SUPPLEMENT TO "DATA-DRIVEN BANDWIDTH SELECTION FOR NONPARAMETRIC NONSTATIONARY REGRESSIONS"

By Federico M. Bandi, Valentina Corradi and Daniel Wilhelm

Johns Hopkins University and Edhec-Risk, University of Warwick and University of Chicago

This supplementary document provides the details and results of a simulation experiment illustrating the performance of the bandwidth selection procedure proposed in the main text.

1. Simulations. In this section, we report a simulation experiment which applies our bandwidth selection procedure as well as the bias correction, both proposed in the main text, and illustrates their finite sample performance. Three different data generating processes are considered. Model I and model III, below, have been simulated in influential recent work on inference for cointegrating regressions (Karlsen and Tjøstheim (2001) and Wang and Phillips (2009)). Model II is a discrete-time counterpart of a popular, in the literature, specification in continuous time (i.e., a square root diffusion). We experimented with an array of different parameter values. The values that are reported are representative of our findings.

Model I As an example of a nonstationary autoregression, as in Karlsen and Tjøstheim (2001) we simulate a unit root process ($\mu(x) = x$ and $\sigma(x) = 1$). We choose $x_0 = 0$, $\mathcal{D}_x = [-5, 5]$ and let u_t be iid $N(0, 1)$.

Model II The discrete-time square-root process is an autoregression with $\mu(x) = (1 - \phi)\theta + \phi x$ and $\sigma(x) = \sigma\sqrt{|x|}$ whose parameters are chosen to be $\theta = 1$, $\phi = 0.8$, $\sigma = 1$ and $\mathcal{D}_x = [0, 4]$. We start the process at its unconditional mean $x_0 = \theta$ and, again, u_t is iid $N(0, 1)$.

Model III To illustrate our procedure in the case of cointegrating regressions, we consider a simulation design similar to the one in Hall and Horowitz (2005) and Wang and Phillips (2009) which specifies $f(x) = \sum_{j=1}^{4}(-1)^{j+1}j^{-2}\sin(j\pi x)$ and $a(x) = 1$, viz. $Y_t = f(X_t) + \epsilon_t$, $X_t = X_{t-1} + u_t$ and $\epsilon_t = \frac{\eta_t + \theta u_t}{\sqrt{1+\theta^2}}$. $(u_t, \epsilon_t, \eta_t)'$ are iid $N(0, I_3)$, I_3 a diagonal matrix of ones, $x_0 = 0$ and $\mathcal{D}_x = [0, 1]$. We consider two scenarios: no

AMS 2000 subject classifications: Primary 62M05, 62M10; secondary 62G05

Keywords and phrases: Data-driven Bandwidth Selection, Nonstationary Autoregression, Nonstationary Cointegration, Recurrence

endogeneity ($\theta = 0$) and strong endogeneity ($\theta = 2$).

To summarize, there are four simulation scenarios: model I, model II, and two versions of model III, each of which is estimated using our point-wise and uniform criteria for selecting the bandwidths. Even though cross-validation has not been formally justified in a nonstationary framework, it is the classical paradigm in empirical work and we therefore consider it here as an important benchmark.

1.1. *Implementation Details.* The conditional moments impose the same requirements on the rate of divergence of the relevant bandwidth sequences. However, the optimization to find

$$h_n(x) = (h_n^\mu(x), h_n^\sigma(x)) = \arg \inf_{h_n \in \mathcal{H}_n(x,\varsigma)} \|\hat{m}_{n,h_n}(x)\|,$$

in the point-wise case and

$$h_n = (h_n^\mu, h_n^\sigma) = \arg \inf_{h_n \in \mathcal{H}_n(\varsigma)} \sup_{x \in \mathcal{D}_x} \|\hat{m}_{n,h_n}(x)\|,$$

in the uniform case is performed with separate bandwidths for the first and the second conditional moment in order to improve finite sample accuracy. Specifically, we implement a search over a grid of 5×5 bandwidths on $[0.01, 1]^2$. The bias correction is instead implemented by virtue of a search over a 100×100 grid on $[0.01, 10]^2$. The supremum over x in the uniform criterion is calculated over a grid of five equally spaced points in $\mathcal{D}_x$. For the point-wise criterion, $\mathcal{D}_x$ is partitioned into five parts of equal size. Five bandwidths are calculated at the center of each of the five subsets of $\mathcal{D}_x$. Since determining the partition depends on the path and introduces extraneous randomness, we choose it to be the same for every simulated path which, in turn, creates issues which are, admittedly, little understood in the literature. For example, it could be the case that a given simulated path does not visit a certain region of the domain at all, or only very few times, so that estimation of a function in that region can be based only on certain paths, but not on all. To minimize these effects, we restrict estimation of the various functions to areas near the processes' points of initialization and, thus, all paths take values in at least some portion of those regions.

The remaining parameters are $R = 200$ and uniform weights $\pi(u) = 1$ over the interval $U = [2, 3]$. Throughout the experiment we use the Tukey-Hanning kernel. The second-stage tests are performed at the 95% confidence level. All results are based on $1,000$ Monte Carlo samples of length 500.

1.2. *Results.* Tables 1 and 2 report the selected bandwidths for models I-III calculated using our point-wise and uniform procedures as well as cross-validation ("CV"). We emphasize that the second-step cross-validated bandwidths have been obtained by applying our bias correction to the original cross-validated bandwidths. In other words, the bias correction is applied *both* to our selected bandwidths and to the classical cross-validated bandwidths. Since the bias correction can, in principle, be applied to any bandwidth, this is simply to give the reader indications about the potential usefulness of bias correcting starting from bandwidths which may not be theoretically minimax optimal (such as the cross-validated bandwidths). Importantly, however, when evaluating the relative performance of cross-validation and our methods, being the bias correction an aspect of what we propose, one should compare the first-step cross-validated bandwidths (those that do not include a bias correction) to our bandwidths (the minimax optimal bandwidths inclusive – or not – of the bias correction).

Table 3 present the bias, standard deviation ("SD") and root mean square error ("RMSE") of the estimated functions, averaged over 20 equally-spaced points in their respective domains $\mathcal{D}_x$.

Figures 1-4 show the corresponding estimates of the first and second conditional moment functions, $\mu(x)$ and $\sigma(x)$ or $f(x)$ and $\alpha(x)$, respectively. Included in the graphs are the true line (thick blue), the line based on our uniform criterion (blue circles) and the cross-validated estimates (red squares) as well as empirical (point-wise) 95% confidence bands. The graphs corresponding to the point-wise criterion, which are similar to the reported ones, are not shown to save space.

Figures 5-8 depict the kernel density estimates of the first conditional moment estimates at the fixed points $x = 0$ for model I, $x = 2$ for model II, and $x = 0.5$ for model III. These values constitute the center of $\mathcal{D}_x$ for all three models. Specifically, we estimated the density of the centered and re-scaled quantities

$$\sqrt{\frac{\hat{h}_n^{\mu}\hat{L}_{n,\hat{h}_n^{\mu}}(x)}{K_2\sigma(x)^2}}\ \left(\hat{\mu}_{n,\hat{h}_n^{\mu}}(x)-\mu(x)\right) \quad \text{and} \quad \sqrt{\frac{\hat{h}_n^{f}\hat{L}_{n,\hat{h}_n^{f}}(x)}{K_2\alpha(x)^2}}\ \left(\hat{f}_{n,\hat{h}_n^{f}}(x)-f(x)\right),$$

respectively, where μ, σ, f and α are the true functions. Again, for brevity, graphs are only shown for the uniform criterion and the first conditional moment. The findings can be summarized as follows:

1. Our combined procedure outperforms cross-validation. In the first and in the third model, the point-wise and the uniform criteria produce comparable (relative to cross-validation), or slightly lower, RMSEs in

both stages. In these two specifications, the second conditional moment is flat and, since cross-validation tends to oversmooth in these models, these are scenarios in favor of a uniform criterion like cross-validation. In model II, however, the nonlinear second conditional moment of the process reveals a dramatic difference in relative performance. The bandwidths selected by cross-validation are much too small leading to a large variance of the resulting estimates and an RMSE which is more than twice as large as the ones produced by our combined procedure.

2. As discussed, the proposed bandwidth procedure optimally balances the estimators' biases and variances. This may, of course, be achieved by choosing relatively large bandwidths h_n^μ and h_n^σ which have the potential to cause some oversmoothing (see, e.g., model III). The reported bias correction is designed to address this issue explicitly since it forces the bandwidths to also satisfy the conditions $(h_n^\mu)^5 \hat{L}_{n,h_n^\mu}(x) \stackrel{a.s.}{\to} 0$ and $(h_n^\sigma)^5 \hat{L}_{n,h_n^\sigma}(x) \stackrel{a.s.}{\to} 0$, which are necessary for a vanishing limiting bias. These conditions require both bandwidths to be small enough. Tables 1 and 2 show significant reductions in the size of the bandwidths after the second-stage procedure is applied. This effect can also be seen by inspecting Figures 5-8 which show that the bias correction successfully re-adjust the distribution of the first moment estimator towards the normal distribution – in model III strikingly so. Consistent with theory, this bias reduction is generally accompanied by increases in the estimators' MSE, particularly when moving away from the minimax optimal solution. Hence, as emphasized, if MSE minimization is the criterion of interest, one should simply find the minimax optimal bandwidths. If a zero bias is the criterion, this goal can be achieved by appropriately reducing the size of the optimal bandwidths at the cost of larger statistical uncertainty.
3. The properties of cross-validated bandwidths in nonstationary frameworks are unknown. However, the results in this section suggest that they may not necessarily perform poorly in such scenarios (see models I and III). Importantly, however, if cross-validated smoothing sequences are used in practice, in light of their tendency to oversmooth, we find that their performance can be further enhanced by applying to them our proposed bias correction.
4. Table 2 and the lower half of Table 3 as well as Figures 3-4, 7-8 all confirm our theoretical results on cointegrating regressions, namely that – whether the regressor and the error are independent or not – the distributions of the first and second conditional moment estimates conform with a zero-mean normal distribution after applying our com-

bined procedure. The results show no difference in performance with or without dependence. Since the presence of this type of endogeneity is common in empirical work, this is an important feature of our proposed method.

REMARK 1. *The fact that, in model II, the second stage adjusts the average cross-validation bandwidths from about* 0.3 *down to about* 0.15 *while some of the average point-wise and uniform bandwidths are not rejected at levels of about* 0.3 *may seem puzzling at first glance. This effect is due to the large variability in the bandwidths selected by cross-validation: it mostly chooses bandwidths much smaller than* 0.3*, but also some huge ones (reflected in the large standard deviation of the first stage). The second stage does not reject the former, but adjusts downwards the latter to values around* 0.3*, which in turn yields an average bandwidth smaller than* 0.3*. On the other hand, the uniform and point-wise criteria tend to select bandwidths between* 0.4 *and* 0.7 *with a small standard deviation so that the second step decreases most of them down to values near* 0.3 *leading to an average of that order of magnitude.*

		MODEL I			
		optimal		bias correction	
		bandwidth	SD	bandwidth	SD
pointwise	h^μ	0.5817	0.3585	0.3253	0.1793
		0.6396	0.3063	0.3624	0.1393
		0.6216	0.3198	0.3617	0.1571
		0.6510	0.3191	0.3666	0.1632
		0.5736	0.3597	0.3210	0.1755
	h^σ	0.6008	0.3541	0.3323	0.1752
		0.6654	0.3051	0.3689	0.1370
		0.5622	0.2974	0.3451	0.1276
		0.6728	0.3111	0.3690	0.1503
		0.5916	0.3547	0.3349	0.1850
uniform	h^μ	0.6971	0.3167	0.3705	0.1604
	h^σ	0.6495	0.3541	0.3463	0.1782
CV	h^μ	0.7634	0.4197	0.3333	0.2199
	h^σ	0.7582	0.4186	0.3295	0.2131
		MODEL II			
		optimal		bias correction	
		bandwidth	SD	bandwidth	SD
pointwise	h^μ	0.4236	0.4429	0.2115	0.2182
		0.4399	0.2109	0.3349	0.1186
		0.6159	0.2429	0.3849	0.1073
		0.7008	0.2626	0.3911	0.1206
		0.7208	0.2609	0.3997	0.1314
	h^σ	0.3956	0.2849	0.2849	0.1340
		0.4372	0.2304	0.3295	0.1214
		0.6010	0.2499	0.3778	0.0988
		0.6976	0.2587	0.3915	0.1005
		0.7334	0.2542	0.3970	0.1039
uniform	h^μ	0.7567	0.2931	0.3933	0.1449
	h^σ	0.5565	0.2344	0.3695	0.0958
CV	h^μ	0.3367	0.4592	0.1494	0.2068
	h^σ	0.3360	0.4584	0.1487	0.2057

TABLE 1
Selected bandwidths and their standard deviation ("SD").

		MODEL III ($\theta = 0$)			
		optimal		bias correction	
		bandwidth	SD	bandwidth	SD
pointwise	h^f	0.5612	0.3124	0.3468	0.1419
		0.5345	0.2678	0.3619	0.1219
		0.5072	0.2812	0.3463	0.1176
		0.5637	0.3149	0.3477	0.1198
		0.6050	0.3000	0.3628	0.1294
	h^a	0.5815	0.3183	0.3516	0.1394
		0.5939	0.2931	0.3574	0.0956
		0.5602	0.2793	0.3617	0.1216
		0.6189	0.2948	0.3701	0.1155
		0.6929	0.3167	0.3689	0.1289
uniform	h^f	0.6292	0.3033	0.3731	0.1419
	h^a	0.7389	0.3200	0.3790	0.1693
CV	h^f	0.5622	0.3533	0.3165	0.1897
	h^a	0.7493	0.4218	0.3293	0.2176
		MODEL III ($\theta = 2$)			
		optimal		bias correction	
		bandwidth	SD	bandwidth	SD
pointwise	h^f	0.5372	0.3198	0.3323	0.1383
		0.5218	0.2651	0.3602	0.1310
		0.5236	0.2817	0.3493	0.1129
		0.5701	0.3134	0.3520	0.1216
		0.6161	0.3087	0.3631	0.1364
	h^a	0.5491	0.3233	0.3394	0.1452
		0.6030	0.2921	0.3616	0.1004
		0.5580	0.2781	0.3637	0.1242
		0.5981	0.2933	0.3614	0.1049
		0.6842	0.3127	0.3736	0.1369
uniform	h^f	0.6280	0.3137	0.3658	0.1448
	h^a	0.7218	0.3293	0.3703	0.1687
CV	h^f	0.5637	0.3537	0.3168	0.1901
	h^a	0.7409	0.4207	0.3293	0.2178

TABLE 2
Selected bandwidths and their standard deviation ("SD").

		MODEL I					
		optimal			bias correction		
		bias	SD	RMSE	bias	SD	RMSE
pointwise	$\mu(x)$	-0.0216	0.5101	0.5110	-0.0166	0.5575	0.5577
	$\mu^{(2)}(x)$	0.1227	2.7806	2.7886	0.1186	3.1204	3.1251
uniform	$\mu(x)$	-0.0096	0.5126	0.5141	-0.0091	0.5650	0.5652
	$\mu^{(2)}(x)$	-0.0249	2.6401	2.6446	-0.0020	3.0054	3.0065
CV	$\mu(x)$	-0.0108	0.5167	0.5202	-0.0136	0.5651	0.5658
	$\mu^{(2)}(x)$	-0.0956	2.7128	2.7204	-0.0341	3.0783	3.0801
		MODEL II					
		optimal			bias correction		
		bias	SD	RMSE	bias	SD	RMSE
pointwise	$\mu(x)$	-0.0302	0.3498	0.3518	-0.0151	0.3993	0.3996
	$\mu^{(2)}(x)$	-0.1035	1.5338	1.5433	-0.0391	1.8316	1.8349
uniform	$\mu(x)$	-0.0423	0.2775	0.2838	-0.0150	0.3386	0.3394
	$\mu^{(2)}(x)$	-0.0756	1.5944	1.6054	-0.0278	1.8256	1.8307
CV	$\mu(x)$	-0.0247	0.8542	0.8547	-0.0053	0.8725	0.8724
	$\mu^{(2)}(x)$	-0.0703	4.2487	4.2506	0.0107	4.3597	4.3593
		MODEL III ($\theta = 0$)					
		optimal			bias correction		
		bias	SD	RMSE	bias	SD	RMSE
pointwise	$f(x)$	-0.1665	0.4886	0.5269	-0.0790	0.5120	0.5226
	$f^{(2)}(x)$	0.0022	0.7907	0.8241	-0.0036	0.9280	0.9420
uniform	$f(x)$	-0.2027	0.4420	0.5028	-0.0887	0.4938	0.5077
	$f^{(2)}(x)$	-0.0193	0.6776	0.7407	-0.0070	0.8837	0.9022
CV	$f(x)$	-0.2072	0.4830	0.5445	-0.0787	0.5245	0.5357
	$f^{(2)}(x)$	-0.0068	0.8164	0.8592	0.0100	0.9862	1.0009
		MODEL III ($\theta = 2$)					
		optimal			bias correction		
		bias	SD	RMSE	bias	SD	RMSE
pointwise	$f(x)$	-0.1321	0.4927	0.5182	-0.0419	0.5243	0.5286
	$f^{(2)}(x)$	0.0582	0.8864	0.9109	0.0753	1.0145	1.0268
uniform	$f(x)$	-0.1818	0.4573	0.5062	-0.0539	0.5139	0.5202
	$f^{(2)}(x)$	0.0347	0.7721	0.8226	0.0682	0.9620	0.9774
CV	$f(x)$	-0.1845	0.4804	0.5292	-0.0523	0.5356	0.5411
	$f^{(2)}(x)$	0.0382	0.9077	0.9427	0.0778	1.0869	1.1009

TABLE 3
Average bias, standard deviation ("SD") and root mean square error ("RMSE") of the respective estimated functions.

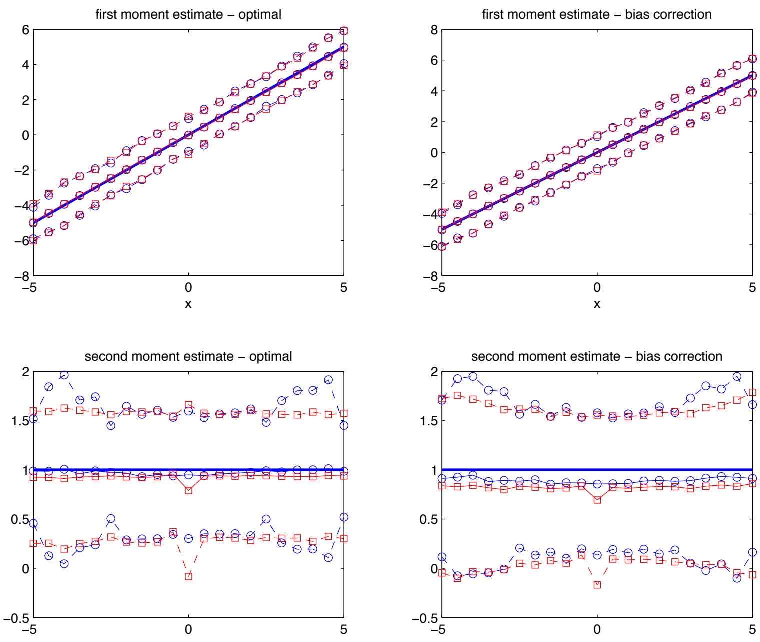

Fig 1. *Model I, estimated moments based on uniform criterion (blue circles), CV (red squares) and the true moments (thick blue lines).*

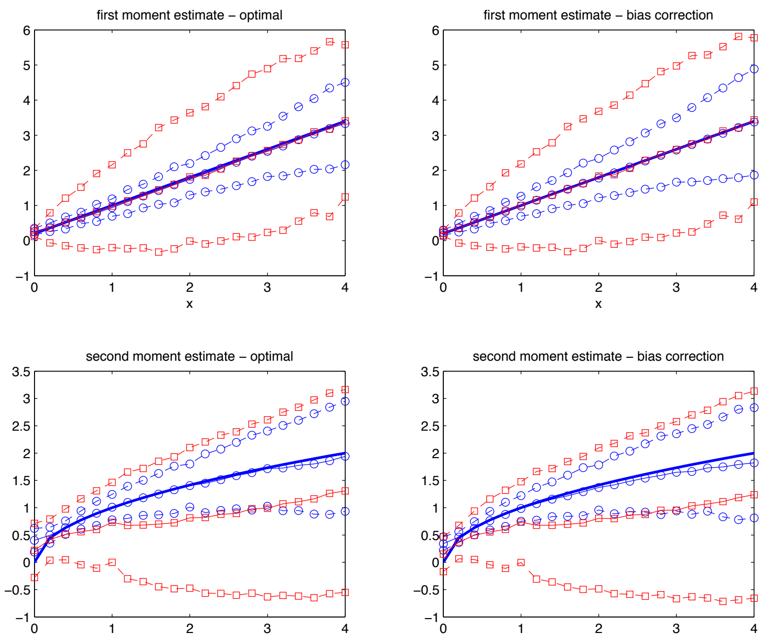

FIG 2. *Model II, estimated moments based on uniform criterion (blue circles), CV (red squares) and the true moments (thick blue lines).*

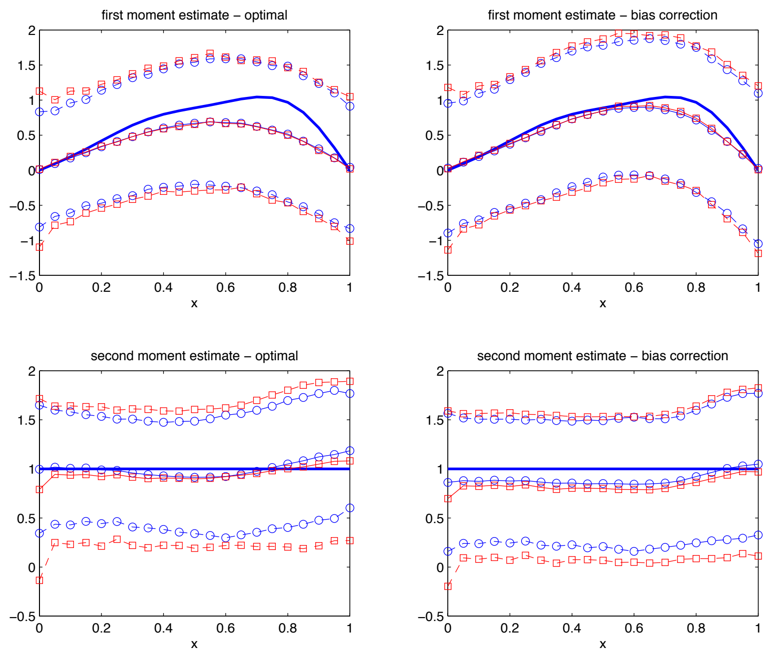

FIG 3. *Model III ($\theta = 0$), estimated moments based on uniform criterion (blue circles), CV (red squares) and the true moments (thick blue lines).*

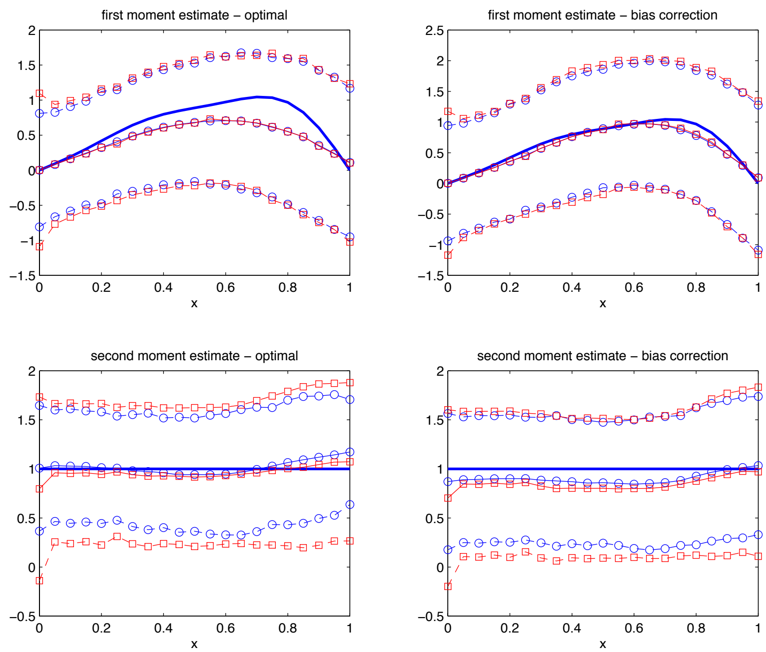

FIG 4. *Model III ($\theta = 2$), estimated moments based on uniform criterion (blue circles), CV (red squares) and the true moments (thick blue lines).*

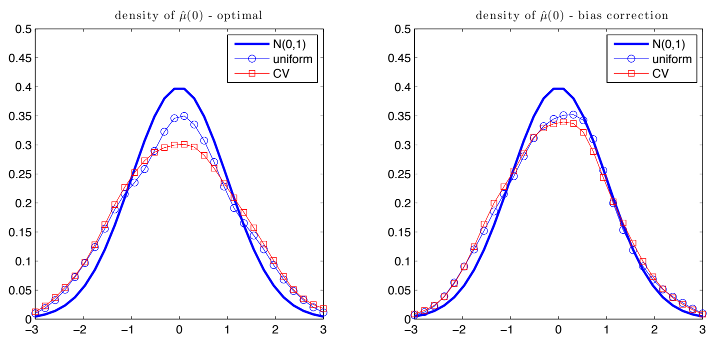

FIG 5. *Model I, distribution of the first moment estimator at $x = 0$, based on uniform bandwidths.*

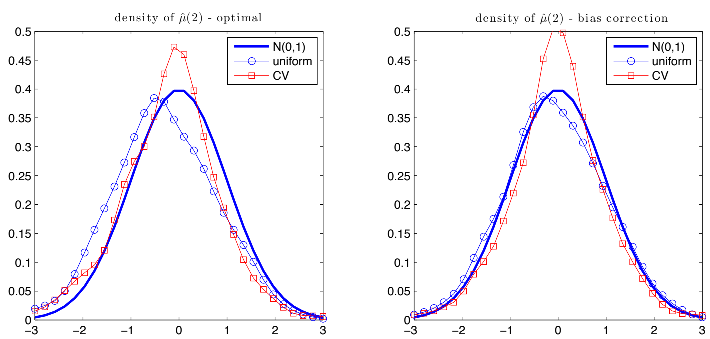

FIG 6. *Model II, distribution of the first moment estimator at $x = 2$, based on uniform bandwidths.*

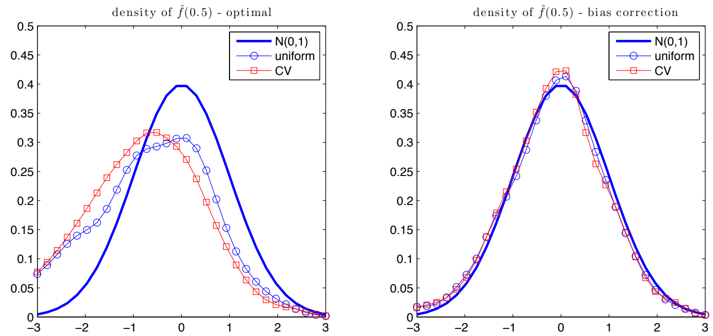

Fig 7. *Model III ($\theta = 0$), distribution of the first moment estimator at $x = 0.5$, based on uniform bandwidths.*

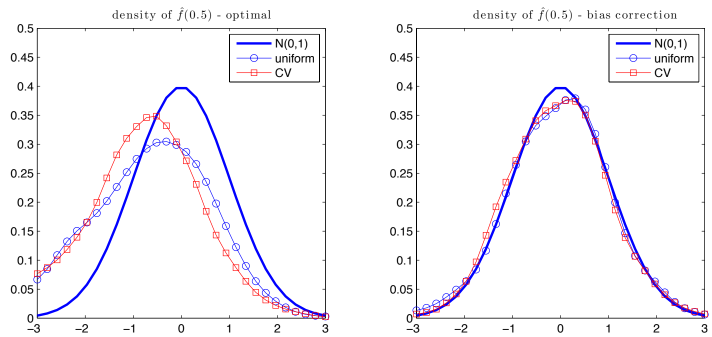

Fig 8. *Model III ($\theta = 2$), distribution of the first moment estimator at $x = 0.5$, based on uniform bandwidths.*

References.

HALL, P. and HOROWITZ, J. L. (2005). Nonparametric Methods for Inference in the Presence of Instrumental Variables. *The Annals of Statistics* **33** 2904-2929.

KARLSEN, H. A. and TJØSTHEIM, D. (2001). Nonparametric Estimation in Null Recurrent Time Series. *The Annals of Statistics* **29** 372-416.

WANG, Q. and PHILLIPS, P. C. B. (2009). Structural Nonparametric Cointegrating Regression. *Econometrica* **77** 1901-1948.

CAREY BUSINESS SCHOOL
JOHNS HOPKINS UNIVERSITY
E-MAIL: fbandi1@jhu.edu

DEPARTMENT OF ECONOMICS
UNIVERSITY OF WARWICK
E-MAIL: v.corradi@warwick.ac.uk

BOOTH SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
E-MAIL: daniel.wilhelm@uchicago.edu