Simpler Bootstrap Estimation of the Asymptotic Variance of U-statistic Based Estimators[*]

Bo E. Honoré[†]   Luojia Hu[‡]

June 2017

Abstract

The bootstrap is a popular and useful tool for estimating the asymptotic variance of complicated estimators. Ironically, the fact that the estimators are complicated can make the standard bootstrap computationally burdensome because it requires repeated re-calculation of the estimator. In this paper, we propose a method which is specific to extremum estimators based on U-statistics. The contribution here is that rather than repeated re-calculation of the U-statistic-based estimator, we can recalculate a related estimator based on single-sums. A simulation study suggests that the approach leads to a good approximation to the standard bootstrap, and that if this is the goal, then our approach is superior to numerical derivative methods.

KEYWORDS: U-statistics; bootstrap; inference; numerical derivatives.

JEL CODE: C10, C18.

[†]Department of Economics, Julis Romo Rabinowitz Building, Princeton, NJ 08544. Email: honore@princeton.edu.

[‡]Economic Research Department, Federal Reserve Bank of Chicago, 230 S. La Salle Street, Chicago, IL 60604. Email: lhu@frbchi.org.

1 Introduction

Applied researchers often use the bootstrap when estimating standard errors. There are a number of advantages of using the bootstrap for estimating standard errors. One is that it is computationally convenient because it requires little additional programming above and beyond what is required for calculating the estimator. A second is that direct estimation of the asymptotic variance of estimators might require choosing tuning parameters such as bandwidths for nonparametric estimation or step-lengths for numerical derivatives. See Hong, Mahajan, and Nekipelov (2015) for a discussion of this. On the other hand, the bootstrap can be computationally demanding because it requires one to recalculate the same estimator many times. This is especially true for estimators that are based on U-statistics such as those in Honoré and Powell (1994) and Alan, Honoré, Hu, and Leth-Petersen (2014).

The contribution of this paper is to point out that for extremum estimators based on U-statistics, one can sometimes estimate the Hessian term of the asymptotic variance by a simpler bootstrap procedure which involves repeated re-estimation of an extremum estimator based on simple averages. This will lower the computational burden of the bootstrap without giving up the aforementioned advantages of the bootstrap.

The proposed approach applies directly to classical rank-estimators of regression models such as those in Adichie (1967) and Sievers (1978) and to estimators of truncated and censored regression models like those in Bhattacharya, Chernoff, and Yang (1983), Honoré and Powell (1994) and Alan, Honoré, Hu, and Leth-Petersen (2014). The approach can also be used to estimate the Hessian-term in the asymptotic variance of rank correlation estimators like those in Han (1987), Cavanagh and Sherman (1998), Abrevaya (1999), Abrevaya (2000), Khan and Tamer (2007) and Bhattacharya (2008).

A small Monte Carlo study suggests that inference based on the proposed method is more reliable than inference based on simple numerical derivatives, and that it performs as well as inference based on an infeasible second-order numerical derivative that uses the (unknown) optimal stepsize.

2 Background and Insight

Honoré and Powell (1994) derive the asymptotic properties for estimators $\widehat{\theta}$ which minimize an mth–order U-statistic of the form

$$U_n(\theta) = \binom{n}{m}^{-1} \sum_c p\left(z_{i_1}, z_{i_2}, \ldots, z_{i_m}, \theta\right) \tag{1}$$

where z_i is the data for observation i and the sum is taken over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \ldots, i_m\}$ from the set $\{1, \ldots, n\}$. The "kernel" function $p\left(\cdot\right)$ is symmetric in its first m arguments and left- and right-differentiable in each component of θ in some open neighborhood of θ_0, the unique minimizer of $E\left[p\left(z_{i_1}, z_{i_2}, \ldots, z_{i_m}, \theta\right)\right]$.

The estimator $\hat{\theta}$ satisfies an approximate first–order condition of the form

$$Q_n\left(\widehat{\theta}\right) \equiv \sqrt{n} \binom{n}{m}^{-1} \sum_c q\left(z_{i_1}, \ldots, z_{i_m}, \widehat{\theta}\right) = o_p(1),$$

where each component of the vector $q(\cdot)$ is a convex combination of the left- and right-partial derivatives of p with respect to the corresponding component of θ. Assuming random sampling of z_i and weak regularity conditions, the resulting asymptotic distribution of $\widehat{\theta}$ has a limiting normal distribution:

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, H_0^{-1} V H_0^{-1}\right), \tag{2}$$

with

$$V \equiv m^2 E\left[r\left(z_i, \theta_0\right) r\left(z_i, \theta_0\right)'\right] \qquad \text{and} \qquad H_0 \equiv \partial E\left[r\left(z_i, \theta_0\right)\right]/\partial\theta'$$

where $r(z_i, \theta) \equiv E[q(z_i, z_{i_2}, \ldots, z_{i_m}, \theta)|z_i]$. See Theorem 2 of Honoré and Powell (1994).

Consistent estimators of the matrices H_0 and V are needed to carry out large–sample inference on θ_0 based on (2). Estimation of V is unproblematic. Define an estimator of $r(z_i, \theta)$ by

$$\hat{r}(z_i, \theta) \equiv \binom{n-1}{m-1}^{-1} \sum_{c'} q(z_i, z_{i_2}, \ldots, z_{i_m}, \theta),$$

where the sum is now taken over the $\binom{n-1}{m-1}$ combinations of $m-1$ distinct elements $\{i_2, \ldots, i_m\}$ from the set $\{1, \ldots, n\} \setminus \{i\}$. An estimator of V which uses $\hat{r}(z_i, \hat{\theta})$ is

$$\widehat{V}_n \equiv \frac{m^2}{n} \sum_{i=1}^n \hat{r}(z_i, \hat{\theta}) \hat{r}(z_i, \hat{\theta})'. \tag{3}$$

Estimation of H_0 is more problematic because the function $q(\cdot, \theta)$ is discontinuous in many interesting examples. When this is the case, the explicit expression for H_0 will involve densities. One possibility is to estimate these nonparametrically. Alternatively, as noted by Pakes and Pollard (1989), one can use a "numerical derivative" estimator of H_0 even when $q(\cdot, \theta)$ is not differentiable. Specifically, an estimator of the ℓth column of H_0 is given by

$$\widehat{H}_{n\ell} \equiv \left(2\hat{h}\right)^{-1} \left(Q_n\left(\hat{\theta} + \hat{h}e_\ell\right) - Q_n\left(\hat{\theta} - \hat{h}e_\ell\right)\right), \tag{4}$$

where e_ℓ is the ℓth standard basis vector and $\hat{h}$ is a (possibly stochastic) sequence of bandwidths satisfying $\hat{h} = o_p(1)$ and $\left(\sqrt{n}\hat{h}\right)^{-1} = O_p(1)$. See also Hong, Mahajan, and Nekipelov (2015).

Although these estimators of H_0 will be consistent, they are somewhat unattractive because of their dependence on tuning parameters such as bandwidths for nonparametric regression and step-lengths for numerical derivatives. In cases like this, many researchers prefer estimating the variance via a bootstrap. Unfortunately, when m is greater than 1, this is often computationally infeasible because of the large number of terms in (1)[1]. We therefore introduce an alternative approach to estimating H_0.

The new insight exploited in this paper is that one can bootstrap the variance of an alternative estimator with the same H-matrix and then use that variance estimator to back out an estimator of the variance of the estimator of interest. Specifically, Honoré and Powell (1994) mention an alternative "split-sample" estimator, where the minimand involves an average of only $n^\star = \text{int}\,(n/m)$ terms using the same kernel evaluated at "non-overlapping" m-tuples of observations; i.e., the minimand is of the form

$$U_n^\star(\theta) \equiv \frac{1}{n^\star} \sum_{i=1}^{n^\star} p\left(z_i, z_{i+n^*}, \ldots, z_{i+(m-1)n^*}, \theta\right). \tag{5}$$

Since the terms in this minimand are *i.i.d.* for each θ, the usual consistency and asymptotic normality arguments for M-estimators apply, and we have

$$\sqrt{n}\left(\widetilde{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, H_1^{-1}\Omega H_1^{-1}\right), \tag{6}$$

---

[1]Subsampling is also unlikely to be feasible in practice unless the subsamples are very small, in which case one would worry about whether the asymptotics provide good finite sample approximations.

with $\Omega \equiv mE\left[q\left(z_i, z_{i+n^*}, \ldots, z_{i+(m-1)n^*}, \theta_0\right) q\left(z_i, z_{i+n^*}, \ldots, z_{i+(m-1)n^*}, \theta_0\right)'\right]$ and, importantly, $H_1 = H_0$ under random sampling. Note that it is again easy to estimate Ω.

Since minimization of (5) only involves n^* terms, it is typically quite simple to estimate the variance of $\widetilde{\theta}$ via the bootstrap. Letting $Avar\left(\widetilde{\theta}\right) = H_0^{-1}\Omega H_0^{-1}$, this suggests estimating H_0^{-1} by solving the equation

$$\widehat{Avar}\left(\widetilde{\theta}\right) = X\widehat{\Omega}_n X \tag{7}$$

for a symmetric and positive definite X.

Specifically, we propose estimating $H_1^{-1} = H_0^{-1}$ by a bootstrap procedure, where each bootstrap sample is first randomly divided into m subsamples of equal size, $\left[\frac{n}{m}\right]$. An estimator, $\widetilde{\theta}$ is then calculated by minimizing $\sum_{i=1}^{[n/m]} p\left(z_{1i}^b, z_{2i}^b, \ldots, z_{mi}^b, \theta\right)$, where z_{ji}^b is the i'th observation in the j'th subsample of the bootstrap sample. The empirical distribution of $\widetilde{\theta}$ yields an estimator [2] of $\widehat{Avar}\left(\widetilde{\theta}\right)$ in (7). Each bootstrap sample also provides an estimator of Ω, $\widehat{\Omega} = m\sum_{i=1}^{[n/m]} q\left(z_{1i}^b, z_{2i}^b, \ldots, z_{mi}^b, \widetilde{\theta}\right) q\left(z_{1i}^b, z_{2i}^b, \ldots, z_{mi}^b, \widetilde{\theta}\right)'$. The average of these can be used as $\widehat{\Omega}_n$ in (7). Alternatively, Ω can be estimated by randomly splitting the original data set and then constructing $\widehat{\Omega}_n = m\sum_{i=1}^{[n/m]} q\left(z_{1i}, z_{2i}, \ldots, z_{mi}, \widetilde{\theta}\right) q\left(z_{1i}, z_{2i}, \ldots, z_{mi}, \widetilde{\theta}\right)'$. Finally, $\widehat{H}_n$, is the solution to (7), and the variance of the original estimator, $\widehat{\theta}$, is estimated by $\widehat{H}_n^{-1}\widehat{V}_n\widehat{H}_n^{-1}$.

Equations like (7) are special cases of what is known as the continuous time algebraic Riccati equation. See Lancaster and Rodman (1995). Ricatti equations play an important role in control theory, and they are know to have unique positive definite solutions. It is also well understood how to solve Riccati equations numerically.[3] See, for example, Ni (2008) and Arnold and Laub (1984).

The approach pursued here could also be used to estimate the "Hessian" for maximum rank correlation-like estimators like those in Han (1987), Cavanagh and Sherman (1998), Abrevaya (1999), Abrevaya (2000), Khan and Tamer (2007) and Bhattacharya (2008). For

[2]Since convergence in distribution does not imply convergence of moments, we recommend using a robust variance estimator like the one we discuss in Section 3.

[3]In Matlab, this is implemented as a special case of the command CARE. Specifically, if B and Q are symmetric and positive definite matrices, then the unique positive definite solution to $XBX = Q$ is obtained from $[X,L,G]=$CARE$(0^*B,$CHOL$(B)',Q)$.

these U-statistics estimators, the naive split-sample estimators would be an estimator similar to the maximum score estimator of Manski (1975). However, this would not be directly fruitful because this estimator is not asymptotically normally distributed (Cavanagh (1987) and Kim and Pollard (1990)). On the other hand, smoothed versions of the maximum score estimator like the one in Horowitz (1992) will be asymptotically normal with the same Hessian as the maximum rank correlation estimator, although the rate of convergence will be different. The reason why we do not pursue this approach here, is that the matrices in the middle of the variance expressions in (2) and (6) (V and Ω) depend on densities. So estimating them would require tuning parameters such as bandwidths or number of nearest neighbors.

More general just-identified method-of-moments estimators also have "sandwich" asymptotic variance of the form $H^{-1}VH'^{-1}$. The main difference is that in this case, H will typically not be symmetric and positive definite. We therefore can not rely on the results about uniqueness of solutions to Riccati equations. Indeed, it is not difficult to see that equations of the form $\widehat{Avar}\left(\widetilde{\theta}\right) = XVX'$ do not have unique solutions. For example, both

$$\begin{pmatrix} \sqrt{1+\sqrt{3/4}} & \frac{1}{2}\left(\sqrt{1+\sqrt{3/4}}\right)^{-1} \\ \frac{1}{2}\left(\sqrt{1+\sqrt{3/4}}\right)^{-1} & \sqrt{1+\sqrt{3/4}} \end{pmatrix}$$

and $\begin{pmatrix} \sqrt{2} & 0 \\ \sqrt{1/2} & \sqrt{3/2} \end{pmatrix}$ will be solutions when $\widehat{Avar}\left(\widetilde{\theta}\right) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. In other words, for method-of-moments estimators, it is not possible to identify the "Hessian"-term, H^{-1}, from the equation $\Sigma = H^{-1}VH'^{-1}$.

3   Monte Carlo Illustration

We illustrate the usefulness of the approach by conducting a small Monte Carlo study of one of the pairwise comparison estimators in Honoré and Powell (1994). In this section, we will implicitly assume that the researcher considers the bootstraps the "gold standard" and that the goal is to compare it with more easily computable alternatives. Of course, in practice, the behavior of test-statistics also depends on how well the finite sample distribution of the estimator is approximated by its limiting normal distribution. It is possible that the estimation error of a poor variance estimator can offset this approximation error and lead to better finite sample performance of the resulting test-statistics. For this reason, we will

think of the bootstrap as the benchmark in the discussion below.

Honoré and Powell (1994) propose estimators of censored and truncated regression models. Here we focus on the censored regression model defined by $y_i = \max(0, x_i'\beta + \varepsilon_i)$ where ε_i is independent of x_i. Honoré and Powell (1994) show that subject to regularity conditions, β is the unique minimizer of $E[s(y_i, y_j, (x_i - x_j)'b)]$ where

$$
s(y_1, y_2, \delta) = \begin{cases}
\Xi(y_1) - (y_2 + \delta)\xi(y_1) & \text{for} & \delta < -y_2 \\
\Xi(y_1 - y_2 - \delta) & \text{for} & -y_2 \leq \delta \leq y_1 \\
\Xi(-y_2) - (\delta - y_1)\xi(-y_2) & \text{for} & y_1 < \delta
\end{cases}
$$

and $\Xi'(d) = \xi(d)$ (for almost all d) for some convex function Ξ. Natural choices for Ξ are $\Xi(d) = d^2$ and $\Xi(d) = |d|$.

They therefore define an estimator of the censored regression model by minimization of

$$
S_n(b) = \binom{n}{2}^{-1} \sum_{i<j} s(y_i, y_j, (x_i - x_j)'b). \tag{8}
$$

Note that $S_n(b)$ is a convex function of b. $S_n(b)$ is differentiable provided that Ξ is. If Ξ is not differentiable at 0, then the points of non–differentiability of $S_n(b)$ occur when $y_i - y_j - (x_i - x_j)'b = 0$ and either $y_i \neq 0$ or $y_j \neq 0$ or both.

This section presents the results from a Monte Carlo experiment designed to investigate the performance of the variance estimator proposed here. We consider the estimator proposed in Honoré and Powell (1994) with $\Xi(d) = d^2$ and $\Xi(d) = |d|$. When $\Xi(d) = d^2$ it is easy to estimate the asymptotic variance of the estimator without having to choose a bandwidth. We include it to illustrate the relative performance of our approach in a setting where we expect other feasible methods to perform well.

There are four explanatory variables. They are all dependent and have the following marginal distributions. x_1 is binary, x_2 is chi-square, x_3 is normal with probability $1/2$ and 0 otherwise, and x_4 is normal. All the variables are normalized to have mean 0 and variance 1 and the correlation structure for the four variables is constructed as follows. First we generate x^* from a four-dimensional normal distribution with means 0, variances 1 and correlations 0.5. We then define $x_1 = 2 \cdot 1(x_1^* > 0) - 1$; $x_2 = (x_2^{*2} - 1)/\sqrt{2}$, $x_3 = \sqrt{2}x_3^*D$ where D is independent of x^* and Bernoulli with $p = \frac{1}{2}$. Finally $x_4 = x_4^*$.

7

The error, ε_i, is independent of the other variables and t-distributed with three degrees of freedom and β is a vector of ones. This gives an R^2 in an uncensored version of the model of approximately 0.48. There is no additional constant in the data generating process. This results in a censoring probability of approximately 51%.

The Monte Carlo experiment has 400 replications, and the results are presented in Tables 1 and 2. All bootstraps are performed using 1000 bootstrap samples. In each replication, we calculate the estimators defined by minimizing (8) with $\Xi(d) = d^2$ (Table 1) and with $\Xi(d) = |d|$ (Table 2). When $\Xi(d) = d^2$, $S_n(b)$ is almost everywhere twice differentiable so there is no particular reason to use the approach proposed here because one can use analytical derivatives. Variance estimation based on the sample analogs of this derivative is labelled "Analyt." We also estimate the variance of $\widehat{\beta}$ by the bootstrap (labelled "Boots.") and the Riccati bootstrap proposed here ("Riccati"). It is well understood that asymptotic normality does not imply that moments are finite. We therefore use a robust estimator for the bootstrap variances. Specifically, with a sample of B bootstrap estimators, $\widetilde{\beta}_1, ..., \widetilde{\beta}_B$, we estimate the variance of the ℓ'th coordinate, $\widetilde{\beta}(\ell)$, by the square of 1.34 times the interquartile range of $\widetilde{\beta}_1(\ell), ..., \widetilde{\beta}_B(\ell)$. To estimate the correlation matrix, we transform the ℓ'th coordinate of $\widetilde{\beta}_b(\ell)$ into $\Phi^{-1}\left(\frac{R_B(\widetilde{\beta}_b(\ell))}{B+1}\right)$ where $R_B\left(\widetilde{\beta}_b(\ell)\right)$ is the rank of $\widetilde{\beta}_b(\ell)$. We then use the sample correlation of $\Phi^{-1}\left(\frac{R_B(\widetilde{\beta}_b(\ell))}{B+1}\right)$ to estimate the correlation matrix of $\widetilde{\beta}$. If the empirical distribution of $\widetilde{\beta}$ is closely approximated by a normal distribution, then this correlation matrix will be very similar to the empirical correlation of $\widetilde{\beta}$. The advantage is that the former is much less sensitive to outliers. The estimators of the variance based on non-robust bootstrap and Riccati bootstrap variance estimation are labelled "Boots.(NR)" and "Ricc.(NR)", respectively. Finally, we estimate H by (two-sided) numerical derivatives of the analytical derivative of $s(y_i, y_j, (x_i - x_j)'b)$ using stepsizes $h_\ell = h/\operatorname{std}(x_\ell)$, for $h = 0.05$, 0.25 and 0.50. The bootstrap draws for the Riccati bootstrap are independent of the draws for the regular bootstrap. The numerical derivative estimator of H is $\frac{1}{2}\left(\hat{H} + \hat{H}'\right)$ where $\hat{H}$ is given in (4).[4]

---

[4]We make the estimator of H symmetric to ensure that $\widehat{H}^{-1}\widehat{V}\widehat{H}^{-1}$ is positive definite (provided that $\widehat{H}$ is invertible). Alternatively we could have constructed $\widehat{H}^{-1}\widehat{V}\left(\widehat{H}^{-1}\right)'$ or $\left(\widehat{H}^{-1}\right)'\widehat{V}\widehat{H}^{-1}$.

Since variance estimation is typically used for inference, we first calculate the "t-test" statistic, T, for whether each coefficient equals its true value using each of the variance estimators. We then construct $\Phi(T)$ where Φ is the normal cumulative distribution function. This corresponds to the p-value for a one-sided test. Since the aim of this paper is to construct a more convenient alternative to the bootstrap variance, we compare the p-values using the different variance estimators.

Tables 1 and 2 report the Monte Carlo average of the absolute difference in percentage points between the p-values based on any pair of the variance estimates averaged over all four parameters when $n = 100$. The results in Table 1 are all expected. The objective function that defines the estimator is almost everywhere twice differentiable, so there is almost no difference between calculating the Hessian and approximating it numerically. This is especially true when the stepsize is small. Both the bootstrap and Riccati bootstrap yield p-values that are within 2 to 3 percentage points of each other and of those from the analytical and numerical derivative estimators. The only outlier is the Riccati bootstrap estimator based on non-robust estimation of the variance. This is not surprising since the estimator based on $\Xi(d) = d^2$ is related to the OLS estimator in a linear regression model and the errors have fairly fat tails.[5] Table 2 is more interesting because $\Xi(d) = |d|$ gives an estimator for which we cannot estimate the Hessian term in the variance by analytical differentiation. In this case, the Ricatti bootstrap does a much better job approximating the bootstrap than the numerical derivative estimators do.

Finally, we investigate whether the relatively poor performance of the numerical derivative estimators is due to unwise choices of stepsizes. The role of the stepsize in numerical differentiation is akin to the role of the bandwidth in nonparametric kernel estimation. This raises two issues. One is that the results can be sensitive to the choice of stepsize, and the other is that it is possible to reduce bias by using higher order numerical derivatives. See for example Hong, Mahajan, and Nekipelov (2015). We therefore calculate the numerical

---

[5]Without censoring, the U-statistic estimator equals the slope of the OLS estimator of y on x and a constant, while the split-sample estimator would be the OLS estimator of the difference in the y on the difference in the x (without a constant). Since the errors are distributed according to a t-distribution with 3 degrees of freedom, these OLS estimators will have poorly-behaved sample moments.

derivative estimator of the Hessian as above using stepsizes $h_\ell = h/\operatorname{std}(x_\ell)$ for a range of h. In addition, using the notation from Section 2, we also consider the second order numerical derivative of Q_n defined by $\frac{1}{2}\left(\widetilde{H}_n + \widetilde{H}_n'\right)$ where the ℓ'th columns of $\widetilde{H}_n$ is

$$\widetilde{H}_{n\ell} \equiv \hat{h}^{-1}\left(-\frac{1}{12}Q_n\left(\hat{\theta}+2\hat{h}e_\ell\right) + \frac{2}{3}Q_n\left(\hat{\theta}+\hat{h}e_\ell\right) - \frac{2}{3}Q_n\left(\hat{\theta}-\hat{h}e_\ell\right) + \frac{1}{12}Q_n\left(\hat{\theta}-2\hat{h}e_\ell\right)\right).$$

Like the higher-order kernel estimators, this construction of $\widetilde{H}_n$ removes the first order term in the bias of $\widehat{H}_n$.

We calculate the average absolute difference between the p-values calculated from the bootstrap and the numerical derivative estimators that use $\widehat{H}_n$ and $\widetilde{H}_n$ with stepsizes $h_\ell = h/\operatorname{std}(x_\ell)$ for h between 0 and 2. These are presented in Figures 1-4 for sample sizes 50, 100, 200 and 400. The horizontal (fatter) line represents the Monte Carlo average of the absolute difference between the p-values calculated from the bootstrap and the Riccati bootstrap, while the thinner solid and the dashed lines represent the Monte Carlo average of the absolute difference between the p-values calculated from the bootstrap and those using $\widetilde{H}_n$ and $\widehat{H}_n$, respectively. The figures suggest that the Riccati bootstrap provides a better approximation to the bootstrap than approaches based on numerical derivative approximations to the Hessian regardless of the choices of stepsize.

4 Conclusion

The bootstrap is a popular and useful tool for estimating the asymptotic variance of complicated estimators. Unfortunately, the standard bootstrap can be computationally burdensome because it requires repeated re-calculation of the estimator. In Honoré and Hu (2017), we proposed a modification to the standard bootstrap that replaces re-estimation of the original parameter vector with re-estimation of one-dimensional estimators. In this paper, we focus on estimators based on minimization of U-statistics. Our contribution here is to replace re-estimation of the original parameter vector with re-estimation of estimators based on minimization of simple averages. We perform a small Monte Carlo experiment for an estimator whose asymptotic variance involves unknown functions. In this case, the common alternative to the standard bootstrap is to estimate the Hessian in the asymptotic variance

by numerical derivatives. We find that our approach performs better than this alternative regardless of the choice of tuning parameter of the numerical derivative.

References

ABREVAYA, J. (1999): "Leapfrog Estimation of a Fixed-Effects Model with Unknown Transformation of the Dependent Variable," *Journal of Econometrics*, 93, 203–228.

——— (2000): "Rank Estimation of a Generalized Fixed Effects Regression Model," *Journal of Econometrics*, 95, 1–23.

ADICHIE, J. N. (1967): "Estimates of Regression Parameters Based on Rank Tests," *The Annals of Mathematical Statistics*, 38(3), pp. 894–904.

ALAN, S., B. E. HONORÉ, L. HU, AND S. LETH-PETERSEN (2014): "Estimation of Panel Data Regression Models with Two-Sided Censoring or Truncation," *Journal of Econometric Methods*, 3(1), 1–20.

ARNOLD, W.F., I., AND A. LAUB (1984): "Generalized eigenproblem algorithms and software for algebraic Riccati equations," *Proceedings of the IEEE*, 72(12), 1746–1754.

BHATTACHARYA, D. (2008): "A Permutation-Based Estimator for Monotone Index Models," *Econometric Theory*, 24(3), pp. 795–807.

BHATTACHARYA, P. K., H. CHERNOFF, AND S. S. YANG (1983): "Nonparametric Estimation of the Slope of a Truncated Regression," *The Annals of Statistics*, 11(2), pp. 505–514.

CAVANAGH, C. (1987): "The Limiting Behavior of Estimators Defined by Optimization," Harvard University.

CAVANAGH, C. L., AND R. P. SHERMAN (1998): "Rank Estimators of Monotonic Index Models," *Journal of Econometrics*, 84, 351–381.

HAN, A. (1987): "Nonparametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

HONG, H., A. MAHAJAN, AND D. NEKIPELOV (2015): "Extremum estimation and numerical derivatives," *Journal of Econometrics*, 188(1), 250 – 263.

HONORÉ, B. E., AND L. HU (2017): "Poor (Wo)man's Bootstrap," *Econometrica*, Forthcoming.

HONORÉ, B. E., AND J. L. POWELL (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64(1-2), 241–78.

HOROWITZ, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

KHAN, S., AND E. TAMER (2007): "Partial rank estimation of duration models with general forms of censoring," *Journal of Econometrics*, 136(1), 251 – 280.

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Ann. Statist.*, 18(1), 191–219.

LANCASTER, P., AND L. RODMAN (1995): *Algebraic Riccati Equations*. Oxford University Press, USA.

MANSKI, C. F. (1975): "The Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

NI, M.-L. (2008): "Existence Condition on Solutions to the Algebraic Riccati Equation," *Acta Automatica Sinica*, 34(1), 85 – 87.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

SIEVERS, G. L. (1978): "Weighted Rank Statistics for Simple Linear Regression," *Journal of the American Statistical Association*, 73(363), pp. 628–631.

Table 1: Difference in p-values in Percentage Points Using Different Variance Estimators. $\Xi(d) = d^2$, $n = 100$.

	Analyt.	Boots.	Riccati	$h = 0.05$	$h = 0.25$	$h = 0.50$	Boots.(NR)	Ricc.(NR)
Analyt.	0.00	1.25	2.77	0.07	0.17	0.36	2.17	6.28
Boots.	1.25	0.00	2.05	1.25	1.23	1.20	1.39	5.37
Riccati	2.77	2.05	0.00	2.77	2.76	2.73	1.79	3.83
$h = 0.05$	0.07	1.25	2.77	0.00	0.15	0.35	2.17	6.27
$h = 0.25$	0.17	1.23	2.76	0.15	0.00	0.23	2.15	6.26
$h = 0.50$	0.36	1.20	2.73	0.35	0.23	0.00	2.10	6.24
Boots(NR)	2.17	1.39	1.79	2.17	2.15	2.10	0.00	4.36
Ricc.(NR)	6.28	5.37	3.83	6.27	6.26	6.24	4.36	0.00

Table 2: Difference in p-values in Percentage Points Using Different Variance Estimators. $\Xi(d) = |d|$, $n = 100$.

	Analyt.	Boots.	Riccati	$h = 0.05$	$h = 0.25$	$h = 0.50$	Boots.(NR)	Ricc.(NR)
Analyt.	NA	NA	NA	NA	NA	NA	NA	NA
Boots.	NA	0.00	1.67	4.66	4.28	4.34	1.10	2.24
Riccati	NA	1.67	0.00	4.05	3.60	3.82	1.64	1.26
$h = 0.05$	NA	4.66	4.05	0.00	2.03	2.25	4.23	3.41
$h = 0.25$	NA	4.28	3.60	2.03	0.00	0.89	3.78	2.78
$h = 0.50$	NA	4.34	3.82	2.25	0.89	0.00	3.72	2.81
Boots.(NR)	NA	1.10	1.64	4.23	3.78	3.72	0.00	1.63
Ricc.(NR)	NA	2.24	1.26	3.41	2.78	2.81	1.63	0.00

Figure 1: Relative Performance of Numerical Derivatives vs. Riccati Bootstrap. $n = 50$
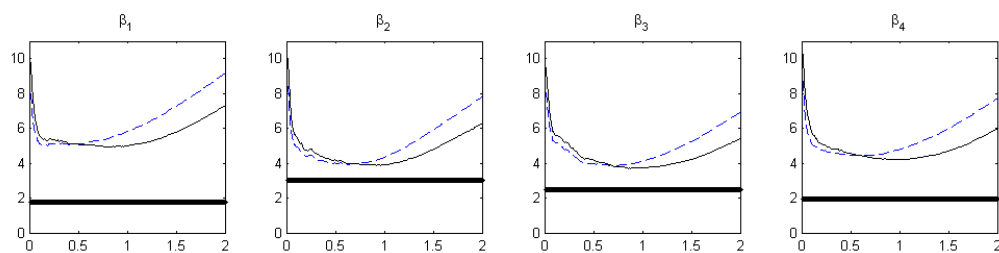

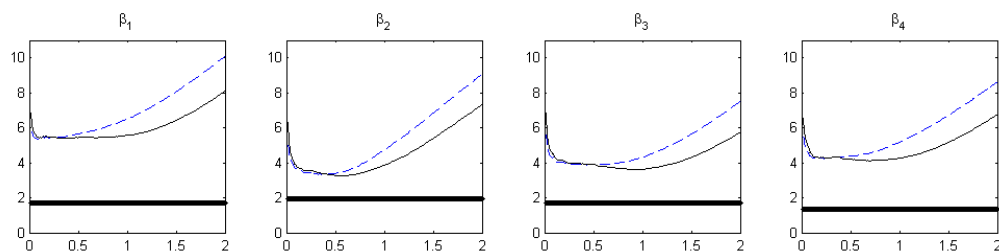Figure 2: Relative Performance of Numerical Derivatives vs. Riccati Bootstrap. $n = 100$

Figure 3: Relative Performance of Numerical Derivatives vs. Riccati Bootstrap. $n = 200$
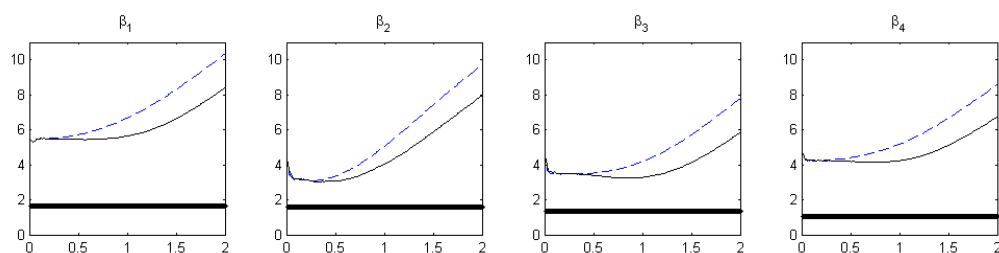

Figure 4: Relative Performance of Numerical Derivatives vs. Riccati Bootstrap. $n = 400$