Robust Inference with Higher Order Influence Functions : Part II

Lingling Li*, Eric Tchetgen*, Aad van der Vaart, James Robins*†,
Departments of Biostatistics* and Epidemiology†, Harvard University
Department of Mathematics Vrije Universiteit.

KEY WORDS Higher Order Influence functions, Weighted average treatment effect, U-statitics, Honest Confidence intervals, Nonparametric inference.

Abstract

In this paper we report the results of a simulation study designed to explore the extent to which the asymptotic results derived in our companion paper are relevant in finite samples.

1. Introduction

In our companion paper (Tchetgen et al. 2005) we introduced novel "honest" $(1-\alpha)$ asymptotic confidence intervals for the weighted average treatment effect (WATE) of a dichotomous treatment in the presence of a high dimensional vector X of confounding factors. In our companion paper we studied the asymptotic properties of our new interval estimators. In the current article, we complement these asymptotic results with an empirical investigation of the finite sample performance of our new CIs for the WATE functional. In section 2, we describe the data generating mechanism used in our simulation studies and the higher order $U-statistics$ that center our CIs. We conclude with a discussion of various theoretical and practical considerations relevant for our new methodology.

1.1 The Weighted Average Treatment Effect (WATE) Functional

We observe $i.i.d$ data $(W_i, A_i, X_i) = O_i, i = 1,...n,$ from a model $\mathcal{M}(\Theta) = \{F(O;\theta), \theta \in \Theta\}$, where W is the outcome of interest, A is a dichotomous exposure and X is a $d-$dimensional vector of continuous covariates. For a fixed scalar ρ, the functional of interest is

$$\psi(\theta) = E_\theta\left((Y - b(X))(A - \omega(X))\right) \quad (1)$$

where $Y = W - \rho A$, $b(X) = E[Y|X]$, $\omega(X) = E[A|X] = \Pr(A = 1|X)$. Denote the density of X by $g(X)$. In our simulations, we choose $\rho = 0$, so $Y = W$. In Part I, we considered a model $\mathcal{M}(\theta)$ that imposed the following known bounds on various L_∞ norms in $\mathcal{M}(\theta)$: $(a.1)$ $|b(X)| \leq K_b$ w.p.1 for some constant $K_b < \infty$, $(a.2)$ $var(Y|X) \leq K_1$ w.p.1 for some constant $K_1 < \infty$, and $(a.3)$ there exist $\delta^*, M > 0$, such that $\delta^* < g(X) < M$ w.p.1. $(a.4)$ $b(\cdot)$, $\omega(\cdot)$, and $g(\cdot)$ belong to given Holder balls $H(\beta_b, C_b), H(\beta_\omega, C_\omega), H(\beta_g, C_g)$. See our companion paper for the definition of a Holder ball. We note that the integrated mean squared error (MSE) and uniform error minimax rates of convergence for estimation of a marginal density or conditional expectation $h(\cdot) \in H(\beta_h, C_h)$ are $O\left(n^{-\frac{\beta_h/d}{2\beta_h/d+1}}\right)$ and $O\left(\left(\frac{n}{\log n}\right)^{-\frac{\beta_h/d}{2\beta_h/d+1}}\right)$, which both depend on β_h and d only through the "effective" smoothness exponent β_h/d.

Our substantive interest is in the case in which the dimension d of X is high (say 10 to 50) and our unknown functions live in Holder balls with $\beta \in (1, 20)$ and $c \in (1, 20)$. However it is much easier to conduct simulations with $d = 1$. Therefore we used $d = 1$ in all simulations, but chose β sufficiently small so that $\beta = \beta/1$ in our simulations matched the ratios β_h/d of actual substantive interest. In future work we plan to conduct simulations with d large.

2. Simulation Design

An important result in functional analysis states that a variety of functional spaces-including Holder and Besov spaces, can be characterized by a set of restrictions on the coefficients of the wavelet expansion of functions belonging to those spaces. In the following, we show how this result may be used to construct a function with a given holder exponent.

Define $\nu(x)$ as the "mother" Daubechies wavelet with Q vanishing moments (denoted $D2Q$)–a function has Q vanishing moments if $\int x^j \nu(x)\,dx = 0,\ j = 0,..,Q-1$; and $\phi(x)$ the corresponding scaling function ("father" wavelet) with support $[0, 2Q-1]$. For a given function $h(x)$, we write $h_{rk}(x) = 2^{r/2}h(2^r x - k)$ as the dilated (by 2^r) and translated (by k) $h(x)$. It is well known that $\{\phi_{0k}(x), \nu_{rk}(x) : r \in \mathbb{Z}^+, k \in \mathbb{Z}\}$ is a complete orthonormal system (ONS) for L_2; therefore, any $\kappa_0(x) \in L_2$ can be written

$$\kappa_0(x) = \sum_{k\in\mathbb{Z}} \alpha_k \phi_{0k}(x) + \sum_{r\in\mathbb{Z}^+}\sum_{k\in\mathbb{Z}} \zeta_{rk}\nu_{rk}(x) \tag{2}$$

where $\alpha_k \in \mathbb{R}, \zeta_{rk} \in \mathbb{R}, r \in \mathbb{Z}^+, k \in \mathbb{Z}$.

By Theorem 9.6 in Hardle. et al. (1997); if (1) $\phi(x)$ is q times weakly differentiable and $\nabla^q \phi(x)$ is bounded; (2) $\forall \beta_{\kappa_0}$ with $0 < \beta_{\kappa_0} < q$ there exists a positive constant L, such that $\sup_k |\alpha_k| \le L$ and $\sup_r \sup_k 2^{r\left(\beta_{\kappa_0}+\frac{1}{2}\right)} |\zeta_{rk}| \le L$, then $\kappa_0(x)$ with support $\mathbb{R}$ has holder exponent β_{κ_0}. Moreover, it can be shown that the restriction $\overline{\kappa}_0(x)$ of $\kappa_0(x)$ to the domain $[0,1]$ lies in a Holder ball $H(\beta_{\kappa_0}, C_{\kappa_0})$ where C_{κ_0} is a non-decreasing function of L. To define functions with Holder exponent below one, we use D6 father and mother wavelets in (2) since D6 wavelets have Holder exponent greater than one. We chose $\beta_b = \beta_\omega = \beta$ and $\forall k \in \mathbb{Z}$, $\alpha_k = 0$, and $\zeta_{rk} = 0$ for $r \notin \Upsilon \equiv \{0, 3, 6, 9, 10, 16\}$. Given constants R_b, R_ω, R_g, we define $b(\cdot), \omega(\cdot)$ and $g(\cdot)$ as: (1) $b(x) = R_b \sum_{r\in\Upsilon,k} 2^{-r(\beta+0.5)} \nu_{rk}(x)$; (2) $h_1(x) = R_\omega \sum_{r\in\Upsilon,k} 2^{-r(\beta+0.5)} \nu_{rk}(x)$, and $\omega(x) = \max\left\{\frac{e^{h_1(x)}}{1+e^{h_1(x)}}, 0.1\right\}$; (3) $g(x) \propto 1 + \exp\left\{R_g \sum_{r\in\Upsilon,k} 2^{-r\left(\beta_f+0.5\right)}\nu_{rk}(x)\right\}$. Next, we sample $2n = 4000$ observations $O_i = (Y_i, A_i, X_i)$ $i = 1, ...4000$, from the following data generating mechanism: $X_i \sim g(x)$, $Y_i|X_i \sim N\left(b(X_i), \sigma^2\right)$, and $A_i|X_i \sim Bernoulli(\omega(X_i))$.

In different simulation experiments we vary R_b, R_g, β_g, β over respective ranges $(1, 100), (1, 1.5), (0.125, 0.6)$, and $(0.125, 0.6)$. with $R_\omega = -2$. By extensive simulations, we verified that in all simulation experiments reported below, conditions $(a.1) - (a.3)$ are satisfied when we choose

$(a.1)$ $|b(X)| \le K_b = 5000$. $(a.2)$ $Var(Y|A=1,X) \le 100$. $(a.3)$ $6^{-3} = \delta_g < g(X) < K_g = 6^3$.

We split the data into a training sample (TS) $\{O_i : i = n+1, ..., 2n\}$ which is used to estimate b, ω and g, and an analysis sample $\{O_i : i = 1, ...n\}$ which we use to construct confidence intervals. We now describe our estimators of b, ω, g which achieve the minimax rate up to $\log n$.

Both $b(x) = E[Y|X=x]$ and $\omega(x) = E[A|X=x]$ were estimated using truncated linear least squares wavelet estimates. Fix $k^* = (n/\log n)^{1/(1+2\beta)}$ and $\widetilde{J} = \log_2(k^* - 4)$, $\widetilde{J}$ may not be a integer. Hence let $J^* = I\left(\widetilde{J} - \lfloor\widetilde{J}\rfloor > 0.5\right) + \lfloor\widetilde{J}\rfloor$ be the nearest integer. $\phi(x)$ is the scaling function of Daubechies wavelet $D6$ whose corresponding mother wavelet has 3 vanishing moments. $\widetilde{b}(x)$ and $\widetilde{\omega}(x)$ are linear least squares estimates on the space spanned by all J^*-level dilated father wavelets whose support intersects $[0,1]$ $\{\phi_{J^*l}(x),\ l = 1, ..., k^*\}$, where $\phi_{J^*l}(x) = 2^{J^*/2}\phi\left(2^{J^*}x - l + 5\right)$. $\widehat{b}(x) = \min(K_b, \max(\widetilde{b}(x), -K_b))$ and $\widehat{\omega}(x) = \min(1, \max(\widetilde{\omega}(x), 0))$ are the estimates we used to calculate the statistics.

>From Theorem 11.3 in Gyorfi. et al. (2002) and Theorem 9.6 in Hardle. et al. (1997),

$$\begin{aligned} E\int \left|\widehat{b}(x) - b(x)\right|^2 g(x)\,dx &= O\left(\frac{n}{\log n}\right)^{-\frac{2\beta}{1+2\beta}} \\ E\int \left|\widehat{\omega}(x) - \omega(x)\right|^2 g(x)\,dx &= O\left(\frac{n}{\log n}\right)^{-\frac{2\beta}{1+2\beta}} \end{aligned}$$

This convergence rate is uniform over b and $\omega's$ Holder balls, and is the minimax rate up to $\log n$. Gyorfi. et al. also discussed alternative complex estimators that achieve the exact MSE minimax rate. For simplicity, we did not use such estimators in our simulation.

$\widetilde{g}(x)$ is the kernel estimator of the density function $g(x)$ using density function $K^\dagger(x) \propto \left(1 - x^2\right)^3$ for $-1 \le x \le 1$ and 0 otherwise, and fixed $h_n = (\log n/n)^{1/\left(1+2\beta_g\right)}$. $K^\dagger(x)$ is symmetric, has bounded support and bounded first two derivatives. Therefore using Theorem 2.3 in Hardle. et al. (1988), we have

$$\sup_{x\in[0,1]} |\widetilde{g}(x) - g(x)| = O_P\left(\left(\frac{n}{\log n}\right)^{-\frac{\beta_g}{1+2\beta_g}}\right).$$

Let $\widehat{g}(x) = \min(\widetilde{g}(x), \delta_g)$. Given the fact $g(x)$ is bounded below by δ_g, $|\widehat{g}(x) - g(x)| \le |\widetilde{g}(x) - g(x)|$ $\forall\, x \in [0,1]$. We will show later our statistics requires $\widehat{g}$ being bounded away from zero. We have assumed (but have no proof) that this convergence is uniform over $g's$ Holder ball.

We now describe the estimator $\widehat{\psi}_{p,m}$ and its estimated variance $\widehat{Var}\left(\widehat{\psi}_{p,m}|\widehat{\theta}\right)$.

We restrict our choice of m, such that $m = 2^M + 4$ for some integer $M > 0$. With $\{\phi_{M,l-5}\}_{l=1}^{m}$ being the set of dilated Mth level father wavelets of $D6$ whose support intersects $[0,1]$, and $\Phi_m(x) = \left(\frac{\phi_{M,-4}(x)}{\sqrt{\widehat{g}(x)}}, \frac{\phi_{M,-3}(x}{\sqrt{\widehat{g}(x)}}), \ldots \frac{\phi_{M,2^M-1}(x)}{\sqrt{\widehat{g}(x)}}\right)^T$, we set $K_m^{\widehat{g}}(x,y) = \Phi_m^T(x)\left(E_{\widehat{g}}\left(\Phi_m\Phi_m^T\right)\right)^{-1}\Phi_m(y)$. By arguing as in the proof of Proposition 6 in the Appendix, one can show that $K_m^{\widehat{g}}(x,y)$ satisfies conditions $(a.4)-(a.7)$ in Sec 3.1 of companion paper. However $\{\phi_{M,l-5}\}_{l=1}^{m}$ is not an ONS since those basis vectors whose support is not fully contained in $[0,1]$ are not orthonormal, e.g., $\int_0^1 \phi_{M,-4}\phi_{M,-3} \neq 0$. However because in our simulation m is always chosen at least as large as sample size n, this small deviation from ONS is negligible as argued below. So we define $\widetilde{K}_m^{\widehat{g}}(x,y) = \Phi_m^T(x)\Phi_m(y)$ and used this as the kernel in our simulations. Let $\Omega = E_{\widehat{g}}\left(\Phi_m\Phi_m^T\right)$. It is easy to show that Ω is a block diagonal matrix, $Diag\left(B_1, I_{m-8}, B_2\right)$, where both B_1 and B_2 are 4×4 matrices with all elements bounded. I_{m-8} is a identity matrix of dimension $(m-8)$. Elsewhere we verify that using $\widetilde{K}_m^{\widehat{g}}(x,y)$ rather than $K_m^{\widehat{g}}(x,y)$ does not change the order of bias and variance as well as other properties of our estimators $\widehat{\psi}_{p,m}$.

Define $\widehat{\epsilon}_i = Y_i - \widehat{b}(X_i)$, $\widehat{\Delta}_i = A_i - \widehat{\omega}(X_i)$. The construction of our estimator $\widehat{\psi}_{p,m}$ is completed by substituting our kernel $\widetilde{K}_m^{\widehat{g}}(x,y)$ in equations presented in Sec 3.2 in our companion paper to obtain

$$\widehat{\psi}_1 = \mathbb{V}_n\left(\widehat{\epsilon}_i\widehat{\Delta}_i\right)$$

$$\widehat{\psi}_{2,m} = \widehat{\psi}_1 - \mathbb{V}_n\left(\widehat{\epsilon}_i\Phi_m^T(X_i)\Phi_m(X_j)\widehat{\Delta}_j\right)$$

$$\widehat{\psi}_{3,m} = \widehat{\psi}_{2,m}$$

$$+\mathbb{V}_n\left(\widehat{\epsilon}_i\Phi_m^T(X_i)\left(\Phi_m(X_s)\Phi_m^T(X_s) - I\right)\Phi_m(X_j)\widehat{\Delta}_j\right)$$

In our simulations, we only considered $\widehat{\psi}_1, \widehat{\psi}_{2,m_{opt}(2)}$, and $\widehat{\psi}_{3,m_{opt}(3)}$. However our formula for $m_{opt}(p,n)$, $p = 2,3$, differs slightly from that in companion paper. First, for $p = 2,3$, we redefine the estimation bias as

$$EB_p\left(\widehat{\psi}_{p,m}\right) = \left(\frac{n}{\log n}\right)^{-\frac{\beta_b}{1+2\beta_b}-\frac{\beta_\omega}{1+2\beta_\omega}-\frac{(g-1)\beta_g}{1+2\beta_g}}$$

where the additional $\log n$ factor results from our use of the truncated linear least squares estimates $\widehat{b}(X)$ and $\widehat{\omega}(X)$. Additionally, we never take $m_{opt}(p,n)$ to be less than the sample size n, because at least in terms of order, this choice has no effect on variance. Therefore $m_{opt}(2) = \max\left(n, n^{\frac{2}{1+4\beta}}, n^2\left(\frac{n}{\log n}\right)^{-\frac{4\beta}{1+2\beta}-\frac{2\beta_g}{1+2\beta_g}}\right)$, and $m_{opt}(3) = \max\left(n, n^{\frac{3}{2+4\beta}}, n^{1.5}\left(\frac{n}{\log n}\right)^{-\frac{2\beta}{1+2\beta}-\frac{2\beta_g}{1+2\beta_g}}\right)$.

Note in contrast to the recommendation in our companion paper we chose $m(p,n)$ equal to (rather than greater than) $m_{opt}(p,n)$ so the orders of bias and standard deviation of the estimator $\widehat{\psi}_{p,m_{opt}(p)}$ $(p=2,3)$ are exactly the same. To understand the implication of this choice, consider the test of the true null hypothesis $H_0 : \psi^{WATE} = 0$ that rejects whenever the confidence interval $C_n^{p,m_{opt}(p)} = \widehat{\psi}_{p,m_{opt}(p)} \pm z_{\alpha/2}\widehat{s.e.}\left(\widehat{\psi}_{p,m_{opt}(p)}\right)$ does not include 0. For $p \in \{2,3\}$ the equivalent test statistic $t_{p,m_{opt}(p)} = \frac{\widehat{\psi}_{p,m_{opt}(p)}}{\widehat{s.e.}\left(\widehat{\psi}_{p,m_{opt}(p)}\right)}$ has an asymptotic distribution $N\left(\mu_p, 1\right)$. Since the order of the bias equals the order of the standard deviation, μ_p is determined by the ratio of the constants that respectively determine the bias and standard deviation. μ_p will be of $O(1)$ under some laws in our model. The coverage rate of the $(1-\alpha)$ CI $C_n^{p,m_{opt}(p)}$, equals $\Phi\left(z_{1-\alpha/2} - \mu_p\right) - \Phi\left(z_{\alpha/2} - \mu_p\right)$, which decreases as μ_p deviates from 0.

In the companion paper we suggested increasing the order of $m(p,n)$ beyond that of $m_{opt}(p,n)$ to make the order of the bias less than the standard deviation. For example we could choose $m(p,n) = m_{opt}(p,n)\,n^{\delta}$. Although asymptotically any $\delta > 0$ will work, nonetheless at our fixed sample size of 2000, there is no guarantee this that a given δ, say .3, will correct the poor coverage of our interval since the bias "constant" may still dominate the somewhat increased standard deviation.

The optimal approach would be to obtain a finite sample bound on the maximal bias under our model. Suppose we could derive a constant C_{bias}, such that $Bias\left(\widehat{\psi}_{2,m_{opt}(2)}\right) \leq C_{bias}n^{-\gamma_1}$, for all $n > 1000$. Then the confidence interval $\widehat{\psi}_{2,m_{opt}(2)} \pm \left(z_{\alpha/2}\widehat{s.e.}\left(\widehat{\psi}_{p,m_{opt}(p)}\right) + C_{bias}n^{-\gamma_1}\right)$ will solve the "constant" problem whenever $n > 1000$. Then for $n > 1000$, $\widehat{\psi}_{2,m_{opt}(2)} \pm$

$\left(z_{\alpha/2}\widehat{s.e.}\left(\widehat{\psi}_{p,m_{opt}(p)}\right)+C_{bias}n^{-\gamma_1}\right)$ will be a honest conservative finite sample $1-\alpha$ confidence interval, provided that $\widehat{\psi}_{p,m_{opt}(p)}$ has nearly converged to its normal limit by size n. Determining C_{bias} as a function of the model (and of our estimators of b,ω,g) is an important open problem. In this paper, we examine the "constant" problem by observing the behavior of our point and interval estimatos as we vary the constants R_b and R_g.

3. Results

The performance of $\widehat{\psi}_1$ and $\widehat{\psi}_p \equiv \widehat{\psi}_{p,m_{opt}(p)}$ $(p=2,3)$ are evaluated in 17 experiments with different combinations of $\left(\beta,\beta_g,R_b,R_g\right)$. R_ω equals -2 in all experiments . In each experiment, we ran 200 monte carlo (MC) replications. Each replication has total sample size of 4000, equally split into analysis and training samples. Results are presented in Tables 1 and 2. In the first 10 experiments we fixed $(R_b,R_g)=(1,0.5)$ and varied $\left(\beta=\beta_b=\beta_\omega,\beta_g\right)$. In the next seven we fixed $\left(\beta,\beta_g\right)=(0.5,0.2)$ and varied (R_b,R_g). For each experiment, we present the empirical MC coverage rate of the nominal 90% CI $C_n^{p,m_{opt}(p)}$, the absolute value of the MC mean and median of $\widehat{\psi}_p$, and the absolute value of the MC median of $t_{p,m_{opt}(p)}$ in Table 1; and present the square-root of the MC MSE of $\widehat{\psi}_p$, the MC interquantile range (IRQ), and the MC mean of $\widehat{s.e.}\left(\widehat{\psi}_{p,m_{opt}(p)}\right)$ in Table 2. The asymptotic standard error of $\widehat{\psi}_1$ is $O\left(n^{-1/2}\right)$ in all experiments while that of $\widehat{\psi}_2$ and $\widehat{\psi}_3$ are $O\left(n^{-\alpha_2}\right)$ and $O\left(n^{-\alpha_3}\right)$ with α_2 and α_3 depending on the experiment. To facilitate comparison of our MC results with our asymptotic predictions, we report (i) in column "Exponent" $\widehat{\psi}_2$ of Table 2 $2000^{-(\alpha_2-1/2)}$ (α_2) and (ii) in column "Exponent" $\widehat{\psi}_3$ $2000^{-(\alpha_3-\alpha_2)}$ (α_3).

Consider now the first 10 experiments. When β is larger than or equal to 0.5, all three estimators attain their nominal coverage rate of 90%. This result is consistent with the asymptotic theory developed in our companion paper. Although $\widehat{\psi}_1$ has greater bias than $\widehat{\psi}_2$ or $\widehat{\psi}_3$, it has the smallest IQR and MSE of the three estimators. As β falls below 0.5, $\widehat{\psi}_1$ starts to break down, as predicted by asymptotic theory. The root-MSE and bias of $\widehat{\psi}_1$ increases and its coverage falls below nominal as β decreases. As expected $\widehat{\psi}_1's$ properties are not affected by β_g. As β decreases from .5 to .125, $\widehat{\psi}_2$ and $\widehat{\psi}_3$ preserve nominal coverage; further the IQR and CI length for $\widehat{\psi}_2$ and $\widehat{\psi}_3$ increase rather dramatically when β falls below .25 to .125, especially for $\beta_g \leq$ .2. Both results are as predicted by asymptotic theory. In particular, for our choices of β_g, the asymptotic rate of convergence of $\widehat{\psi}_3$ first falls below $n^{1/2}$ when β falls below .25. In contrast when $\beta_g=.2$, $\widehat{\psi}_2$ already has convergence rate slower than $n^{1/2}$ when $\beta=.25$.

One striking discrepancy with the predictions of asymptotic theory is that the relative efficiency of $\widehat{\psi}_3$ compared to $\widehat{\psi}_2$ is much better than expected when $\beta_g=0.2$. For example when $\beta=0.3$, $\widehat{\psi}_3's$ IQR is only 60% of $\widehat{\psi}_2's$ when $\beta_g=.2$ but is approximately 100% when $\beta_g=.3$ or $\beta_g=.4$. Asymptotic theory, however, would predict no dependence on β_g. To understand this discrepancy, we calculated the Monte Carlo correlation between $U_{3,3}\left(\widehat{\theta}\right)$ and $U_{2,2}\left(\widehat{\theta}\right)$ with $U_{j,j}\left(\widehat{\theta}\right)\equiv U_{j,j;\overline{\psi}_{m_{opt}(j)}}\left(\widehat{\theta}\right)$ and found it to be -0.85, which fully explained the smaller variability of $\widehat{\psi}_3$. Since $U_{j,j}\left(\widehat{\theta}\right)$ is a degenerate U-statistic under $\widehat{\theta}$, $U_{3,3}\left(\widehat{\theta}\right)$ and $U_{2,2}\left(\widehat{\theta}\right)$ are uncorrelated under $\widehat{\theta}$ and have a correlation of $O\left(\|\widehat{g}-g\|\right)$ under the actual data generating mechanism θ. We were quite surprised that our estimate was as large as -0.85. Presumably it reflects the facts that (i) at sample size 2000 with β_g only 0.2, $\widehat{g}(x)$ estimates $g(x)$ poorly and (ii) $\widehat{g}(x)$ occurs in the denominator of the kernel $\widetilde{K}_m^{\widehat{g}}(x,y)$ and $\widehat{g}(x)$ is rather close to zero for certain x.

Turn now to experiments 11-17 in which $\left(\beta,\beta_g,\sigma\right)=(0.5,0.2,1)$. We examine the "constant problem" by increasing (R_b,R_g) and thereby increasing the radius of the smallest Holder ball in which b and g respectively live. Asymptotic theory predicts that the variance of our estimators should not vary with R_b. It is easy to prove that the actual finite sample bias of all three estimators should increase linearly with R_b. Thus, we might expect that the MC coverage of all three CIs would fall below nominal as R_b increased. This conjecture was born out for $\widehat{\psi}_1$ but, remarkably, not for $\widehat{\psi}_2$ or $\widehat{\psi}_3$, implying these estimators have even better finite sample robustness than we had reason to suppose. How can we account for these MC results? Turn first to bias and consider experiments with $R_g = .5$. The MC median of $\widehat{\psi}_1$ did increase linearly. However MC medians of $\widehat{\psi}_p$ $(p = 2, 3)$ did not increase as R_b went from 1 to 10, only increasing finally at $R_b = 100$. Now asymptotic theory predicts the the bias of $\widehat{\psi}_p$ $(p = 2, 3)$ is less than that of $\widehat{\psi}_1$. Morever it must be that the actual bias of $\widehat{\psi}_p$ $(p = 2, 3)$ for R_b of 10 or less is so small that the apparent MC bias is actually not bias but sampling variability. Only when R_b is 100 are the MC medians of $\widehat{\psi}_2$ and $\widehat{\psi}_3$ accurate estimates of the true bias. Turn next to the variance. For all three estimators R_b enters the variance only through $E\left[\widehat{\epsilon}^2\right] = E\left[\left(Y - \widehat{b}(X)\right)^2\right] = \sigma^2 + E\left[\left(b(X) - \widehat{b}(X)\right)^2\right]$. Asymptotic theory ignores the second term and so predicts no dependence of the variance on R_b. However, in finite samples, the second term will dominate if R_b is large enough. In our simulations we find that, for all three estimators, the IQR did not change with R_b for $R_b \leq 10$, but that at $R_b = 100$, the IQR increased by a factor of more than 6. Once R_b is large enough that the second term dominates, the finite sample standard error of all three estimators will scale linearly with increasing R_b, exactly like the bias. Suppose, as for $\widehat{\psi}_2$ and $\widehat{\psi}_3$ in our simulations, the bias is much less than the standard error at the critical value of R_b where the standard error begins to scale linearly with R_b. Then the finite sample coverage of the associated Wald intervals can remain at or above the nominal level even as R_b becomes very large.

4. Discussion/Conclusion

In this paper we have demonstrated that the asymptotic results derived in our companion paper provide accurate guidance to finite sample performance, provided that "constants" determined by the radii of our Holder balls are not too large. We found, via simulation, that even when these "constants" become large, interval estimators based on our higher order U-statistics can, but usual first order interval estimators cannot, continue to cover at their nominal level.

REFERENCES

Bickel P. , Klassen C., Ritov Y., Wellner J. (1993) " Efficient and Adaptive Estimation for Semiparametric Models". Springer.

Bickel, P, Ritov, Y. (1988). "Estimating Integrated Squared Density Derivatives: Sharp Best Order of Convergence Estimates". Sankhya Ser. A 50 381-393

Cohen A., Daubechies I., Vial P. (1998) "Wavelets on the Interval and Fast Wavelet Transforms". Applied and Computational Harmonic Analysis, 1, 54-81.

Frazier M., Jawerth B., Weiss G. (1991) "Littlewood-Paley Theory and the Study of Function Spaces". NSF-CBMS Regional Conf. Ser in Mathematics, 79, American Mathematical Society, Providence, RI.

Gyorfi L., Kohler M., Krzyzak A. Walk H. "A Distribution-Free Theory of Nonparametric Regression". Springer Series in Statistics.

Hardle W., Janssen P., Serfling R (1988). "Strong Uniform Consistency Rates for EStimators of Conditional Functionals". The Annals of Statistics, Vol. 16, No. 4, 1428-1449

Hardle W., Kerkyacharian G., Picard D., Tsybakov A. (1997) "Wavelets, Approximation and Statistical Applications". Seminar Berlin-Paris.

Huang JZ (2003). "Local Asymptotics for Polynomial Spline Regression". The Annals of Statistics, Vol. 31, No. 5, 1600-1635.

Kerkyacharian G., Picard D. (1996) "Estimating Nonquadratic Functionals of a Density Using Haar Wavelets" Ann. Statist. 24, no. 2, 485–507.

Lindsay R, Waterman B. (1996) "Projected Score Methods for Approximating Conditional Scores". Biometrika 83(1):1-13.

Mallat SG. (1998) "A Wavelet Tour of Signal Processing". Acedemic Press.

McLeish C. ,Small D. (1994) "Hilbert Space Methods in Probability and Statistical Inference".

Newey W.K., Hsieh,F., Robins,J.(1991) "Undersmoothing and Bias Corrected Functional Estimation" Working Paper, MIT

Pfanzagl J. (1990) "Estimation in Semiparametric Models: Some Recent Developments". Lecture Notes in Statistics 31, Springer-Verlag, Berlin 1985, 505 pages

Powell, M.J.D.(1981) "Approximation Theory and Methods". Cambridge University Press, Cambridge, UK.

Robins J., Ritov, Y (1997). "Toward a Curse-of-dimensionality Appropriate (CODA) Asymptotic Theory for Semiparametric Models". Statistics in Medicine 16 285-319.

Robins J. , van der Vaart, AW (2004). "A Unified Approach to Estimation in Non- and Semiparametric Higher Order Estimating Equations". Technical Report.

Robins JM. (2004) "Optimal Structural Nested Models for Optimal Sequential Decisions". in DY Lin and P. Heagerty (Eds.) Proceedings of the second seattle symposium in biostatistics. New York Springer.

van der Vaart, AW (1991). "On Differentiable Functionals". Ann. Statist. 19 178-204.

5. Appendix

Proposition 1 *Assume that conditions* $(a.1)-(a.7)$ *hold. In addition, suppose that* $(a.8.1)$ *either* $\widehat{\omega}(\cdot)\in H(\beta_\omega,C_\omega),\widehat{b}(\cdot)\in H(\beta_b,C_b)$ *or* $\widehat{\omega}(\cdot),\widehat{b}(\cdot)\in L_m$, *and* $\widehat{g}(\cdot)\in H(\beta_g,C_g)$, *and* $(a.8.2)$ $\widehat{\omega}(\cdot),\widehat{b}(\cdot)$ *are* L_2-risk *minimax rate optimal nonparametric estimators, and* $\widehat{g}(\cdot)$ *is an* $L_\infty-risk$ *minimax rate optimal nonparametric density estimator. Then the following is true for* $p\geq 1$ *and* $\psi\in\left\{\psi^{WATE},\psi^{CAR}\right\}$:

$$E\left(\widehat{\psi}_p-\psi|\widehat{\theta}\right)=I(p\geq 2)O\left(TB_m\right)+O\left(EB_p\right)$$
$$var\left(\widehat{\psi}_p|\widehat{\theta}\right)=O_P\left(\frac{1}{n}\max\left(1,\left(\frac{m}{n}\right)^{p-1}\right)\right)$$

for TB_m *and* EB_p *defined in our companion paper.*

Proof. We restrict the proof to the $WATE$ model; the proof for the CAR model is essentially the same. Write $\delta_{b,i}=(b(X_i)-\widehat{b}(X_i))$, $\delta_{\omega,i}=\left(\frac{\omega(X_i)}{\widehat{\omega}(X_i)}-1\right)$, $\delta_{g,i}=(g(X_i)-\widehat{g}(X_i))$; all expectations are conditional on the training sample. The bias of $\widehat{\psi}_1^{WATE}$ is given by:

$$\begin{aligned}
& E\left(\delta_{b,i}\times\delta_{\omega,i}\right)\\
=\;& E\left[\delta_{b,1}K_m^g\left(X_1,X_2\right)\delta_{b,2}\right]\\
& +E\left[\begin{array}{c}\left(\delta_{b,2}-E\left(K_m^g\left(X_1,X_2\right)\delta_{b,1}|X_2\right)\right)\\ \times\left(\delta_{\omega,2}-E\left(K_\omega^g\left(X_1,X_2\right)\delta_{b,1}|X_2\right)\right)\end{array}\right]\\
=\;& EB_1^{WATE}+TB_m^{WATE}
\end{aligned}$$

By the Cauchy-Schwarz inequality, $(a.5)$ and $(a.8.1)$, $TB_m^{WATE}=O\left(m^{-\frac{\beta_b+\beta_\omega}{d}}\right)$. Moreover, using the fact that $\mathbf{K}_m^g(\cdot)$ is an orthogonal projection operator, we have: $EB_1\leq ||\delta_{b,1}||_g||\delta_{\omega,1}||_g\leq M^2O\left(n^{-\left(\frac{\beta_b}{d+2\beta_b}+\frac{\beta_\omega}{d+2\beta_\omega}\right)}\right)$, where we used $(a.3)$ and $(a.8.2)$. In addition, using $(a.7)$, the conditional estimation bias EB_2^{WATE} of $\widehat{\psi}_2^{WATE}$ is given by :

$$\begin{aligned}
& EB_1+E\left(U_{2,2}^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right)\\
=\;& E\left[\delta_{b,1}K_m^g\left(X_1,X_2\right)\delta_{\omega,2}\right]-\\
& E\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_2\right)\delta_{\omega,2}\right]\\
=\;& E\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_2\right)\delta_{\omega,2}\right]O\left(||\delta_g||_\infty\right)\\
=\;& \widehat{E}\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_2\right)\delta_{\omega,2}\right]O\left(||\delta_g||_\infty\right)\\
& +o_p(1)\\
\leq\;& ||\delta_{b,1}||_{\widehat{g}}||\delta_{\omega,1}||_{\widehat{g}}O\left(||\delta_g||_\infty\right)+o_p(1)\\
=\;& O\left(\left(\frac{\log n}{n}\right)^{\frac{\beta_g}{d+2\beta_g}}n^{-\left(\frac{\beta_b}{d+2\beta_b}+\frac{\beta_\omega}{d+2\beta_\omega}\right)}\right)
\end{aligned}$$

where $\widehat{E}\left(a(X)\right)=\int\widehat{g}(X)a(X)dX$ and we again used the Cauchy-Schwarz inequality. For $p=3$,

$$\begin{aligned}
& EB_2+E\left(U_{3,3}^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right)\\
=\;& E\left[\delta_{b,1}K_m^g\left(X_1,X_2\right)\delta_{\omega,2}\right]\\
& -E\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_2\right)\delta_{\omega,2}\right]\\
& +E\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_3\right)K_m^{\widehat{g}}\left(X_3,X_2\right)\delta_{\omega,2}\right]\\
& -E\left[\delta_{b,1}K_m^{\widehat{g}}\left(X_1,X_2\right)\delta_{\omega,2}\right]
\end{aligned}$$

Using $(a.4)$,

$$\begin{aligned}
& EB_2+E\left(U_{3,3}^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right)\\
=\;& E\left\{\left(\widehat{E}-E\right)\left[\begin{array}{c}\delta_{b,1}\times\delta_{g,3}K_m^{\widehat{g}}\left(X_1,X_3\right)\\ \times\left(\begin{array}{c}K_m^g\left(X_3,X_2\right)\\ -K_m^{\widehat{g}}\left(X_3,X_2\right)\end{array}\right)\\ \times\delta_{\omega,2}|X_1,X_2\end{array}\right]\right\}\\
\leq\;& ||\delta_{b,1}||_{\widehat{g}}||\delta_{\omega,1}||_{\widehat{g}}O\left(||\delta_g||_\infty^2\right)+o_p(1)\\
=\;& O\left(\left(\frac{\log n}{n}\right)^{\frac{2\beta_g}{d+2\beta_g}}n^{-\left(\frac{\beta_b}{d+2\beta_b}+\frac{\beta_\omega}{d+2\beta_\omega}\right)}\right)
\end{aligned}$$

The proof for the general case EB_p, $p>3$ is lengthy and will be reported elsewhere. Next, we consider

$var\left(\widehat{\psi}_p^{WATE}\right) = var\left(U_1^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right) + \sum_{j=2}^{p} var\left(U_{p,p}^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right)$; now, because $var\left(U_p^{WATE}\left(\widehat{\theta},\widehat{\psi}\right)\right) = var\left(U_p^{WATE}\left(\theta,\psi\left(\theta\right)\right)\right)\left(1+o_p\left(1\right)\right)$, it is enough to consider

$$var\left(U_1^{WATE}\left(\theta,\psi\right)\right) = O\left(\frac{1}{n}\right)$$

and $var\left(\mathbb{V}_n U_{p,p}^{WATE}\left(\theta,\psi\right)\right) =$

$$\begin{aligned}
& O\left(\frac{1}{n^p}E\left(\begin{array}{c} E\left(\epsilon_{i_1}^2|X_{i_1}\right)E\left(\Delta_{i_2}^2|X_{i_2}\right) \\ \times\left(\prod_{s=1}^{p-1}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2 \end{array}\right)\right) \\
\leq & \; O\left(\frac{K_1}{n^p}E\left(\prod_{s=1}^{p-1}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2\right) \\
= & \; O\left(\frac{m^{p-1}}{n^p}\right)
\end{aligned}$$

so that we get the desired result. ■

Proof. (Proposition 6 in Tchetgen et al 2005) Condition $(a.4)$ is easily verified since $K_m^g\left(x,y\right)$ is a well defined orthogonal projection kernel in $L_2\left(g\right)$. For sufficiently "regular" wavelets (Mallat,1998), condition $(a.5)$ is satisfied since by assumption $(a.8)$ there exists $\left\{\widetilde{a}_l, l=1,...,m\right\}$,such that

$$\begin{aligned}
& \int\left(h(x)\widehat{g}(x)^{1/2}-\sum_{l=1}^{m}\widetilde{a}_l\phi_{M,l}(x)\right)^2 dx \\
= & \; O\left(m^{-\frac{2\beta_h}{d}}\right)
\end{aligned}$$

so that for $\mathbf{K}_m^g h(x) = \sum_{l=1}^{m} a_l^*\phi_{M,l}(x)/\sqrt{\widehat{g}(x)}$,

$$\begin{aligned}
& \int\left(h(x)-\mathbf{K}_m^g h(x)\right)^2 g(x)dx \\
= & \int\left(h(x)-\frac{\sum_{l=1}^{m}a_l^*\phi_{M,l}(x)}{\sqrt{\widehat{g}(x)}}\right)^2 g(x)dx \\
\leq & \left|\left|\frac{g}{\widehat{g}}\right|\right|_\infty \int\left(h(x)\widehat{g}(x)^{1/2}-\sum_{l=1}^{m}\widetilde{a}_l\phi_{M,l}(x)\right)^2 dx \\
= & \; O\left(m^{-\frac{\beta_h}{d}}\right)
\end{aligned}$$

where we use the fact that $\mathbf{K}_m^g$ is a least squares projection in $L_2(g)$.We proceed to prove that condition $(a.6)$ holds. For $p=2$,

$$\begin{aligned}
& E\left[\left(K_m^{\widehat{g}}\left(X_{i_1},X_{i_2}\right)\right)^2\right] \\
= & \; E\left(\Phi_m^T\left(X_{i_1}\right)\Phi_m\left(X_{i_2}\right)\Phi_m^T\left(X_{i_2}\right)\Phi_m\left(X_{i_1}\right)\right) \\
\leq & \left|\left|\frac{g}{\widehat{g}}\right|\right|_\infty^2 \sum_{l}^{m}\int\phi_{M,l}(x)\phi_{M,l}(x)dx = O\left(m\right)
\end{aligned}$$

where the last step uses the fact that $\left\{\phi_{M,l}: l=1...m\right\}$ is an orthonormal basis with respect to Lebesgue measure on $L_2[0,1]$. For any $p\geq 3$, $E\left[\left(\prod_{s=1}^{p-1}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2\right]$

$$\begin{aligned}
= & \; E\left(\begin{array}{c} \left(\prod_{s=1}^{p-2}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2 \\ \times\Phi_m^T\left(X_{i_{p-1}}\right) \\ \times E\left(\Phi_m\left(X_{i_p}\right)\Phi_m^T\left(X_{i_p}\right)\right) \\ \times\Phi_m\left(X_{i_{p-1}}\right) \end{array}\right) \\
\leq & \left|\left|\frac{g^{1/2}}{\widehat{g}}\right|\right|_\infty^2 \times E\left[\begin{array}{c} \left(\prod_{s=1}^{p-2}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2 \\ \times\sum_l^m\phi_{M,l}(X_{i_{p-1}})\phi_{M,l}(X_{i_{p-1}}) \end{array}\right] \\
= & \; O\left(m\right)E\left[\left(\prod_{s=1}^{p-2}K_m^{\widehat{g}}\left(X_{i_s},X_{i_{s+1}}\right)\right)^2\right]
\end{aligned}$$

which leads to the desired result. In the latest proof, we use the fact that by having compact support, $\phi_{M,l}$ has overlapping support with $\phi_{M,l'}$ if and only if $|l-l'|<r$, for a constant r which only depends on the regularity of the wavelet being used and therefore is independent of $m.(Mallat, 1998)$. Moreover, $|\sum_l^m\phi_{M,l}^2(x)| = O\left(m\right)$ due to the dilation of the father wavelet. Finally, we prove $(a.7)$:

$$\begin{aligned}
& K_m^g\left(x,y\right)-K_m^{\widehat{g}}\left(x,y\right) \\
= & \; \Phi_m^T(x)\left[\begin{array}{c}\left(E_g\left(\Phi_m(X)\Phi_m^T(X)\right)\right)^{-1} \\ -I_{m\times m}\end{array}\right]\Phi_m(y) \\
= & \left(\begin{array}{c}\Phi_m^T(x)\left(E_g\left(\Phi_m(X)\Phi_m^T(X)\right)\right)^{-1} \\ \times\left[I-E_g\left(\Phi_m(X)\Phi_m^T(X)\right)\right]\Phi_m(y)\end{array}\right) \\
= & \; \Phi_m^T(x)\left(E_g\left(\Phi_m(X)\Phi_m^T(X)\right)\right)^{-1} \\
& \left[\int\widehat{g}(x)\left(1-\frac{g(x)}{\widehat{g}(x)}\right)\Phi_m(x)\Phi_m^T(x)dx\right]\Phi_m(y) \\
\leq & \; -O_P\left(||\delta_g||_\infty\right)K_m^g\left(x,y\right)
\end{aligned}$$

which leads to the desired result. ■

Table 1: Coverage, Mean and Median of Bias, and Median of Test Statistic across Scenarios

	$Coverage$			$100 \times \lvert Mean\left(\widehat{\psi}_p\right) \rvert$			$100 \times \lvert Median\left(\widehat{\psi}_p\right) \rvert$			$100 \times \lvert Median\left(t_p\right) \rvert$		
(β, β_g)	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$
$(0.6, 0.2)$	0.905	0.9	0.91	0.21	0.01	0.03	0.24	0.02	0.098	0.25	0.017	0.05
$(0.5, 0.2)$	0.89	0.9	0.95	.26	0.08	0.03	0.26	0.04	0.18	0.28	0.02	0.09
$(0.3, 0.6)$	0.43	0.91	0.92	1.75	0.09	0.17	1.78	0.11	0.22	1.81	0.06	0.11
$(0.3, 0.4)$	0.39	0.85	0.96	1.76	0.1	0.2	1.79	0.02	0.20	1.86	0.01	0.1
$(0.3, 0.2)$	0.475	0.9	1	.65	0.3	0.3	1.65	0.34	0.24	1.7	0.11	0.07
$(0.25, 0.4)$	0.1	0.91	0.96	2.85	0.15	0.46	2.8	0.26	0.46	2.81	0.11	0.16
$(0.25, 0.2)$	0.135	0.9	0.99	2.87	0.03	0.37	2.9	0.19	0.35	2.95	0.06	0.1
$(\frac{1}{8}, 0.4)$	0	0.85	0.93	10.28	1.18	0.08	10.22	1.2	0.37	8.73	0.23	0.04
$(\frac{1}{8}, 0.2)$	0	0.85	0.93	10.02	0.78	2.09	9.94	1.35	0.79	8.44	0.21	0.08
$(\frac{1}{8}, \frac{1}{8})$	0	0.87	0.95	10.09	0.51	0.78	10	0.65	0.11	8.44	0.06	0.01
(R_b, R_g)												
$(3, 0.5)$	0.76	0.89	0.92	0.79	0.15	0.07	0.75	0.13	0.04	0.81	0.07	0.01
$(5, 0.5)$	0.58	0.91	0.96	1.47	0.02	0.08	1.44	0.01	0.02	1.38	0	0.01
$(10, 0.5)$	0.29	0.93	0.95	2.77	0	0.23	2.75	0.01	0.19	2.18	0	0.06
$(10, 0.75)$	0.3	0.91	0.95	2.74	0.04	0.065	2.78	0.03	0.05	2.18	0.01	0.01
$(10, 1)$	0.31	0.92	0.93	2.67	0.08	0.39	2.69	0.13	0.21	2.16	0.05	0.08
$(100, 0.5)$	0.02	0.96	0.94	29.87	1.15	4.42	28	1.57	2.62	3.31	0.09	0.13
$(100, 1.5)$	0.18	0.91	0.81	20.99	3.87	1.76	21	3.21	7.43	2.65	0.19	0.34

Table 2: Root MSE, Interquantile Range (IQR), Exponent of Convergence Rate, and Length of CI across Scenarios

	$100 \times \sqrt{\widehat{MSE}}$			$100 \times IQR$			$Exponent$		100×Length of CI		
(β, β_g)	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_2$	$\widehat{\psi}_3$	$\widehat{\psi}_1$	$\widehat{\psi}_2$	$\widehat{\psi}_3$
$(0.6, 0.2)$	0.95	1.39	1.76	1.36	1.96	2.57	1 (0.5)	1 (0.5)	0.92	1.34	1.79
$(0.5, 0.2)$	0.95	1.62	1.81	1.20	2.12	2.28	1 (0.5)	1 (0.5)	0.93	1.68	2.09
$(0.3, 0.6)$	2.03	1.66	1.86	1.44	2.46	2.68	1 (0.5)	1 (0.5)	0.96	1.71	2.07
$(0.3, 0.4)$	2.01	1.72	1.59	1.47	2.27	2.12	1 (0.5)	1 (0.5)	0.96	1.70	2.05
$(0.3, 0.2)$	1.89	3.04	2.12	1.21	4.11	2.49	1 (0.5)	1 (0.5)	0.96	3.05	3.37
$(0.25, 0.4)$	3.05	2.16	1.97	1.38	3.07	2.73	1 (0.5)	1 (0.5)	0.98	2.28	2.56
$(0.25, 0.2)$	3.05	3.03	1.88	1.26	4.23	2.81	1.2 (.476)	0.83 (.5)	0.98	3.11	3.36
$(\frac{1}{8}, 0.4)$	10.35	5.14	7.03	1.53	7.22	8.84	3.56 (.33)	1.29 (.3)	1.18	4.92	7.31
$(\frac{1}{8}, 0.2)$	10.10	7.51	9.61	1.5	10.24	9.8	3.56 (.33)	1.29 (.3)	1.18	6.89	9.38
$(\frac{1}{8}, \frac{1}{8})$	10.17	10.47	9.98	1.83	13.23	10.81	4.57 (.3)	1 (.3)	1.21	9.79	11.85
(R_b, R_g)											
$(3, 0.5)$	1.27	1.74	2.37	1.27	2.26	2.36	1 (0.5)	1 (0.5)	0.96	1.73	2.17
$(5, 0.5)$	1.93	1.99	2.12	1.32	2.48	2.84	1 (0.5)	1 (0.5)	1.07	1.88	2.29
$(10, 0.5)$	3.05	2.0	2.39	1.58	2.53	3.18	1 (0.5)	1 (0.5)	1.27	2.26	2.78
$(10, 0.75)$	3.06	2.27	3.32	1.72	3.17	2.86	1 (0.5)	1 (0.5)	1.26	2.27	3.16
$(10, 1)$	2.97	2.25	11.69	1.59	2.81	3.33	1 (0.5)	1 (0.5)	1.24	2.27	3.15
$(100, 0.5)$	31.03	13.49	19.34	10.4	18.73	19.0	1 (0.5)	1 (0.5)	8.8	15.71	19.48
$(100, 1.5)$	22.48	16.49	468	11.5	20.35	28.91	1 (0.5)	1 (0.5)	8.16	15.94	94.35