, \dots, l_d, n} = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_1(z_i, z_j; b_n),$$

with (recall that $s = l_1 + \dots + l_d$)

$$u_1(z_i, z_j; b) = b^{-(d+1+s)} \left(\frac{\partial^s}{\partial x_1^{l_1} \dots \partial x_d^{l_d}} \dot{k}(x) \Big|_{x=(x_i-x_j)/b} \right) (y_i - y_j).$$

First, change of variables and integration by parts give

$$\mathbb{E}[u_1(z_i, z_j; b_n) | z_i] = \int_{\mathbb{R}^d} k(u) \left(\frac{\partial^s}{\partial x_1^{l_1} \dots \partial x_d^{l_d}} \dot{f}(x) \Big|_{x=x_i-ub_n} y_i - \frac{\partial^s}{\partial x_1^{l_1} \dots \partial x_d^{l_d}} \dot{e}(x) \Big|_{x=x_i-ub_n} \right) du.$$

Second, a Taylor series expansion gives $\mathbb{E}[u_1(z_i, z_j; b_n)] = \vartheta_{l_1, \dots, l_d} + O(b_n^{\min(R, S)})$. Next, letting $\hat{\vartheta}_n = \hat{\vartheta}_{l_1, \dots, l_d, n}$ to save notation, a Hoeffding decomposition gives $\mathbb{V}[\hat{\vartheta}_n] = \mathbb{V}[\hat{\vartheta}_{1,n}] + \mathbb{V}[\hat{\vartheta}_{2,n}]$, where

$$\hat{\vartheta}_{1,n} = \frac{1}{n} \sum_{i=1}^n 2 [\mathbb{E}[u_1(z_i, z_j; b_n) | z_i] - \mathbb{E}[u_1(z_i, z_j; b_n)]] ,$$

and

$$\hat{\vartheta}_{2,n} = \binom{n}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n [u_1(z_i, z_j; b_n) - \mathbb{E}[u_1(z_i, z_j; b_n) | z_i] - \mathbb{E}[u_1(z_i, z_j; b_n) | z_j] + \mathbb{E}[u_1(z_i, z_j; b_n)]] .$$

Finally, using standard calculations, $\mathbb{V}[\hat{\vartheta}_{1,n}] = O(n^{-1})$ and $\mathbb{V}[\hat{\vartheta}_{2,n}] = O(n^{-2} b_n^{-(d+2+2s)})$, and the conclusion follows by Markov's Inequality.

For part (ii), note that $\hat{\delta}_n$ is also a n -varying U -statistic, given by

$$\hat{\delta}_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_2(z_i, z_j; b_n), \quad u_2(z_i, z_j; b) = b^{-d} k\left(\frac{x_j - x_i}{b}\right) (y_i - y_j)^2.$$

First, change of variables gives

$$\mathbb{E}[u_2(z_i, z_j; b_n) | z_i] = \int_{\mathbb{R}^d} k(u) (y_i^2 f(x_i - ub_n) + v(x_i - ub_n) f(x_i - ub_n) - 2y_i e(x_i - ub_n)) du.$$

Second, a Taylor's expansion gives $\mathbb{E}[\hat{\delta}_n] = 2\mathbb{E}[\sigma^2(x) f(x)] + O(b_n^{\min(R, s+1+S)})$. Next, a Hoeffding decomposition gives $\mathbb{V}[\hat{\delta}_n] = \mathbb{V}[\hat{\delta}_{1,n}] + \mathbb{V}[\hat{\delta}_{2,n}]$, where

$$\hat{\delta}_{1,n} = \frac{1}{n} \sum_{i=1}^n 2 [\mathbb{E}[u_2(z_i, z_j; b_n) | z_i] - \mathbb{E}[u_2(z_i, z_j; b_n)]] ,$$

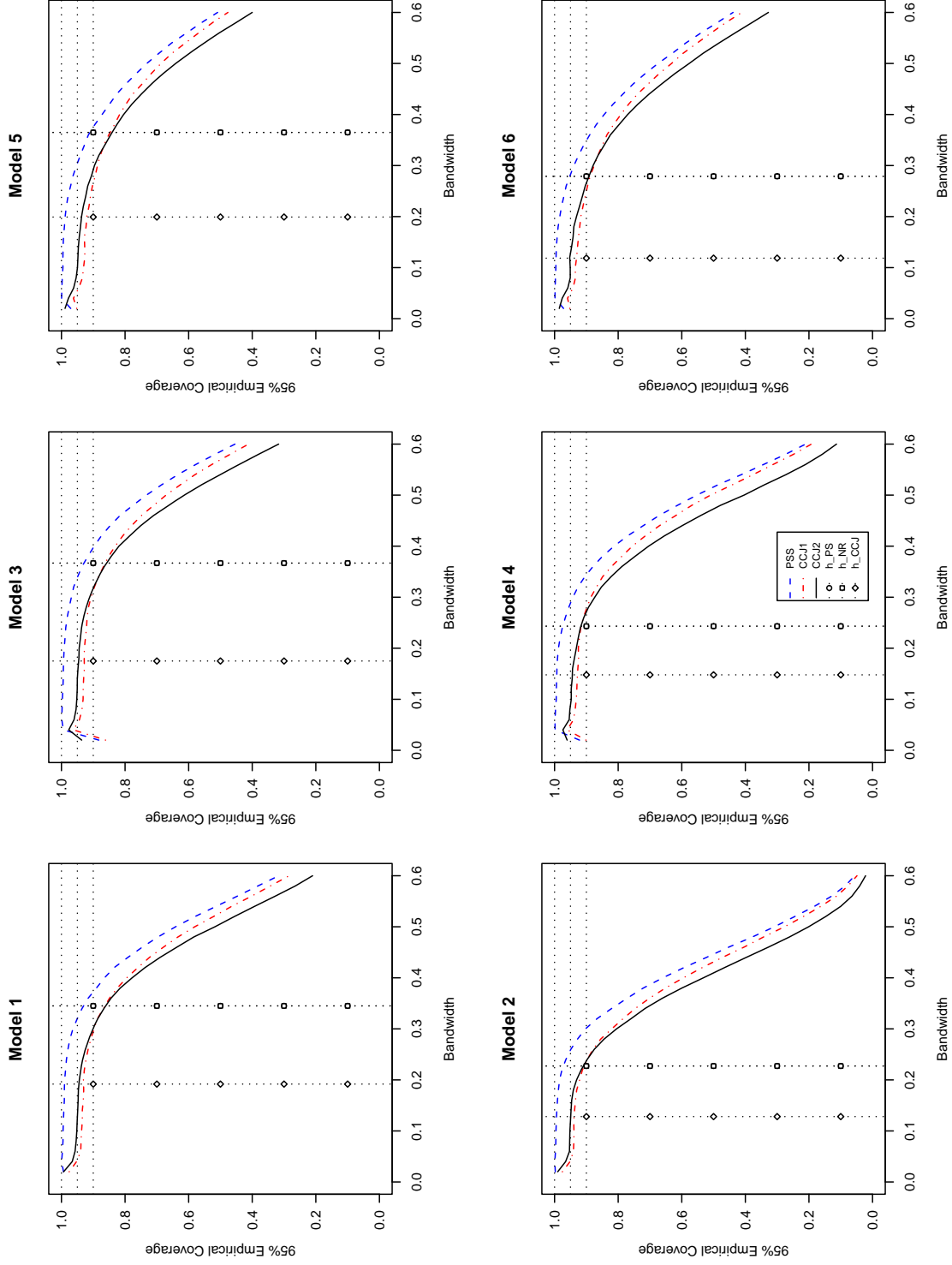
and

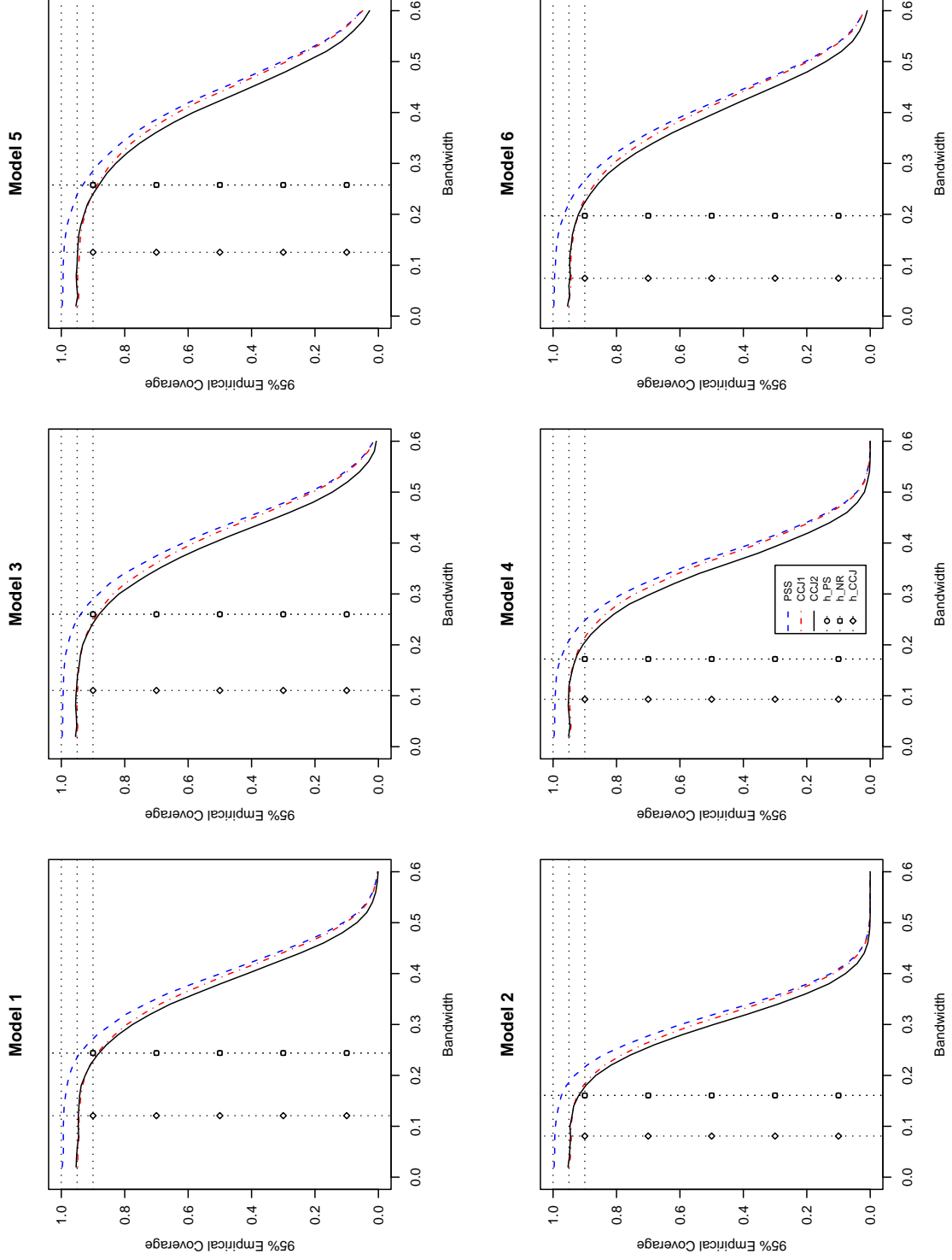
$$\hat{\delta}_{2,n} = \binom{n}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n [u_2(z_i, z_j; b_n) - \mathbb{E}[u_2(z_i, z_j; b_n) | z_i] - \mathbb{E}[u_2(z_i, z_j; b_n) | z_j] + \mathbb{E}[u_2(z_i, z_j; b_n)]] .$$

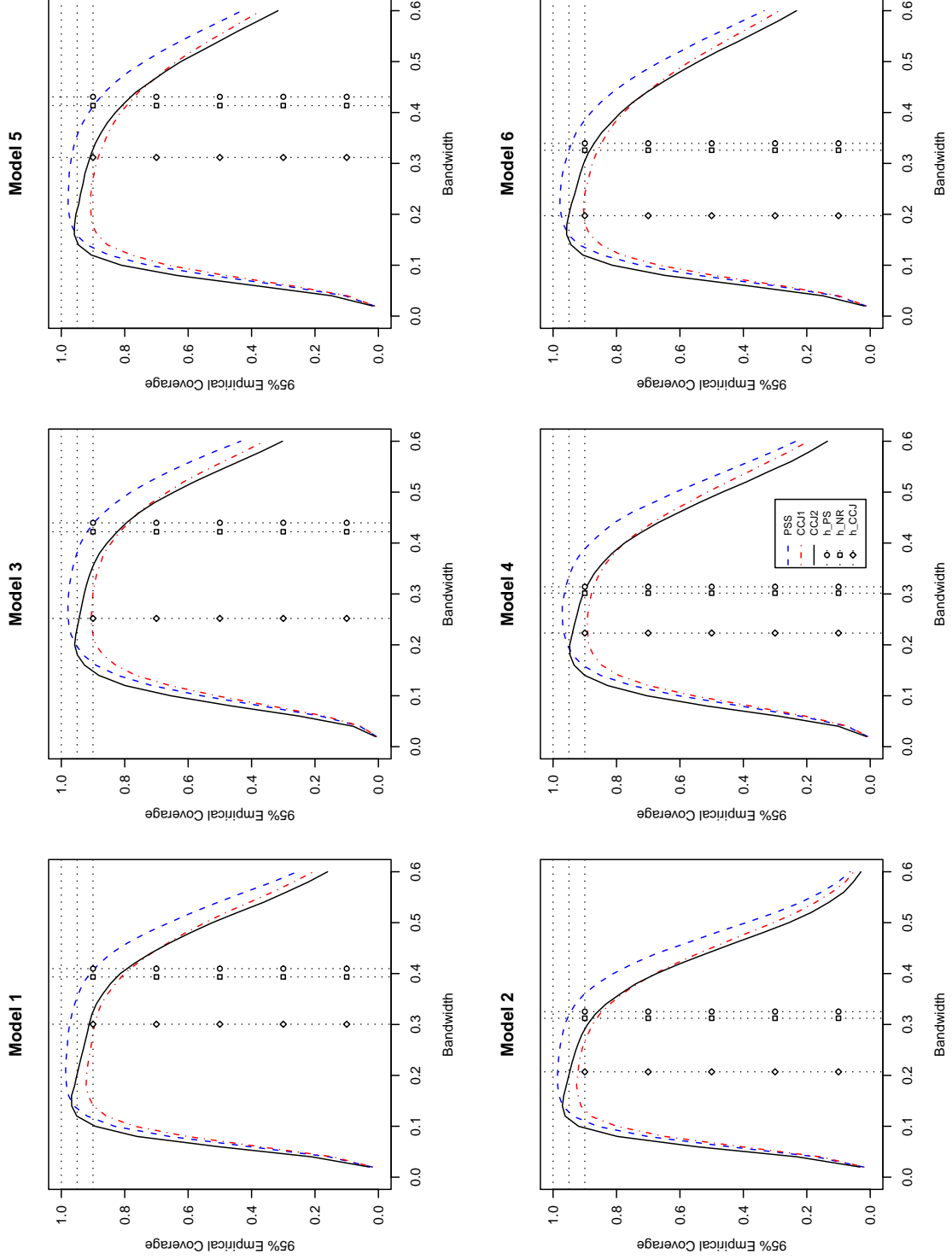
Finally, using standard calculations, $\mathbb{V}[\hat{\delta}_{1,n}] = O(n^{-1})$ and $\mathbb{V}[\hat{\delta}_{2,n}] = O(n^{-2}b_n^d)$, and the conclusion follows by Markov's Inequality. \blacksquare

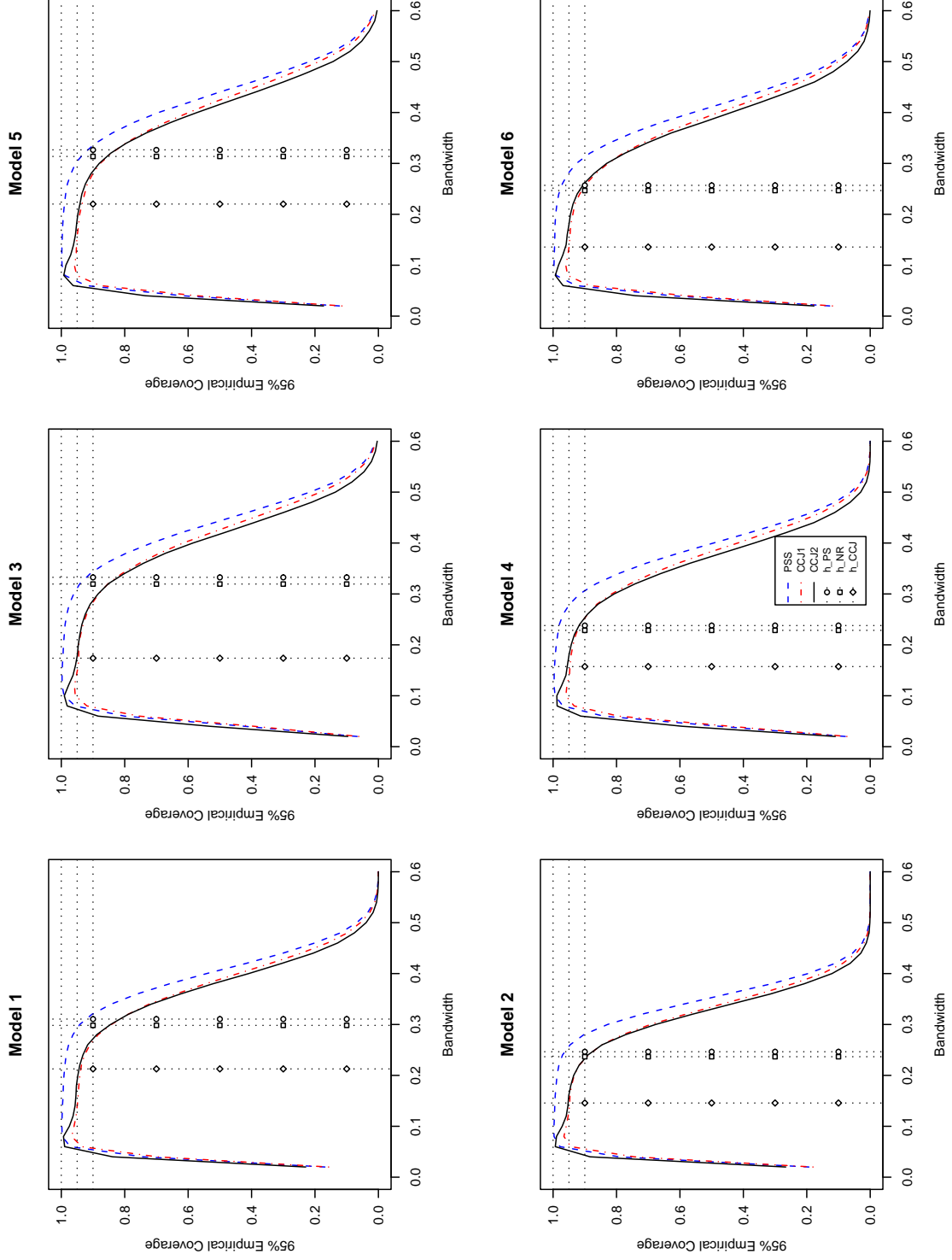
REFERENCES

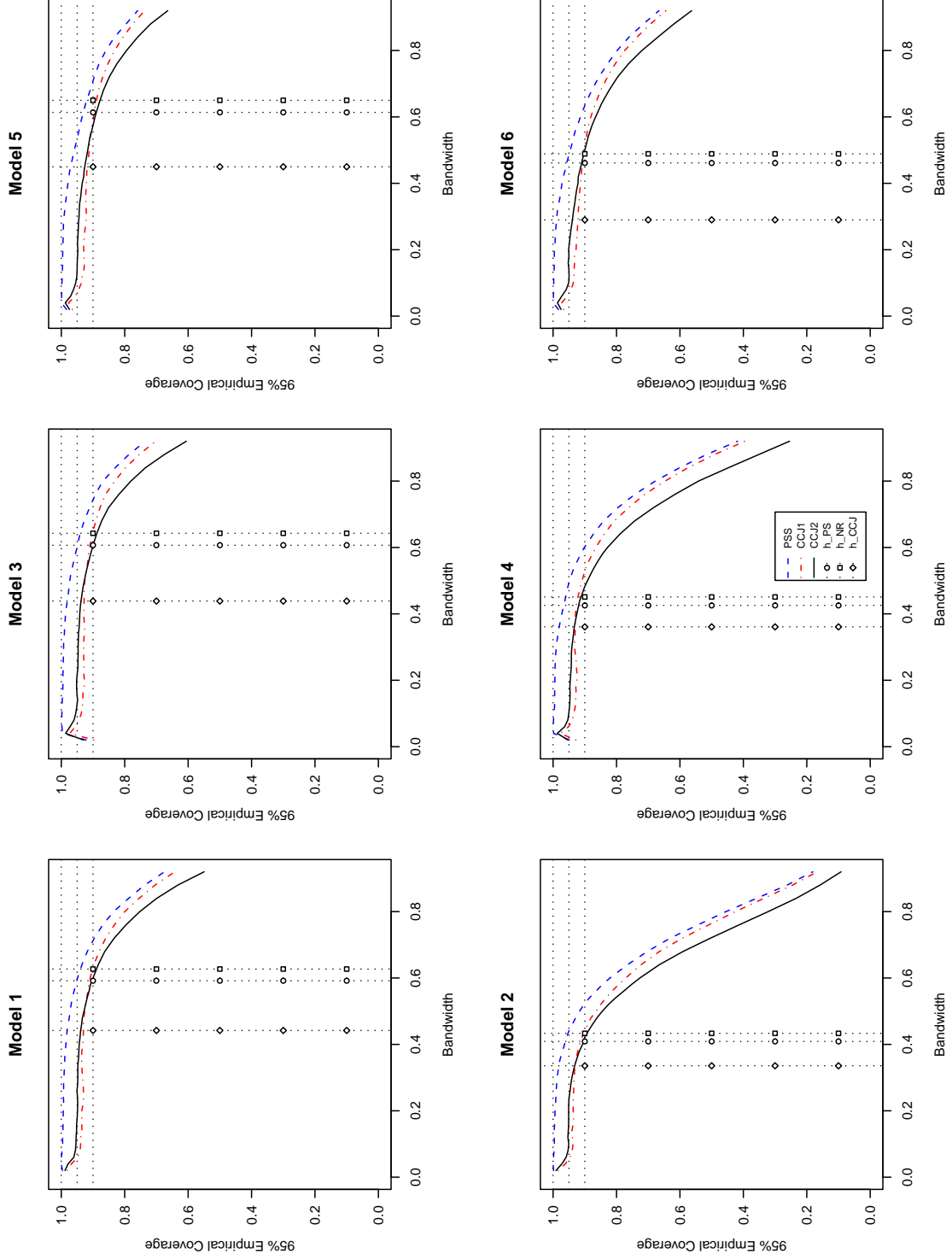
- CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2009): “Small Bandwidth Asymptotics for Density-Weighted Average Derivatives,” working paper.
- DEATON, A., AND S. NG (1998): “Parametric and Nonparametric Approaches to Price and Tax Reform,” *Journal of the American Statistical Association*, 93, 900–909.
- HÄRDLE, W., J. HART, J. MARRON, AND A. TSYBAKOV (1992): “Bandwidth Choice for Average Derivative Estimation,” *Journal of the American Statistical Association*, 87, 218–226.
- HÄRDLE, W., AND T. STOKER (1989): “Investigating Smooth Multiple Regression by the Method of Average Derivatives,” *Journal of the American Statistical Association*, 84, 986–995.
- HÄRDLE, W., AND A. TSYBAKOV (1993): “How Sensitive are Average Derivatives?,” *Journal of Econometrics*, 58, 31–48.
- HOROWITZ, J., AND W. HÄRDLE (1996): “Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates,” *Journal of the American Statistical Association*, 91, 1632–1640.
- ICHIMURA, H., AND P. E. TODD (2007): “Implementing Nonparametric and Semiparametric Estimators,” in *Handbook of Econometrics, Volume VI*, ed. by R. F. Engle, and D. L. McFadden, pp. 5370–5468. Elsevier Science B.V.
- NEWKEY, W. K. (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62(6), 1349–1382.
- NEWKEY, W. K., F. HSIEH, AND J. M. ROBINS (2004): “Twicing Kernels and a Small Bias Property of Semiparametric Estimators,” *Econometrica*, 72, 947–962.
- NISHIYAMA, Y., AND P. M. ROBINSON (2000): “Edgeworth Expansions for Semiparametric Averaged Derivatives,” *Econometrica*, 68, 931–979.
- (2001): “Studentization in Edgeworth Expansions for Estimates of Semiparametric Index Models,” in *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. L. Powell, pp. 197–240. Cambridge University Press, New York.
- (2005): “The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives,” *Econometrica*, 73, 197–240.
- POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): “Semiparametric Estimation of Index Coefficients,” *Econometrica*, 57(6), 1403–1430.
- POWELL, J. L., AND T. M. STOKER (1996): “Optimal Bandwidth Choice for Density-Weighted Averages,” *Journal of Econometrics*, 75, 291–316.
- ROBINSON, P. M. (1995): “The Normal Approximation for Semiparametric Averaged Derivatives,” *Econometrica*, 63, 291–216.
- STOKER, T. M. (1986): “Consistent Estimation of Scaled Coefficients,” *Econometrica*, 54(6), 1461–1481.
- WAND, M., AND M. JONES (1995): *Kernel Smoothing*. Chapman & Hall/CRC, Florida.

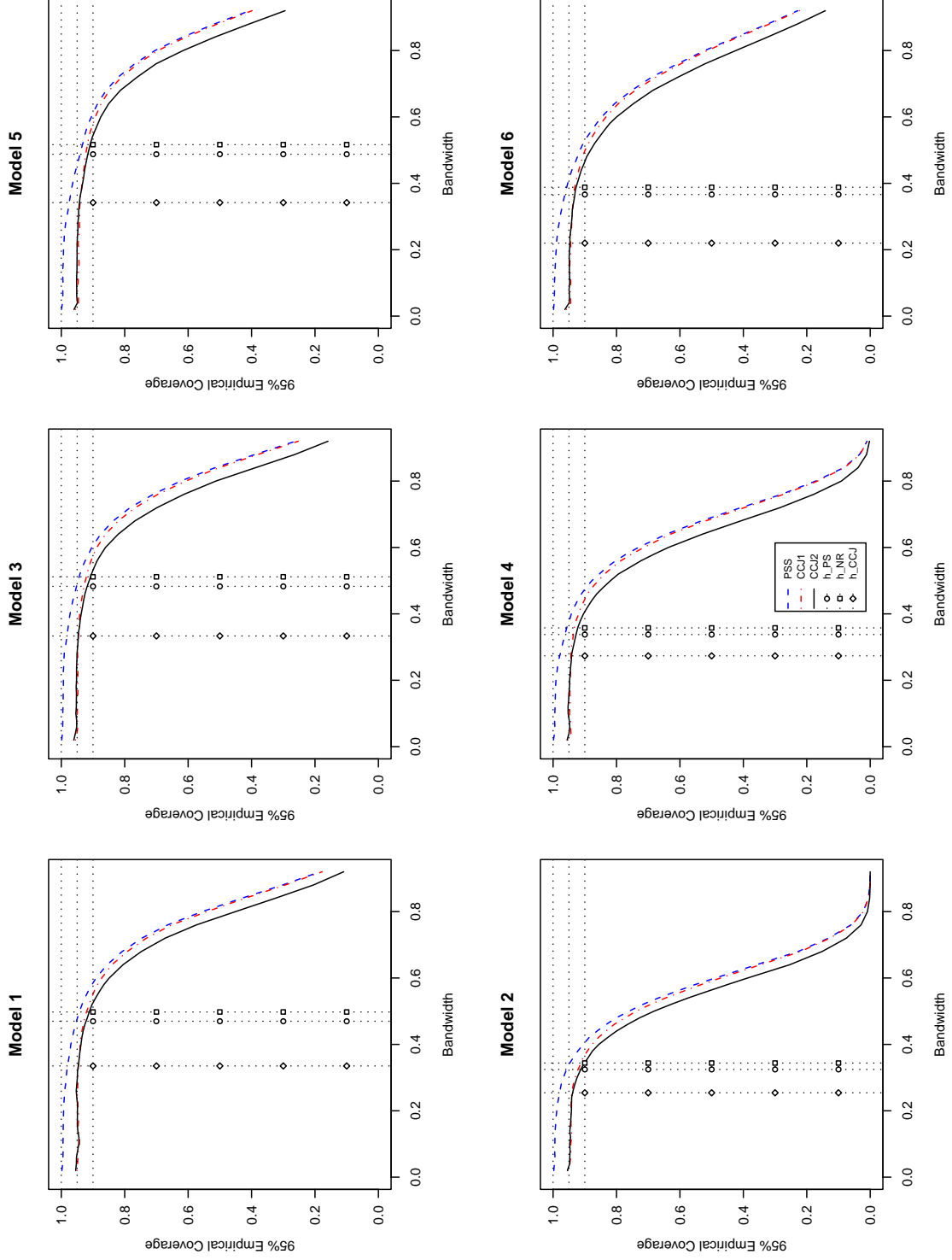
Figure 1: Empirical Coverage Rates for 95% Confidence Intervals: $P = 2$, $d = 2$, $n = 100$

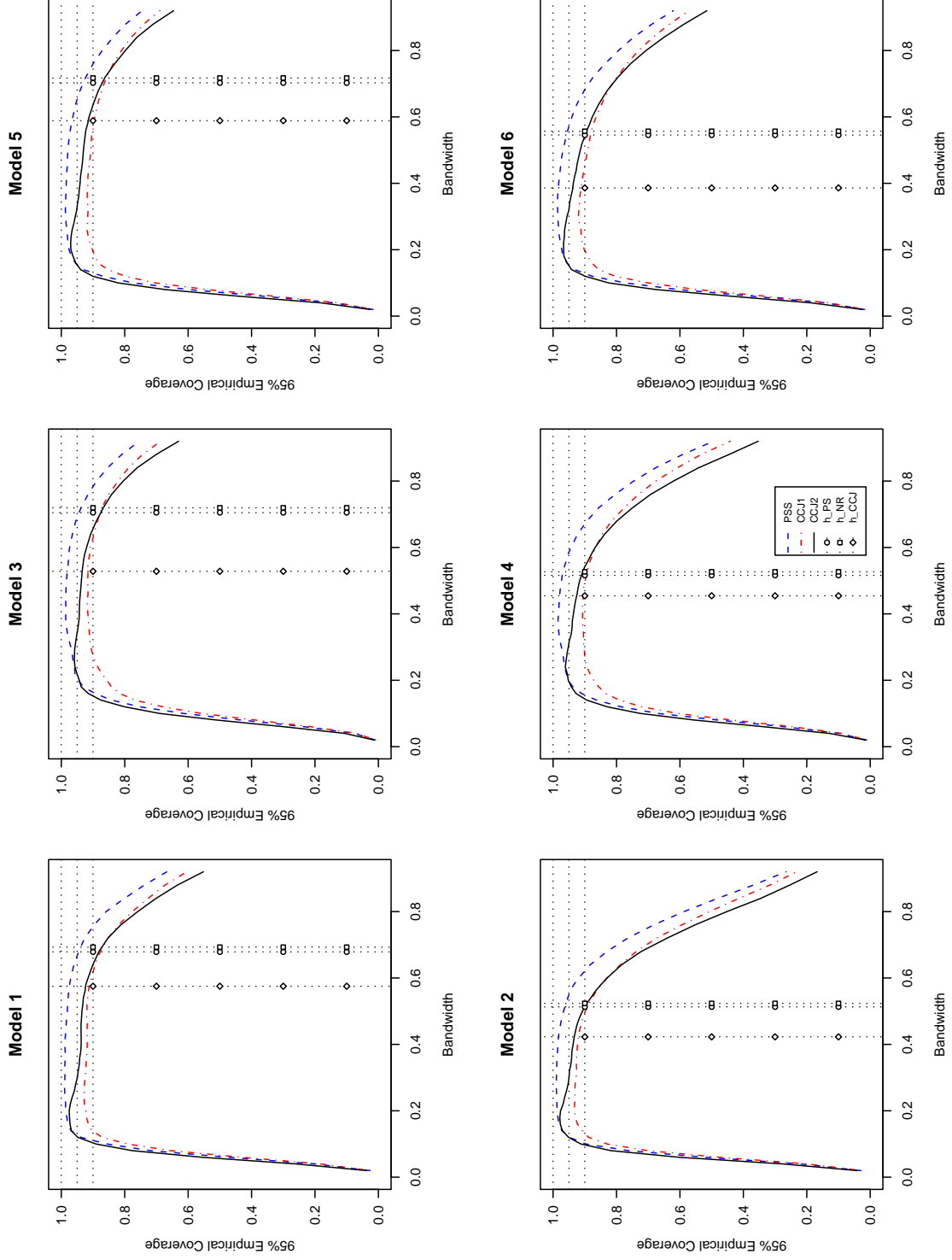
Figure 2: Empirical Coverage Rates for 95% Confidence Intervals: $P = 2$, $d = 2$, $n = 400$

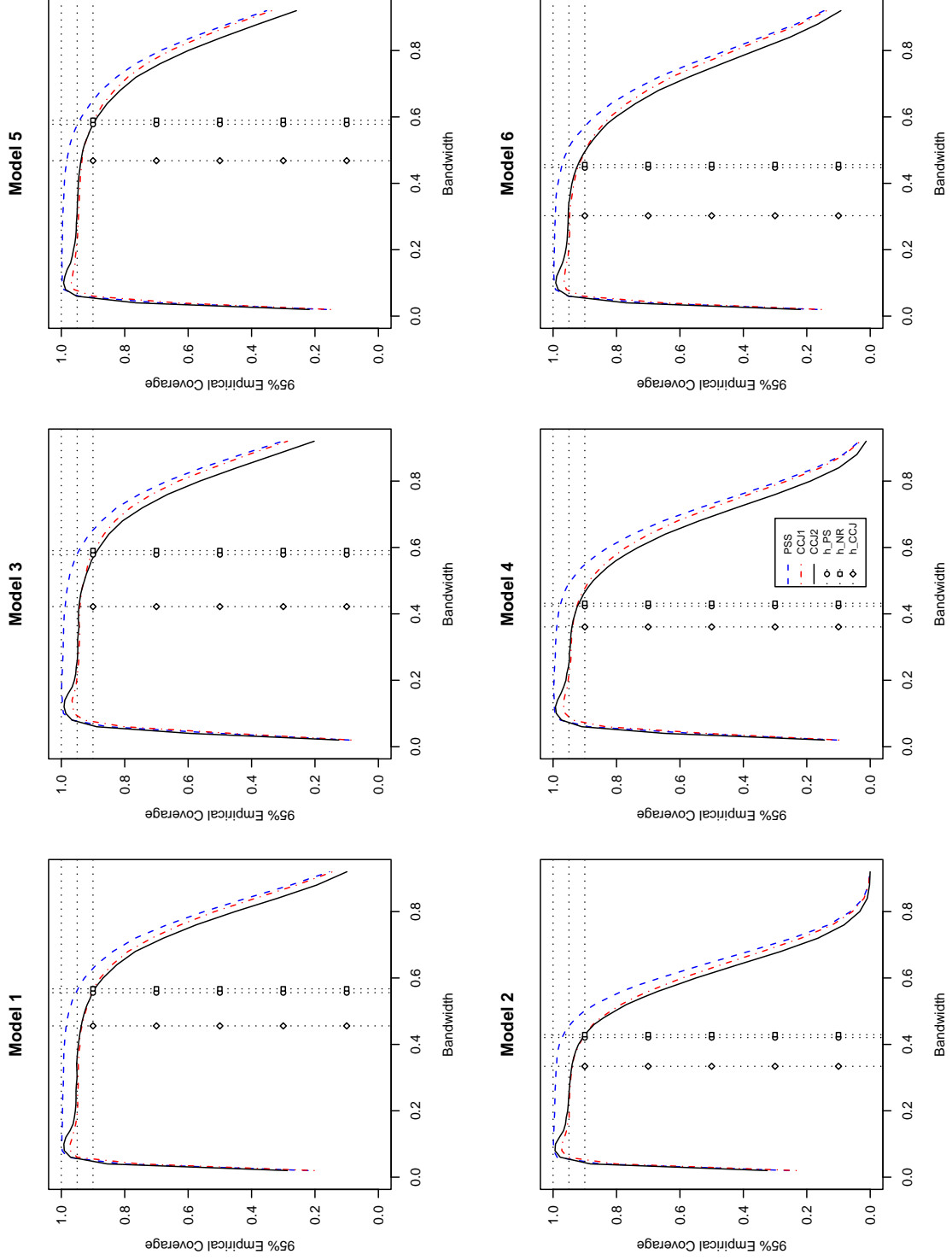
Figure 3: Empirical Coverage Rates for 95% Confidence Intervals: $P = 2$, $d = 4$, $n = 100$

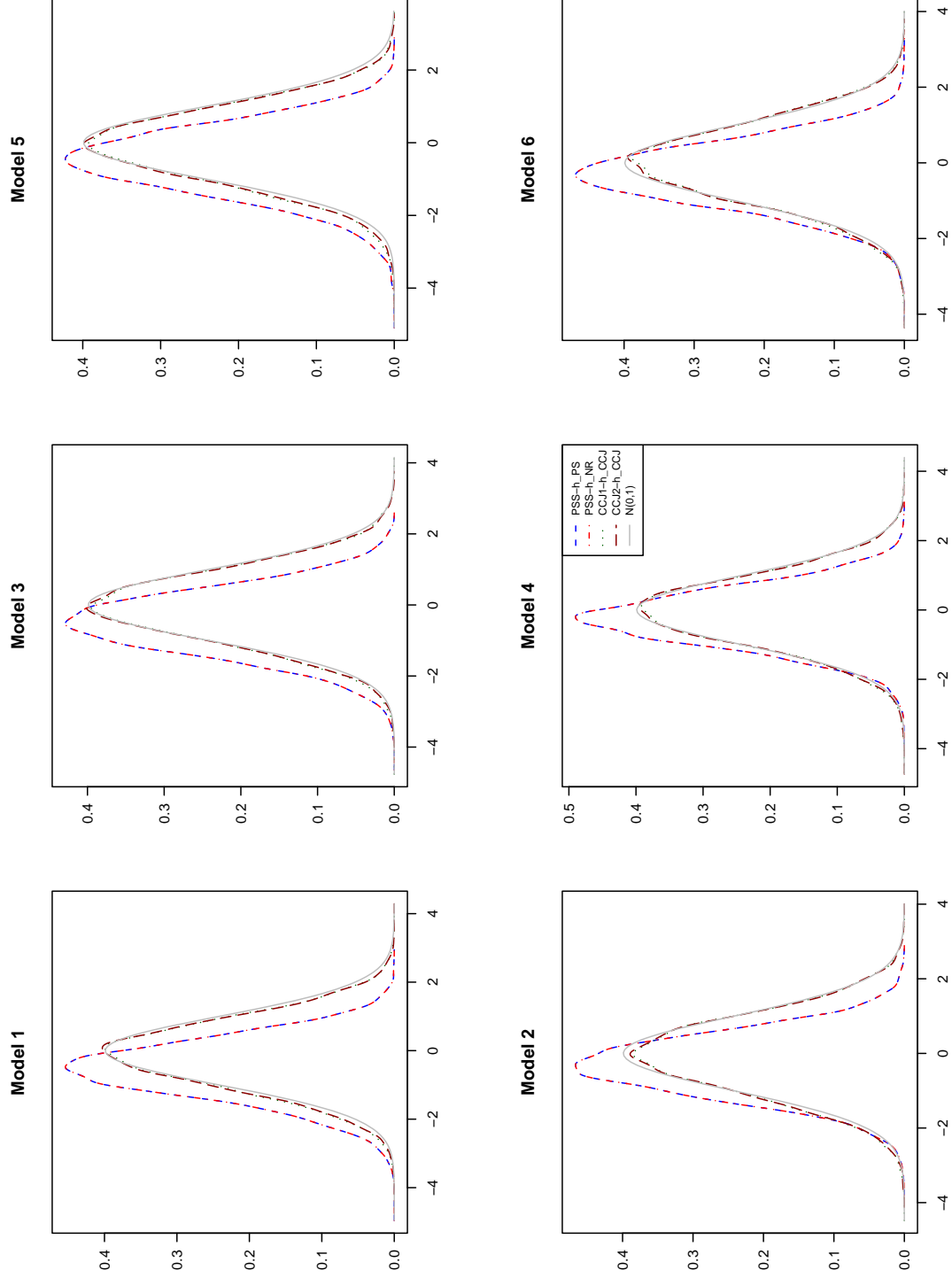
Figure 4: Empirical Coverage Rates for 95% Confidence Intervals: $P = 2$, $d = 4$, $n = 400$

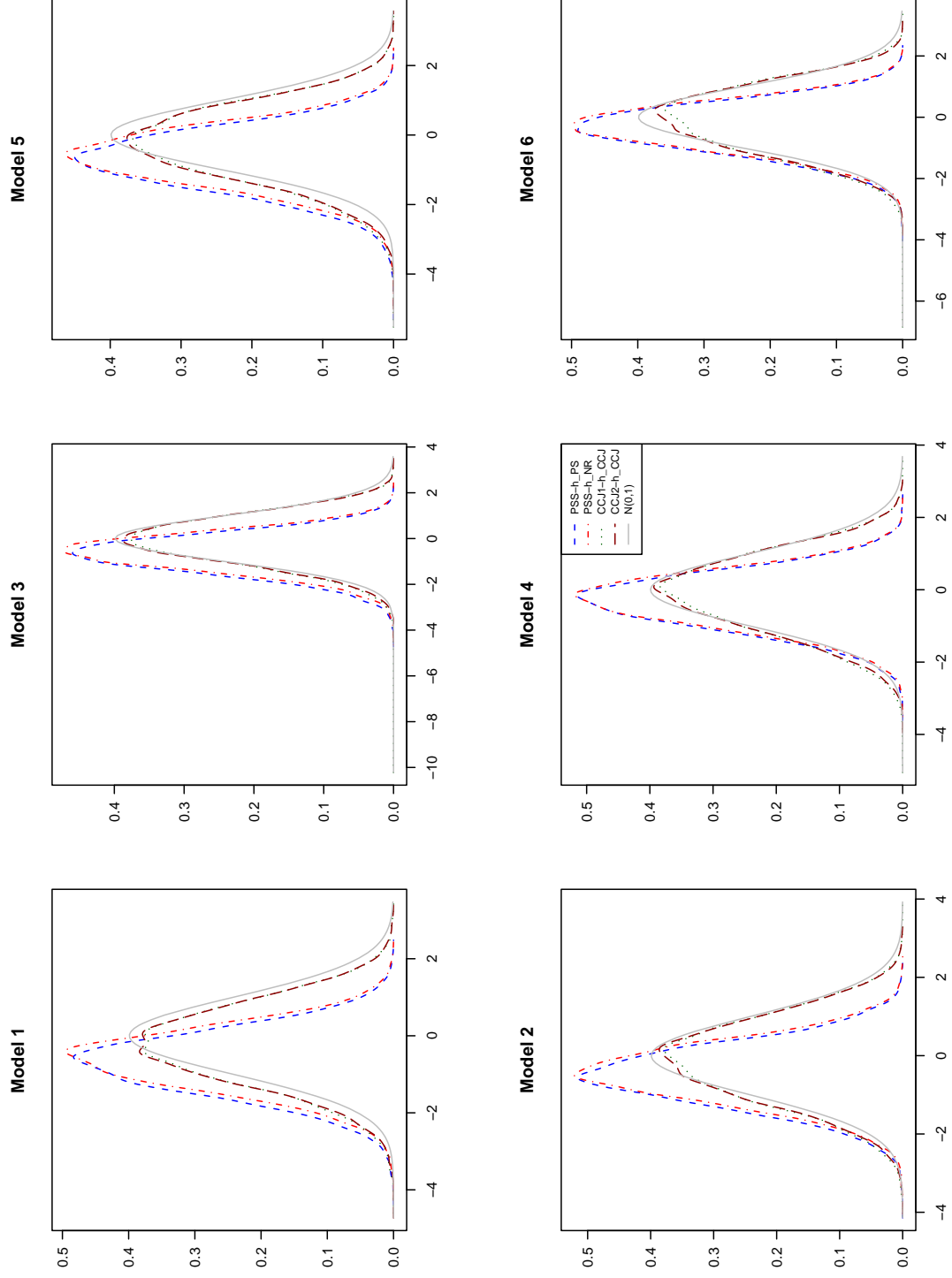
Figure 5: Empirical Coverage Rates for 95% Confidence Intervals: $P = 4$, $d = 2$, $n = 100$

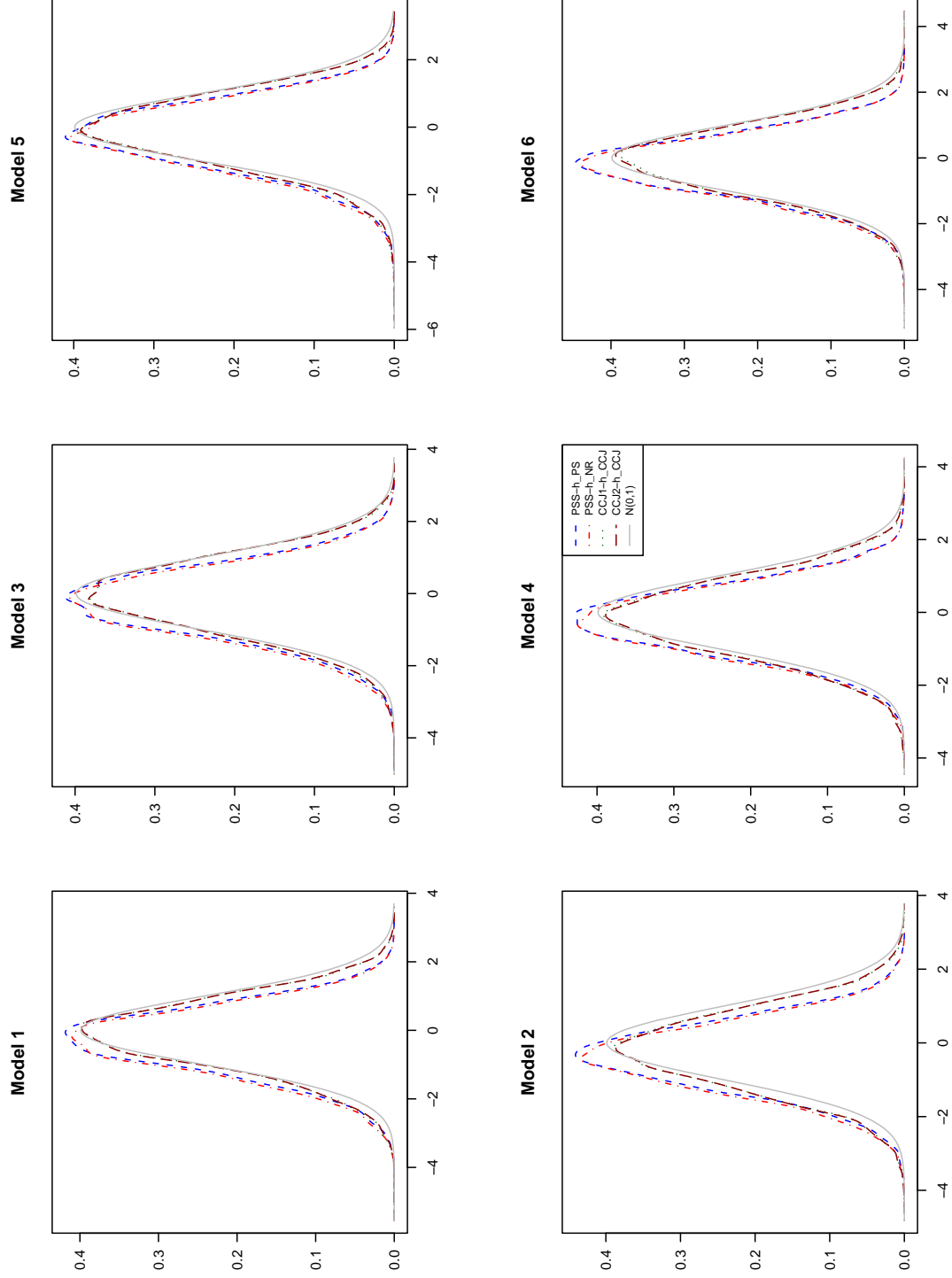

 Figure 6: Empirical Coverage Rates for 95% Confidence Intervals: $P = 4$, $d = 2$, $n = 400$

Figure 7: Empirical Coverage Rates for 95% Confidence Intervals: $P = 4$, $d = 4$, $n = 100$

Figure 8: Empirical Coverage Rates for 95% Confidence Intervals: $P = 4$, $d = 4$, $n = 400$


 Figure 9: Empirical Gaussian Approximation with Population Bandwidth: $P = 2$, $d = 2$, $n = 400$

Figure 10: Empirical Gaussian Approximation with Population Bandwidth: $P = 2$, $d = 4$, $n = 400$


 Figure 11: Empirical Gaussian Approximation with Population Bandwidth: $P = 4$, $d = 2$, $n = 400$

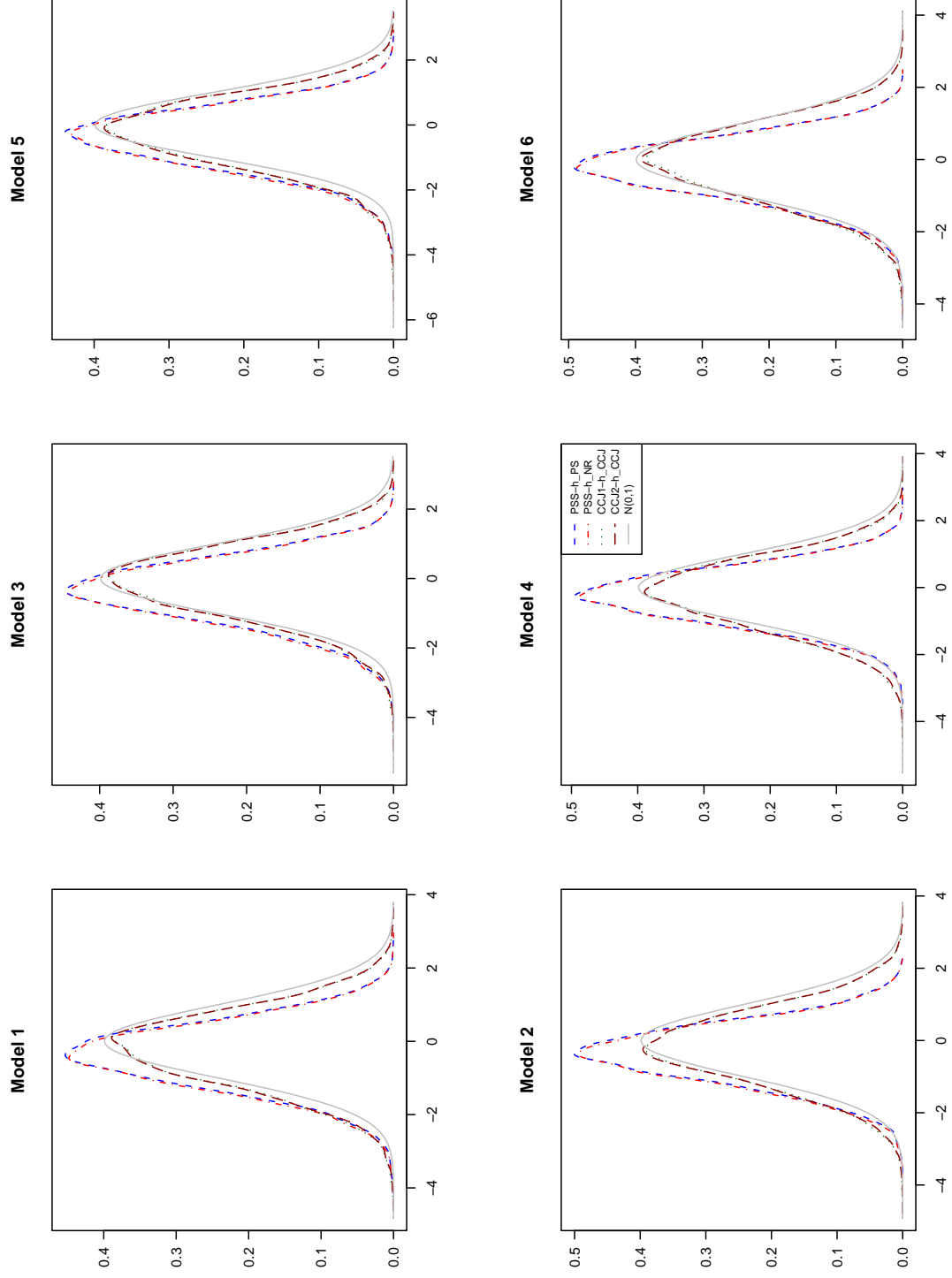
Figure 12: Empirical Gaussian Approximation with Population Bandwidth : $P = 4$, $d = 4$, $n = 400$

Table 1: Empirical Coverage Rates of 95% Confidence Intervals: $P = 2$ and $d = 2$.

		Model 1				Model 3				Model 5			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.345	0.928	0.854	0.851	0.367	0.929	0.858	0.853	0.365	0.906	0.844	0.842
	h_{NR}^*	0.345	0.928	0.854	0.851	0.367	0.929	0.858	0.853	0.365	0.906	0.844	0.842
	h_{CCJ}^*	0.192	0.991	0.928	0.943	0.175	0.994	0.933	0.949	0.199	0.988	0.914	0.934
	\hat{h}_{PS}	0.327	0.884	0.808	0.805	0.330	0.897	0.815	0.810	0.331	0.888	0.823	0.822
	\hat{h}_{NR}	0.327	0.884	0.808	0.805	0.330	0.897	0.815	0.810	0.331	0.888	0.823	0.822
	\hat{h}_{CCJ}	0.182	0.971	0.916	0.927	0.198	0.972	0.906	0.924	0.194	0.968	0.900	0.919
$n = 400$	h_{PS}^*	0.244	0.931	0.878	0.876	0.260	0.939	0.887	0.881	0.258	0.929	0.885	0.880
	h_{NR}^*	0.244	0.931	0.878	0.876	0.260	0.939	0.887	0.881	0.258	0.929	0.885	0.880
	h_{CCJ}^*	0.121	0.994	0.948	0.952	0.110	0.995	0.947	0.954	0.125	0.993	0.947	0.951
	\hat{h}_{PS}	0.248	0.870	0.817	0.809	0.255	0.883	0.819	0.809	0.252	0.887	0.833	0.823
	\hat{h}_{NR}	0.248	0.870	0.817	0.809	0.255	0.883	0.819	0.809	0.252	0.887	0.833	0.823
	\hat{h}_{CCJ}	0.113	0.980	0.937	0.940	0.132	0.976	0.932	0.932	0.120	0.981	0.938	0.941
$n = 700$	h_{PS}^*	0.212	0.940	0.892	0.889	0.226	0.942	0.895	0.889	0.224	0.936	0.900	0.895
	h_{NR}^*	0.212	0.940	0.892	0.889	0.226	0.942	0.895	0.889	0.224	0.936	0.900	0.895
	h_{CCJ}^*	0.100	0.993	0.949	0.952	0.091	0.994	0.951	0.952	0.104	0.993	0.947	0.949
	\hat{h}_{PS}	0.220	0.871	0.820	0.811	0.225	0.886	0.834	0.825	0.224	0.892	0.846	0.838
	\hat{h}_{NR}	0.220	0.871	0.820	0.811	0.225	0.886	0.834	0.825	0.224	0.892	0.846	0.838
	\hat{h}_{CCJ}	0.095	0.984	0.939	0.941	0.108	0.982	0.944	0.946	0.098	0.985	0.947	0.949
		Model 2				Model 4				Model 6			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.227	0.972	0.910	0.911	0.243	0.977	0.916	0.918	0.279	0.957	0.895	0.900
	h_{NR}^*	0.227	0.972	0.910	0.911	0.243	0.977	0.916	0.918	0.279	0.957	0.895	0.900
	h_{CCJ}^*	0.128	0.992	0.935	0.946	0.148	0.993	0.932	0.946	0.119	0.995	0.930	0.949
	\hat{h}_{PS}	0.278	0.864	0.789	0.775	0.282	0.908	0.835	0.820	0.284	0.919	0.838	0.840
	\hat{h}_{NR}	0.278	0.864	0.789	0.775	0.282	0.908	0.835	0.820	0.284	0.919	0.838	0.840
	\hat{h}_{CCJ}	0.167	0.961	0.898	0.905	0.182	0.968	0.907	0.915	0.170	0.976	0.919	0.934
$n = 400$	h_{PS}^*	0.161	0.970	0.921	0.916	0.172	0.978	0.935	0.931	0.197	0.968	0.920	0.919
	h_{NR}^*	0.161	0.970	0.921	0.916	0.172	0.978	0.935	0.931	0.197	0.968	0.920	0.919
	h_{CCJ}^*	0.081	0.994	0.944	0.946	0.093	0.993	0.947	0.949	0.074	0.995	0.946	0.950
	\hat{h}_{PS}	0.201	0.858	0.796	0.780	0.208	0.903	0.851	0.838	0.212	0.920	0.860	0.854
	\hat{h}_{NR}	0.201	0.858	0.796	0.780	0.208	0.903	0.851	0.838	0.212	0.920	0.860	0.854
	\hat{h}_{CCJ}	0.104	0.972	0.916	0.919	0.119	0.973	0.929	0.930	0.105	0.986	0.943	0.946
$n = 700$	h_{PS}^*	0.140	0.970	0.919	0.914	0.150	0.977	0.934	0.932	0.172	0.969	0.927	0.926
	h_{NR}^*	0.140	0.970	0.919	0.914	0.150	0.977	0.934	0.932	0.172	0.969	0.927	0.926
	h_{CCJ}^*	0.067	0.993	0.946	0.947	0.077	0.994	0.951	0.954	0.062	0.994	0.951	0.952
	\hat{h}_{PS}	0.178	0.849	0.796	0.783	0.182	0.902	0.849	0.836	0.185	0.922	0.869	0.862
	\hat{h}_{NR}	0.178	0.849	0.796	0.783	0.182	0.902	0.849	0.836	0.185	0.922	0.869	0.862
	\hat{h}_{CCJ}	0.088	0.973	0.926	0.925	0.100	0.978	0.938	0.936	0.085	0.989	0.951	0.952

Note: Column BW reports population bandwidths and sample mean of estimated bandwidths, respectively.

Table 2: Empirical Coverage Rates of 95% Confidence Intervals: $P = 2$ and $d = 4$.

		Model 1				Model 3				Model 5			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.410	0.900	0.788	0.796	0.440	0.888	0.783	0.788	0.431	0.876	0.764	0.773
	h_{NR}^*	0.393	0.921	0.815	0.828	0.422	0.912	0.811	0.817	0.414	0.898	0.790	0.804
	h_{CCJ}^*	0.300	0.977	0.903	0.922	0.252	0.976	0.898	0.939	0.312	0.968	0.882	0.908
	\hat{h}_{PS}	0.297	0.948	0.856	0.882	0.297	0.950	0.865	0.903	0.297	0.944	0.850	0.884
	\hat{h}_{NR}	0.286	0.957	0.869	0.898	0.285	0.955	0.876	0.917	0.285	0.950	0.859	0.898
	\hat{h}_{CCJ}	0.218	0.971	0.912	0.950	0.222	0.943	0.879	0.944	0.224	0.955	0.887	0.935
$n = 400$	h_{PS}^*	0.311	0.926	0.820	0.820	0.333	0.920	0.826	0.822	0.327	0.910	0.811	0.813
	h_{NR}^*	0.298	0.945	0.847	0.853	0.319	0.940	0.851	0.852	0.314	0.932	0.838	0.846
	h_{CCJ}^*	0.213	0.992	0.940	0.946	0.174	0.995	0.948	0.952	0.220	0.990	0.929	0.936
	\hat{h}_{PS}	0.230	0.960	0.886	0.893	0.230	0.970	0.903	0.909	0.230	0.967	0.893	0.904
	\hat{h}_{NR}	0.221	0.967	0.900	0.906	0.221	0.975	0.914	0.921	0.220	0.973	0.905	0.916
	\hat{h}_{CCJ}	0.151	0.988	0.955	0.962	0.152	0.989	0.957	0.974	0.155	0.989	0.953	0.966
$n = 700$	h_{PS}^*	0.278	0.932	0.830	0.832	0.298	0.926	0.831	0.826	0.292	0.922	0.831	0.832
	h_{NR}^*	0.267	0.950	0.860	0.862	0.286	0.945	0.859	0.859	0.281	0.943	0.858	0.861
	h_{CCJ}^*	0.185	0.992	0.941	0.944	0.155	0.996	0.948	0.951	0.192	0.992	0.938	0.944
	\hat{h}_{PS}	0.205	0.963	0.892	0.896	0.207	0.971	0.907	0.910	0.207	0.971	0.905	0.911
	\hat{h}_{NR}	0.197	0.970	0.905	0.910	0.199	0.977	0.916	0.920	0.199	0.976	0.916	0.921
	\hat{h}_{CCJ}	0.130	0.992	0.956	0.962	0.132	0.992	0.961	0.970	0.135	0.991	0.956	0.968

		Model 2				Model 4				Model 6			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.325	0.946	0.844	0.859	0.314	0.960	0.871	0.892	0.339	0.946	0.848	0.868
	h_{NR}^*	0.312	0.955	0.862	0.878	0.302	0.966	0.878	0.902	0.326	0.952	0.858	0.878
	h_{CCJ}^*	0.207	0.986	0.919	0.946	0.223	0.966	0.892	0.939	0.197	0.972	0.903	0.948
	\hat{h}_{PS}	0.256	0.953	0.866	0.891	0.256	0.953	0.876	0.918	0.257	0.958	0.875	0.914
	\hat{h}_{NR}	0.246	0.959	0.877	0.904	0.246	0.955	0.881	0.927	0.247	0.960	0.883	0.923
	\hat{h}_{CCJ}	0.192	0.968	0.908	0.951	0.197	0.932	0.868	0.939	0.201	0.957	0.889	0.945
$n = 400$	h_{PS}^*	0.246	0.962	0.874	0.877	0.238	0.980	0.913	0.915	0.257	0.970	0.894	0.898
	h_{NR}^*	0.237	0.971	0.894	0.895	0.228	0.983	0.920	0.924	0.247	0.976	0.904	0.910
	h_{CCJ}^*	0.146	0.995	0.946	0.952	0.157	0.996	0.946	0.953	0.136	0.996	0.949	0.960
	\hat{h}_{PS}	0.190	0.969	0.904	0.911	0.189	0.980	0.917	0.926	0.191	0.980	0.916	0.927
	\hat{h}_{NR}	0.182	0.974	0.913	0.922	0.181	0.984	0.926	0.937	0.183	0.984	0.925	0.938
	\hat{h}_{CCJ}	0.128	0.990	0.957	0.969	0.130	0.988	0.959	0.981	0.135	0.989	0.957	0.976
$n = 700$	h_{PS}^*	0.220	0.963	0.874	0.875	0.213	0.984	0.925	0.925	0.230	0.975	0.902	0.904
	h_{NR}^*	0.212	0.972	0.894	0.894	0.204	0.986	0.930	0.932	0.221	0.980	0.912	0.914
	h_{CCJ}^*	0.127	0.995	0.954	0.957	0.137	0.995	0.949	0.954	0.121	0.997	0.949	0.954
	\hat{h}_{PS}	0.169	0.970	0.908	0.910	0.168	0.982	0.925	0.930	0.168	0.982	0.925	0.930
	\hat{h}_{NR}	0.162	0.975	0.917	0.922	0.161	0.986	0.932	0.936	0.161	0.985	0.931	0.938
	\hat{h}_{CCJ}	0.110	0.992	0.964	0.967	0.111	0.992	0.968	0.979	0.114	0.993	0.963	0.973

Note: Column BW reports population bandwidths and sample mean of estimated bandwidths, respectively.

Table 3: Empirical Coverage Rates of 95% Confidence Intervals: $P = 4$ and $d = 2$.

		Model 1				Model 3				Model 5			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.592	0.949	0.909	0.901	0.607	0.951	0.907	0.899	0.614	0.933	0.900	0.892
	h_{NR}^*	0.627	0.936	0.898	0.885	0.643	0.939	0.898	0.884	0.650	0.924	0.892	0.880
	h_{CCJ}^*	0.442	0.979	0.930	0.936	0.439	0.984	0.932	0.938	0.450	0.971	0.919	0.926
	\hat{h}_{PS}	0.373	0.980	0.905	0.917	0.374	0.982	0.903	0.921	0.374	0.970	0.897	0.911
	\hat{h}_{NR}	0.395	0.976	0.903	0.912	0.397	0.978	0.902	0.914	0.397	0.966	0.896	0.907
	\hat{h}_{CCJ}	0.271	0.992	0.928	0.948	0.284	0.991	0.918	0.940	0.280	0.988	0.916	0.940
$n = 400$	h_{PS}^*	0.470	0.949	0.926	0.920	0.483	0.951	0.930	0.921	0.488	0.941	0.925	0.918
	h_{NR}^*	0.498	0.940	0.920	0.912	0.512	0.943	0.925	0.912	0.517	0.935	0.918	0.910
	h_{CCJ}^*	0.335	0.978	0.942	0.943	0.333	0.981	0.945	0.945	0.342	0.975	0.940	0.941
	\hat{h}_{PS}	0.290	0.978	0.921	0.924	0.290	0.980	0.922	0.923	0.290	0.979	0.923	0.926
	\hat{h}_{NR}	0.308	0.975	0.921	0.922	0.307	0.977	0.921	0.921	0.308	0.975	0.921	0.922
	\hat{h}_{CCJ}	0.187	0.993	0.949	0.953	0.198	0.994	0.948	0.954	0.192	0.995	0.949	0.954
$n = 700$	h_{PS}^*	0.428	0.951	0.930	0.925	0.438	0.950	0.934	0.927	0.444	0.948	0.933	0.927
	h_{NR}^*	0.453	0.942	0.924	0.916	0.464	0.944	0.930	0.920	0.471	0.941	0.928	0.921
	h_{CCJ}^*	0.299	0.978	0.946	0.946	0.298	0.978	0.943	0.942	0.306	0.976	0.944	0.945
	\hat{h}_{PS}	0.261	0.980	0.923	0.924	0.261	0.981	0.931	0.930	0.260	0.978	0.925	0.926
	\hat{h}_{NR}	0.276	0.976	0.923	0.922	0.277	0.976	0.929	0.928	0.276	0.974	0.925	0.926
	\hat{h}_{CCJ}	0.161	0.993	0.952	0.956	0.171	0.994	0.953	0.955	0.164	0.994	0.952	0.954
		Model 2				Model 4				Model 6			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.409	0.966	0.912	0.905	0.426	0.971	0.926	0.918	0.461	0.960	0.918	0.918
	h_{NR}^*	0.433	0.955	0.903	0.893	0.451	0.964	0.922	0.911	0.489	0.954	0.914	0.911
	h_{CCJ}^*	0.336	0.982	0.929	0.933	0.361	0.983	0.932	0.933	0.290	0.988	0.925	0.944
	\hat{h}_{PS}	0.322	0.970	0.898	0.903	0.322	0.980	0.910	0.920	0.323	0.980	0.900	0.914
	\hat{h}_{NR}	0.341	0.962	0.892	0.894	0.341	0.977	0.907	0.913	0.342	0.976	0.898	0.910
	\hat{h}_{CCJ}	0.246	0.987	0.922	0.934	0.253	0.990	0.922	0.939	0.246	0.991	0.921	0.943
$n = 400$	h_{PS}^*	0.325	0.951	0.917	0.907	0.338	0.964	0.938	0.927	0.366	0.962	0.936	0.931
	h_{NR}^*	0.344	0.940	0.909	0.897	0.358	0.958	0.933	0.922	0.388	0.956	0.931	0.926
	h_{CCJ}^*	0.254	0.977	0.940	0.939	0.273	0.982	0.945	0.943	0.220	0.990	0.945	0.949
	\hat{h}_{PS}	0.239	0.975	0.912	0.911	0.241	0.981	0.925	0.925	0.241	0.986	0.922	0.925
	\hat{h}_{NR}	0.254	0.967	0.908	0.906	0.255	0.976	0.925	0.921	0.256	0.981	0.919	0.921
	\hat{h}_{CCJ}	0.166	0.991	0.942	0.945	0.175	0.993	0.943	0.948	0.164	0.995	0.951	0.958
$n = 700$	h_{PS}^*	0.296	0.947	0.916	0.909	0.307	0.964	0.937	0.931	0.333	0.962	0.936	0.934
	h_{NR}^*	0.313	0.937	0.907	0.896	0.326	0.956	0.935	0.926	0.353	0.956	0.934	0.929
	h_{CCJ}^*	0.227	0.976	0.937	0.936	0.245	0.980	0.947	0.944	0.195	0.990	0.950	0.952
	\hat{h}_{PS}	0.213	0.970	0.918	0.916	0.212	0.983	0.928	0.928	0.212	0.984	0.927	0.928
	\hat{h}_{NR}	0.225	0.963	0.915	0.913	0.225	0.980	0.928	0.924	0.225	0.982	0.926	0.927
	\hat{h}_{CCJ}	0.143	0.991	0.948	0.949	0.151	0.991	0.947	0.950	0.138	0.995	0.950	0.954

Note: Column BW reports population bandwidths and sample mean of estimated bandwidths, respectively.

Table 4: Empirical Coverage Rates of 95% Confidence Intervals: $P = 4$ and $d = 4$.

		Model 1				Model 3				Model 5			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.679	0.951	0.878	0.882	0.705	0.942	0.875	0.875	0.703	0.928	0.858	0.863
	h_{NR}^*	0.693	0.946	0.873	0.875	0.719	0.935	0.867	0.864	0.717	0.919	0.850	0.853
	h_{CCJ}^*	0.575	0.977	0.913	0.925	0.528	0.983	0.913	0.931	0.589	0.966	0.896	0.912
	\hat{h}_{PS}	0.374	0.987	0.908	0.936	0.374	0.979	0.902	0.943	0.373	0.980	0.898	0.934
	\hat{h}_{NR}	0.382	0.986	0.907	0.933	0.381	0.980	0.902	0.938	0.381	0.980	0.896	0.932
	\hat{h}_{CCJ}	0.324	0.989	0.925	0.957	0.332	0.974	0.905	0.958	0.330	0.980	0.908	0.953
	h_{PS}^*	0.556	0.955	0.905	0.902	0.579	0.947	0.899	0.892	0.578	0.946	0.899	0.895
	h_{NR}^*	0.567	0.949	0.897	0.894	0.591	0.941	0.894	0.884	0.590	0.940	0.892	0.888
	h_{CCJ}^*	0.456	0.984	0.936	0.939	0.422	0.990	0.940	0.943	0.468	0.980	0.934	0.939
$n = 400$	\hat{h}_{PS}	0.291	0.994	0.931	0.938	0.291	0.994	0.937	0.944	0.290	0.995	0.932	0.941
	\hat{h}_{NR}	0.297	0.993	0.929	0.936	0.297	0.994	0.934	0.942	0.296	0.994	0.930	0.939
	\hat{h}_{CCJ}	0.233	0.996	0.953	0.961	0.237	0.998	0.955	0.966	0.239	0.996	0.953	0.965
	h_{PS}^*	0.514	0.955	0.914	0.911	0.535	0.952	0.912	0.904	0.532	0.949	0.910	0.907
	h_{NR}^*	0.524	0.950	0.908	0.904	0.546	0.948	0.906	0.896	0.543	0.943	0.906	0.901
	h_{CCJ}^*	0.416	0.986	0.938	0.941	0.387	0.990	0.944	0.946	0.427	0.982	0.937	0.941
	\hat{h}_{PS}	0.260	0.993	0.931	0.936	0.262	0.994	0.936	0.941	0.262	0.994	0.930	0.938
	\hat{h}_{NR}	0.266	0.992	0.928	0.934	0.268	0.994	0.934	0.939	0.267	0.994	0.928	0.937
	\hat{h}_{CCJ}	0.203	0.996	0.953	0.958	0.208	0.998	0.956	0.961	0.209	0.996	0.956	0.961
		Model 2				Model 4				Model 6			
		BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2	BW	PSS	CCJ1	CCJ2
$n = 100$	h_{PS}^*	0.513	0.966	0.893	0.904	0.516	0.969	0.895	0.908	0.545	0.963	0.886	0.902
	h_{NR}^*	0.523	0.963	0.889	0.897	0.527	0.966	0.892	0.903	0.557	0.961	0.881	0.899
	h_{CCJ}^*	0.423	0.983	0.917	0.932	0.454	0.980	0.909	0.924	0.386	0.983	0.911	0.937
	\hat{h}_{PS}	0.323	0.987	0.911	0.937	0.322	0.975	0.903	0.950	0.322	0.980	0.906	0.944
	\hat{h}_{NR}	0.330	0.986	0.909	0.933	0.328	0.976	0.903	0.947	0.329	0.981	0.905	0.941
	\hat{h}_{CCJ}	0.286	0.987	0.921	0.956	0.293	0.971	0.902	0.957	0.293	0.978	0.906	0.956
	h_{PS}^*	0.421	0.970	0.904	0.903	0.423	0.979	0.923	0.922	0.447	0.974	0.916	0.919
	h_{NR}^*	0.429	0.965	0.898	0.897	0.432	0.977	0.922	0.919	0.456	0.971	0.914	0.914
	h_{CCJ}^*	0.334	0.990	0.938	0.939	0.361	0.989	0.938	0.939	0.302	0.993	0.943	0.947
$n = 400$	\hat{h}_{PS}	0.240	0.994	0.935	0.942	0.240	0.996	0.939	0.946	0.240	0.995	0.934	0.945
	\hat{h}_{NR}	0.245	0.993	0.934	0.939	0.245	0.996	0.936	0.943	0.245	0.995	0.931	0.942
	\hat{h}_{CCJ}	0.199	0.997	0.954	0.964	0.201	0.997	0.959	0.971	0.203	0.998	0.953	0.969
	h_{PS}^*	0.388	0.964	0.902	0.900	0.390	0.978	0.925	0.921	0.413	0.975	0.923	0.924
	h_{NR}^*	0.396	0.959	0.896	0.894	0.399	0.975	0.922	0.917	0.422	0.972	0.921	0.920
	h_{CCJ}^*	0.304	0.989	0.940	0.941	0.328	0.989	0.944	0.943	0.279	0.993	0.944	0.946
	\hat{h}_{PS}	0.213	0.994	0.941	0.944	0.213	0.996	0.938	0.943	0.213	0.996	0.935	0.940
	\hat{h}_{NR}	0.217	0.994	0.938	0.940	0.217	0.995	0.935	0.940	0.217	0.995	0.933	0.937
	\hat{h}_{CCJ}	0.171	0.997	0.957	0.963	0.174	0.998	0.958	0.966	0.176	0.998	0.957	0.964

Note: Column BW reports population bandwidths and sample mean of estimated bandwidths, respectively.

- 2009-33: Peter Christoffersen, Redouane Elkamhi, Bruno Feunou and Kris Jacobs: Option Valuation with Conditional Heteroskedasticity and Non-Normality
- 2009-34: Peter Christoffersen, Steven Heston and Kris Jacobs: The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well
- 2009-35: Peter Christoffersen, Jeremy Berkowitz and Denis Pelletier: Evaluating Value-at-Risk Models with Desk-Level Data
- 2009-36: Tom Engsted and Thomas Q. Pedersen: The dividend-price ratio does predict dividend growth: International evidence
- 2009-37: Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis
- 2009-38: Frank S. Nielsen: Local Whittle estimation of multivariate fractionally integrated processes
- 2009-39: Borus Jungbacker, Siem Jan Koopman and Michel van der Wel: Dynamic Factor Models with Smooth Loadings for Analyzing the Term Structure of Interest Rates
- 2009-40: Niels Haldrup, Antonio Montañés and Andreu Sansó: Detection of additive outliers in seasonal time series
- 2009-41: Dennis Kristensen: Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models
- 2009-42: Ole Eiler Barndorff-Nielsen and Robert Stelzer: The multivariate supOU stochastic volatility model
- 2009-43: Lasse Bork, Hans Dewachter and Romain Houssa: Identification of Macroeconomic Factors in Large Panels
- 2009-44: Dennis Kristensen: Semiparametric Modelling and Estimation: A Selective Overview
- 2009-45: Kim Christensen, Silja Kinnebrock and Mark Podolskij: Pre-averaging estimators of the ex-post covariance matrix
- 2009-46: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Robust Data-Driven Inference for Density-Weighted Average Derivatives