December 28, 2000 **Draft - Comments Welcome – Do not quote without permission**
For JEP Econometrics Symposium

The bootstrap and multiple imputations*

David Brownstone
Department of Economics
University of California
Irvine, CA 92697-5100
Email: dbrownst@uci.edu

and

Robert Valletta
Federal Reserve Bank of San Francisco
and OECD
2, rue André-Pascal
75775 Paris Cedex 16 France
Email: rob.valletta@oecd.org

* The views expressed in this paper are solely those of the authors and are not in any way endorsed by the Federal Reserve or the OECD. The authors are also solely responsible for any errors or omissions.

1 Introduction

This article describes two relatively new techniques for calculating confidence bands and critical values for econometric estimators and test statistics. The bootstrap and multiple imputations are both computationally intensive methods that have been developed by statisticians in the last twenty years. After a ten-year lag, the bootstrap has become quite popular in applied econometrics. Unfortunately, not all users of bootstrap methodology are aware of its limitations or the best way to implement bootstrap procedures. We will try to clarify which bootstrap procedures are best for particular models as well as point out situations where the bootstrap fails to provide consistent inference.

The multiple imputation technique was developed by Rubin (1987) as a general method for inference with missing data. Since formally measurement error is just a case of missing values on the uncontaminated data, multiple imputations also can be used with measurement error models. Econometricians traditionally have modeled measurement error and missing data as simultaneous equations with latent variables. These models can be very difficult to estimate and therefore are rarely applied by non-specialists. By contrast, multiple imputations typically can be implemented in standard econometric software packages, thereby providing a much simpler approach for consistent inference in these difficult problems.

Both multiple imputations and the bootstrap are techniques for deriving confidence bands and critical values for test statistics, although each also can be used to remove biases in estimators. If they are applied to bad estimators or tests, then they will give bad results. Like many econometric techniques, the validity of these methods is based on asymptotic approximations, and these underlying approximations may be inaccurate for a particular data and model. In spite of these caveats, both of these techniques are going to be important parts of applied econometricians' tool kits.

This paper will concentrate on the standard linear model, $y = X\mathbf{b} + \mathbf{e}$, using the least squares estimator $\hat{\mathbf{b}} = \left(X'X\right)^{-1}X'y$. We will assume that X and $\mathbf{b}$ are $(T\times K)$ and $(K\times 1)$ matrices of full rank with $T>K$. $\mathbf{e}$ satisfies $E(\mathbf{e}|X) = 0$, the mean of $\mathbf{e}$ is zero, and its variance is finite. The linear model is the most commonly used model in applied econometrics, although both the bootstrap and multiple imputations have been applied with more general models.

2 Bootstrap

Efron's (1979 and 1982) bootstrap is a repeated application of the analogy (or plug-in) principle long used to motivate method of moment estimators. If F is the population distribution generating an observed random sample of size T, then both the estimator $\hat{\mathbf{b}}$ and its sampling distribution, G, can be considered functions of F. Efron suggests replacing F by a consistent estimate, $\hat{F}$. Most applications and theoretical work use the empirical distribution function that places mass $1/T$ on each observation in the random sample. If $\boldsymbol{b}$ is a moment of *F*, then this analogy corresponds to $\hat{\boldsymbol{b}}$ being given by the corresponding sample moment. Since it is rarely possible to directly evaluate the estimated sampling distribution as a function of the empirical distribution function, Efron further proposed evaluating the bootstrap distribution, $G(\hat{F})$, by Monte Carlo simulation. When applied to estimating the population mean from a random sample, Efron's bootstrap consists of repeatedly resampling with replacement from the observed sample, computing the sample mean from each such "bootstrap" sample, and then using the distribution of the resulting resampled means to approximate the sampling distribution of the original sample mean.

The standard asymptotic approximation used in econometrics uses the central limit theorem to approximate the distribution of sums of independent random variables that frequently occur in econometric estimators. For example, the standard asymptotic distribution of the least squares estimator is based on the asymptotic normality of $X'\boldsymbol{e}/\sqrt{T}$. This is different from the direct approximation of *F* used in the bootstrap, but both the bootstrap and the standard method are only asymptotic approximations. Horowitz (1997) points out that the different nature of the bootstrap approximation provides a rough check on the adequacy of the asymptotic theory for cases where both the bootstrap and standard asymptotic inferences are consistent. If the asymptotic theory is accurate, then both the bootstrap and standard methods should yield the same answers. If there are significant differences, then this is evidence that one or both of the methods is not reliable for the particular problem and data. Of course, without more detailed analysis we can't tell which method, if any, is appropriate when they disagree. It is also possible that both methods agree and both are inaccurate.

There is nothing unique about the bootstrap method described above, and there are generally many possible implementations for a particular model. For example, if we further assume that *X* in the standard linear model is fixed and that $\boldsymbol{e}$ are independent but not identically distributed, then there are at least four reasonable ways to implement the bootstrap to approximate the sampling distribution of the least squares estimator, $\hat{\boldsymbol{b}}$. Note that in the rest of this section, the superscript "*" denotes something calculated from the approximate bootstrap distribution $\hat{F}$.

1. XY or paired bootstrap; Randomly sample pairs (y_i, X_i) with replacement and recompute the least squares estimator $\hat{\boldsymbol{b}}^*$ for each sample. This corresponds to $\hat{F}$ being the empirical distribution function of the pairs (y_i, X_i).
2. Nonparametric residual bootstrap; Randomly sample with replacement from the least squares residuals $e = y - X\hat{\boldsymbol{b}}$ to yield e^*. Compute the bootstrap sample $y^* = X\hat{\boldsymbol{b}} + e^*$ and recompute the least squares estimator $\hat{\boldsymbol{b}}^*$ for each sample. This corresponds to $\hat{F}$ being the empirical distribution function of the least squares residuals.
3. Parametric residual bootstrap; Generate a resampled residual vector e^* by independently sampling from a normal distribution with mean zero and variance

$s^2 = (y - X\hat{\boldsymbol{b}})'(y - X\hat{\boldsymbol{b}})/(T-K)$. Generate bootstrap samples y^* as in 2 above. This corresponds to $\hat{F}$ being the normal distribution function corresponding to the usual least squares estimates.

4. Wild Bootstrap (originally called the weighted bootstrap and due to Wu, 1986); For each observation *i* in the least squares residual vector generate the two-point distribution $\hat{F}_i$ which equals $(1-\sqrt{5})e_i/2$ with probability $(1+\sqrt{5})/(2\sqrt{5})$ and $(1+\sqrt{5})e_i/2$ with probability $1-(1+\sqrt{5})/(2\sqrt{5})$ (note that the mean of $\hat{F}_i$ is zero, the variance is e_i^2, and the third moment equals e_i^3). Now generate a resampled residual vector e_i^* by sampling from $\hat{F}_i$ and continue as in 2 above to compute the bootstrap samples. The $\hat{F}_i$ are similar to the approximate distribution of the true residuals $\boldsymbol{e}_i$ used to derive Eicker-White robust standard errors.

If the errors, $\boldsymbol{e}$, are homoskedastic, then the paired, nonparametric residual, and wild bootstrap methods will generate consistent estimates of the sampling distribution of the least squares estimator. Parametric residual bootstrapping is only consistent if the errors follow a normal distribution, in which case it will yield the most accurate estimates.

If the errors are heteroskedastic and the heteroskedasticity depends on *y* and/or *X*, then both parametric and nonparametric residual bootstrapping will be inconsistent since the bootstrapped residuals will not reflect this heteroskedasticity. Both paired bootstrapping and wild bootstrapping will provide consistent inference with heteroskedasticity, but wild bootstrapping will generally be more accurate. The wild bootstrap imposes the moment restriction E(e^* |*X*) = 0 on each bootstrap sample. Paired bootstrap samples satisfy E(e^* | X^*) = 0, where X^* denotes the resampled rows of the original *X* matrix. Since the linear model is identified by the assumption that E($\boldsymbol{e}$ | X) = 0, the wild bootstrap is imposing the correct moment restriction. Mammen (1992, Chapter 8) shows that the distribution of the wild bootstrap converges faster than the paired bootstrap. Section 5.2 in Horowitz (2001) gives Monte Carlo evidence supporting the superiority of the wild bootstrap for carrying out a *t* test for the least squares estimator in the heteroskedastic linear model.

2.1 Key results for consistency of bootstrap estimates

In addition to the intuitive requirement that the bootstrap distribution, $\hat{F}$, be a consistent estimator of the underlying population distribution, the consistency of a bootstrap approximation also depends on the continuity of the mapping between the population distribution and the distribution of the statistic we are examining. Horowitz (2001, Section 2.1) gives precise results, and these continuity conditions require that small changes around the population distribution, *F*, do not lead to large changes in *G* considered as a function of *F*.

If we are interested in the slope parameter of a least squares regression, then this mapping is generated by the linear relationship between $\hat{b}$ and e and therefore satisfies the continuity conditions. However, if we add the restriction that $b \geq 0$, then Andrews (1997) shows that all of the bootstrap methods described above will fail. The problem here is that the mapping between the population distribution *F* and the distribution of the restricted least squares estimator is discontinuous at $b = 0$. Andrews' counterexample applies to many situations where the true parameter lies on the boundary of the parameter space. A similar problem occurs when bootstrapping a possibly non-stationary first order autoregressive process. The sampling distribution of the least squares estimator in this case changes discontinuously as the population autocorrelation crosses one. This causes the bootstrap distribution to be inconsistent even though the least squares residuals are consistently estimated.

Bootstrapping has been applied to shrinkage estimation and model selection problems where the continuity conditions for consistent bootstrap inference are violated. Monte Carlo studies reviewed in Brownstone and Kazimi (2000) show that bootstrap methods perform reasonably well in spite of their theoretical shortcomings. For both of these problems there are no useful asymptotic alternatives, which is why investigators have been applying the bootstrap. However, since the conditions for consistency are violated there is no reason to believe that the bootstrap will perform well outside the limited range investigated by the Monte Carlo studies.

For almost all cases where the bootstrap fails due to violations of the continuity conditions, bootstrap subsampling restores consistency of bootstrap inference. Bootstrap subsampling takes bootstrap sample sizes smaller than the original sample size. Horowitz (2001, Section 2.2) reviews the theoretical literature on subsampling, but both theory and some Monte Carlo evidence on the performance of bootstrap subsampling in Brownstone and Kazimi (2000) suggest that bootstrap subsampling is less accurate than regular bootstrapping when both are valid.

Finally, the bootstrap approximation is different from the approximations inherent in usual first-order asymptotic theory. Asymptotic theory typically approximates the unknown distribution *G* with a known function G_∞ while bootstrap methods approximate the population distribution *F*. The conditions for the consistency of these methods are also different, although for a wide variety of applied econometric models both methods yield consistent inference. Horowitz (1997) points out that the different nature of the bootstrap approximation provides a rough check on the adequacy of the asymptotic theory for cases where both the bootstrap and standard asymptotic inferences are consistent. If the asymptotic theory is accurate, then both the bootstrap and standard methods should yield the same answers. If there are significant differences, then this is evidence that one or both of the methods is not reliable for the particular problem and data. Of course, without more detailed analysis we can't tell which method, if any, is appropriate when they disagree. It is also possible that both methods agree and are both inaccurate.

2.2 Asymptotic refinements

The conditions required to establish the consistency of the bootstrap do not allow us to examine the efficiency of bootstrap methods. Horowitz (2001, Section 3) shows that if the statistic b is a smooth function of moments of *F* (called the Smooth Function Model or SFM) then it is possible to derive asymptotic efficiency results for bootstrap estimates. Horowitz points out that if the likelihood function or moment conditions have sufficiently many derivatives, then maximum likelihood and generalized-method-of-moments estimators satisfy the SFM.

With some additional assumptions, Hall (1992) shows that bootstrap approximations converge at the rate $\sqrt{T}$. This means that the difference between the bootstrap confidence bands and the true confidence bands goes to zero at the same rate as $1/\sqrt{T}$. Since this is the same rate that standard asymptotic approximations converge, we should not expect bootstrap methods to improve the rate of convergence. However, Hall (1992) shows that if the bootstrap is applied to asymptotically pivotal statistics, then the rate of convergence of the bootstrap is increased to *T* for one-sided distributions and to $T^{3/2}$ for symmetrical distributions. An asymptotically pivotal statistic has the property that its asymptotic distribution doesn't depend on the unknown population distribution *F*. This is the theoretical justification for the claim that the bootstrap provides better small sample performance than traditional asymptotic inference procedures.

Many test statistics used in applied econometrics, including most likelihood ratio, Wald, and Lagrange multiplier tests, are asymptotically pivotal. Theory, and a fair amount of Monte Carlo evidence and empirical examples, suggest that bootstrapping these statistics will provide improvements equivalent to using second-order Edgeworth expansions in standard asymptotic theory. In most cases, bootstrapping asymptotically pivotal statistics is much easier than implementing inference based on second-order asymptotic expansions. For example, to test the null hypothesis that the first component of b in the standard linear model, $b_1 = 0$, bootstrapping the *t*-statistic: $\hat{b}_1 \Big/ \sqrt{s^2 (X'X)^{-1}_{1,1}}$ will provide asymptotic refinements over just bootstrapping $\hat{b}_1$. Note that although $\hat{b}$ is computed on the observed data, the bootstrap samples should be drawn from a model estimated under the null hypothesis that $b_1 = 0$.

Unfortunately, implementing a higher-order Edgeworth expansion does not always guarantee better finite-sample performance. There are important applications in econometrics where asymptotically pivotal statistics don't exist or the higher order terms in the Edgeworth expansion dominate the initial terms. Shrinkage, or Stein-rule, estimators are an example of the latter problem. Kazimi and Brownstone (1999) show that bootstrapping *t* statistics for shrinkage estimators is worse than simply bootstrapping the parameter values. A similar problem can occur if the asymptotically pivotal statistic is not a smooth function of the population distribution *F*.

2.3 Bootstrap confidence regions and critical values

Since the bootstrap produces a large sample from the sampling distribution of a statistic, the simplest way to generate confidence intervals for this statistic is to take quantiles from this sample. Efron (1987) shows that this method, called the simple percentile interval, generates intervals with correct asymptotic coverage which are invariant to monotone transformations. To derive the simple percentile interval for the first component of b, b_1, in the standard linear model we would first obtain *R* ordered bootstrap replications { $\hat{b}_{11}^{*}, \ldots, \hat{b}_{1R}^{*}$ } using any consistent bootstrap resampling method. The level a percentile confidence interval would then be given by [$\hat{b}_{1(R a/2)}^{*}, \hat{b}_{1(R(1-a/2))}^{*}$]. Note that unlike standard asymptotic confidence intervals, these bootstrap percentile intervals will not generally be symmetric around the underlying parameter estimate. This feature is useful for situations where the true sampling distribution may not be symmetric.

Confidence regions for more than one parameter can frequently be based on bootstrapped critical values from the Wald test statistic for the null hypothesis that $b = \hat{b}$. For the linear model this would involve generating *R* ordered bootstrap replications of the Wald statistic: $W^{*} = \left(\hat{b}^{*} - \hat{b}\right)' V^{*-1}\left(\hat{b}^{*} - \hat{b}\right)$ where *V* is an estimate of the covariance of $\hat{b}$. The level a confidence ellipsoid is then given by: $\left\{ b : \left(\hat{b} - b\right)' V^{-1}\left(\hat{b} - b\right) \leq W_{R(1-a)}^{*} \right\}$. If b is a scalar, then this region is just Hall's (1992) "percentile-*t*" interval.

The elliptical confidence regions generated by this procedure may be inaccurate if the true distribution of the estimator is not approximately multivariate normal. However, in many econometric applications the Wald statistic is asymptotically pivotal. Hall's results above then imply that the bootstrap elliptical intervals will converge faster than the typical first order asymptotic intervals based on the Chi-squared distribution. Horowitz (2001, Sections 3.4 and 3.5) stresses the practical importance of using asymptotically pivotal statistics, and he provides some Monte Carlo evidence for their improved performance.

Practical application of confidence regions based on asymptotically pivotal Wald statistics require accurate covariance estimators, *V*. If a good variance estimator is not available, then the covariance can be estimated using a second level bootstrap. This requires drawing another set of bootstrap resamples based on each realization, $\hat{b}^{*}$, in the original bootstrap sample. This nested bootstrap procedure can become computationally intractable for large applications with complicated estimators requiring numerical optimization techniques.

2.4 Cross Section

Bootstrapping has been applied to many estimators used with cross-section data. Since cross-section data are independent across observations, the paired bootstrap is almost always appropriate. In many models, residual bootstrap methods also can be used. For example, all of the bootstrapping methods introduced for the linear regression model in the beginning of this section can be applied to the nonlinear regression model. The residuals need to be recentered to have mean zero before residual bootstrap methods can be used. The standard asymptotic inference for nonlinear regression begins by linearizing the regression function, but bootstrap methods evaluate the appropriate transformations by Monte Carlo simulation. Therefore we would expect bootstrap methods to be more accurate as this linearization becomes less accurate. Horowitz (2001, Section 5.3) verified this in a small Monte Carlo study of a Box-Cox regression model. He found that asymptotic critical values for *t* tests for a slope coefficient were strongly biased, but the bootstrap critical values were essentially correct.

Quantile regression techniques were introduced to econometricians by Koenker and Bassett (1978), and they have recently been used by Buchinsky (1994) and Chamberlain (1994) to examine changes in the U.S. wage distribution using Current Population Survey (CPS) data. Buchinsky and Chamberlain used a censored quantile estimator given by Powell (1986) to account for topcoding in the wage data. They used paired bootstrapping to estimate the covariances of their slope parameter and quantile estimators. Buchinsky (1995) carries out a large Monte Carlo study of various covariance estimators for quantile estimator parameters under heteroskedasticity. Buchinsky's Monte Carlo design is based on one year of the CPS data used in Buchinsky (1994) and Chamberlain (1994), and he concludes that the paired bootstrap performs well (and better than the other methods in the study) even when the errors are homoskedastic. Note that the wild bootstrap is not immediately applicable to quantile estimation. The wild bootstrap matches moments of the observed residuals, but the existence of these moments are not required for the consistency of quantile regression estimators.

Horowitz (2001, Section 4.3.1) shows that the smooth function model doesn't apply to quantile regression estimators. He develops smoothed versions of these estimators, and he shows that paired bootstrapping of pivotal statistics yield asymptotic refinements. Horowitz (2001, Section 4.3.2) also develops a smoothed version of Manski's Maximum Score estimator for discrete choice models which satisfies the smooth function model.

Kernel density and nonparametric mean regression estimators are not smooth functions of sample moments and the estimators converge more slowly than $\sqrt{T}$. Therefore the asymptotic refinements from proper bootstrap methods will be of different orders, and there are subtle issues of how to implement bootstrap methods (see Horowitz, 2001, Section 4.2). For example, proper bootstrap methods must remove the bias typically present in kernel density estimators. This bias is best removed by resampling from an undersmoothed kernel density estimator. With some assumptions on the smoothness of the underlying density or mean regression function, either paired or residual sampling from this undersmoothed kernel density will yield asymptotic refinements when applied to asymptotically pivotal statistics.

Brownstone and Kazimi (2000, Section 3) survey applications of bootstrap inference in applied cross-section econometrics (such as DEA frontier estimates) where there is no obvious asymptotic alternative. While it is certainly possible that a good theoretical econometrician could develop a non-bootstrap inference procedure for these cases, one of the strengths of the bootstrap is the possibility of avoiding unpleasant mathematical statistics. Unfortunately the bootstrap can be applied in situations where it is not consistent, and checking the consistency conditions given in Section 2.1 can be difficult. Most of the applications verified the validity of the bootstrap using a small Monte Carlo study.

2.5 Time Series

The bootstrapping methods for the linear model described earlier are only appropriate if the e are not serially correlated. Since this assumption is violated in most econometric applications with time series data, we need to explicitly consider methods for bootstrapping time series data. The simplest approach, called model-based bootstrapping, is to assume that the time series can be fit by an appropriate ARMA model with white noise (or at least serially uncorrelated) residuals. These estimated residuals can then be bootstrapped using any of the methods described at the beginning of this section, except that the method of generating the bootstrap samples from these bootstrapped residuals needs to be modified to account for the dynamic nature of the time series model.

The simplest example of model-based bootstrapping is Bickel and Freedman's (1983) recursive bootstrap. To implement this method on a stationary AR(p) model, $y_t = \sum_{i=1}^{p} r_i y_{t-i} + e_t$ (where e_t are white noise residuals), first compute the least squares residuals, e_t, and then resample them using the nonparametric residual bootstrap to get bootstrap residuals e_t^*. The bootstrap time series values are then recursively computed according to: $y_t^* = \sum_{i=1}^{p} \hat{r}_i y_{t-i}^* + e_t^*$, where $\hat{r}_i$ are the least squares estimates. The wild bootstrap could also be used to draw the e_t^* if heteroskedasticity is suspected. The recursive bootstrap can be generalized to univariate and vector ARMA models (see Berkowitz and Kilian, 1997, Li and Maddala, 1996, and Shao and Tu, 1995, Chapter 9) and cointegrated regression models (Li and Maddala, 1997 and 1996, Section 7.1).

The recursive bootstrap is not appropriate for models such as Generalized Method of Moments (GMM) which do not fully specify the time series properties of the residuals. The Moving Block Bootstrap divides the data into B (overlapping or non-overlapping) blocks of length $L = T / B$ contiguous observations and then resamples the blocks. If the autocorrelations are negligible for all lengths greater than L, then this Moving Block Bootstrap will yield bootstrap samples with approximately the same autocorrelation structure as the original series. Unfortunately, this moving block scheme does not preserve some important features of the original series such as stationarity. The problem is that observations in the same block are dependent in the bootstrap samples, but observations in different blocks are independent. Either changing the resampling method or altering the statistics to remove the biases can alleviate these problems.

Hall and Horowitz (1996) show that naïve application of moving block bootstrapping to GMM inferences yields inconsistent inferences. However, they show that by first recentering the moment conditions in the bootstrap samples and then rescaling usual GMM *t* statistic for overidentifying restrictions consistency can be restored. The recentering insures that the moment conditions defining the particular GMM estimator are imposed on the bootstrap samples. They further show that this method yields asymptotic improvements over standard asymptotic inference. Ziliak (1997) examined the performance of Hall and Horowitz's GMM bootstrap procedure to a panel data model of lifecycle labor supply under uncertainty. Ziliak's Monte Carlo results support the need for recentering residuals when bootstrapping GMM models.

Politis and Romano (1994) propose another resampling scheme, called the stationary bootstrap, which preserves stationarity and removes some of the biases associated with the moving block bootstrap methods. The stationary bootstrap resamples blocks of random length, where the length is sampled from an independent geometric distribution. Politis and Romano also specify that the original series should be "wrapped" to fill blocks that go past the last observation. Therefore a block of length two beginning at the last observation would yield the block: $\{ y_T ,\ y_1 \}$.

All of the block bootstrap resampling schemes requires some choice of (expected) block size. Theoretical considerations suggest that the (expected) block size should increase at the rate $T^{1/3}$, and Hall, Horowitz and Jing (1995) derive optimal rules which are a function of the autocovariance function of the time series. Unfortunately Politis and Romano (1994) show that the bootstrap variance estimates are quite sensitive to choice of block length for a simple moving average process. The stationary bootstrap is somewhat less sensitive to the choice of expected block length since the actual block lengths vary considerably over stationary bootstrap repetitions. However, Horowitz (2001, Section 4.1.1) points out that the stationary bootstrap is asymptotically inefficient relative to the moving block bootstrap.

Bühlmann (1997) has proposed an alternative bootstrapping technique which may be less sensitive to block size and still provide asymptotic refinements. If the data generation process can be represented by an infinite-order autoregressive model, then Bühlmann's sieve bootstrap approximates this with a finite-order autoregressive model with the order of the autoregression increasing at a suitable rate. Bootstrap samples are then generated by the recursive bootstrap applied to the approximate model.

Kilian (1998a, 1998b) considers bootstrapping VAR models, with emphasis on the generation of confidence intervals for impulse response functions. If the innovations in a VAR system are not normally distributed, standard methods of generating confidence intervals, such as Lütkepohl's delta method (1990) and Sims and Zha's (1995) Monte Carlo integration method, produce unsatisfactory results. Kilian develops a bootstrapping technique that accounts for the non-normality of VAR innovations by adjusting for the bias in the OLS coefficient estimates of the VAR system. Kilian's technique ("bias-corrected" bootstrap intervals) is similar in spirit to Beran's (1988) double bootstrap technique. Instead of drawing a second-level bootstrap within each bootstrap repetition to correct for bias in the original bootstrap, Kilian estimates the bias in the OLS estimates and uses a re-centered OLS estimate to draw the bootstrap. The bias in the original OLS estimator can be approximated by the following procedure:

1) draw an initial bootstrap sample using the OLS estimator
2) calculate a bootstrap OLS estimator for each bootstrap sample
3) estimate the bias of the original OLS estimator by comparing the original OLS estimator to the mean of the bootstrap OLS estimators.

Kilian (1998b) shows through extensive Monte Carlo evidence that confidence intervals for impulse responses generated by this bootstrap-followed-by-bootstrap procedure are much more accurate than standard asymptotic confidence intervals. Kilian (1998a) applies this technique to examine impulse response functions in a VAR model of the international effects of monetary policy. Kilian’s empirical VAR model includes equations for the log of industrial production, log of the consumer price index excluding shelter, the log of commodity prices, and the federal funds rate. Data spans the period from January 1965 to December 1993. Kilian investigates the impact of an (unexpected) one percent increase in the federal funds rate. Previous econometric analysis using standard error bands concluded that tightening monetary policy lead to declines in output and price level. Kilian’s bias corrected bootstrap results confirm the decline in output, but indicate that price level may not change (e.g. confidence intervals include zero) after tighter monetary policy.

2.6 Panel Data

Consider the linear panel data (or cross-section, time-series) model:

$$y_{it} = X_{it}\boldsymbol{b} + \boldsymbol{e}_{it}\text{ , } i = 1,\ldots,N \text{ and } t = 1,\ldots T \qquad (1)$$

Typically i indexes cross-section units and t indexes time. As long as the residuals, $\boldsymbol{e}_{it}$, are independent across cross-section units, then resampling the pairs ($y_{i\bullet}$, $X_{i\bullet}$) will yield bootstrap samples that preserve the heteroskedasticity and dependence properties of the original data. Ziliak (1997) used this method together with the recentering technique described in Horowitz (2001, Section 3.7) in his study of GMM panel data estimation. Note that these methods can also be applied to seemingly unrelated regression and reduced form simultaneous equations models.

If the vector of residuals, $\boldsymbol{e}_{i\bullet}$, are also homoskedastic, then residual resampling can be carried out by resampling blocks of the least squares residuals, $e_{i\bullet}$. This residual bootstrap also preserves time dependence, but it should be more accurate than the above paired resampling because it imposes the correct moment restriction E(e^{*} |X) = 0. This block residual sampling scheme is appropriate for the standard fixed effects model estimated by differencing the data over time. Even if the e_{it} are independent, the differencing process induces time dependence in the differenced residuals, and this time dependence is usually ignored in applied work.

The wild bootstrap also can be generalized to generate residual bootstrap samples that preserve the cross-section heteroskedasticity and time dependence of the original data. The wild bootstrap works by multiplying the least squares residuals by an independent random variable, *Z*, with mean zero and second and third moments equal to one. The implementation of this method at the beginning of this section used the unique 2-point distribution satisfying these properties (which equals $(1-\sqrt{5})/2$ with probability $(1+\sqrt{5})/(2\sqrt{5})$ and $(1+\sqrt{5})/2$ with probability $1-(1+\sqrt{5})/(2\sqrt{5})$). To implement this method with panel data draw *N* independent values Z_i from this distribution and multiply the vector of least squares residuals, $e_{i\bullet}$, by the scalar Z_i. The resulting bootstrap samples will also impose the correct moment restriction E(e^* |*X*) = 0 and should yield more accurate inference than the paired bootstrap method discussed at the beginning of this section (see also Horowitz, 2001, Section 5.2).

2.7 Empirical Example

We illustrate these ideas with a simple empirical example. Table 1 presents results from standard male earnings equations. The data are drawn from the Panel Study of Income Dynamics (PSID) for the years 1981-90. We restricted the sample to male household heads, aged 21-64 and employed in the private sector (excluding self-employed individuals). The dependent variable is the log of yearly earnings, deflated by the GDP deflator for personal consumption expenditure. The explanatory variables include years of formal education, potential experience (and its square), and tenure at the current firm, along with dummy variables indicating whether the respondent is a union member, black, married, in a blue-collar job, an hourly worker, has children, and lives in an MSA.

Panel A of Table 1 presents results from the pooled cross-section model (which also includes a complete set of year dummies as explanatory variables). Panel B presents results from a fixed-effects model, which uses the same data as in Panel A but includes a complete set of individual effects; the resulting parameter estimates are based on "within" variation over time, rather than variation across individuals at a point in time.

In Table 1, the estimated coefficients are listed in the second column. The subsequent four columns list 95 percent confidence intervals estimated several different ways. The first two sets are obtained analytically, through the standard least squares formula and through "robust" regression techniques. The third set of confidence intervals are obtained through application of a paired (x,y) bootstrap procedure (using the percentile method for the confidence intervals), and for the final column we used the Wild Bootstrap to obtain confidence intervals based on bootstrapped (asymptotically pivotal) t-statistics. The bootstrap techniques are applied using the panel data methods described in Section 2.6.

Table 1: Bootstrap Results
Regression Equation, Log(Yearly Earnings)
Male Heads, PSID, 1981-90 (N=13,050)

Panel A: Pooled Cross-Section

Variable	Coefficient	95% Confidence Intervals			
		Least squares	**Robust Regression**	**(x,y) bootstrap**	Wild bootstrap
Education	.060	.055, .064	.049, .071	.049, .071	.049, .071
Potential Experience	.039	.035, .042	.031, .045	.031, .046	.031, .045
(Pot. Exp.)^2/100	-.067	-.074, -.060	-.083, -.051	-.083, -.051	-.082, -.051
Tenure	.014	.013, .015	.012, .017	.012, .017	.012, .017
Union	.332	.308, .357	.285, .380	.283, .382	.283, .379
Black	-.307	-.351, -.263	-.418, -.196	-.426, -.204	-.426, -.200
Married	.131	.105, .158	.074, .189	.074, .188	.073, .185
Blue collar	-.181	-.207, -.155	-.225, -.136	-.224, -.139	-.224, -.137
Hourly	-.291	-.317, -.265	-.340, -.242	-.339, -.239	-.336, -.242
Children present	.044	.021, .067	.002, .086	.000, .084	.004, .086
MSA	.114	.094, .135	.074, .155	.075, .153	.076, .155

The sample is restricted to male heads aged 21-64 and employed in the private sector. Other variables controlled for include a constant and 9 year dummies.

(continued)

Table 1: Bootstrap Results (continued)

Panel B: Fixed Effects

Variable	**Coefficient**	**95% Confidence Intervals**			
		Least squares	**Robust Regression**	**(x,y) bootstrap**	**Wild bootstrap**
Education	.020	.010, .030	.006, .034	.004, .038	.004, .037
Potential Experience	.058	.053, .063	.051, .065	.049, .067	.049, .068
(Pot. Exp.)^2/100	-.079	-.090, -.069	-.094, -.065	-.097, -.063	-.098, -.062
Tenure	.012	.010, .014	.010, .015	.009, .016	.009, .016
Union	.146	.113, .179	.102, .190	.096, .202	.093, .192
Married	.102	.071, .133	.060, .144	.056, .151	.058, .153
Blue collar	-.062	-.088, -.035	-.094, -.030	-.098, -.027	-.098, -.027
Hourly	-.065	-.093 -.037	-.103, -.028	-.109, -.024	-.105, -.025
Children present	.035	.012, .059	.008, .062	.001, .066	.004, .069
MSA	.021	-.009, .050	-.016, .057	-.028, .070	-.029, .066

The sample is restricted to male heads aged 21-64 and employed in the private sector. The equation also includes a constant.

Comparison of the results across the four columns suggests that with a large data set and well-behaved estimating equation such as we have, neither bootstrap procedure appears to provide asymptotic improvements. The robust confidence intervals, which account for arbitrary heteroskedasticity in the model, are noticeably wider than the least-squares confidence intervals. However, the two sets of bootstrapped confidence intervals are nearly identical to the robust confidence intervals. Identifying conditions under which the robust and bootstrapped confidence intervals will differ is a topic for further work.

3 Multiple Imputations

The multiple imputation technique was developed by Rubin (1987, 1996) as a general method for inference with missing data. This methodology also can be used for consistent inference with imputed values for erroneous observations, which are treated as having missing values for the correct data. If the imputed values are somehow produced to match the first two moments of the correct unobserved values, then standard estimation methods that treat the imputed values as if they are correct will yield consistent parameter estimates. Unfortunately the standard errors produced by this approach will be inconsistent and downward biased because they ignore the errors introduced by the imputation process. Rubin proposed solving this problem by independently drawing multiple imputed values. The component of variance due to the imputation error is then estimated by the variability of the estimates across the different imputed data sets. Typically, drawing these multiple imputed values is the hard part of this methodology, so we will first describe Rubin's methods for combining results from multiply imputed data. Although he developed the theoretical properties of this methodology for Bayesian models, Rubin (1996 and 1987, Chapter 4) showed that these results apply asymptotically to classical statistical models.

Suppose we are interested in estimating an unknown parameter vector $\boldsymbol{q}$. If no data are missing or measured with error, then we would use the estimator $\hat{\boldsymbol{q}}$ and its associated covariance estimator $\tilde{\Omega}$. If we have a model for predicting the missing (or erroneous) values conditional on all observed data, then we can use this model to make independent simulated draws for the missing data. If *m* independent sets of missing data are drawn and *m* corresponding parameter and covariance estimators, $\tilde{\boldsymbol{q}}_j$ and $\tilde{\Omega}_j$, are computed, then Rubin's Multiple imputation estimators are given by

$$\hat{\boldsymbol{q}} = \sum_{j=1}^{m} \tilde{\boldsymbol{q}}_j \Big/ m$$

$$\hat{\Sigma} = U + \left(1 + m^{-1}\right) B, \qquad (2)$$

where

$$B = \sum_{j=1}^{m} \left(\tilde{\theta}_j - \hat{\theta}\right)\left(\tilde{\theta}_j - \hat{\theta}\right)' \Big/ (m-1)$$

$$U = \sum_{j=1}^{m} \tilde{\Omega}_j \Big/ m.$$

Note that *B* is an estimate of the covariance among the *m* parameter estimates for each independent simulated draw for the missing data, and U is an estimate of the covariance of the estimated parameters given a particular draw. B can also be interpreted as a measure of the covariance caused by the nonresponse (or measurement error) process.

Rubin (1987) shows that for a fixed number of draws, $m \geq 2$, $\hat{\theta}$ is a consistent estimator for θ and $\hat{\Sigma}$ is a consistent estimator of the covariance of $\hat{\theta}$. Of course, *B* will be better estimated if the number of draws is large, and the factor $(1 + m^{-1})$ compensates for the effects of small *m*. Rubin (1987) shows that as *m* gets large, then the Wald test statistic for the null hypothesis that $\theta = \theta^0$, $\left(\theta - \theta^0\right)' \hat{\Sigma}^{-1}\left(\theta - \theta^0\right)$, is asymptotically distributed according to an F distribution with *K* (the number of elements in θ) and $\boldsymbol{n}$ degrees of freedom. The value of $\boldsymbol{n}$ is given by:

$$\nu = (m - 1)(1 + r_m^{-1})^2 \text{ and} \qquad (3)$$
$$r_m = (1 + m^{-1}) \text{Trace}(BU^{-1})/K .$$

This suggests increasing m until ν is large enough (e.g. 100) so that the standard asymptotic Chi-squared distribution of Wald test statistics applies. Meng and Rubin (1992) show how to perform likelihood ratio tests with multiply-imputed data. Their procedures are useful in high-dimensional problems where it may be impractical to compute and store the complete covariance matrices required for the Wald test statistic.

The key to successful implementation of multiple imputation is to use a *proper* imputation procedure. The full definition of a proper imputation procedure is given in Rubin (1987, pp. 118-119). Loosely speaking, if the estimates computed with the true values of the missing data ($\ddot{\theta}$ and $\ddot{\Omega}$) are treated as fixed, then $\hat{\theta}$ and U must be approximately unbiased estimators of $\ddot{\theta}$ and $\ddot{\Omega}$. In addition B must be an approximately unbiased estimator of the variation in $\hat{\theta}$ caused by the non-response mechanism. The safest way to generate proper imputation procedures is to draw explicitly from the (Bayesian) posterior predictive distribution of the missing values under a specific model. Meng (2000) shows that the process of drawing proper imputations is identical to Bayesian data augmentation (Tanner and Wong, 1987). There are other proper multiple imputation procedures that require no explicit Bayesian calculations, and one such is described below. Any proper imputation procedure must condition on all observed data, and different sets of imputed values must be drawn independently so that they reflect all sources of uncertainty in the response process.

The multiple imputations technique requires specifying a model for the missing (or erroneous) data process, but joint estimation of this model and the substantive model is not required. The computations required can frequently be done in standard econometric packages with relatively minimal programming. Perhaps the largest advantage of multiple imputations is that it allows the imputations to be made once and then used for a variety of analyses. This allows agencies or researchers who collect data to represent the uncertainty in their data by including multiply imputed values for key variables. The U.S. Federal Reserve Board now provides multiply imputed income and wealth variables in the public release of its Survey of Consumer Finances, and the U.S. Bureau of Labor Statistics is experimenting with multiply imputing income and durable expenditures in its Consumer Expenditure Survey. Note that these imputations can take advantage of confidential information (such as precise location) which are not normally released in public use data sets.

3.1 Measurement Error in Wages

Brownstone and Valletta (1996) applied multiple imputation to correct for measurement error in the dependent variable in an earnings regression equation. Estimation of earnings equations typically proceeds under the assumption that measurement error in the dependent variable is “classical” – that is, normally distributed with mean zero and constant variance, uncorrelated with true earnings and the explanatory variables, and uncorrelated over time for a given individual. However, research based on data sources that provide validated earnings data – through company personnel files or social security records – has shown that measurement error in earnings is large and not classical in nature (see for example Bound et al 1994, Bound and Krueger 1991, Duncan and Hill 1985). Nonclassical measurement error in earnings is likely to bias estimated coefficients and sampling statistics from regression models in which error-prone earnings serve as the dependent variable.

Multiple imputation can be applied to correct for measurement error in earnings and improve the estimates from earnings equations. In particular, the technique enables researchers to combine information on the measurement error process from validated data with the structure of earnings estimated from a more general data set (referred to below as the "main" data). The technique consists of three steps.

> (1) Estimate a model of the measurement error process using validated data. In this stage, the true (validated) value of earnings is estimated as a function of the error-prone survey earnings variable and a set of related covariates. Although many econometricians are suspicious of including the endogenous survey earnings variable in this imputation equation, it is crucial to condition on all available data to make proper imputations. The purpose of this model is to approximate the distribution of the true earnings conditional on the survey earnings and other covariates.
> (2) Use the estimated error model to provide multiply imputed estimates of true earnings in the main data set (in which only the error-prone survey response is available).
> (3) Combine the multiply-imputed estimates of true earnings with observed values of the explanatory variables to estimate the earnings equation. Efficient estimation requires that the validation and main samples be pooled. If the structure of earnings is different in the two samples, pooling can be enabled through inclusion of interaction terms that capture the structural differences between the two data sets.

We illustrate the technique by providing results from corrected earnings equations; these equations represent a slight simplification of the results discussed by Brownstone and Valletta (1996). We estimated earnings equations similar to those used in our bootstrapping example above, with data confined to the years 1983 and 1987. We used the PSID Validation Study (PSIDVS) to estimate true earnings conditional on interview earnings. The PSIDVS was applied to several hundred employees from a large Detroit-area manufacturing firm in 1983 and 1987. The resulting data set matches standard PSID survey responses with company personnel records on earnings and other variables. The company records on earnings are highly accurate and are treated as error-free variables in our analysis.

We first estimated a standard cross-section earnings equation using the 1983 data. Our imputation equation included a quadratic in interview earnings in addition to the same variables used in the final earnings equations. The results are displayed in Panel A of Table 2. The first column displays results from the model uncorrected for measurement error. The second column displays results that account for measurement error in earnings using multiple imputations. For most variables, the estimated coefficients are quite similar across the two columns. The standard errors are increased by the multiple imputation procedure in many cases, because the procedure correctly incorporates additional sampling error associated with measurement error in the dependent variable. In terms of substantive results, the estimated effect of union membership is reduced about 15 percent by error correction, and the negative effect of working in a blue-collar occupation falls by about 40 percent.

Panel B of Table 2 displays results from a difference equation, which models the change in log earnings between 1983 and 1987 as a function of changes in the explanatory variables (and a few level variables). As in the 1983 cross-section (Panel A), the effect of union status on earnings is reduced. In addition, the negative effect of blue-collar status on the change in earnings is eliminated. Brownstone and Valletta (1996) found similar results and also found that measurement error in earnings appears to pose more significant problems during recessionary periods than expansionary periods, probably because errors in reported earnings are based largely on misperceptions of hours worked.

Table 2: Multiple Imputation Results
Regression Equation, Log(Yearly Earnings)
Pooled PSID and PSID Validation Study (PSIDVS)

Panel A: 1983 Cross-Section

Variable	Coefficient (Standard Error)	Corrected Coefficient (Standard Error)[1]
Schooling	0.057	0.051
	(0.008)	(0.011)
Potential Experience	0.040	0.040
	(0.005)	(0.007)
(Pot. Exp.)^2/100	-0.072	-0.069
	(0.010)	(0.014)
Tenure	0.011	0.013
	(0.002)	(0.003)
Union	0.270	0.225
	(0.037)	(0.100)
Black	-0.187	-0.132
	(0.060)	(0.059)
Married	0.120	0.122
	(0.036)	(0.038)
Blue Collar	-0.192	-0.112
	(0.036)	(0.043)
Hourly	-0.300	-0.312
	(0.040)	(0.106)
PSIDVS	-0.075	-0.071
	(0.217)	(0.195)
PSIDVS*Schooling	-0.016	-0.011
	(0.015)	(0.014)
PSIDVS*Blue Collar	0.225	0.179
	(0.071)	(0.068)

[1] Corrected estimates obtained using the multiple imputation technique described in the text.

The sample is restricted to males aged 21-64 and employed in the private sector. The estimating equation also includes a constant.

Table 2: Multiple Imputation Results (continued)

Panel B: Difference Equation (1987 – 1983)

Variable	Coefficient (Standard Error)	Corrected Coefficient (Standard Error)[1]
DSchooling	0.008 (0.026)	-0.014 (0.032)
DPotential Experience	0.005 (0.021)	0.011 (0.025)
Potential Experience 1983	-0.008 (0.001)	-0.010 (0.002)
DTenure	0.010 (0.003)	0.015 (0.016)
DUnion	0.203 (0.050)	0.156 (0.073)
DMarried	0.040 (0.043)	-0.034 (0.045)
DBlue Collar	-0.071 (0.041)	0.061 (0.046)
Blue Collar 1983	-0.036 (0.028)	0.056 (0.037)
DHourly	-0.039 (0.047)	-0.009 (0.045)

[1] Corrected estimates obtained using the multiple imputation technique described in the text.

The sample is restricted to males aged 21-64 and employed in the private sector. The estimating equation also includes a constant.

4 Recommendations

Efron and Tibshirani (1993) and Vinod (1998) envision the bootstrap as an integral part of a strategy to find universally applicable methods for estimation and inference. Manski (1996) argues that this vision is flawed, but in any case the current state of the art in bootstrap methods can not support such schemes. This review has indicated many areas where the bootstrap can be very valuable for applied econometricians, but generally the bootstrap method needs to be tailored to the specific model(s) under consideration. The bootstrap currently can not replace econometric theory, but econometric theory can frequently help choose appropriate bootstrap techniques.

For cross-section linear and nonlinear regression models, the wild bootstrap appears to dominate other methods. It imposes the correct conditional moment condition, and it produces inferences that are robust to residual heteroskedasticity. If the residuals are homoskedastic, then the wild bootstrap usually is very close to nonparametric residual bootstrapping. Finally, the wild bootstrap is quite easy to implement in current econometric software packages. (Although bootstrapping would be much easier if software developers provide tools and "shell" programs for drawing bootstrap samples and keeping track of the estimates calculated from each sample).

The situation is less clear for time series models. Model-based residual bootstrapping is clearly more accurate when it can be applied (typically VARMA models). Models such as GMM, which don't fully specify the residual distribution, require some type of block bootstrap resampling. In practice, all block sampling (or subsampling) techniques can be very sensitive to block size, so it is necessary to search over a range of block sizes in any real application.

For statistics with well-behaved Edgeworth expansions, bootstrapping pivotal statistics and/or iterated bootstrapping can yield substantial improvements in small sample behavior. Unfortunately, these techniques can perform very badly when Edgeworth expansions are not well-behaved (e.g. for instrumental variables with poor instruments). In practice, comparing the results of these second-order correct bootstrap methods with those from bootstrap subsampling may indicate whether the second-order correct method is appropriate. An even safer approach would be to conduct a Monte Carlo study of the particular bootstrap methodology based as closely as possible on the same data and model used in the original application.

Applied economists should consider multiple imputation methods to help alleviate problems caused by survey non-response and missing data. Multiple imputation is like an adjustable wrench - it is rarely the ideal tool for any particular job, but it works well for a wide variety of problems. The example in Section 3.1 show that multiple imputation can be successfully implemented for real applied problems using existing software packages. Furthermore, Brownstone and Valletta's (1996) application shows that using this methodology can make a substantial difference in the qualitative conclusions. Brownstone et. al. (1999) and Clogg et. al. (1991) show that multiple imputations can also be implemented with discrete choice models.

Manski (1995) shows that missing data and measurement error causes serious problems with identification and inference from even simple models. The best way to circumvent these problems is to put more resources into reducing response biases during survey administration. The next best solution is to collect external validation data that allow identification of the non-response process. If these validation data become more widely available, then the multiple imputation methods presented in this chapter provide an easy and consistent way for researchers to incorporate this information into their modeling and forecasting efforts.

5 References

Andrews, D. W. K. (1997), A simple counterexample to the bootstrap, Cowles foundation Discussion Paper No. 1157, Yale University, August, 1997.

Beran, R. (1988): Prepivoting test statistics: a bootstrap view of asymptotic refinements, Journal of the American Statistical Association 83(403): 687-697.

Berkowitz, J. and L. Kilian (1997), Recent developments in bootstrapping time series, Board of Governors of the Federal Reserve System, Finance and Economics, Discussion Series 94/95.

Bickel, P. J. and Freedman, D. A. (1983). Bootstrapping regression models with many parameters, in: P. Bickel, K. Docksum, and J. L. Hodges, eds., A Festschrift for Erich Lehmann. Wadsworth, Belmont, California: 28-48.

Bound, J., C. Brown, G.J. Duncan, and W.L. Rogers (1994), "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," Journal of Labor Economics 12: 345-368.

Bound, J., and A.B. Krueger (1991), "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," Journal of Labor Economics 9: 1-24.

Brownstone, D. and C. Kazimi (2000), Applying the Bootstrap, Dept. of Economics Working Paper, University of California, Irvine.

Brownstone, D., and R.G. Valletta (1996), "Modeling Earnings Measurement Error: A Multiple Imputations Approach," Review of Economics and Statistics 78: 705-717.

Brownstone, D., T.F. Golob and C. Kazimi (1999). Modeling non-ignorable attrition and measurement error in panel surveys: an application to travel demand modeling. Presented at the International Conference on Survey Nonresponse, October 28-31, Portland, OR. To appear in Survey Nonresponse, New York: Wiley.

Buchinsky, M. (1994), Changes in the U.S. wage structure 1963-1987: application of quantile regression, Econometrica 62(2): 405-458.

Buchinsky, M. (1995), Estimating the asymptotic covariance matrix for quantile regression models: A monte carlo study, Journal of Econometrics 68:303-338.

Bühlmann, P. (1997), Sieve bootstrap for time series, Bernoulli 3: 123-148.

Chamberlain, G. (1994), Quantile regression, censoring, and the structure of wages, in: C. Sims and J. J. Laffont, eds., Proceedings of the Sixth World Congress of the Econometric Society, Barcelona, Spain, Cambridge University Press (New York).

Clogg, Clifford C., Donald B. Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman, "Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression," Journal of the American Statistical Association 86 (March 1991), 68-78.

Duncan, G.J., and D.H. Hill (1985), "An Investigation of the Extent and Consequences of Measurement Error in Labor-economics Survey Data," Journal of Labor Economics 3: 508-532.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, Annals of Statistics, 7: 1-26.

Efron, B. (1982), The Jacknife, the Bootstrap and Other Resampling Plans, SIAM, Philadelphia, Pennsylvania.

Efron, B. and R. Tibshirani (1993): An Introduction to the Bootstrap, (Chapman & Hall, New York).

Hall, P. (1992), The Bootstrap and Edgeworth Expansion (Springer Verlag, New York).

Hall, P. and J. Horowitz (1996), Bootstrap critical values for tests based on generalized method of moments estimators, Econometrica 64(4): 891-916.

Hall, P. , J. Horowitz and B. Jing (1995), On blocking rules for the block bootstrap with dependent data, Biometrika 82(3): 561-574.

Horowitz, J. L. (1997), Bootstrap methods in econometrics: Theory and numerical performance, in: R. Kreps and K. Wallis, eds., Advances in Economics and Econometrics: Seventh World Congress of the Econometric Society, vol. 3 (Cambridge University Press) 188-222.

Horowitz, J. L. (2001), The Bootstrap, in E. E. Leamer and J. J. Heckman, eds., Handbook of Econometrics, Vol 5, Elsevier, Amsterdam, forthcoming.

Kazimi, C. and D. Brownstone (1999), Bootstrap confidence bands for shrinkage estimators, Journal of Econometrics 90(1): 99-127.

Killian, L. (1998a), Small-sample confidence intervals for impulse response functions, Review of Economics and Statistics 80(2): 218-230.

Killian, L. (1998b), Confidence intervals for impulse responses under departures from normality, Econometric Reviews 17(1): 1-29.

Koenker, R. and G. Bassett (1978), Regression quantiles, Econometrica 46: 33-50.

Li, H. and G. S. Maddala (1996), Bootstrapping time series models, Econometric Reviews 15(2): 115-158.

Li, H. and G. S. Maddala (1997), Bootstrapping cointegrating regressions, Journal of Econometrics 80: 297-318.

Lütkepohl, H. (1990), Asymptotic distributions of impulse response functions and forecast error variance decomposition of vector autoregressive models, The Review of Economics and Statistics, 72: 116-125.

Mammen, E. (1992), When Does Bootstrap Work? Asymptotic Results and Simulations (Springer-verlag, New York.)

Manski, C.F. (1995), Identification problems in the social sciences, Harvard University Press, Cambridge, Massachusetts.

Manski, C. F. (1996) Review of an introduction to the bootstrap by B. Efron and R. Tibshirani, Journal of Economic Literature 34: 1340-1342.

Meng, X-L (2000), Missing data: dial M for ???, Journal of the American Statistical Association, 95, 1325-1330.

Meng, Xiao-li and Donald B. Rubin , "Performing Likelihood-Ratio Tests with Multiply-Imputed Data Sets," Biometrika 79 (Jan. 1992), 103-111.

Politis, D. and Joseph P. Romano (1994), The stationary bootstrap, Journal of the American Statistical Association 89(428): 1303-1313.

Rubin, Donald B. (1987), Multiple Imputation for Nonresponse in Surveys (New York: John Wiley, 1987).

Rubin, Donald B. (1996), "Multiple Imputation After 18+ Years," Journal of the American Statistical Association 91 (June 1996), 473-489.

Shao, J. and D. Tu (1995), The Jackknife and Bootstrap (Springer-Verlag, New York.)

Sims, C. A. and R. Zha (1995). Error Bands for Impulse Responses, Federal Reserve Bank of Atlanta Working Paper: 95-96

Tanner, M. A. and W. H. Wong (1987), The calculation of posterior distributions by data augmentation (with discussion), Journal of the American Statistical Association, 82, 528-550.

Vinod, H. D. (1998), Foundations of statistical inference based on numerical roots of robust pivot functions, Journal of Econometrics 86: 387-396.

Wu, C. F. J. (1986), Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics: 14(4) 1261-1295.

Ziliak, J. P. (1997), Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators, Journal of Business and Economic Statistics 15(4): 419-431.