# BOOTSTRAPPING AN ECONOMETRIC MODEL: SOME EMPIRICAL RESULTS

DAVID A. FREEDMAN AND STEPHEN C. PETERS

TECHNICAL REPORT NO. 21
MARCH 1983
REVISED
AUGUST 1983

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY

RESEARCH PARTIALLY SUPPORTED
BY THE
ENERGY INFORMATION ADMINISTRATION
THROUGH THE
OAK RIDGE NATIONAL LABORATORY

August, 1983
Prepared for JBES

# Bootstrapping an Econometric Model: Some Empirical Results

David A. Freedman
Statistics Department
University of California, Berkeley

Stephen C. Peters
Statistics Department
Stanford University

## Abstract

The bootstrap, like the jackknife, 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 bootstrap will be applied to an econometric model describing the demand for capital, labor, energy, and materials. The model is fitted by three-stage least squares. In sharp contrast with previous results, the coefficient estimates and the estimated standard errors perform very well. However, the model's forecasts show serious bias and large random errors.

Keywords and phrases: Regression, generalized least squares, two-stage least squares, three-stage least squares, econometric models, forecasting, standard errors.

Running head: Bootstrapping an econometric model

Acknowledgements

We would like to thank Douglas Hale, Edwin Kuh, and Thomas Rothenberg for useful discussions and other support. Computer facilities were provided by the Center for Computational Research in Economics and Management Science at M.I.T. The project was funded by the Office of Statistical Standards, Energy Information Administration, Department of Energy, through the Oak Ridge National Laboratory.

# 1. Introduction

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 in order to construct "pseudo-data" on which the estimator of interest is exercised. More specifically, the theoretical distribution of an unobservable disturbance term is approximated by the empirical distribution of an observable 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, generalized least squares with estimated covariance matrices, and developments like two-stage or three-stage least squares. This paper, a sequel to Freedman and Peters (1983), investigates the performance of widely used econometric estimators in an example consisting of a set of simultaneous equations. In the example studied here, for two-stage or three-stage least squares, the coefficient estimates and the conventional estimates of standard error perform very well. This is in marked contrast to the RDFOR example reported in Freedman and Peters (1983), and the model of the California canning-tomato industry discussed in Daggett and Freedman (1983). In the RDFOR example, the conventional estimates of standard error were too small by a factor of three. In the tomato example, there were similar difficulties with the standard errors; in addition, there was appreciable small-sample bias in the coefficient estimates. Elsewhere, we hope to explore the reasons for such differences.

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. Sections 3 and 4 give even briefer reviews of generalized least squares, two-stage, and three-stage least squares. These sections pinpoint the technical issue to be addressed by the bootstrap. Section 5 applies these ideas to the Berndt-Wood (1975) model. The bootstrap is used to attach standard errors to forecasts in Section 6. As it turns out, the Berndt-Wood model does well in policy-analysis mode, but quite badly in straight forecasting. Conclusions are drawn in Section 7.

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, 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. In effect, the bootstrap is just a simulation experiment, where the parameters are fixed at the values estimated from the real data, and the error distribution is chosen to be the empirical distribution of the residuals from the real data. These choices are by no means critical. In the context of "seemingly unrelated nonlinear regressions," Gallant (1975) reports a Monte Carlo experiment with artificial data-sets which demonstrates that the conventional estimates of standard error mildly underestimate actual coefficient standard error; he does not address forecasting performance.

## 2. The Bootstrap

The bootstrap is described by Efron (1979, 1982). Related papers are Bickel and Freedman (1981, 1983), Freedman (1981), and Freedman and Peters (1983). The bootstrap is a procedure for estimating standard errors by resampling 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 model is tested against its own assumptions.

A coherent simulation world is obtained by assuming the model, with the parameters fixed at their previously estimated values; the errors are assumed to follow the empirical distribution of the residuals. The resampling generates "pseudo-data," and the model can be re-fitted to the pseudo-data obtaining new parameter estimates. 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 (Section 5) and the forecasts (Section 6).

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

$$\underset{1\times q}{Y_t} = \underset{1\times q}{Y_t}\ \underset{q\times q}{A} + \underset{1\times p}{X_t}\ \underset{p\times q}{B} + \underset{1\times q}{\epsilon_t} \tag{1}$$

Linearity is assumed to simplify the exposition; the method is easily adapted to cover nonlinear models, although the computational costs may be prohibitive. Lagged variables are easily handled, but would only complicate the exposition. For an explicit treatment of dynamics, see Freedman and Peters (1983).

In the equation, $A$ and $B$ 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 $\epsilon_t$ 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 $\epsilon$, exogenous variables are

independent of $\epsilon$. The system (1) has $q$ equations in $q$ endogenous variables; there are $p$ exogenous variables. The matrix $I - A$ is assumed invertible; identifying restrictions are imposed on $A$ and $B$.

To be more specific about the stochastic assumptions:

$$\text{The } \epsilon_t \text{ are independent and identically distributed in time.} \tag{2a}$$

$$E(\epsilon_t) = 0 \text{ and } \operatorname{cov}(\epsilon_t) = V \text{ for all } t. \tag{2b}$$

$$\text{The } X\text{'s are independent of the } \epsilon\text{'s.} \tag{2c}$$

Here, $V$ is a $q \times q$ positive definite matrix. As is conventional, we normalize $A$ so $A_{jj} = 0$ for all $j$. Write $Y_{tj}$ for the $j^{th}$ component of the vector $Y_t$. Then the $j^{th}$ equation in the system (1) explains $Y_{tj}$ in terms of the other endogenous variables, and the exogenous variables:

$$Y_{tj} = Y_t A_j + X_t B_j + \epsilon_{tj}$$

Here $A_j$ is the $j^{th}$ column of $A$ and $B_j$ is the $j^{th}$ column of $B$.

Data are available for $t = 1, \ldots, n$. The coefficient matrices are estimated as $\hat{A}$ and $\hat{B}$ by some well-defined statistical procedure, for example three-stage least squares (Section 4). Coming now to the bootstrap, when $\hat{A}$ and $\hat{B}$ are computed, residuals are defined:

$$\hat{\epsilon}_t = Y_t - Y_t\hat{A} - X_t\hat{B} \tag{3}$$

These are estimates for the true disturbances $\epsilon_t$ in the model (1). Let $\mu$ 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) on centering. Also some inflation of the residuals may prove desirable, to compensate for the deflation 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; but this is not done here. It is unlikely that inflation or the BLUS transformation would have much impact in the present setting.

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

- Set the coefficients at $\hat{A}$ and $\hat{B}$ respectively.
- Make the disturbance terms independent, with common distribution $\mu$.

The exogenous $X$'s are kept fixed. Using this simulation model, pseudo-data can be generated for the past, namely periods $t = 1, \ldots, n$. These pseudo-data will be denoted by stars. The mechanism is suggested by solving (1) for $Y$:

$$Y_t^* = (X_t\hat{B} + \epsilon_t^*)(I - \hat{A})^{-1}, \tag{4}$$

where the $\epsilon^*$'s are independent with the common distribution $\mu$. It is assumed that $I - \hat{A}$ is invertible.

Now pretend the pseudo-data (4) 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 $\hat{A}^*$ and $\hat{B}^*$. The distribution of the pseudo-errors $\hat{A}^* - \hat{A}$ and $\hat{B}^* - \hat{B}$ can be computed, and used to approximate the distribution of the real errors $\hat{A} - A$ and $\hat{B} - B$. This approximation is the bootstrap. It is emphasized that the calculation assumes the validity of the model (1). The calculation also depends to some extent on the choice of parameter values and error distribution; although experience shows this dependence is slight. In case of doubt, various combinations of parameter values and error distributions can be tried.

The distribution of the pseudo-errors can be computed, for instance by Monte Carlo, simply repeating the procedure some number of times and seeing what happens. The distribution of the pseudo-errors is of interest only as an approximation to the distribution of the real errors.

## 3. Generalized Least Squares

Consider the model

$$Y = X\beta + \epsilon, \quad E(\epsilon) = 0, \quad \text{cov}(\epsilon) = \Sigma \tag{5}$$

For historical reasons, $X$ is called "the design matrix." With $\Sigma$ known, the generalized least squares (*gls*) estimate is

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

As usual,

$$E(\hat{\beta}_{gls}) = \beta \tag{7}$$

$$\text{cov}(\hat{\beta}_{gls}) = (X^T\Sigma^{-1}X)^{-1} \tag{8}$$

When $\Sigma$ is unknown, statisticians routinely use (6) and (8) with $\Sigma$ replaced by some estimate $\hat{\Sigma}$. Iterative procedures are often used, as follows. Let $\hat{\beta}^{(0)}$ be some initial estimate for $\beta$, 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

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

Let $\hat{\Sigma}_k$ be an estimator for $\Sigma$, based on $\hat{e}^{(k)}$. Then

$$\hat{\beta}^{(k+1)}_{gls} = (X^T\hat{\Sigma}_k^{-1}X)^{-1}X^T\hat{\Sigma}_k^{-1}Y \tag{9}$$

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

The covariance matrix for $\hat{\beta}^{(k+1)}_{gls}$ is usually estimated from (8), with $\hat{\Sigma}_k$ put in for $\Sigma$:

$$\widehat{cov}^{(k+1)} = (X^T\hat{\Sigma}_k^{-1}X)^{-1} \tag{10}$$

This may be legitimate, asymptotically. In finite-sample situations, all depends on whether $\hat{\Sigma}_k$ is a good estimate for $\Sigma$ or not. If $\hat{\Sigma}_k$ is a poor estimate for $\Sigma$, the standard errors estimated from (10) may prove to be unduly optimistic, and approximate *gls* estimators are often used when there is too little data to offer any hope of estimating $\Sigma$ with reasonable accuracy: an example is given in Freedman and

Peters (1983). In such circumstances, the bootstrap is a useful diagnostic, and in many cases 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 some examples, further iteration seems to make the coefficient estimates better, but also exaggerates the optimism of the standard error estimates. In other examples, iteration actually makes the coefficient estimators worse. The effects of additional iteration are considered in Peters (1983).

In econometric work, it is usual to constrain $\beta$ to fall in some linear space $\Lambda$; these are the "identifying restrictions." Typically, many elements of $\beta$ are constrained to vanish, and others are constrained to equal each other. The constraints are often incorporated by re-expressing the model in terms of linearly independent parameters: this involves linear manipulations of the columns of the design matrix $X$. A more elegant solution is to make unconstrained estimates, as indicated above, and then to project the unconstrained estimator $\hat{\beta}_u = \hat{\beta}_{gls}^{(k)}$ into the constraint space $\Lambda$. However, the projection must take into account the covariance structure of $\hat{\beta}_u$, *i.e.*, the constrained estimator $\hat{\beta}_c$ is the element of $\Lambda$ minimizing the distance

$$(\hat{\beta}_u - \hat{\beta}_c)^T \hat{W}^{-1} (\hat{\beta}_u - \hat{\beta}_c) \quad . \tag{11}$$

where $\hat{W}$ is the estimated covariance matrix of $\hat{\beta}_u$. Thus, the constrained *gls* estimator is found by two applications of unconstrained *gls*. See Chapter 6 of Theil (1971).

## 4. Two-stage and Three-stage Least Squares

The object of this section is to review two-stage least squares (2SLS) and three-stage least squares (3SLS). These are estimation techniques designed to overcome the problems created by the presence of endogenous variables, *i.e.*, explanatory variables correlated with the disturbance term. The present exposition is self-contained but terse. The idea is to bring out the connection with work reported in Freedman and Peters (1983). For a fuller account, see Theil (1971).

We return to the model (1). By conditioning on the exogenous $X$'$s$, we may suppose them nonrandom: see (2). Multiply (1) by $X_t^T$ and sum:

$$\underset{p\times q}{R} = \underset{p\times q}{R}\ \underset{q\times q}{A} + \underset{p\times p}{S}\ \underset{p\times q}{B} + \underset{p\times q}{\Delta} \tag{12}$$

where

$$R = \sum_{t=1}^{n} X_t^T Y_t \qquad S = \sum_{t=1}^{n} X_t^T X_t \qquad \Delta = \sum_{t=1}^{n} X_t^T \epsilon_t \tag{13}$$

Notice that the $j^{th}$ column of (12) corresponds to the $j^{th}$ equation in (1). The manipulation leading to (12) can be motivated by analogy to *ols*.

In applications, $[A, B]$ is constrained to fall in some linear space $\Lambda$ of dimension at most $pq$. Then $A$ and $B$ can be estimated from (12) by some variant of constrained least squares, as in the previous section. Notice that $S$ is constant (non-random) since $X$ is. It is conventional to treat $R$ on the right side of (12) as constant. This may be legitimate asymptotically, but is false in any finite sample. Indeed, $R$ is correlated with $\Delta$, and this is the source of "small-sample bias" in 2SLS and 3SLS estimators.

The matrix of errors $\Delta$ on the right hand side of (12) has covariance structure, so generalized least squares is the procedure of choice. To make contact with the standard format of (5), we stack the columns in (12): column #1 on top of column #2, ..., on top of column #$q$. In the stack, information corresponding to the first equation in model (1) comes first, information about the last equation comes last.

The parameter vector $\beta$ in (5) is obtained by stacking $A$ and $B$: column #1 of $A$, followed by column #1 of $B$, ..., followed by column #$q$ of $A$, followed by column #$q$ of $B$. The design matrix is obtained by writing $R$ and $S$ down the diagonal, and padding with zeroes.

The left hand side $Y$ vector in (5) consists of the stacked $R$ matrix; the error $\epsilon$ vector consists of the stacked $\Delta$ matrix. The full system of equations (12) is layed out in stacked form below, with

$R_j$ being the $j^{th}$ column of the matrix $R$, and likewise for the other matrices.

$$\begin{bmatrix} R_1 \\ \vdots \\ R_q \end{bmatrix} = \begin{bmatrix} R\ S\ 0\ 0 \ldots 0\ 0 \\ 0\ 0\ R\ S \ldots 0\ 0 \\ \vdots\ \vdots \qquad\qquad \vdots\ \vdots \\ 0\ 0\ 0\ 0 \ldots R\ S \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ \vdots \\ A_q \\ B_q \end{bmatrix} + \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_q \end{bmatrix} \tag{14}$$

At this point, the design matrix is highly singular, having dimension $pq \times (pq + q^2)$. Usually the elements of $\beta$ known *a priori* to vanish are suppressed, and the design matrix is adjusted accordingly by deleting the corresponding columns. An alternative approach is to use generalized inverses: see Chapter 6 of Theil (1971), or Maddala (1977, p. 447). By (2), the covariance matrix of the error vector (the stacked $\Delta$ matrix) is the Kronecker product

$$\Sigma = V \otimes S = \begin{bmatrix} V_{11}S & V_{12}S & \ldots & V_{1q}S \\ V_{21}S & V_{22}S & \ldots & V_{2q}S \\ \vdots & \vdots & & \vdots \\ V_{q1}S & V_{q2}S & \ldots & V_{qq}S \end{bmatrix} \tag{15}$$

We can now give a brief description of two-stage least squares, focussing on the connection with generalized least squares, and sharpen~ning the idea that 2SLS is a "single-equation" method. Consider each column of (12) in isolation. Take column $j$, corresponding to equation $j$ in system (1):

$$\underset{p\times 1}{R_j} = \underset{p\times q}{R}\ \underset{q\times 1}{A_j} + \underset{p\times p}{S}\ \underset{p\times 1}{B_j} + \underset{p\times 1}{\Delta_j} \tag{16}$$

The 2SLS procedure amounts to estimating (16) by *gls*, treating $R$ on the right as a constant. The constraints specific to the $j^{th}$ equation would be imposed, but not the cross-equation constraints. The covariance matrix of $\Delta_j$ is required. Plainly, $\operatorname{cov} \Delta_j = V_{jj}S$, where $S = \sum_{t=1}^{n} X_t^T X_t$ by (13). Thus $S$ is computable from the data; $V_{jj}$ is unknown, but enters only as a constant of proportionality, and its value is immaterial. With large enough samples, this procedure is preferable to *ols*, because it takes account of the correlation between $\epsilon$ and $Y$ on the right side of (1): this correlation would make *ols* inconsistent.

Let $\hat{A}_{II}$ and $\hat{B}_{II}$ denote the 2SLS estimators. To estimate their covariances, let

$$\hat{\epsilon}_t = Y_t - Y_t\hat{A}_{II} - X_t\hat{B}_{II} \tag{17a}$$

$$\hat{V} = \frac{1}{n}\sum_{t=1}^{n} \hat{\epsilon}_t^T\hat{\epsilon}_t \tag{17b}$$

The $\hat{\epsilon}$'s are the residuals, and $\hat{V}$ is an empirical covariance matrix which estimates $V$ in (2b) and (15). And the covariance matrix of $[\hat{A}_{II}, \hat{B}_{II}]$ can be estimated by the *gls* formula (8) as $(X^T\hat{\Sigma}^{-1}X)^{-1}$, where $X$ denotes the relevant design matrix, and $\Sigma = \text{cov}\,\Delta_j = V_{jj}S$ is estimated as $\hat{\Sigma} = \hat{V}_{jj}S$. It is conventional, for the purpose of estimating $\text{cov}[\hat{A}_{II}, \hat{B}_{II}]$ only, to inflate $\hat{V}_{jj}$ by $n/n-r$, where $r$ is the number of variables actually coming into the $j^{th}$ equation.

Turn now to 3SLS. Having completed the 2SLS estimation, we have in hand the estimator $\hat{V}$ for $V$. Approximate generalized least squares can now be applied to the full system (12), putting $\hat{V}$ in for $V$ at (15); the entries are not inflated: and we are pretending that $R$ on the right in (12) is constant. All constraints are imposed. This gives the three-stage least squares estimates, to be denoted $\hat{A}$ and $\hat{B}$. The covariance of these estimators can be estimated as before; a new and much larger design matrix $X$ is involved; and $\hat{\Sigma} = \hat{V} \otimes S$ estimates the covariance matrix $\Sigma = V \otimes S$ of the stacked $\Delta_j$'s: see (14) and (15). The discussion indicates how 3SLS is a "system-wide" method, and brings out the connection with *gls*. With large samples, 3SLS improves on 2SLS, by taking into account all the constraints and all the covariance structure.

## 5. Bootstrapping the Berndt-Wood Model

The Berndt-Wood (1975) model explains the shares in year $t$ of capital $S_{tK}$, labor $S_{tL}$, energy $S_{tE}$, and materials $S_{tM}$, in terms of their prices $p_{tK}$, $p_{tL}$, $p_{tE}$, and $p_{tM}$. There are 11 exogenous variables (including a constant) whose values at time $t$ are written as the $1 \times 11$ row vector $X_t$, with $X_{t1} \equiv 1$. (The exogenous variables are defined in the Berndt-Wood paper.) Data runs from 1947 through 1971. For a review of the model, see Kuh and Welsch (1980); also see Berndt-Wood (1979). The equations of the model may be written as follows:

$$S_{tK} = \alpha_K + \gamma_{KK} \log \frac{p_{tK}}{p_{tM}} + \gamma_{KL} \log \frac{p_{tL}}{p_{tM}} + \gamma_{KE} \log \frac{p_{tE}}{p_{tM}} + \epsilon_{tK} \tag{18a}$$

$$S_{tL} = \alpha_L + \gamma_{LK} \log \frac{p_{tK}}{p_{tM}} + \gamma_{LL} \log \frac{p_{tL}}{p_{tM}} + \gamma_{LE} \log \frac{p_{tE}}{p_{tM}} + \epsilon_{tL} \tag{18b}$$

$$S_{tE} = \alpha_E + \gamma_{EK} \log \frac{p_{tK}}{p_{tM}} + \gamma_{EL} \log \frac{p_{tL}}{p_{tM}} + \gamma_{EE} \log \frac{p_{tE}}{p_{tM}} + \epsilon_{tE} \tag{18c}$$

The $\gamma$-matrix is constrained by Berndt and Wood to be symmetric:

$$\gamma_{KL} = \gamma_{LK} \qquad \gamma_{KE} = \gamma_{EK} \qquad \gamma_{LE} = \gamma_{EL} \tag{19}$$

To complete the model for simulation purposes, we adjoin three more equations, explaining the log price ratios in terms of the exogenous variables. This was necessary to construct simulated data-sets for the bootstrap. The 3SLS estimates of the full system of equations give values for $\alpha$ and $\gamma$ identical to the 3SLS estimates for the original three-equation system (18). However, the coefficient estimates for the three adjoined equations differ from the *ols* estimates. For an outline of proofs, see Theil (1971, p. 514).

To make contact with (1), we write the complete system as

$$\underset{1\times 6}{Y_t} = \underset{1\times 6}{Y_t} \underset{6\times 6}{A} + \underset{1\times 11}{X_t} \underset{11\times 6}{B} + \underset{1\times 6}{\epsilon_t} \tag{20}$$

The assumptions (2) on the $\epsilon$'s are in force. Here

$$\begin{aligned} Y_{t1} &= S_{tK} & Y_{t2} &= S_{tL} & Y_{t3} &= S_{tE} \\ Y_{t4} &= \log \frac{p_{tK}}{p_{tM}} & Y_{t5} &= \log \frac{p_{tL}}{p_{tM}} & Y_{t6} &= \log \frac{p_{tE}}{p_{tM}} \\ \epsilon_{t1} &= \epsilon_{tK} & \epsilon_{t2} &= \epsilon_{tL} & \epsilon_{t3} &= \epsilon_{tE} \end{aligned}$$

The other components of $\epsilon$ do not appear explicitly in (18). The matrix $A$ is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_{KK} & \gamma_{LK} & \gamma_{EK} & 0 & 0 & 0 \\ \gamma_{KL} & \gamma_{LL} & \gamma_{EL} & 0 & 0 & 0 \\ \gamma_{KE} & \gamma_{LE} & \gamma_{EE} & 0 & 0 & 0 \end{bmatrix}$$

The symmetry constraints (19) are imposed. Notice that $I - A$ and $I - \hat{A}$ are automatically invertible, due to the zeroes in $A$ and $\hat{A}$. The matrix $B$ is as follows:

$$B_{11} = \alpha_K \quad B_{12} = \alpha_L \quad B_{13} = \alpha_E$$
$$B_{ij} = 0 \quad \text{for } 2 \leq i \leq 11 \text{ and } 1 \leq j \leq 3$$
$$B_{ij} \quad \text{is free for } 1 \leq i \leq 11 \text{ and } 4 \leq j \leq 6$$

We fit the model (20) by 3SLS. The coefficient estimates ($\alpha$ and $\gamma$ only) are shown in the first column of Table 1; the nominal standard errors, computed as in Section 4, in the second column. Our results differ slightly from those in Berndt-Wood (1975), since they fit by "iterated" 3SLS: in terms of the notation developed in Section 3, we use the (*gls*,1) estimators, and they iterate to convergence.

Table 1. Bootstrap experiment for equation (18). Estimation is by 3SLS. There are 400 bootstrap replications. The $t$-test is for the difference between the "mean" and the assumed value shown in the "estimate" column.

| | 3SLS | | Bootstrap | | | |
|---|---|---|---|---|---|---|
| | (1) Estimate | (2) Nominal SE | (3) Mean | (4) SD | (5) $t$ | (6) RMS Nominal SE |
| $\alpha_K$ | .0565 | .00162 | .0568 | .00151 | −3.58 | .00125 |
| $\alpha_L$ | .2540 | .00199 | .2537 | .00210 | 3.05 | .00184 |
| $\alpha_E$ | .0443 | .00112 | .0443 | .000950 | .632 | .000821 |
| $\gamma_{KK}$ | .0264 | .00837 | .0288 | .00679 | −6.89 | .00592 |
| $\gamma_{KL}$ | .000551 | .00424 | .00000360 | .00414 | 2.64 | .00353 |
| $\gamma_{KE}$ | −.00994 | .00449 | −.0105 | .00383 | 2.72 | .00362 |
| $\gamma_{LL}$ | .0734 | .00618 | .0743 | .00702 | −2.48 | .00593 |
| $\gamma_{LE}$ | −.00439 | .00287 | −.00429 | .00257 | −.778 | .00228 |
| $\gamma_{EE}$ | .0200 | .00833 | .0183 | .00563 | 5.97 | .00530 |

To get started on the bootstrap, let $\hat{A}$ and $\hat{B}$ be the 3SLS parameter estimates partially reported in Table 1. Consider the residuals

$$\hat{\epsilon}_t = Y_t - Y_t\hat{A} - X_t\hat{B}$$

Let $\mu$ be the empirical distribution of $\{\hat{\epsilon}_t : t = 1947, \ldots, 1971\}$. Note that $\mu$ has mean 0, because (20) has intercepts. Now simulate the equation (20), where all the ingredients are known:

- The exogenous variables are held fixed.
- The coefficient matrices are set at their estimated values $\hat{A}$ and $\hat{B}$.
- The disturbance terms are independent with common distribution $\mu$.

More specifically, let $\{\epsilon_t^* : t = 1947, \ldots, 1971\}$ be the results of twenty-four draws made at random with replacement from the set of twenty-four 6-vectors $\{\hat{\epsilon}_t : t = 1947, \ldots, 1971\}$. Thus $\hat{\epsilon}_{1947}$ may be drawn twice, but $\hat{\epsilon}_{1948}$ not at all. On the other hand, the pattern of the disturbances across equations does not change. Thus, the simulation preserves the key stochastic assumptions: the disturbances are independent and identically distributed in time but show a pattern across equations. The pseudo-data can now be built up for $t = 1947, \ldots, 1971$:

$$Y_t^* = (X_t\hat{B} + \epsilon^*)(I - \hat{A})^{-1}.$$

Bootstrap parameter estimates $\hat{A}^*$ and $\hat{B}^*$ can now be obtained by applying 3SLS to the starred data.

This procedure was repeated 400 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 in the original model the sample mean and sample standard deviation (SD) for these 400 starred estimates. These SDs are the bootstrap estimates of variability in the parameter estimates.

Comparing columns 1 and 3 in Table 1 indicates the presence of a significant small-sample bias in the coefficient estimates, due to the fact that $R$ in (12) is correlated with $\Delta$. For example, in our simulation, $\gamma_{KL}$ is set at $5.51 \times 10^{-4}$; the 3SLS estimate averages $3.60 \times 10^{-6}$. This difference is very unlikely to be due to variability in the Monte-Carlo, as shown by the $t$-statistic in column 5. With 400 replications the standard error of the mean is

$$4.14 \times 10^{-3}/\sqrt{400} = 2.07 \times 10^{-4}$$

so

$$t = (5.51 \times 10^{-4} - 3.60 \times 10^{-6})/(2.07 \times 10^{-4}) = 2.64$$

The difference between the true value of $\gamma_{KL}$ and the mean of the estimated values is very hard to explain by sampling error ($p < .005$). The other entries in column 5 may be interpreted in a similar way. Small-sample bias is statistically significant in four of the estimates; however, the practical significance may be small, because the means are so close to the assumed values. On the whole, 3SLS is performing well: bias in the coefficient estimates is small in practical terms, and the nominal standard errors are close to right.

Column 6 in Table 1 shows the RMS of the nominal SEs computed by applying formula (10) to the starred data set. Take, for example, $\gamma_{KL}$. In the simulation world of the experiment, the real variability of the 3SLS estimate for $\gamma_{KL}$ is .00414 from column 4. But the apparent variability, from the conventional asymptotics, is in a typical run only .00353, from column 6. This is

$$\sqrt{\frac{1}{400}\sum_{i=1}^{400} \mathrm{SE}_i^2}$$

where $\mathrm{SE}_i$ is the nominal SE for $\gamma_{KL}$ computed from (10) applied to the $i^{th}$ starred data-set. In a typical starred data-set, the formula (10) is off by about 10 percent. The other entries in column 6 may be interpreted in a similar way. The nominal SEs are too small by 10 to 20 percent, but the bias is not nearly as bad as it was in RDFOR; see Freedman and Peters (1983). Much of the bias may be due to the impact of fitting, in making the residuals smaller than the disturbance terms. On the whole, the conventional formulae seem to be doing very well. (Whether or not they have moments, the bootstrap distributions of the coefficient estimates are well approximated by normal distributions with the parameters indicated in Table 1.)

Table 2 shows similar results for 2SLS estimation. The pseudo-data are built up as before, *i.e.*, $\hat{A}$ and $\hat{B}$ are the 3SLS estimates based on the original data. However, the parameter estimates $\hat{A}^*_{II}$ and $\hat{B}^*_{II}$ are computed from the pseudo data using 2SLS, and their nominal SEs are computed using the appropriate conventional formulae. Again, small-sample bias in the coefficient estimates is statistically significant, and in several cases ($\gamma_{KL}, \gamma_{LK}, \gamma_{LE}$) perhaps of practical significance. There is no pattern of bias in the nominal SEs: some are a bit high, some a bit low. The correction factor $n/n-r$ discussed near the end of Section 4 seems to help. (We used 3SLS estimated coefficients and residuals for the simulation, as these are considered the most reliable.)

Table 2. Bootstrap experiment for equation (18). Estimation is by 2SLS. There are 400 bootstrap replications.

| | 3SLS | 2SLS Bootstrap | | | |
|---|---|---|---|---|---|
| | (1) Estimate | (2) Mean | (3) SD | (4) $t$ | (5) RMS Nominal SE |
| $\alpha_K$ | .0565 | .0572 | .00210 | -6.86 | .00213 |
| $\alpha_L$ | .254 | .253 | .00307 | 4.36 | .00347 |
| $\alpha_E$ | .0443 | .0445 | .00114 | -3.68 | .00121 |
| $\gamma_{KK}$ | .0264 | .0292 | .00792 | -7.02 | .00821 |
| $\gamma_{KL}$ | .000551 | -.000820 | .00523 | 5.24 | .00541 |
| $\gamma_{KE}$ | -.00994 | -.0130 | .00168 | 3.59 | .0161 |
| $\gamma_{LK}$ | .000551 | .00441 | .0138 | -5.59 | .0133 |
| $\gamma_{LL}$ | .00734 | .0752 | .00777 | -4.76 | .00879 |
| $\gamma_{LE}$ | -.00439 | .00697 | .0263 | -8.66 | .0261 |
| $\gamma_{EK}$ | -.00994 | -.00922 | .00452 | -3.19 | .00464 |
| $\gamma_{EL}$ | -.00439 | -.00481 | .00293 | 2.87 | .00307 |
| $\gamma_{EE}$ | .0200 | .0185 | .00897 | 3.26 | .00915 |

To indicate some of the difficulty in assessing standard errors, we repeat the bootstrap experiments for the three equations adjoined to (18). These equations explain the log price ratios in terms of the exogenous variables $X_t$. Only the intercept terms will be considered. Table 3 presents the results of an experiment using the 3SLS estimator. The nominal SEs in column 5 are more than 30 percent smaller than the actual variability observed in the simulation experiments, the SDs shown in column 4. This is substantially worse than the situation reported in Table 1. The reason may be that the nominal SEs for the adjoined equations depend on all of the 6 by 6 estimated covariance matrix $\hat{V}$; while for the original three-equation system, only the leading 3 by 3 submatrix of $\hat{V}$ comes in. With more covariances to estimate, the asymptotics may be less reliable.

Table 3. Bootstrap experiment for the equations adjoined to (18). Only the intercepts are reported. Estimation is by 3SLS. There are 400 bootstrap replications.

| | 3SLS | | Bootstrap | | |
|---|---|---|---|---|---|
| | (1) Estimate | (2) Nominal SE | (3) Mean | (4) SD | (5) RMS Nominal SE |
| $B_{1,4}$ | -3.59 | 1.22 | -3.74 | 1.36 | .875 |
| $B_{1,5}$ | -1.13 | .274 | -1.10 | .393 | .255 |
| $B_{1,6}$ | .316 | .563 | .370 | .878 | .551 |

With a discrete error distribution, the bootstrap distribution of coefficient estimators will usually have moments of all orders. With a continuous distribution like the normal, moments may not exist. Still, there is good evidence to show that the bootstrap distribution can be well approximated by a normal distribution, whose first and second moments can be estimated by the method indicated here.

All the computations described in this report were performed on a VAX 11/750 with the floating point accelerator. The 2SLS and 3SLS estimators were coded in FORTRAN using the LINPACK library. The computations described in this Section took about one hour of CPU time.

## 6. Forecasting

The object of this section is to indicate how the bootstrap can be used to develop standard errors of forecasts. The idea is that in the simulation world of the bootstrap, we can generate both the actual value and the predicted value in the forecast year, say 1995; hence we can observe the difference, which is the forecast error. This idea will be illustrated on equation (18), with a historical period of 1947–71. (The absence of dynamics makes the exercise a bit tame.) There will be two flavors: "straight forecasting" and "policy analysis." In "straight forecasting mode," the exogenous variables are forecast by some procedure external to the equation. The forecast values will be denoted by hats. The values of the exogenous variables, for the forecast year and the historical period, will be held fixed in all the simulations described below. In particular, the component of variance due to forecast error in the exogenous variables is not treated here. (In the simulations, the exogenous variables were forecast to grow from 1971 to 1995 at the compound annual growth rate observed from 1947 to 1971.)

The equation (18) is used to forecast the endogenous variable as follows:

- Estimate the coefficients using the historical data from 1947 to 1971.
- Set the disturbances $\epsilon_t$ to their expected value of zero for the forecast year 1995, and solve the equation:

$$\hat{Y}_{1995} = \hat{X}_{1995}\hat{B}(I - \hat{A})^{-1}$$

To get started on the bootstrap, estimate the parameters in the model (18) using the data from 1947 to 1971, obtaining the 3SLS estimates $\hat{A}$ and $\hat{B}$, and calculate the residuals $\hat{\epsilon}_t = Y_t - Y_t\hat{A} - X_t\hat{B}$. As usual, let $\mu$ be the empirical distribution of $\{\hat{\epsilon}_t : t = 1947, \ldots, 1971\}$. Make twenty-five independent draws $\epsilon_t^*$ for $t = 1947, \ldots, 1971$ and for $t = 1995$ from $\mu$. Construct a starred data set with the resampled residuals:

$$Y_t^* = (X_t\hat{B} + \epsilon_t^*)(I - \hat{A})^{-1}$$

For $t = 1947, \ldots, 1971$ , the $Y_t^*$ are simulated historical data. For $t = 1995$, the $Y_t^*$ are simulated future "actuals;" $X_{1995}$ should be replaced by the forecast $\hat{X}_{1995}$. Now make the forecasts by the standard procedure, but using the simulated historical data instead of the real data. In particular, the parameters $\hat{A}^*$ and $\hat{B}^*$ are re-estimated from the starred historical data. As usual,

$$\hat{Y}_{1995}^* = \hat{X}_{1995}\hat{B}^*(I - \hat{A}^*)^{-1}$$

Table 4. Bootstrap forecast experiment for equation (18) in "straight forecasting mode." Estimation is by 3SLS. There are 400 replications.

| | (1) Sample Mean Actuals $Y^*_{1995}$ | (2) Sample Mean Forecasts $\hat{Y}^*_{1995}$ | (3) Standard Deviation $Y^*_{1995} - \hat{Y}^*_{1995}$ | (4) The RMS G-N-O Standard Error |
|---|---|---|---|---|
| $S_K$ | .0207 | .00995 | .0647 | .0417 |
| $S_L$ | .301 | .301 | .0370 | .0257 |
| $S_E$ | .0117 | .0179 | .0344 | .0275 |
| $\log(p_K/p_M)$ | -2.38 | -2.56 | 1.65 | 1.08 |
| $\log(p_L/p_M)$ | .500 | .492 | .456 | .314 |
| $\log(p_E/p_M)$ | -2.71 | -2.74 | 1.02 | .678 |

The result is a set of simulated actuals $Y^*_t$ for $t = 1947, \ldots, 1971$ as well as $t = 1995$, and a simulated forecast $\hat{Y}^*_{1995}$ for the year 1995. Note that *e.g.* $\hat{A} \neq \hat{A}^*$, and $Y^*_t$ incorporates $\epsilon^*_t$ while $\hat{Y}^*_t$ does not, so $Y^*_t \neq \hat{Y}^*_t$. The difference between the future $Y^*_t$ and the forecast $\hat{Y}^*_t$ is the forecast error. This procedure can be repeated, getting new starred disturbances on each repetition, to develop the distributions of

- the simulated actual $Y^*_{1995}$
- the simulated forecast $\hat{Y}^*_{1995}$
- the simulated forecast error $Y^*_{1995} - \hat{Y}^*_{1995}$.

Table 4 summarizes 400 replications of the bootstrap forecasting experiment just described. Parameters were estimated by 3SLS. The table displays the sample mean of the 400 simulated actuals, the sample mean of the 400 simulated forecasts, and the standard deviation of the 400 simulated forecast errors. The standard deviations in column 3 are the bootstrap measures of random error in the forecasts. Take for example, the energy factor share $S_E$. The forecasts tend to be 50 percent too high: column 2 vs. column 1. In addition, the random error in the forecast is about three times the size of the quantity being forecast: column 3 vs. column 1. For the other quantities, bias does not seem to be a problem; but the random errors are large. The model does not do well in straight forecasting. (With our discrete error distribution, moments of all orders usually exist. With a continuous error distributions, the bootstrap distribution of forecast error may fail to have moments, but still be close to normally distributed, with the parameters indicated by the bootstrap.)

Column 4 of Table 4 shows what happens when the Goldberger-Nagar-Odeh (1961) method is used to approximate the random error of forecasts. This is a "delta-method" based on Taylor expansions. As can be seen by comparison with column 3, it substantially underestimates the error. Furthermore, it ignores bias completely.

We turn now from forecasting to "policy analysis," in effect using the model to answer questions about the reaction of the economy to policy initiatives. The Berndt-Wood model is really a set of demand equations. These may be solved against some (unspecified) set of supply equations to get prices and quantities. The Berndt-Wood paper is not very specific on this point, but the endogeneity of prices may be ascribed to this implicit simultaneity. In policy analysis mode, consider a policy which sets the prices of capital, energy, labor, and materials at pre-assigned values, without changing the distribution of the $\epsilon$'s. What happens to the factor shares? (Whether this is a sensible question or not is a very deep question, which we will ignore.) The equation (18) gives the factor shares corresponding to the assumed prices.

To predict the factor shares from (18), we set the parameters at the values estimated from the historical data, and drop the $\epsilon$'s. The likely size of the chance error in this prediction may be assessed by the bootstrap. Historical starred data and the starred parameter estimates are built up as before. But the starred actuals $Y^*_{1995}$ are derived from (18), the starred disturbance terms $\epsilon^*_{tK}$, $\epsilon^*_{tL}$, $\epsilon_{tM}$. being the first three components of $\epsilon^*_{1995}$ drawn from $\mu$. For this experiment, the prices in 1995 for capital, labor, and materials were taken to be the same as those in 1971, the price of energy was doubled from its 1971 value. The results are shown in Table 5 below. From this point of view, the model seems to be doing fine. To sum up, in policy analysis mode the forecasts seem quite good. The straight forecasts, however, are subject to serious bias and large chance errors.

Table 5. Bootstrap forecast experiment for equation (18) in "policy analysis mode." Estimation is by 3SLS. There are 400 replications.

| | Sample Mean Actuals $Y^*_{1995}$ | Sample Mean Forecasts $\hat{Y}^*_{1995}$ | Standard Deviation $Y^*_{1995} - \hat{Y}^*_{1995}$ |
|---|---|---|---|
| $S_K$ | .0706 | .0723 | .00866 |
| $S_L$ | .233 | .233 | .00699 |
| $S_E$ | .0494 | .0480 | .00557 |

# 7. Conclusions

This paper demonstrates the use of the bootstrap to attach standard errors to coefficient estimates and forecasts in a static simultaneous-equation energy demand model fitted by three-stage least squares. In contrast to previous findings, the asymptotic formulae for coefficient standard errors performed reasonably well. At present, there does not seem to be any good rule of thumb for deciding when such formulae will give acceptable results. Developing such a rule would be a good focus for future research. In the present example, the usual asymptotic methods for estimating standard errors of forecast do not seem satisfactory: They miss the appreciable bias, and substantially underestimate the random error component. Of course, the bootstrap procedure has troubles of its own: see Bickel and Freedman (1983) or Freedman and Peters (1983) or Peters (1983).

# References


BERNDT, E. L. , and WOOD, D. O. (1975), "Technology, prices, and the derived demand for energy," Review of Economics and Statistics, 57, 259–268.

BERNDT, E. L. , And WOOD, D. O. (1979), "Engineering and econometric interpretation of energy–capital complementarity," American Economic Review, 69, 342–354.

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.

DAGGETT, R. S. , and FREEDMAN, D. A. (1983), "Econometrics and the law: a case study in the proof of antitrust damages," Technical Report, Department of Statistics, University of California, Berkeley.

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.

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

FREEDMAN, D. A. , and PETERS, S. C. (1983), "Bootstrapping a regression equation: some empirical results," to appear in JASA.

GALLANT, A. R. (1975), "Seemingly unrelated nonliner regressions," Journal of Econometrics, 3, 35–50.

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.

INTRILLIGATOR, M. D. (1978), Econometric Models, Techniques and Applications, Englewood Cliffs, New Jersey: Prentice Hall.

KUH, E. , and WELSCH, R. E. (1980), "Econometric models and their assessment for policy: some new diagnostics applied to translog energy demand in manufacturing," in Validation and Assessment Issues of Energy Models, ed. Saul Gass, Washington, D.C.: National Bureau of Standards.

MADDALA, G. S. (1977), Econometrics, New York: McGraw Hill.

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

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