DEPARTMENT OF ECONOMICS

Working Paper

BETTER CONFIDENCE INTERVALS FOR THE POPULATION MEAN BY USING TRIMMED MEANS AND THE ITERATED BOOTSTRAP?

Viggo Høst

Working Paper No. 1997-14
Centre for Non-linear Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

BETTER CONFIDENCE INTERVALS FOR THE POPULATION MEAN BY USING TRIMMED MEANS AND THE ITERATED BOOTSTRAP?

Viggo Høst

Working Paper No. 1997-14

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

BETTER CONFIDENCE INTERVALS FOR THE POPULATION MEAN BY USING TRIMMED MEANS AND THE ITERATED BOOTSTRAP?
A Monte Carlo Study on Data from the Danish Wage Statistics

Viggo Høst*
Department of Economics
University of Aarhus
E-mail: vhoest@eco.aau.dk

August 1997

Key words: trimmed means, pps-sampling, iterated bootstrap, confidence intervals, wage statistics

JEL code: C13, C15, C42, J30

Abstract

In the quarterly Danish sample based wage statistics firms within a given industry are randomly selected with probabilities proportional to size. The hourly mean wage per employee is registered within each selected firm. Then the population mean per employee is estimated by the pps-estimator, i.e. the simple mean of the selected firm means per employee. I raise the question: Is it possible to produce better confidence intervals for the population mean by using trimmed means instead of the simple mean when the iterated bootstrap is used? Monte Carlo experiments indicate that it is the case.

* I wish to thank Torben Poulsen, Danish Employers' Confederation, for help with data and Niels Lund

for able research assistance.

Introduction

A few years ago the producer of Danish wage statistics, the Danish Employers' Confederation, began a totally new production of wage statistics for Denmark. The new statistics cover all kinds of employees and is based on new international standard nomenclatures and new wage concepts. The wage statistics are covered by an annual census, with detailed information, and a quarterly sample statistics containing more aggregate results. The quarterly sample statistics cover 13 branches.

It is earlier shown, Høst, Lund and Poulsen (1994), that within a given branch the simple mean from a pps-sample (probability proportional to size) is a very efficient estimator of the population mean. This is, in fact, the estimator which the producer of the quarterly sample based wage statistics decided to use and still uses. It is also shown, Høst (1997), that the use of the robust estimator, the trimmed mean, *without* application of the iterated bootstrap only gives possibility for minor improvement of the confidence intervals.

The purpose of this paper is to study if application of the iterated bootstrap to the trimmed mean will improve the confidence intervals. In order to do this we consider three different branches, Machinery & Electrical Equipment, Wholesale Trade and Construction, and Monte Carlo experiments are performed on population data from the year 1995 for these branches.

Sampling Design and Data

The firms are sampling units and within the firms all individual hourly wages are calculated and aggregated to the firm level by their sum Y_i. The firm sizes M_i defined by the number of employees are known a priori. Due to the fact that Y_i and M_i are strongly positively correlated , see Figure 1, an auxiliary variable (M_i) estimation procedure of the population mean $R = \sum_i Y_i / \sum_i M_i$ (the population ratio) is appropriate. Both the ratio estimator and the pps-estimator are estimators using auxiliary information of M_i.

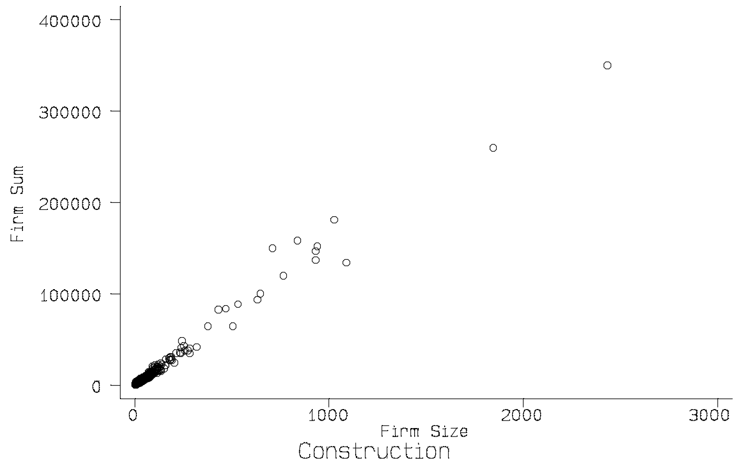

Figure 1. Scatterplots of firm sum (Y_i) versus firm size (M_i)

As mentioned earlier the pps-estimator of R is a very good choice, and it was also found that it performs better than the ratio estimator in these cases.

Due to the experience of relatively low quality of data from small firms and relative higher cost of sampling the producer has decided to disregard firms with less than 5 employees. On the other hand it is decided to include all large firms in the sample.

Consequently, the population frame of firms (with $M_i \geq 5$) is divided into two strata according to firm size (i.e. the number of employees), see Table 1.

Table 1. Stratification of the Frame		
Stratum	Definition	Method
S	$5 \leq M_i < a$	Sampling
C	$a \leq M_i$	Census

A possible bias in the estimation of R introduced by excluding firms with less than 5 employees is considered as a problem outside this paper.[1]

[1] Anyway the bias will be negliable, see Tables A2, B2 and C2 in the appendices.

The stratum limit *a* which defines the distinction between smaller and larger firms may take different values between branches. The value of *a* is chosen such that the variance of the pps-estimator is low and such that the variance is not very sensitive to the choice of the limit, i.e. the stratum limit is chosen in an interval where the variance function of the stratum limit *a* is low and flat. Then the stratum limit will be stable over time. In this study we examine three branches based on data from 1995, see Table 2, which contains the number of firms with at least 5 employees and the number of employees.

Table 2. The Three Branches 1995

Branch	No. of firms	No. of employees
Machinery and Electrical Equipment	402	51778
Wholesale Trade	413	21583
Construction	1726	51930

Let us consider a given branch within the sampling stratum S. Then let N be the number of firms (we omit the index S). Let R_i be hourly mean wage per employee in firm no. i and M_i the number of employees of firm no. i. If M is the total number of employees in the stratum, then the hourly mean wage per employee in the stratum is given by

$$R = \Sigma_{\varkappa=\Theta}^{\phi} R_{\varkappa} M_{\varkappa} / M$$

If we add subscripts S and C we can define the hourly mean wage per employee in the frame T as

$$R_{\omega} = \frac{M_{\Omega} R_{\Omega} + M_P R_P}{M_{\Omega} + M_P}$$

This is the population characteristic which we want to estimate.

Estimators and Sampling Distributions

A pps-sample from stratum S is generated as follows: *n firms are randomly selected with replacement with probabilities P_i proportional to firm size M_i*, i.e.

$$P_{\varkappa} = M_{\varkappa} / M \qquad i = 1, 2, \ldots, N$$

The sample values (with subscript S omitted) are given by

$$r_\Theta, r_\theta, \ldots, r_\varphi$$

Then the sample mean (we call it the pps-estimator)

$$\bar{r}_\varphi = \Sigma_\varkappa^\varphi r_\varkappa / n$$

has the following properties, Cochran (1977):

a. $\bar{r}_\varphi$ is an unbiased estimator of R

b. $\bar{r}_\varphi$ is consistent

c. $\bar{r}_\varphi$ is asymptotically normal, $\bar{r}_\varphi \xrightarrow{\ddot{\upsilon}} N(R, Var(\bar{r}_\varphi))$, where

$$Var(\bar{r}_\varphi) = \Sigma_{\varkappa=\Theta}^{\phi} P_\varkappa (R_\varkappa - R)^\theta / n$$

which can be estimated consistently by

$$s^\theta(\bar{r}_\varphi) = \Sigma_{\varkappa=\Theta}^{\varphi} (r_\varkappa - \bar{r}_\varphi)^\theta / n(n-1)$$

Most branches have distributions with excess kurtosis, i.e. they have heavy tails. Consequently there is room for considering robust estimators of the mean, R. In this connection we introduce an *asymmetric* (α, β)*-trimmed mean* $\bar{r}_n^{\alpha,\beta}$, defined as

$$\boxed{\bar{r}_\varphi^{\alpha\epsilon\beta} = \frac{1}{n - ([n\alpha] + [n\beta])} \sum_{\varkappa = 'I\varphi\alpha\ddot{I} + \Theta}^{\varphi - 'I\varphi\beta\ddot{I}} r_{\gamma\Delta}} \qquad (1)$$

where α, β are positive values such that $\alpha + \beta < 1$, $[\cdot]$ represents the greatest-integer function, and $r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(n)}$ are the order statistics. If $\alpha = \beta = 0$, then $\bar{r}_n^{\alpha,\beta}$ reduces to the mean $\bar{r}_n$, i.e. $\bar{r}_n^{0,0} = \bar{r}_n$. The values α and β are chosen such that the corresponding population functional $R^{\alpha,\beta}$ is as close as possible to the population mean R, that is

$$R^{\alpha\epsilon\beta} = \frac{1}{1 - \alpha - \beta} \sum_{\varkappa = 'I\phi\alpha\ddot{I} + \Theta}^{\phi - 'I\phi\beta\ddot{I}} R_{\gamma\Delta} P_{\psi\gamma\Delta} \doteq R,$$

where $R_{(1)} \leq R_{(2)} \leq \ldots \leq R_{(N)}$ are the ordered population values and $P_{R(1)}, P_{R(2)}, \ldots, P_{R(N)}$ are the corresponding selection probabilities. Recall that $P_i = M_i/M$.

It follows that the special population functional $R^{0,0}$ is equal to the mean R:

$$R^{0,0} = \Sigma_{i=1}^{N} R_{(i)} P_{R(i)} = \Sigma_{i=1}^{N} R_i P_i = R.$$

The reason for using an *asymmetric* trimmed mean instead of the symmetric trimmed mean $\bar{r}_n^{\alpha}$ defined by

$$\bar{r}_n^{\alpha} = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} r_{(i)}$$

is that the symmetric trimmed mean presupposes a symmetric population distribution. This is, however, not the case for our data. On the contrary, the population distributions are skew to the right.

The estimator $\bar{r}_n^{\alpha,\beta}$ is consistent and asymptotically normal distributed. Its variance is estimated by means of the bootstrap method. This is explained in the next section.

After estimating R (or R_S) by means of $\hat{R}$ (or $\hat{R}_S$) which is either $\bar{r}_n^{\alpha,\beta}$ or $\bar{r}_n = \bar{r}_n^{0,0}$ we estimate the frame mean, R_T, by

$$\hat{R}_T = \frac{M_S \hat{R}_S + M_C R_C}{M_S + M_C}$$

The variance of $\hat{R}_T$ is easily seen to be

$$\mathrm{Var}(\hat{R}_T) = \left(\frac{M_S}{M_S + M_C} \right)^2 \mathrm{Var}(\hat{R}_S)$$

The purpose in the following of this paper is to compare the statistical performance of the trimmed means $\bar{r}_n^{\alpha,\beta}$ with the simple mean $\bar{r}_n$. This is done by comparing the mean squared errors, MSE of these estimators, and coverage probabilities of the corresponding confidence intervals. We therefore estimate the variance $\bar{r}_n^{\alpha,\beta}$ for suitable values of α and β. This is done by means of the bootstrap technique, see next section. Finally we are looking for values of the sum $\alpha+\beta$ such that coverage probabilities are best, i.e. closest to the confidence coefficient 95% and such that the mean squared error of the estimate is low or minimum.

Bootstrap Estimation of the Variance of $\bar{r}^{\alpha,\beta}$

Consider a pps-sample

$$\mathbf{r} = (r_1, r_2, ..., r_n).$$

Then B bootstrap samples of size n,

$$\mathbf{r}_1^*, \mathbf{r}_2^*, ..., \mathbf{r}_B^*$$

are generated by sampling with replacement from **r** with equal probability 1/n, and the B trimmed means $\bar{r}_{nb}^{\alpha,\beta *}$ are computed. For simplicity we replace $\bar{r}_n^{\alpha,\beta}$ by $\hat{R}_n$ and $\bar{r}_n^{\alpha,\beta *}$ by $\hat{R}_n^*$ in the following. The values $\hat{R}_{n1}^*, ..., \hat{R}_{nB}^*$ then constitute the conditional distribution of the trimmed mean $\hat{R}_n^*$. The bootstrap estimate of the variance of $\hat{R}_n$, $s_{boot}^2(\hat{R}_n)$ is defined by

$$\boxed{s_{boot}^2(\hat{R}_n) = \frac{1}{B-1} \Sigma_{b=1}^{B} (\hat{R}_{nb}^* - \hat{R}_{n(\cdot)}^*)^2,} \tag{2}$$

where

$$\hat{R}_{n(\cdot)}^* = \frac{1}{B} \Sigma_{b=1}^{B} \hat{R}_{nb}^*$$

Confidence Interval for R Based on Normal Approximation

The straightforward 95% confidence interval for R based on the normal distribution and the trimmed mean $\hat{R}_n$ is then given by

$$\boxed{\hat{R}_n \pm z_{.975}\, s_{boot}(\hat{R}_n)} \tag{3}$$

where $z_{.975}$ is the .975-quantile of the standard normal distribution.

Another very simple and useful confidence interval which is based on Efron's *backward* bootstrap method is introduced in the next section.

Confidence Interval for R Based on Percentile Method

Consider the conditional distribution of $\hat{R}_n^*$ as defined above. Then the 95 percent percentile (or backward) confidence interval for R is given by

$$\boxed{(b^*_{n.025}, b^*_{n.975})} \qquad (4)$$

where b_{np}^* is the p-quantile of the conditional distribution of $\hat{R}_n^*$, Efron and Tibshirani (1993).

In order to produce better confidence intervals for R we introduce the procedure "*first Studentize*, *then bootstrap*", Hall (1988) and the so-called *Prepivoting* method, Beran (1987). Both methods presuppose the use of the iterated bootstrap. The procedure is called *iterated*, because bootstrap samples are created in two steps: Consider the pps-sample $\mathbf{r} = (r_1, r_2, ... r_n)$. In the first step we create B_1 bootstrap samples, $\mathbf{r}_b^*$, b = 1, 2, ..., B_1. In the second step B_2 bootstrap samples $\mathbf{r}_{b1}^*$, $\mathbf{r}_{b2}^*$, ..., $\mathbf{r}_{bB2}^*$ are created by resampling from each bootstrap sample $\mathbf{r}_b^*$ from the first step. Consequently, the total number of bootstrap samples is B_1B_2.

Confidence Interval for R Based on First Studentize, then Bootstrap

The Studentized statistic T of the trimmed mean $\hat{R}_n$ based on bootstrap estimation of the variance of the estimator is given by

$$T = \frac{\hat{R}_n - R}{s_{boot}(\hat{R}_n)}$$

If T is approximately normally distributed, then the normal confidence interval for R given in (3) can be used.

If, however, the normality assumption does not hold, then the distribution function of T $G_n(x) = P(T \leq x)$ is estimated by means of the bootstrap replica

$$G_n^*(x) = P^*\left(\frac{\hat{R}_n^* - \hat{R}_n}{s^*_{boot}(\hat{R}_n^*)} \leq x \right),$$

where $s^*_{boot}(\hat{R}_n^*)$ is estimated in the second step of the iterated bootstrap procedure.

That is

$$G_n^*(x) = P^*\left(\frac{\hat{R}_n^* - \hat{R}_n}{s_{boot}^*(\hat{R}_n^*)} \le x \right) \approx G_n(x) = P\left(\frac{\hat{R}_n - R}{s_{boot}(\hat{R}_n)} \le x \right)$$

To summarize: For every given sample **r** B_1 values of $\hat{R}_n^*$ are computed in the first step and B_1 values of $s_{boot}^*(\hat{R}_n^*)$ are computed in the second step.

Let c_{np} be the p-quantile of the Studentized statistic T, i.e. $c_{np} = G_n^{-1}(p)$ and let the corresponding bootstrap estimate be $c_{np}^* = G_n^{*-1}(p)$. Then the bootstrap confidence interval for R is given by

$$\boxed{\left(\hat{R}_n - c_{n.975}^* s_{boot}(\hat{R}_n), \quad \hat{R}_n - c_{n.025}^* s_{boot}(\hat{R}_n)\right)} \tag{5}$$

which is found by isolating R in the following probability statement

$$P\left(c_{n.025} \le \frac{\hat{R}_n - R}{s_{boot}(\hat{R}_n)} \le c_{n.975} \right) = 0.95,$$

and replace c_{np} by c_{np}^*.

Confidence Interval for R Based on the Prepivoting Method

The rationale behind Beran's prepivoting technique is the well-known theorem based on the probability integral transform, which states that if X is a continuous random variable with a strictly increasing distribution function F, then Y = F(X) has a uniform distribution on (0,1).

Let G_n be the distribution function of the statistic $\sqrt{n}(\hat{R}_n - R)$ and let G_n^* be its bootstrap approximation, i.e. the conditional distribution function of $\sqrt{n}(\hat{R}_n^* - \hat{R}_n)$, then the distribution of $G_n(\sqrt{n}(\hat{R}_n - R))$ is uniform on (0,1) and consequently $G_n^*(\sqrt{n}(\hat{R}_n^* - \hat{R}_n))$ is expected to be close to the uniform distribution on (0,1) also. If H_n^* denotes the known distribution of $G_n^*(\sqrt{n}(\hat{R}_n^* - \hat{R}_n))$, then the 95% confidence interval for R is found by solving the following two equations

$$H_n^*\left(\sqrt{n}(\hat{R}_n - R)\right) = 0.025$$

$$H_n^*\left(\sqrt{n}(\hat{R}_n - R)\right) = 0.975\,.$$

Let d_{np}^* be the p-quantile of the stochastic variable $G_n^*(\sqrt{n}(\hat{R}_n^* - \hat{R}_n))$, that is $d_{np}^* = H_n^{*-1}(p)$, then the confidence interval is given by

$$\boxed{\left(\hat{R}_n - \frac{d_{n.975}^*}{\sqrt{n}},\quad \hat{R}_n - \frac{d_{n.025}^*}{\sqrt{n}}\right)} \tag{6}$$

The distribution function H_n^* is determined by using the iterated bootstrap in the following way: For a given sample **r**, B_1 values of the statistic $\sqrt{n}(\hat{R}_n^* - \hat{R}_n)$ are computed in the first step and B_2 values of $\sqrt{n}(\hat{R}_n^{**} - \hat{R}_n^*)$ are computed in the second step. Consider a given bootstrap sample in the first step. Then the corresponding value of the distribution function of $\sqrt{n}(\hat{R}_n^* - \hat{R}_n)$ is estimated from the second step bootstrap samples by the formula

$$G_n^*\left(\sqrt{n}(\hat{R}_n^* - \hat{R}_n)\right) = \Sigma_{b=1}^{B_2} 1\left\{\sqrt{n}(\hat{R}_{nb}^{**} - \hat{R}_n^*) \le \sqrt{n}(\hat{R}_n^* - \hat{R}_n)\right\}/B_2,$$

where 1{ } is the indicator function. The B_1 values of $G_n^*(\sqrt{n}(\hat{R}_n^* - \hat{R}_n))$ then determines the bootstrap estimate of $H_n^*(\sqrt{n}(\hat{R}_n^* - \hat{R}_n))$.

Results of Monte Carlo Experiments

The number of replications is 2000 and the number of bootstrap samples in the first step B_1 is 350 and in the second step B_2 is 200. The values of B_1 and B_2 are kept relatively low in order to reduce computer time, but still the total number of samples including bootstrap samples to be generated in one experiment is as big as $1.4\ 10^8$. The total sample size n_T for Machinery & Electrical Equipment is set to 100 and 125 which means that the corresponding values of n_S are 23 and 48, respectively. For the two other branches, WholesaleTrade and Construction, n_T is chosen to be 75 and 100 such that n_S is equal to 18 and 43 for Wholesale Trade, and 22 and 47 for Construction. The bootstrap sample size is equal to the original sample size n_S. I have chosen the total trimmed percentage, TTP = $(\alpha+\beta)$ 100 to be 0, 10, 20 and 30. The reason for these rather few values of TTP is again the overwhelming computer time which is necessary to perform the described Monte Carlo experiments. The actual chosen values of TTP are also consistent with the results from earlier experiments, Høst (1997), which indicate that improvements of confidence intervals are most likely for values of TTP around 10-20. Coverage probabilities are based on the four confidence intervals described above with confidence coefficients equal to 95 percent. The results for the three branches appear in the appendices, see Tables A1, B1, C1 and C3. A few important population characteristics are described in the appendices too, see Tables A2, B2 and C2, and Figures A1, B1 and C1.

Machinery & Electrical Equipment (Appendix A)

From Table A1 follows that coverage probabilities are improved when trimmed means are used instead of the simple mean. It also follows that the method based on *first Studentize, then bootstrap* gives the best results. It should be noted that the improvements in coverage probability are paid by a slightly increase in variance and bias, and therefore also in mean squared error, MSE.

Wholesale Trade (Appendix B)

For this branch, Table B1, we get almost similar results as for the former branch with the exception that for the relatively big sample size (n_T = 100) and the biggest TTP (30) *Prepivoting* is even better than *first Studentize, then bootstrap*.

Construction (Appendix C)

For this branch it follows, Table C1, that both variance and bias increase with the TTP. This may be explained by the extreme values of skewness and kurtosis for this branch. But again we find coverage probability improves when trimmed means are used instead of the simple mean and confidence intervals are based on the iterated bootstrap, i.e. either *first Studentize, then bootstrap* or *Prepivoting*.

In order to reduce the considerable size of bias, I have also made Monte Carlo experiments where the original α-values of the trimmed means, see (1) p. 5, are reduced and the corresponding β-values incrased such that their sums are unchanged. These results are shown in Table C3. The bias is now closer to zero and not especially increasing with TTP and at the same time the variance is also reduced. I do not, however, claim that the chosen values of α and β are “optimal”, since bias is not totally eliminated, but it follows that reduction of MSE is at least possible by adjusting the α and β-values. It now follows that *first Studentize, then bootstrap* gives the best confidence intervals for both sample sizes and for all the selected TTP-values greater than zero. According to MSE a value of TTP around 20 is best.

Conclusion

The three branches have different population distributions, and of most interest, different coefficients of skewness and kurtosis. Consequently it is not surprising that we do not find exactly the same solutions for three branches. But for all the studied branches it can be concluded that better confidence intervals for the population mean can be found by using trimmed means instead of the simple mean when the iterated bootstrap technique is used.

References

Beran, R.J., 1987, Prepivoting to reduce level error of confidence sets, *Biometrica*, *74*, 171-200.

Cochran, G., 1977, *Sampling Techniques*, Wiley, New York.

Efron, B. and R. Tibshirani, 1993, *An introduction to the bootstrap*, Chapman & Hall, New York.

Hall, P., 1988, Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics*, *16*, 927-81.

Høst, V., N. Lund and T. Poulsen, 1994, Kan resamplingprocedurer betale sig i den nye stikprøvebaserede lønstatistik?, S. Boelskifte (ed.), *Symposium i Anvendt Statistik*, UNI C, Copenhagen.

Høst, V., 1997, Trimmed Means and PPS-sampling. A Monte Carlo Study on Data from the Danish Wage Statistics, Holst, H. et al. (eds.), *Symposium i Anvendt Statistik*, Technical University of Denmark, Copenhagen.

APPENDIX A: MACHINERY & ELECTICAL EQUIPMENT 1995

Table A1. Results

	TTP	Variance	Bias	MSE	Coverage Probability (Percentage) Normal	Percentile	Student	Prepivot
$n_T = 100$	00	1.92	0.01	1.92	91.5	91.3	91.8	90.8
$n_S = 23$	10	2.05	0.42	2.22	92.4	92.3	94.3	92.3
	20	2.34	0.50	2.60	92.8	92.8	95.4	93.0
	30	2.42	0.25	2.48	92.5	93.3	95.3	93.0
$n_T = 125$	00	0.95	0.02	0.95	93.3	93.5	93.3	92.3
$n_S = 48$	10	1.04	0.26	1.11	94.3	94.0	95.3	94.0
	20	1.13	0.45	1.34	94.0	93.0	94.8	94.6
	30	1.15	0.05	1.15	93.2	94.5	95.0	94.5

Table A2. Population Characteristics

Population characteristic	$M_i < 5$ (deleted)	$5 \le M_i < 120$ (S)	$120 \le M_i$ (C)	Frame (T)	Total population
N	58	325	77	402	460
M	152	11425	40353	51778	51930
R	152.39	142.50	138.08	139.05	139.09

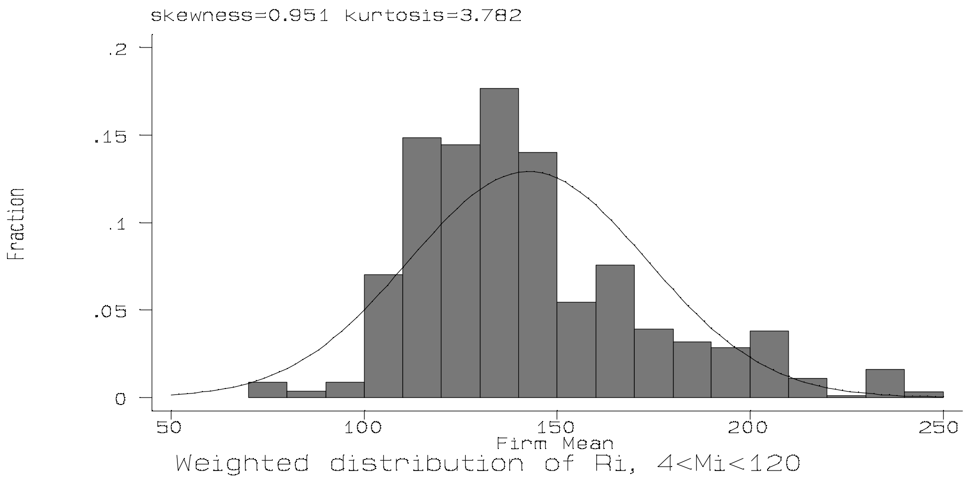

Figure A1.

APPENDIX B: WHOLESALE TRADE 1995

Table B1. Results

	TTP	Variance	Bias	MSE	Coverage Probability (Percentage) Normal	Percentile	Student	Prepivot
$n_T = 75$	00	11.63	-0.12	11.64	91.7	91.3	93.7	90.5
$n_S = 18$	10	12.52	1.07	13.67	92.4	91.6	94.8	91.3
	20	12.40	-0.56	12.72	92.0	92.5	94.8	90.4
	30	13.14	0.11	13.15	93.4	93.8	95.7	92.9
$n_T = 100$	00	5.12	0.08	5.13	92.3	92.0	93.0	91.9
$n_S = 43$	10	5.29	0.29	5.37	93.8	93.5	94.8	93.7
	20	5.05	-0.24	5.11	92.5	93.0	93.3	93.6
	30	5.09	-0.14	5.10	93.2	93.5	93.8	95.8

Table B2. Population Characteristics

Stratum characteristic	$M_i < 5$ (deleted)	$5 \le M_i < 90$ (S)	$90 \le M_i$ (C)	Frame (T)	Total population
N	75	356	57	413	488
M	193	9506	12077	21583	21776
R	152.36	149.88	155.62	153.09	153.08

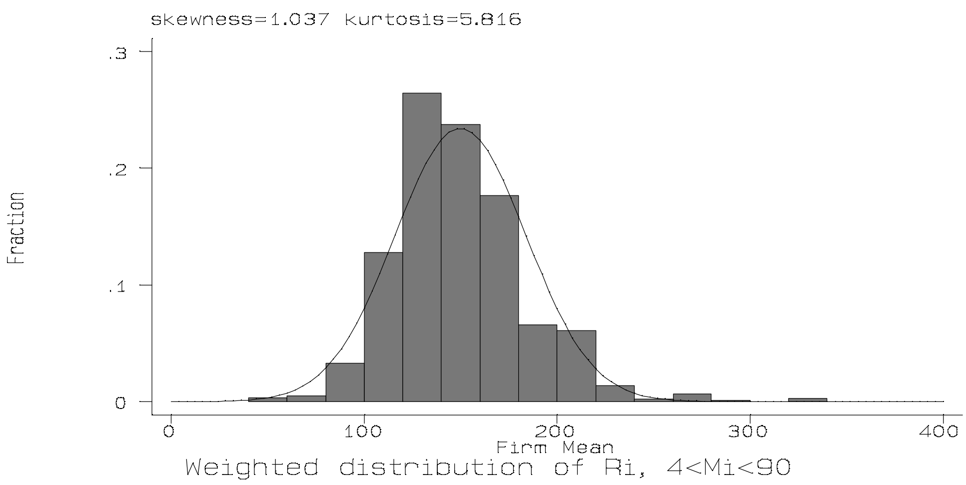

Figure B1.

APPENDIX C: CONSTRUCTION 1995

Table C1. Results

	TTP	Variance	Bias	MSE	Coverage Probability (Percentage) Normal	Percentile	Student	Prepivot
$n_T = 75$	00	9.28	0.07	9.29	91.5	90.8	91.9	90.5
$n_S = 22$	10	10.00	0.95	10.90	92.2	91.6	94.6	92.3
	20	10.97	1.12	12.22	93.8	93.0	95.3	94.0
	30	11.74	1.92	15.42	92.9	91.1	94.9	93.8
$n_T = 100$	00	4.51	-0.03	4.51	92.1	91.6	92.0	92.2
$n_S = 47$	10	4.73	0.38	4.87	94.5	93.6	94.3	94.0
	20	4.94	1.16	6.27	93.8	92.4	93.5	94.8
	30	5.07	1.59	7.59	91.8	89.9	91.5	94.8

Table C2. Population Characteristics

Population characteristic	$M_i < 5$ (deleted)	$5 \leq M_i < 120$ (S)	$120 \leq M_i$ (C)	Frame (T)	Total population
N	1061	1673	53	1726	2787
M	2540	30149	21781	51930	54470
R	126.87	136.64	153.75	143.82	143.03

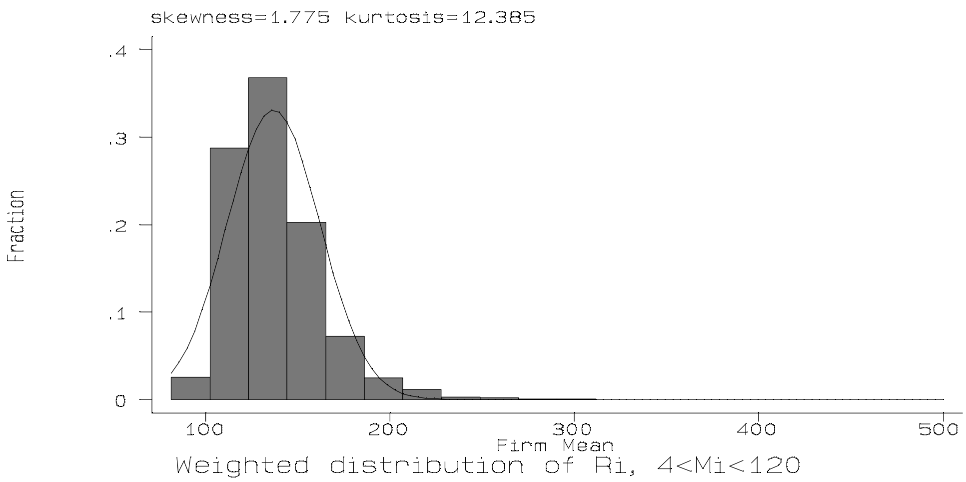

Figure C1.

Table C3. Results

	TTP	Variance	Bias	MSE	Coverage Probability (Percentage) Normal	Percentile	Student	Prepivot
$n_T = 75$	00	9.47	0.04	9.47	92.7	92.3	93.3	91.3
$n_S = 22$	10	10.21	0.91	11.04	93.5	93.0	95.5	92.5
	20	10.36	0.24	10.42	94.4	94.3	95.8	93.2
	30	10.69	0.50	10.94	94.0	94.5	95.7	94.0
$n_T = 100$	00	4.49	0.02	4.49	94.2	94.0	93.9	93.8
$n_S = 47$	10	4.57	0.01	4.57	94.5	94.9	95.6	94.3
	20	4.51	0.15	4.53	95.0	95.3	95.6	95.0
	30	4.71	0.65	5.13	95.2	94.7	95.5	96.3

Working Paper

1997-1 Tom Engsted, Jesus Gonzalo and Niels Haldrup: Testing for Multicointegration.

1997-2 Lykke E. Andersen: Modelling the Relationship between Government Policy, Economic Growth, and Deforestation in the Brazilian Amazon.

1997-3 Alvaro Forteza: Overinsurance in the Welfare State.

1997-4 Alvaro Forteza: Multiple Equilibria in the Welfare State with Costly Policies.

1997-5 Torben M. Andersen and Morten Stampe Christensen: Contract Renewal under Uncertainty.

1997-6 Jan Rose Sørensen: Do Trade Unions Actually Worsen Economic Performance?

1997-7 Luca Fanelli: Estimating Multi-Equational LQAC Models with I(1) Variables: a VAR Approach.

1997-8 Bo Sandemann Rasmussen: Long Run Effects of Employment and Payroll Taxes in an Efficiency Wage Model.

1997-9 Bo Sandemann Rasmussen: International Tax Competition, Tax Cooperation and Capital Controls.

1997-10 Toke S. Aidt: Political Internalization of Economic Externalities. The Case of Environmental Policy in a Politico-Economic Model with Lobby Groups.

1997-11 Torben M. Andersen and Bo Sandemann Rasmussen: Effort, Taxation and Unemployment.

1997-12 Niels Haldrup: A Review of the Econometric Analysis of I(2) Variables.

1997-13 Martin Paldam: The Micro Efficiency of Danish Development Aid.

1997-14 Viggo Høst: Better Confidence Intervals for the Population Mean by Using Trimmed Means and the Iterated Bootstrap?

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Non-linear Modelling in Economics:

1994-28	Lykke Eg Andersen: Neural Networks as an Exploratory Statistical Technique.
1995-5	David Easley, Nicholas M. Kiefer and Maureen O'Hara: Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow.
1995-6	David Easley, Nicholas M. Kiefer, Maureen O'Hara and Joseph B. Paperman: Liquidity, Information, and Infrequently Traded Stocks.
1995-13	Tom Engsted, Jesus Gonzalo and Niels Haldrup: Multicointegration and Present Value Relations.
1996-1	Tom Engsted and Niels Haldrup: Estimating the LQAC Model with I(2) Variables.
1996-2	Peter Boswijk, Philip Hans Franses and Niels Haldrup: Multiple Unit Roots in Periodic Autoregression.
1996-3	Clive W.J. Granger and Niels Haldrup: Separation in Cointegrated Systems, Long Memory Components and Common Stochastic Trends.
1996-4	Morten O. Ravn and Martin Sola: A Reconsideration of the Empirical Evidence on the Asymmetric Effects of Money-Supply shocks: Positive vs. Negative or Big vs. Small?
1996-13	Robert F. Engle and Svend Hylleberg: Common Seasonal Features: Global Unemployment.
1996-14	Svend Hylleberg and Adrian R. Pagan: Seasonal Integration and the Evolving Seasonals Model.
1997-1	Tom Engsted, Jesus Gonzalo and Niels Halrup: Testing for Multicointegration.
1997-7	Luca Fanelli: Estimating Multi-Equational LQAC Models with I(1) Variables: a VAR Approach.
1997-12	Niels Haldrup: A Review of the Econometric Analysis of I(2) Variables.
1997-14	Viggo Høst: Better Confidence Intervals for the Population Mean by Using Trimmed Means and the Iterated Bootstrap?