BOOTSTRAPPING A REGRESSION EQUATION:
SOME EMPIRICAL RESULTS

BY

DAVID A. FREEDMAN AND STEPHEN C. PETERS

TECHNICAL REPORT NO. 10

JULY 1982

REVISED

AUGUST 1983

DEPARTMENT OF STATISTICS

UNIVERSITY OF CALIFORNIA, BERKELEY

Bootstrapping a Regression Equation:
Some Empirical Results[†]

David A. Freedman
Statistics Department
University of California, Berkeley

Stephen C. Peters
Statistics Department
Stanford University

Abstract

The bootstrap, like the jack-knife, is a technique for estimating standard errors. The idea is to use Monte-Carlo simulation, based on a non-parametric estimate of the underlying error distribution. The main object of this paper is to present the bootstrap in the context of an econometric equation describing the demand for energy by industry. As it turns out, the conventional asymptotic formulae for estimating standard errors are too optimistic by factors of nearly three, when applied to a particular finite-sample problem. In a simpler context, this finding can be given a mathematical proof.

Keywords and phrases: Regression, generalized least squares, seemingly unrelated equations, econometric models, forecasting, standard errors.

Running head: Bootstrapping a regression equation

Table of Contents

1. Introduction

This paper is mainly concerned with estimating standard errors for regression coefficients obtained by constrained generalized least squares with an estimated covariance matrix. Existing methods are largely asymptotic, and may not apply with finite samples. We use "the bootstrap," a computer-based methodology, to check the accuracy of the asymptotics and make alternative estimates of the standard errors which are more reliable. This paper is the first application of the bootstrap to generalized least squares.

The bootstrap is a relatively new statistical technique, which permits the assessment of variability in an estimate using just the data at hand; see Efron (1979). The idea is to resample the original observations in a suitable way, to construct "pseudo-data" on which the estimator of interest is exercised. More specifically, the theoretical distribution of a disturbance term is approximated by the empirical distribution of a set of residuals. Measures of variability, confidence intervals, and even estimates of bias may then be calculated.

In the regression case, the bootstrap is useful for investigations when mathematical analysis can give only asymptotic results. Within the scope of the bootstrap are: non-normal errors, lag structures, and generalized least squares with estimated covariance matrices. This paper compares the performance of conventional asymptotic estimates of standard error to the performance of a bootstrap procedure in the setting of a single econometric equation. The main finding is that for generalized least squares with estimated covariance matrices, the asymptotic formulae for standard errors can be too optimistic, sometimes by quite large factors. The bootstrap procedure is appreciably better than the conventional asymptotics, when applied to the finite-sample situation. For a partial explanation, see Beran (1983) or Singh (1981).

This study is mainly empirical; however, in very simple contexts, a mathematical reason for the findings is given (Section 5). As a simple illustration of those results, take for instance the one-way analysis of variance model, with Gaussian errors, equal numbers of observations per cell, but different variances. Constrain the theoretical cell means to equality. If the variances are known, the generalized least squares estimator $\hat{\alpha}_{gls}$ for the common theoretical mean weights the sample means by the reciprocals of the cell variances; $\text{var}\,\hat{\alpha}_{gls}$ is proportional to the harmonic mean of these variances. If the variances are unknown, they can be estimated by the sample variances, leading to

the approximate *gls* estimator $\hat{\alpha}_{agls}$; the a in the subscript stands for "approximate." The variance of $\hat{\alpha}_{agls}$ would be estimated as proportional to the harmonic mean of the sample variances. Call this estimated variance $\widehat{var}$. Then $\widehat{var}$ is systematically too small:

$$\text{var } \hat{\alpha}_{agls} > \text{var } \hat{\alpha}_{gls} > E(\widehat{var}).$$

An extension is made to the general multivariate linear model.

The balance of this paper is organized as follows. Section 2 gives a brief review of the bootstrap idea, in the context of linear econometric models. Section 3 gives an even briefer review of generalized least squares, and pinpoints the technical issue to be addressed by the bootstrap. Section 4 applies these ideas to an econometric model and presents a simulation experiment to assess the validity of the bootstrap. Some mathematical results are presented in Section 5, while Section 6 reports some computational details, and discusses estimates of the stability of the Monte-Carlo results. Finally, Section 7 reports a bootstrap experiment on a formula of Srivastava and Dwivedi (1979).

The approach may be distinguished from the classical work of Brown (1954), or Goldberger, Nagar, and Odeh (1961): the bootstrap uses simulation rather than asymptotics based on Taylor series. The work of Fair (1979 and 1980) is closer in spirit to the bootstrap, but somewhat different in detail: Fair assumes that the disturbance terms follow a multivariate normal distribution, and that the parameter estimates follow their multivariate normal limiting distribution. The bootstrap is distribution-free, and develops the appropriate finite-sample behavior for the estimates. Of course, the bootstrap has problems of its own, as will be seen below. The bootstrap can also be used to attach standard errors to multi-period forecasts, and to choose among competing forecasting equations; it can also be applied to simultaneous equation models. These extensions will be discussed elsewhere. In other models, not on the face of things too dissimilar from the one studied here, the conventional asymptotics do rather well. We hope later to explore the reasons for such differences.

The generalized least squares procedure we study is often called the two-stage Aitken estimator (2SAE). For a particular class of models — "seemingly unrelated regression equations" — Zellner (1962) shows that 2SAE is asymptotically valid. Maddala (1971) studies the asymptotics in a more general setting. Some theoretical results for finite samples have been obtained for special cases. Zellner (1963) analyzes two "seemingly unrelated regression equations" where the "independent variables" are

assumed to be orthogonal by equations, and obtains exact first and second moments for the 2SAE in finite samples. Some of his work will be summarized in Section 5.

Phillips (1977) develops Edgeworth expansions for the distribution of the two-stage estimator in the "seemingly unrelated regressions" model with many equations. Taylor (1977) derives a second-order approximation to the covariance matrix of the two-stage estimator in finite samples. These investigations do not focus on the validity of the approximate standard error formulae in finite samples, or the sensitivity of the theoretical results to departures from assumptions.

The bias in the SEs is demonstrated in this paper by a simulation experiment, where the parameters are fixed at estimates from a real data-set, and the error distribution is chosen to be the empirical distribution of the residuals. These choices are by no means critical, and normal errors could be used. Fiebig and Theil (1983), for example, have results similar to ours, for demand equations with normal errors. Theil, Finke, and Rosalsky (1983) also have such results, for maximum likelihood estimates, the asymptotic standard errors being computed from the information matrix in the usual way. These two papers have useful reviews of previous work. Mikhail (1975) also reports bias in standard errors, but only in the range from 5 percent to 30 percent. For additional details on the material in the present paper, and other related results, see Peters (1983b).

2. The bootstrap

The bootstrap is described by Efron (1979, 1982). Related papers are by Bickel and Freedman (1981, 1983) and Freedman (1981, 1982). The bootstrap is a procedure for estimating standard errors by re-sampling the data in a suitable way. This idea can be applied to econometric models, where the technical difficulties include simultaneity, correlated errors, heteroscedasticity, and dynamics. First, an informal overview of the idea. In brief, the model has been fitted to data, by some statistical procedure; and there are residuals, namely the difference between observed and fitted values. Some stochastic structure was imposed on the stochastic disturbance terms, explicitly or implicitly, in the fitting. The key idea is to resample the residuals, preserving this stochastic structure, so the standard errors are generated using the model's own assumptions. Assuming the model and the estimated parameters to be right, the resampling generates "pseudo-data." Now the model can be re-fitted to the pseudo-data. In this artificial world, the errors in the parameter estimates are directly observable. The Monte-Carlo distribution of such errors can be used to approximate the distribution of the unobservable errors in the real parameter estimates. This approximation is the bootstrap: it gives a measure of the statistical uncertainty in the parameter estimates.

A more explicit, but still informal, description is as follows. Consider a dynamic linear model, of the form

$$Y_t = Y_{t-1} B + X_t C + \epsilon_t \qquad (1)$$
$$\underset{1 \times q}{} \quad \underset{1 \times q}{} \underset{q \times q}{} \quad \underset{1 \times p}{} \underset{p \times q}{} \quad \underset{1 \times q}{}$$

In this equation, B and C are coefficient matrices of unknown parameters, to be estimated from the data, subject to identifying restrictions; Y_t is the vector of "endogenous" variables at time t; X_t is the vector of "exogenous" variables at time t; and ϵ_t is the vector of disturbances at time t. The endogenous variables are determined within the model, the exogenous variables by some external process: technically, endogenous variables may be correlated with ϵ, exogenous variables are not correlated with ϵ. The following standard condition is imposed on the error distribution: given the X's, the ϵ's are independent and identically distributed with mean 0. Linearity is assumed to simplify the exposition; the method is easily adapted to cover nonlinear models, although the computational costs may be prohibitive. The form (1) is general enough to cover the case of "seemingly unrelated regressions;" see Zellner (1962).

Data are available for $t = 1, \ldots, n$ and Y_0 is available too. The coefficient matrices are

estimated as $\hat{B}$ and $\hat{C}$ by some well-defined statistical procedure, like generalized least squares (Sections 3 and 4). When $\hat{B}$ and $\hat{C}$ are computed, residuals are defined:

$$\hat{\epsilon}_t = Y_t - Y_{t-1}\hat{B} - X_t\hat{C} \tag{2}$$

These are estimates for the true disturbances ϵ_t in the model (1). Let μ be the empirical distribution of the residuals, assigning mass $1/n$ to each of $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$. To avoid trivial complications, assume the equations have intercepts. See Freedman (1981, pp. 1220 and 1224) on centering.

Some inflation of the residuals may prove desirable, to compensate for the deflation of the residuals in fitting. However, there is no generally valid rule, except in the case of a standard regression model with homoscedastic errors where the factor $\sqrt{n/(n-p)}$ is appropriate. The residuals are linearly dependent, again due to the fitting. It may be appropriate to transform the residuals as in Theil's (1971, pp. 205–206) BLUS procedure. This is not done here.

Consider next a model like (1), but where all the ingredients are known:

► Set the coefficients at $\hat{B}$ and $\hat{C}$ respectively.

► Make the disturbance terms independent, with common distribution μ.

The exogenous X's are kept as before, as is Y_0. Using this simulation model, pseudo-data can be generated. These will be denoted by stars: $Y_0^*, \ldots, Y_n^*$. The construction is iterative: $Y_0^* = Y_0$, and for all $t = 1, \ldots, n$,

$$Y_t^* = Y_{t-1}^*\hat{B} + X_t\hat{C} + \epsilon_t^*$$

the ϵ^*'s being independent with the common distribution μ.

Now pretend the pseudo-data $Y_0^*, \ldots, Y_n^*$ come from a model like (1), with unknown coefficient matrices. Using the previous estimation procedures, estimate these coefficients from the pseudo-data; denote the estimates by $\hat{B}^*$ and $\hat{C}^*$. The distribution of the pseudo-errors $\hat{B}^* - \hat{B}$, $\hat{C}^* - \hat{C}$ can be computed, and used to approximate the distribution of the real errors $\hat{B} - B$, $\hat{C} - C$. This approximation is the bootstrap. It is emphasized that the calculation assumes the validity of the model (1). The distribution of the pseudo-errors can be computed, e.g., by Monte Carlo, simply repeating the procedure some number of times and seeing what happens. This paper will give experimental evidence to show the approximation is good; for other experimental evidence, see Efron (1979, 1982). For asymptotic results, see Freedman (1981, 1982).

3. Generalized least squares

Consider the model

$$Y = X\beta + \epsilon, \quad E(\epsilon) = 0, \quad \text{cov}(\epsilon) = \Sigma \tag{3}$$

With Σ known, the generalized least squares (gls) estimate is

$$\hat{\beta}_{gls} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y \tag{4}$$

As usual,

$$E(\hat{\beta}_{gls}) = \beta \tag{5}$$

$$\text{cov}(\hat{\beta}_{gls}) = (X^T \Sigma^{-1} X)^{-1} \tag{6}$$

When Σ is unknown, statisticians routinely use (4) and (6) with Σ replaced by some estimate $\hat{\Sigma}$. Iterative procedures are often used, as follows. Let $\hat{\beta}^{(0)}$ be some initial estimate for β, typically from a preliminary ordinary least squares (ols) fit. There are residuals $\hat{e}^{(0)} = Y - X\hat{\beta}^{(0)}$. Suppose the procedure has been defined through stage k, with residuals

$$\hat{e}^{(k)} = Y - X\hat{\beta}_{gls}^{(k)}$$

Let $\hat{\Sigma}_k$ be an estimator for Σ, based on $\hat{e}^{(k)}$: an example will be given below assuming a block diagonal structure for Σ. Then

$$\hat{\beta}_{gls}^{(k+1)} = (X^T \hat{\Sigma}_k^{-1} X)^{-1} X^T \hat{\Sigma}_k^{-1} Y \tag{7}$$

This procedure can be continued for a fixed number of steps, or until $\hat{\beta}_{gls}^{(k)}$ settles down. Indeed, a convexity argument shows that $\hat{\beta}_{gls}^{(k)}$ converges to the maximum likelihood estimate for β, assuming ϵ is independent of X and multivariate Gaussian with mean 0.

The covariance matrix for $\hat{\beta}_{gls}^{(k+1)}$ is usually estimated from (7), with $\hat{\Sigma}_k$ put in for Σ:

$$\hat{\text{cov}}^{(k+1)} = (X^T \hat{\Sigma}_k^{-1} X)^{-1} \tag{8}$$

This may be legitimate, asymptotically. In finite-sample situations, all depends on whether $\hat{\Sigma}_k$ is a good estimate for Σ or not. If $\hat{\Sigma}_k$ is a poor estimate for Σ, the standard errors estimated from (8) may

prove to be unduly optimistic: an example is given below. Unfortunately, approximate *gls* estimators are often used when there is too little data to offer any hope of estimating Σ with reasonable accuracy. In such circumstances, the bootstrap is a useful diagnostic, and in cases like the present one it gives a more realistic estimate of the standard errors.

To ease notation, $\hat{\beta}_{gls}^{(k)}$ will be refered to as the (*gls,k*)-estimator. This paper only considers the (*gls*,1) estimator, which in many situations has full asymptotic efficiency; see Cox and Hinkley (1974, p. 308). In our example, further iteration seems to make the coefficient estimates better, but also exaggerates the optimism of the standard error estimates. In other cases, the approximate *gls* coefficient estimators may prove to be worse than *ols* estimators, due to the variability of $\hat{\Sigma}_k$: so iteration can hurt.

4. Bootstrapping RDFOR

The object of this section is to illustrate the bootstrap procedure for determining the variability of parameter estimates in a real example. The main experimental finding is that the conventional asymptotics can be off by factors of nearly three. The example is the Regional Demand Forecasting Model (RDFOR). This is a system of econometric equations designed to forecast demand for energy through 1995. It is a component of the Midterm Energy Forecasting System (MEFS). MEFS was the principal energy model used by the Department of Energy to make midterm forecasts for its Annual Report to Congress, through 1981. MEFS was a development of PIES, the Project Independence Evaluation System. RDFOR forecasts what demand would be in a future year for various fuel types by consumption sector and geographical region, as a function of prices and other exogenous variables. The focus here is on that part of the model concerned with the industrial sector demand for fuel. For more detailed discussions of RDFOR, see Freedman, Rothenberg, and Sutch (1983) or Kuh *et al* (1982).

The Department of Energy (DOE) distinguishes ten geographical regions, indexed here by r. The equation for total demand by the industrial sector in geographical region $r = 1, \ldots, 10$ and year $t = 1961, \ldots, 1978$ is taken as

$$q_{rt} = a_r + b\,c_{rt} + c\,h_{rt} + d\,p_{rt} + e\,q_{r,t-1} + f\,v_{rt} + \epsilon_{rt} \tag{9}$$

where in region r and year t: q_{rt} is the log of an index of fuel consumption, c_{rt} is the log of cooling degree days, h_{rt} is the log of heating degree days, p_{rt} is the log of a fuel price index, v_{rt} is the log of value added in manufacturing, ϵ_{rt} is a stochastic disturbance term, and a_r, b, c, d, e, f are parameters to be estimated. This particular equation is the one reported by Kuh *et al* (1982). The equation is dynamic in the sense that the lagged endogenous variable $q_{r,t-1}$ appears on the right hand side. Notice that the coefficients b, c, d, e, f are constant across regions; however, the intercepts a_r are region-specific. The constraint that b, c, d, e, f be constant across regions is a significant technical complication, not usually encountered in treatments of Zellner's method.

The assumptions on the stochastic disturbance terms ϵ_{rt} are as follows:

$$E(\epsilon_{rt}) = 0 \text{ for all } r \text{ and } t. \tag{10a}$$

The ϵ_{rt} are stochastically independent of the c_{rt}, h_{rt}, p_{rt}, and v_{rt}. $\qquad$ (10b)

The vectors $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{10,t})$ are independent and identically (10c)

distributed in time.

This model is outside the framework of standard regression theory because of the dynamics: q_{rt} is correlated with $\epsilon_{r,t-1}$. It is outside the framework of standard multivariate theory because the coefficients are constrained to equality across regions. However, equation (9) does fit into the framework (1) with $q = 10$ and $p = 5 \times 10 + 1 = 51$; the matrices are subject to numerous constraints.

Historical data for estimating this regression relation were taken from the SEDS (State Energy Data System) data base. SEDS was previously called FEDS. This data base is reviewed in Freedman, Rothenberg, and Sutch (1983). This data base contains the annual data required for the period 1960 through 1978. The fitting period, however, runs from 1961 to 1978: a year of data is lost due to the lag term.

Consider the one-step *gls* estimator, $\hat{\beta}_{gls}^{(1)}$ in the notation above, starting from the *ols* estimator $\hat{\beta}^{(0)} = \hat{\beta}_{ols}$. The first column in Table 1 below displays this (*gls*,1) fit to the model. The standard errors (SEs) are obtained from the conventional formula (8) using $\hat{\Sigma}_0$; these are shown in the second column of Table 1, and will be called the "nominal" SEs. The computation of $\hat{\Sigma}_0$ may be described as follows. For the model (9–10) the distribution of $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{10,t})$ has an unknown inter-regional covariance matrix K; this 10-by-10 matrix is assumed constant over time. The covariance matrix Σ for all the disturbances is a 180-by-180 block diagonal matrix with K repeated on the diagonal. Let $\hat{\epsilon}_t^{(k)}$ denote the 10-vector $(\hat{\epsilon}_{1,t}^{(k)}, \ldots, \hat{\epsilon}_{10,t}^{(k)})$ of residuals at the kth stage of *gls* iteration; $k = 0$ corresponds to *ols*. Let $\hat{K}_k$ be the sample covariance matrix of these eighteen 10-vectors, with r, s entry given by

$$\frac{1}{18} \sum_{t=1961}^{1978} \hat{\epsilon}_{rt}^{(k)} \hat{\epsilon}_{st}^{(k)}$$

(11)

Then the estimate $\hat{\Sigma}_k$ is the 180-by-180 block-diagonal matrix with $\hat{K}_k$ repeated on the diagonal.

The validity of the nominal standard errors shown in Table 1 is open to serious question, because $\hat{\Sigma}_0$ is not an accurate estimate of Σ. This is because there are only 18 years of data and 10 regions, from which must be estimated 10 intercepts, 5 coefficients, and the 10-by-10 variance-covariance matrix K. The bootstrap gives an alternative method for approximating the standard errors, and a program for assessing the validity of the nominal standard errors.

Table 1. Bootstrap experiment for equation (9). Estimation is
by one-step *gls*. There are 100 bootstrap replications.

| | | GLS | | Bootstrap | | | |
		(1) Estimate	(2) Nominal SE	(3) Mean	(4) SD	(5) RMS Nominal SE	(6) RMS Boot SE
	a_1	−.95	.31	−.94	.54	.19	.43
	a_2	−1.00	.31	−.99	.55	.19	.43
	a_3	−.97	.31	−.95	.55	.19	.43
	a_4	−.92	.30	−.90	.53	.18	.41
	a_5	−.98	.32	−.96	.55	.19	.44
	a_6	−.88	.30	−.87	.53	.18	.41
	a_7	−.95	.32	−.94	.55	.19	.44
	a_8	−.97	.32	−.96	.55	.19	.44
	a_9	−.89	.29	−.87	.51	.18	.40
	a_{10}	−.96	.31	−.94	.54	.19	.42
c.d.d.	b	.022	.013	.021	.025	.0084	.020
h.d.d.	c	.10	.031	.099	.052	.019	.043
price	d	−.056	.019	−.050	.028	.011	.022
lag	e	.684	.025	.647	.042	.017	.034
v.a.	f	.281	.021	.310	.039	.014	.029

To get started on the bootstrap, let $\hat{a}_r$, $\hat{b}$, $\hat{c}$, $\hat{d}$, $\hat{e}$, and $\hat{f}$ be the (*gls*,1)-parameter estimates reported in Table 1. Consider the residuals

$$\hat{\epsilon}_{rt} = q_{rt} - \hat{a}_r - \hat{b}c_{rt} - \hat{c}h_{rt} - \hat{d}p_{rt} - \hat{e}q_{r,t-1} - \hat{f}v_{rt}$$

Let $\hat{\epsilon}_t$ be the 10-vector $(\hat{\epsilon}_{1,t}, \ldots, \hat{\epsilon}_{10,t})$ of residuals for year t. Let μ be the empirical distribution of $\{\hat{\epsilon}_t : t = 1961, \ldots, 1978\}$. Note that μ has mean 0, because (9) has region-specific intercepts. Now simulate the equation (9), where all the ingredients are known:

▶ $q_{r,1960}$ and the exogenous variables are held fixed.

▶ The parameters are set at their estimated values $\hat{a}_r$, $\hat{b}$, $\hat{c}$, $\hat{d}$, $\hat{e}$, and $\hat{f}$.

▶ The disturbance terms are independent with common distribution μ.

More specifically, let $\{\epsilon_t^* : t = 1961, \ldots, 1978\}$ be the results of 18 draws made at random with replacement from the set of eighteen 10-vectors $\{\hat{\epsilon}_t : t = 1961, \ldots, 1978\}$. Thus $\hat{\epsilon}_{1961}$ may be drawn

twice, but $\hat{\epsilon}_{1962}$ not at all. On the other hand, the regional pattern of the disturbances does not change. Thus, the simulation preserves the key stochastic assumptions: the disturbances are independent and identically distributed in time but show a geographic pattern.

The pseudo-data can now be built up iteratively year-by-year: $q^*_{r,1960} = q_{r,1960}$ and for $t = 1961,\ldots,1978$,

$$q^*_{rt} = \hat{a}_r + \hat{b} c_{rt} + \hat{c} h_{rt} + \hat{d} p_{rt} + \hat{e} q^*_{r,t-1} + \hat{f} v_{rt} + \epsilon^*_{rt} \tag{12}$$

Here ϵ^*_{rt} denotes the rth component of the 10-vector ϵ^*_t. The bootstrap parameter estimates $\hat{a}^*_r$, $\hat{b}^*$, $\ldots$, $\hat{f}^*$ can now be obtained from the $(gls,1)$ regression of q^*_{rt} on c_{rt}, h_{rt}, p_{rt}, $q^*_{r,t-1}$, and v_{rt}.

This procedure was repeated 100 times. On each repetition, a new set of starred disturbances was generated, hence a new set of pseudo-data, and therefore a new set of starred parameter estimates. Columns 3 and 4 in Table 1 show for each parameter the sample mean and sample standard deviation (SD) for these 100 starred estimates. These SDs are the bootstrap estimates of variability in the parameter estimates. They are appreciably larger than the nominal SEs.

It will now be shown that the nominal SEs are substantially too small. To do the bootstrap, we have set up a fully-defined simulation model, where the parameters and the distribution of the disturbances are all known. In this world, the variability of the $(gls,1)$ estimates was determined empirically, as reported in the "SD" column of Table 1. In the same world, how good are the nominal standard errors? The answer is, they are much too small, as shown in column 5 of Table 1. This column may be explained as follows. At each of the 100 repetitions, the nominal SE for each $(gls,1)$ estimate is computed using (8) on the starred data set. The root mean square of these SEs is shown in the table.

Take, for example, the coefficient d of the price term. In the simulation world of this experiment, the "real" variability of the $(gls,1)$ estimate for this parameter is .028, from column 4. But the apparent variability, from the conventional asymptotics, is in a typical run only .011, from column 5: this is

$$\sqrt{\frac{1}{100} \sum_{i=1}^{100} SE_i^2}$$

where SE_i is the nominal SE for the price coefficient computed from formula (8) applied to the ith starred data-set. Typically, the conventional formula is off by a factor of nearly three. The other entries in column 5 may be interpreted in a similar way. This finding cannot be explained by "specification error." The specification is built into the simulation procedure. The explanation was noted before: There are not enough data to estimate the parameters and the covariance matrix with any reasonable accuracy. Thus, an asymptotic formula has been misused in a finite-sample situation. (In Table 1, we do not recommend comparing columns 2 and 5; the underlying models have different parameters, and different error structures. We believe the comparison between columns 4 and 5 shows that column 2 is too small; this is an inductive step.)

The shapes of the bootstrap distributions may be of some interest. The coefficient estimates like $\hat{f}^*$ are close to normally distributed, as may be anticipated. A bit more surprising: the nominal SEs are close to normal too, and not especially variable. Take value-added, for example. Let SE_i be the nominal SE from the ith starred data-set. A histogram for these 100 numbers is close to the normal curve, with mean .014 and an SD of .004.

A sidelight is the bias in the *gls* coefficient estimates. For a simple auto-regression it is well known that the least squares coefficient estimates are biased; see Hurwicz (1950). The estimates in the more complicated dynamic model considered here also exhibit significant bias, for a similar reason. Compare columns 1 and 3 in Table 1. For instance, the coefficient f for value-added was set to the estimated value .281 in the construction of the pseudo-data. However, the 100 coefficients $\hat{f}^*$ had a sample average of .310. The discrepancy is .029. A standard error for the discrepancy can be calculated from the standard deviation of the $\hat{f}^*$ divided by the square root of the number of replications, $.039/\sqrt{100} = .0039$. The t-value is $.029/.0039 = 7.4$ on 99 degrees of freedom, so the bias is significant. The coefficient for the lag term is also significantly biased; the remaining coefficients, less so. When the lag is removed from the model (1), the bias in the *gls* coefficient estimates subsides. The usual argument to show (gls,k) estimates are unbiased depends on the assumption that ϵ has a symmetric distribution given the design matrix. When the lag term is dropped, this is approximately so.

More interesting for present purposes: when the lag term is dropped, the conventional estimates of standard errors are still too optimistic, by factors like those in Table 1. Thus the bias in

the conventional asymptotics is not due to the autoregressive structure. Likewise, Table 1 can be re-run using a multivariate Gaussian distribution for the errors, with mean 0 and covariance matrix equal to the empirical covariance matrix $\hat{K}_1$ of the residuals. This covariance matrix is displayed in Table 2. Again, the results do not change much. Thus, the bias in the conventional asymptotics is not due to the discreteness of the error distribution.

Table 2. $\hat{K}_1$, the inter-regional covariance matrix estimated from the one-step *gls* residuals in equation (9). Entries have been scaled up by 10^3. The 10×10 matrix is symmetric; only the upper half is reported.

2.031	1.186	1.087	1.008	.942	1.064	1.359	.881	.565	.695
	2.989	.993	1.118	.569	1.161	.937	.297	.355	.208
		1.184	.831	.643	.650	.905	.392	.490	.337
			1.064	.630	.594	.672	.420	.491	.367
				.580	.394	.805	.554	.363	.311
					1.302	.433	.0243	.144	.227
						1.906	.824	.717	.257
							1.567	.367	−.125
								1.049	−.0079
									1.086

We also redid the simulation experiments, using Gaussian errors, eliminating the lag, and selectively removing the weather and price variables; we had 3, 6, and 10 regions; we had i.i.d. errors, as well as errors with covariance. The results were somewhat surprising:

▶ The condition number of the design matrix does not indicate the probable magnitude of the bias in the conventional standard errors.

▶ Decreasing the number of regions sometimes increased the bias.

▶ So did the change from correlated to i.i.d. errors.

The quality of the bootstrap estimates of standard error will now be checked by a simulation experiment. It will be shown that these estimates are much better than the conventional ones, but are still biased downwards. The details may be a bit complicated, but the main idea is straightforward. We will check the bootstrap by trying it out in a simulation world where we know the answers. Going back

to Table 1, the "SD" column shows the "real" variability in the (gls,1) estimates, in the simulation world of the bootstrap. The "RMS Nominal SE" column shows the variability indicated by the conventional formulae. The last column in the table (developed from the procedure described below) shows the variability indicated by the bootstrap.

The experiment involves a nested iteration: at the "outer loop" starred data sets are built up one after another, and presented to the "inner loop" bootstrap for an estimate of the standard errors. Here are the details. The outer loop is just the bootstrap procedure described above: ϵ_t^*, $\hat{q}_{rt}^*$, $\hat{a}_r^*$, etc. are as previously defined. Let $\hat{\epsilon}_{rt}^*$ be the residuals:

$$\hat{\epsilon}_{rt}^* = q_{rt}^* - \hat{a}_r^* - \hat{b}^* c_{rt} - \hat{c}^* h_{rt} - \hat{d}^* p_{rt} - \hat{e}^* q_{r,t-1}^* - \hat{f}^* v_{rt}$$

Let $\hat{\epsilon}_t^*$ be the 10-vector $(\hat{\epsilon}_{1,t}^*, \ldots, \hat{\epsilon}_{10,t}^*)$ of residuals for year t. Let μ^* be the empirical distribution of $\{\hat{\epsilon}_t^* : t = 1961, \ldots, 1978\}$. So μ^* will change on each pass through the outer loop.

On each pass through the inner loop generate ϵ_t^{**} for $t = 1961, \ldots, 1978$ as eighteen independent draws from μ^*. Let ϵ_{rt}^{**} denote the rth component of ϵ_t^{**}. Construct a doubly-starred data set: $q_{rt}^{**} = q_{r,1960}$ and for $t = 1961, \ldots, 1978$

$$q_{rt}^{**} = \hat{a}_r^* + \hat{b}^* c_{rt} + \hat{c}^* h_{rt} + \hat{d}^* p_{rt} + \hat{e}^* q_{r,t-1}^{**} + \hat{f}^* v_{rt} + \epsilon_{rt}^{**}$$

Obtain the doubly-starred parameter estimates $\hat{a}_r^{**}$, $\hat{b}^{**}$, $\ldots$, $\hat{f}^{**}$ by the (gls,1) regression of q_{rt}^{**} on c_{rt}, h_{rt}, p_{rt}, $q_{r,t-1}^{**}$ and v_{rt}.

The "outer loop" may be repeated to develop the distribution of these bootstrap standard errors. Column 6 of Table 1 summarizes an experiment with 100 passes through the outer loop, and at each pass there were 100 passes through the inner loop. Column 6 gives the root mean square of the 100 bootstrap estimates for the standard error, each such estimate being itself the standard deviation of 100 doubly-starred estimates. Consider, for example, the coefficient d of the price term. Let i index the outer loop, and j index the inner loop. On pass i through the outer loop and pass j through the inner loop, a doubly-starred parameter estimate $\hat{d}^{**}$ is computed; call this value $\hat{d}_{ij}$. On pass i, the bootstrap standard error is the standard deviation of the 100 numbers $\{\hat{d}_{ij} : j = 1, \ldots, 100\}$: call this SD_i. Then the last column of Table 1 reports

$$\sqrt{\frac{1}{100} \sum_{i=1}^{100} SD_i^2} \approx .022$$

This is the typical standard error for d estimated by the bootstrap method, in the simulation world. The "real" (gls,1) parameter variability is displayed in column 4 and is .028. Column 6 is uniformly smaller than column 4, indicating the bias in the bootstrap procedure. But the bootstrap is closer to the mark than the conventional asymptotics, shown in column 5. Indeed, the bootstrap is off by 20 to 30 percent; the conventional asymptotics, by factors ranging from 1.5 to 3.

One problem, both for the bootstrap and for the conventional asymptotics, is that the residuals $\hat{\epsilon}$ tend to be smaller than the disturbance term ϵ, due to the effect of fitting. In some designs, $e.g.$ the standard regression model, there is an easy fix, namely scaling up the residuals by $\sqrt{n/(n-p)}$. This fix is not appropriate here. Due to the inter-regional constraints, the bias in $\hat{\Sigma}$ turns out to depend in a complicated way on the design matrix and Σ. However, the bootstrap can be used as a bias-correction device for Σ, and this reduces the bias in the bootstrap SEs to below 10 percent.

5. Some mathematics

Why are the nominal SEs so badly biased in RDFOR? The main reason is that the true *gls* estimator depends on Σ; the approximate *gls* estimators replace Σ by an estimate $\hat{\Sigma}$, and this source of error is ignored by the conventional asymptotics. More particularly:

▶ The conventional formula $(X^T\hat{\Sigma}^{-1}X)^{-1}$ is a concave function of $\hat{\Sigma}$, and this creates a downwards bias, which is severe when $\hat{\Sigma}$ is quite variable — even if $\hat{\Sigma}$ were an unbiased estimate of Σ.

▶ In fact, the conventional estimate $\hat{\Sigma}$ for Σ is biased downward in RDFOR, due to the constraints.

The object of this section is to give a mathematical treatment of the concavity issue, in settings much simpler than RDFOR. The bias in $\hat{\Sigma}$ will not be discussed here.

Consider first the one-way analysis of variance model

$$Y_{rt} = \alpha + \epsilon_{rt}, \quad E(\epsilon_{rt}) = 0, \quad \operatorname{var}\epsilon_{rt} = \sigma_r^2$$

where α is an unknown location parameter to be estimated; the ϵ_{rt} are independent for $r = 1, \ldots, R$ and $t = 1, \ldots, T$; for each r, they are identically distributed, but this distribution may depend on r. Suppose $R \geq 2$ and $T \geq 2$. When $\sigma_1^2, \ldots, \sigma_R^2$ are known, the *gls* estimate for α is

$$\hat{\alpha}_{gls} = \frac{\sum_{r=1}^{R} \overline{Y}_r/\sigma_r^2}{\sum_{r=1}^{R} 1/\sigma_r^2} \tag{13}$$

where $\overline{Y}_r = \frac{1}{T}\sum_{t=1}^{T} Y_{rt}$. Of course,

$$\operatorname{var}\hat{\alpha}_{gls} = 1/[\sum_{r=1}^{R} T/\sigma_r^2] \tag{14}$$

If the σ_r^2 are unknown, consider (13) and (14) with σ_r^2 replaced by the unbiased estimate

$$\hat{\sigma}_r^2 = \frac{1}{T-1}\sum_{t=1}^{T}(Y_{rt} - \overline{Y}_r)^2$$

Namely,

$$\hat{\alpha}_{agls} = \frac{\sum_{r=1}^{R} \overline{Y}_r/\hat{\sigma}_r^2}{\sum_{r=1}^{R} 1/\hat{\sigma}_r^2} \tag{15}$$

and

$$\widehat{var} = 1/[\sum_{r=1}^{R} T/\hat{\sigma}_r^2] \tag{16}$$

This $\hat{\alpha}_{agls}$ is a (gls,1) estimate. And $\widehat{var}$ is a variability estimate analogous to the conventional formula (8) considered previously. The next theorem shows that the random variable $\widehat{var}$ tends to be too small. This is a finite-sample result: asymptotically, gls and $agls$ are equivalent.

Theorem 1. *If the ϵ_{rt} are normally distributed, then*

$$\text{var } \hat{\alpha}_{agls} > \text{var } \hat{\alpha}_{gls} > E(\widehat{var}) \tag{17}$$

That is, provided the errors are normal, the true variability of $\hat{\alpha}_{agls}$ must exceed the variability of $\hat{\alpha}_{gls}$, and this in turn exceeds the expected value of $\widehat{var}$.

Proof: To verify the first inequality, notice that when the ϵ_{rt} are normally distributed, $\overline{Y}_r$ and $\hat{\sigma}_r^2$ are independent random variables. Condition on $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2$. Clearly,

$$E(\overline{Y}_r | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2) = \alpha \quad \text{and} \quad \text{var}(\overline{Y}_r | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2) = \sigma_r^2/T$$

Then the conditional minimum variance unbiased linear estimator for α is still the $\hat{\alpha}_{gls}$ defined by (13). So with probability one,

$$\text{var}(\hat{\alpha}_{agls} | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2) > \text{var}(\hat{\alpha}_{gls} | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2) = \text{var } \hat{\alpha}_{gls}$$

But

$$\text{var } \hat{\alpha}_{agls} = E\{\text{var}(\hat{\alpha}_{agls} | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2)\} + \text{var}\{E(\hat{\alpha}_{agls} | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2)\}$$

The first term on the right is greater than var $\hat{\alpha}_{gls}$. The second term is zero, since $E(\hat{\alpha}_{agls} | \hat{\sigma}_1^2, \ldots, \hat{\sigma}_R^2) = \alpha$. This establishes the first inequality in (17). For the second inequality, $\hat{\sigma}_r^2$ is an unbiased estimate of σ_r^2, and these variables are independent; also, $1/\sum_{i=1}^{R} T/\xi_i$ is strictly concave in each of its arguments: now use Jensen's inequality. ∎

When the normality assumption is not satisfied, (17) may fail to hold. For example, set $\alpha = 0$, $R = 2$, $T = 2$, and let the ϵ_{rt} be independent and identically distributed with a common distribution μ. Take μ as a mixture of two normal distributions

$$\mu = (1 - \eta)\Phi_{0,1} + \eta\Phi_{0,\sigma^2}$$

-- 18 --

where Φ_{0,σ^2} is the normal distribution with mean 0 and variance σ^2. Thus, μ is symmetric and unimodal. Let η be small and σ^2 large so $\eta\sigma^2$ is moderately large. Then

$$\text{var}\,\hat{\alpha}_{agls} < \text{var}\,\hat{\alpha}_{gls}$$

In effect, $\hat{\alpha}_{agls}$ is a trimmed mean. With this example, however, $E(\widehat{var}) < \text{var}\,\hat{\alpha}_{agls}$. We do not know what happens in general.

The next result is a fairly straightforward extension of Theorem 1 to the general multivariate model. To state the result, consider the model

$$\underset{1\times q}{Y_i} = \underset{1\times p}{X_i}\ \underset{p\times q}{C} + \underset{1\times q}{\epsilon_i} \qquad \text{for } i = 1,\dots,n$$

where X_i is nonrandom, and the coefficient matrix C will be constrained to fall in the linear space Λ. Linear constraints of this sort are common in econometric work; for example, components of C may be constrained to vanish. The Gaussian disturbances ϵ_i have mean zero; they are independent and identically distributed in i, but have an arbitrary positive definite covariance matrix $\text{cov}(\epsilon_i) = K$. Suppose $q \geq 2$, $n \geq p + q$, and $S = \sum_{i=1}^{n} X_i^T X_i$ is nonsingular.

In this circumstance, unconstrained *ols* and *gls* estimators for C coincide, see Schmidt (1976, p. 78), or Theil (1971, p. 309); call them $\hat{C}_0$. Stack the q columns of $\hat{C}_0$ to form a $pq \times 1$ vector denoted $\text{vec}[\hat{C}_0]$. The covariance matrix of $\text{vec}[\hat{C}_0]$ is $K \otimes S^{-1}$, and

$$(K \otimes S^{-1})^{-1} = K^{-1} \otimes S = \begin{bmatrix} k^{11}S & k^{12}S & \dots & k^{1q}S \\ k^{21}S & k^{22}S & \dots & k^{2q}S \\ \vdots & \vdots & & \vdots \\ k^{q1}S & k^{q2}S & \dots & k^{qq}S \end{bmatrix}$$

where k^{ij} is the ij entry of K^{-1}. For a discussion, see Anderson (1958, Section 8.2.2).

Let $\hat{\epsilon}_i = Y_i - X_i\hat{C}_0$. Let $\hat{K}$ be the empirical covariance matrix of the $\hat{\epsilon}_i$, scaled by $n/n - p$ to be an unbiased estimate of K. Let $\hat{C}_{gls}$ be the true *gls* estimator of C constrained to fall in Λ, with K known. Let $\hat{C}_{agls}$ be the approximate *gls* estimator of C constrained to fall in Λ, with K unknown but estimated by $\hat{K}$. As usual, $\hat{C}_{gls}$ is obtained by projecting $\hat{C}_0$ into Λ relative to $K^{-1} \otimes S$, while $\hat{C}_{agls}$ is obtained by projecting $\hat{C}_0$ into Λ relative to $\hat{K}^{-1} \otimes S$. The covariance matrix for $\hat{C}_{gls}$ is a function of K, obtained as in Section 3; and the estimated covariance matrix $\widehat{cov}$ is obtained by substituting $\hat{K}$ for K. Suppose *ols* and *gls* differ, when the constraints are imposed.

Theorem 2. $\operatorname{cov} \widehat{C}_{agls} > \operatorname{cov} \widehat{C}_{gls} > E(\widehat{cov})$, *where* $M > N$ *means* $M - N$ *is nonnegative definite and* $M \neq N$.

This theorem is proved much like Theorem 1, because $\widehat{K}$ and $\widehat{C}_0$ are independent; $\widehat{K}$ is distributed like the empirical covariance matrix of $n - p$ independent draws from a multivariate Gaussian distribution, with mean 0 and covariance matrix K. In general, $M - N$ need not be strictly positive definite. This is because for some contrasts *ols* and *gls* may coincide, even though they differ on other contrasts. The following inequality is used in proving Theorem 2; Ylvisaker (1964).

Theorem 3. *Let* X *be an* $n \times p$ *matrix and* Σ *a* $p \times p$ *positive definite matrix. Then* $(X^T \Sigma^{-1} X)^{-1}$ *is a weakly concave function of* Σ.

Arnold Zellner (private communication) has considered the "seemingly unrelated" regression problem for two "regions:"

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

where Y_r is $T \times 1$, X_r is $T \times k_r$ nonrandom of full rank, β_r is $k_r \times 1$ and ϵ_r is $T \times 1$. He assumes $X_1^T X_2 = 0$, $E(\epsilon_r) = 0$, $E(\epsilon_r \epsilon_s^T) = \sigma_{rs} I_T$, with $\Sigma = \{\sigma_{rs}\}$ positive definite, and the ϵ's multivariate Gaussian. In this model, Zellner can compute the finite-sample covariance matrix for the approximate *gls* estimator; the asymptotic covariance matrix is biased downward, by $(k_1 + k_2 + 2)/T$. Srivastava and Dwivedi (1979) survey other such developments in the estimation of seemingly unrelated regressions. The inter-regional constraints in models like RDFOR seem to make this sort of calculation difficult: but see Section 7. In the one-way analysis of variance model with two regions and $\sigma_1^2 = \sigma_2^2$, we can compute the exact bias: it is $1/T$. For three regions, or unequal regional variances, or unequal numbers of observations per region, our computation fails.

6. Computational details

This section gives additional details about data, algorithms, and the stability of the Monte-Carlo experiments. All of the computer work reported here was performed using the TROLL econometric modeling system running on an IBM 370/168 at M.I.T. The cost for the simple bootstrap experiments reported in this section was $10. Validating the bootstrap was more expensive, about $120, but this procedure would not be routinely used in practice. The SEDS data base is installed in the TROLL file-system as a collection of single-precision data series. Listings of the relevant data-series and of the TROLL functions used to construct the divisia index and to aggregate to the ten DOE regions are available on request.

The bootstrap experiments were conducted within the BOOTMOD subsystem of TROLL; see Peters (1983a). In this program, numerical linear algebra used for *ols* and *gls* fitting relies on the LINPACK library, described in Dongarra *et al* (1981); double-precision is maintained for all the fitting. Uniformly distributed pseudo-random numbers are obtained from the McGill University random number package "Super-Duper," described in Marsaglia *et al* (1976). This random number generator combines a congruential sequence with a shift register procedure and has very high quality. The uniform variates are used to select at random with replacement from the eighteen 10-vectors of residuals. The seeds (1073,12345) were used for the experiments reported here. The results of Table 1 were replicated in an unreported experiment using the seeds (31415,14121).

Turn now to the stability of the Monte-Carlo experiments. The bootstrap SEs obtained from the simulation experiments are random variables subject to sampling error. To get a rough idea of their stability, an approximation to the variance of the bootstrap SEs was calculated. The approximation is developed as follows. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with mean μ and variance σ^2. Let $\hat{\sigma}^2$ denote the sample variance of the X's,

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

. The object is to approximate $\operatorname{var} \hat{\sigma}^2$. To first order,

$$\operatorname{var} \hat{\sigma}^2 \doteq \frac{1}{n} \operatorname{var}\{(X_1 - \mu)^2\}$$

$$= \frac{1}{n}[E(X_1 - \mu)^4 - \sigma^4]$$

This last expression may be estimated from the sample by

$$\widehat{var}(\hat{\sigma}^2) = \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4 - \hat{\sigma}^4\right] \tag{18}$$

Of course,

$$\widehat{SE}(\hat{\sigma}^2) = \sqrt{\widehat{var}(\hat{\sigma}^2)}$$

Finally, an approximate standard error for $\hat{\sigma}$ is

$$\widehat{SE}(\hat{\sigma}) = \frac{1}{2}\frac{\widehat{SE}(\hat{\sigma}^2)}{\hat{\sigma}} \tag{19}$$

because

$$\sqrt{\hat{\sigma}^2 \pm \widehat{SE}(\hat{\sigma}^2)} = \sqrt{\hat{\sigma}^2\left[1 \pm \frac{\widehat{SE}(\hat{\sigma}^2)}{\hat{\sigma}^2}\right]} \approx \hat{\sigma}\left[1 \pm \frac{1}{2}\frac{\widehat{SE}(\hat{\sigma}^2)}{\hat{\sigma}^2}\right] = \hat{\sigma} \pm \frac{1}{2}\frac{\widehat{SE}(\hat{\sigma}^2)}{\hat{\sigma}}$$

For the bootstrap, identify X_i with the ith replicate of a starred parameter estimate, e.g. $\hat{f}_i^*$ in the ith starred data set. Then an estimate for the approximate variability of the bootstrap SE is easily calculated from (18–19), by accumulating fourth moments.

Table 3 shows the bootstrap SEs from column 4 of Table 1. Alongside stand the values calculated from (19). These are in the natural units for comparison: column 2 gives a rough standard error for column 1. For example, the bootstrap estimate for the SE of $\hat{f}$ is .039 from column 4 of Table 1. This estimate is based on a sample of size 100, namely, the bootstrap replications. How much does sampling error affect this estimate? The answer is given by the estimated approximate standard error of .0019, shown in the second column of Table 3; this is computed from (18–19). The entries in the second column are between 5 and 10 percent as large as those in the first column. The uncertainties are not large enough to change any conclusions that have been drawn. An approximation for the variability of the RMS Nominal SE (column 5 of Table 1) can be developed along similar lines, with similar results.

Table 3. Stability assessment for the bootstrap SEs. Estimation is
by one-step *gls*. There are 100 bootstrap replications.

		Bootstrap SE (from Table 1)	Approximate SE for Bootstrap SE
	a_1	.54	.046
	a_2	.55	.046
	a_3	.55	.046
	a_4	.53	.046
	a_5	.55	.047
	a_6	.53	.046
	a_7	.55	.048
	a_8	.55	.046
	a_9	.51	.044
	a_{10}	.54	.046
c.d.d.	b	.025	.0023
h.d.d.	c	.052	.0038
price	d	.028	.0026
lag	e	.042	.0024
v.a.	f	.039	.0019

7. On a formula of Srivastava and Dwivedi

To demonstrate the bias in the nominal SEs very clearly, consider a model like (9) with no lag term, normal errors, and region-specific parameters. This is precisely in the form of a "seemingly unrelated regression problem;" it is not a standard multivariate regression problem, because, e.g., the fuel price in region r does not appear in the equation for region s: as is usually said, "not all variables appear in all equations." Srivastava and Dwivedi (1979, p. 18) give an asymptotic expansion for the SEs, whose first term is the conventional "large sample" formula, and whose second term is a "finite sample" correction. Table 4 below show the results of a bootstrap experiment in this context; for simplicity of presentation, the table gives the root mean square of the indicated ratios across all 10 regions. As can be seen, the large-sample formula is off by as much as a factor of about two, and the Srivastava-Dwivedi formula is only somewhat better.

Table 4. Bootstrap results for a "seemingly unrelated regression;"
$\gamma.m.s.$ ~~averaged~~ across 10 regions. There are 100 bootstrap
replications.

		RMS Nominal SE Bootstrap SD	RMS Sriv.-Dwiv. SE Bootstrap SD
constant	a	.577	.714
c.d.d.	b	.515	.639
h.d.d.	c	.625	.772
price	d	.831	.980
v.a.	f	.677	.796

References

ANDERSON, T. W. (1958), An Introduction to Multivariate Statistical Analysis,
New York: John Wiley.

BELSLEY, D. A., KUH, E., and WELSCH, R. E. (1980), Regression Diagnostics,
New York: John Wiley.

BERAN, R. (1983), "Estimated sampling distributions: the bootstrap and competitors,"
Annals of Statistics, 10, 212–225.

BICKEL, P. J. , and FREEDMAN, D. A. (1981), "Some asymptotic theory for the bootstrap,"
Annals of Statistics, 9, 1196–1217.

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

BROWN, T. M. (1954), "Standard errors of forecast of a complete econometric model,"
Econometrica, 22, 178–192.

COX, D. R. , and HINKLEY, D. V. (1974), Theoretical Statistics, London: Chapman and Hall.

DONGARRA, J. J., BUNCH, J. R., MOLER, C. B., and STEWART, G. W. (1979), LINPACK Users' Guide,
Philadelphia: Society for Industrial and Applied Mathematics.

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

EFRON, B. (1982), The jackknife, the bootstrap, and other resampling plans, CBMS-NSF Regional
Conference Series in Applied Mathematics, Monograph 38, Philadelphia: Society for Industrial
and Applied Mathematics.

FAIR, R. (1979), "An analysis of the accuracy of four macro-economic models,"
Journal of Political Economy, 87, 701–718.

FAIR, R. (1980), "Estimating the expected predictive accuracy of econometric models,"
International Economic Review, 21, 355–378.

FIEBIG, D. G., and THEIL, H. (1983), "The two perils of symmetry-constrained estimation of demand systems," Technical report, Graduate School of Business Administration, University of Florida.

FREEDMAN, D. A. (1981), "Bootstrapping regression models," Annals of Statistics, 9, 1218–1228.

FREEDMAN, D. A. (1982), "On bootstrapping instrumental-variables least-squares estimation in stationary linear models," Technical report, Department of Statistics, University of California, Berkeley.

FREEDMAN, D. A. and PETERS, S. C. (1983a), "Bootstrapping an econometric model: some empirical results," Technical report, Department of Statistics, University of California, Berkeley.

FREEDMAN, D. A. and PETERS, S. C. (1983b), "Using the bootstrap to evaluate forecasting equations," Technical report, Department of Statistics, University of California, Berkeley.

FREEDMAN, D. A. , ROTHENBERG, T. , and SUTCH, R. (1983), "On energy policy models," Journal of Business and Economic Statistics, 1, 24–36.

GOLDBERGER, A. , NAGAR, A. L. , and ODEH, H. S. (1961), "The covariance matrices of reduced-form coefficients and of forecasts for structural econometric models," Econometrica, 29, 556–573.

HURWICZ, L. (1950), "Least-squares bias in time series," in Statistical Inference in Dynamic Economic Models, ed. T. C. Koopmans, New York: John Wiley.

KUH, E. , LAHIRI, S. , MINKOFF, A. , SWARTZ, S. , and WELSCH, R. (1982), "Analysis of the validity of the coefficient estimates and forecasting properties of the RDFOR models — a summary report," Technical report, M.I.T. Center for Computational Research in Economics and Management Science, Cambridge, Massachusetts.

MADDALA, G. S. (1971), "Generalized least squares with an estimated covariance matrix," Econometrica, 39, 22–33.

MARSAGLIA, G., ANANTHANARAYANAN, K., and PAUL, N. J. (1976), "Improvements on fast methods for generating normal random variables," Information Processing Letters, 5, 27–30.

MIKHAIL, W. M. (1975), "A comparative Monte-Carlo study of the properties of econometric estimators," JASA, 70, 94–104.

PETERS, S. C. (1983a), "TROLL experimental programs: BOOTMOD," M.I.T. Center for Computational Research in Economics and Management Science, Cambridge, Massachusetts.

PETERS, S. C. (1983b), Bootstrapping a Regression Equation: Some Empirical Results, Ph.D. dissertation, Department of Statistics, Stanford University.

PHILLIPS, P. C. B. (1977), "An approximation to the finite sample distribution of Zellner's seemingly unrelated regression estimator," Journal of Econometrics, 6, 147–164.

SCHMIDT, P. (1976), Econometrics, New York: Marcel Dekker.

SINGH, K. (1981), "On the asymptotic accuracy of Efron's bootstrap," Annals of Statistics, 9, 1187–1195.

SRIVASTAVA, V. K., and DWIVEDI, T. D. (1979), "Estimation of seemingly unrelated regression equations," Journal of Econometrics, 10, 15–32.

TAYLOR, W. E. (1977), "Small sample properties of two stage Aitken estimators," Econometrica, 45, 497–508.

THEIL, H. (1971), Principles of Econometrics, New York: John Wiley.

THEIL, H., FINKE, R., and ROSALSKY, M. C. (1983), "Verifying a demand equation by simulation," Technical report, Graduate School of Business Administration, University of Florida.

YLVISAKER, N. D. (1964), "Lower bounds for minimum covariance matrices in time series regression problems," Annals of Mathematical Statistics, 35, 362–368.

ZELLNER, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," JASA, 57, 348–368.

ZELLNER, A. (1963), "Estimators for seemingly unrelated regressions: some exact finite sample results," JASA, 58, 977–992.