Journal of Business & Economic Statistics, October 1985, Vol. 3, No. 4

Estimation and Inference in Two-Step Econometric Models

Kevin M. Murphy
Robert H. Topel
Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637

A commonly used procedure in a wide class of empirical applications is to impute unobserved regressors, such as expectations, from an auxiliary econometric model. This two-step (T-S) procedure fails to account for the fact that imputed regressors are measured with sampling error, so hypothesis tests based on the estimated covariance matrix of the second-step estimator are biased, even in large samples. We present a simple yet general method of calculating asymptotically correct standard errors in T-S models. The procedure may be applied even when joint estimation methods, such as full information maximum likelihood, are inappropriate or computationally infeasible. We present two examples from recent empirical literature in which these corrections have a major impact on hypothesis testing.

1. INTRODUCTION

The estimation of models that contain unobservable, though estimable, variables is now common in several areas of applied econometrics. As examples, these unobservables have included (rational) expectations of future forcing variables in models of dynamic optimization (e.g., Crawford 1978, Sargent 1978, Topel 1982, Flavin 1982, Blanchard 1983), "unanticipated" components of aggregate money growth in macroeconometric models (Barro 1977, 1978, Mishkin 1983, Lilien 1982), or equilibrium unemployment probabilities in spatial models of labor market equilibrium (Abowd and Ashenfelter 1981). In these and other econometric studies, unobserved variables are either replaced by their predicted values from an auxiliary statistical model when estimating the relationship of interest, or they are estimated jointly with that model. This article provides simple but theoretically correct procedures for statistical inference in this class of econometric models when joint procedures are infeasible or inappropriate.

There are two main approaches to estimating this type of model. The first, which we will refer to as the two-step (T-S) estimator, will be our main interest in this article. The T-S procedure simply replaces the unobserved components with their estimated or predicted values from the auxiliary model. Critically, in most applications these values are then treated as if they are known for purposes of estimation and inference in the second-step model, which is usually the model of interest. It is well known that T-S procedures yield consistent estimates of second-stage parameters under fairly general conditions. It is also well known that the second-step estimated standard errors and related test statistics based on these procedures are incorrect. In some cases this feature of the T-S estimator is acknowledged (e.g., Topel 1982, Blanchard 1983), but more commonly the problem is ignored, and no correction of second-step test statistics is attempted. (Some authors who use the T-S procedure write as if the problem were merely one of efficiency. This conclusion is false, as our subsequent analysis will show.)

Following the suggestion of Liederman (1979), the usual alternative to the T-S procedure is to estimate the first- and second-step models via some joint method, such as full information maximum likelihood (FIML). Under appropriate assumptions, FIML procedures yield efficient estimators and asymptotically correct estimates of standard errors. Unfortunately, in many situations FIML estimation is both computationally complex and costly to implement, facts that account in part for its limited use. For example, in defense of implementing a T-S procedure, Blanchard (1983) argued that

> The reason for using [the T-S method] is simply a reason of cost: the cost of maximizing the [joint] likelihood function with respect to 20 coefficients rather than four in the second step of the two-step method is very high. The asymptotic standard errors and test statistics reported in this section are derived under the assumption that A is known, ignoring that $\hat{A}$ is substituted for A in the second step. (p. 307)

Although these cost constraints are a commonly stated explanation for using the T-S procedure, FIML methods may be unattractive for other reasons as well. For example, in some applications the researcher may be reluctant to hypothesize a specific joint distribution for the random components of the unobservables in the first- and second-step models. Alternatively, the estimated model for the unobservables may come from different data or may even be based on parameter estimates obtained by another researcher (e.g., Lilien 1982). In these cases, joint maximum likelihood procedures are infeasible.

This article presents a simple yet general method of calculating the correct asymptotic covariance matrix for the T-S estimation procedure. The basic idea is as follows. The T-S method fails to account for the fact that the unobservable regressors have been estimated in calculating second-step coefficients and standard errors. The imputed unobservables applied in the second step are therefore measured with sampling error. We assume that the auxiliary model for the unobservables produces consistent estimates of both first-step parameters and their asymptotic covariance matrix. Thus the sampling error of the unobservables vanishes in the limit, so second-step parameters are consistently estimated. Moreover, under fairly general conditions the estimated limiting distribution of this error may be used to consistently estimate the variances of the second-step parameter estimates.

The remainder of the article is organized as follows. Section 2 describes the class of models that we analyze and provides two examples, based on the research of Barro (1977) and Topel (1984), that apply the standard error corrections we derive. In these examples, the T-S correction has substantial effects on the second-step hypothesis tests: in both cases the corrected standard errors are uniformly larger and are often more than double their uncorrected levels. Section 3 begins the formal analysis, in which we derive the correct asymptotic distribution of the T-S estimator when the two models are statistically independent. The assumptions and calculations required to estimate the correct asymptotic covariance matrix for the second-step estimator are stated in Theorem 1. In this case, standard T-S procedures unambiguously underestimate standard errors of the consistent second-step estimates. Section 4 extends these results to the case in which the two models are estimated from the same sample and the independence assumption is not imposed. Results for this case are stated in Theorem 2. We show that commonly advocated instrumental-variables procedures for T-S models are a special case of this theorem. Section 5 provides two extensions to the basic results, based on the structure of models that arise frequently in applied research. Section 6 contains concluding remarks. Details of proofs are in the Appendix.

2. TWO ILLUSTRATIVE EXAMPLES

To motivate the subsequent analysis, we begin with two illustrative examples based on recent applied research. The first is drawn from Barro (1977) (see also Barro 1978; Lilien 1982), who estimated the impact of "unanticipated money growth" on unemployment in the United States. Unanticipated monetary disturbances are defined to be the residuals from a first-step linear regression of proportional growth in the M1 definition of money (DM_t) on two annual lags of this variable, current deviations of real Federal expenditures from their normal level, and lagged unemployment. These residuals, DMR_t, are then entered as regressors in a second-step linear regression fitted to data on aggregate unemployment. More precisely, using annual data for the years 1941–1973, the model is

$$DM_t = X_{1t}\theta + DMR_t \qquad (1)$$

$$\ln(UN_t/1 - UN_t) = X_{2t}\beta + \gamma(L)DMR_t + u_t, \qquad (2)$$

where X_{1t} contains the variables just mentioned, UN_t is the annual average unemployment rate, X_{2t} contains controls for the level of the minimum wage and a measure of military conscriptions, and $\gamma(L)$ is a second-order polynomial in the lag operator, L. The two-step strategy forms the estimated residuals from (1), $D\hat{M}R_t$, based on the estimated parameter vector $\hat{\theta}$, which are then used to replace the unobserved true values of DMR_t in estimating β and $\gamma(L)$ in the second-step equation, (2). Estimated standard errors from this procedure ignore the fact that $\hat{\theta}$ is estimated in the first stage.

For our purposes, there are two noteworthy features of the structure (1)–(2). First, by construction the residuals from the first- and second-step equations must be orthogonal, even though the models are estimated from contemporaneous data. Second, unlike other two-step procedures for linear models, such as two-stage least squares or instrumental variables, X_{2t} is not contained in X_{1t}. Under these conditions we may apply Theorem 1 in Section 3 to estimate asymptotically correct standard errors for this model.

Table 1 reports parameter estimates and corrected and uncorrected standard errors for this model. The parameter estimates themselves are identical to those reported by Barro (1977) and do not concern us here. Note that the variance corrections are quite substantial: corrected standard errors are larger than their uncorrected counterparts for all parameter estimates (as we show must occur under the independence assumption), in some cases by a factor of two or more. The average estimated t ratio (excluding the intercept) is overestimated by 60% in column 2 relative to column 3, and two of the five variables in the model are no longer significant by conventional standards. Thus in this example the correct two-step adjustments to standard errors that we advocate have an appreciable impact on statistical inference in the second-stage model.

As we noted, the usual alternative to a T-S procedure is FIML. For comparison, we also report FIML estimates for this model under the assumption of normally distributed errors. In general, the estimated coefficients

Table 1. Barro's Model of Money Growth and Unemployment Dependent Variable = $\log(u_t/1 - u_t)$

	Two-step procedure			FIML	
Variable	Estimate	Original standard error	Corrected standard error	Estimate	Standard error
Intercept	−3.078	.146	.371	−2.845	.201
DMR_t	−5.825	2.071	2.960	−4.840	1.942
DMR_{t-1}	−12.088	1.878	2.150	−11.774	1.794
DMR_{t-2}	−4.167	1.869	2.356	−3.193	2.028
Military enlistment	−4.637	.791	1.580	−5.751	.998
Minimum wage	.969	.456	.988	.298	.527

NOTE: Number of observations = 33; R^2 =. 78. For variable definitions, see Barro 1977.

and associated standard errors from FIML estimation are similar to those obtained from the T-S procedure. In this case the reduced-form restrictions imposed by the two-step estimator allow efficiency close to FIML. In addition, the incorrect but commonly reported T-S standard errors overestimate the precision of even the more efficient FIML estimates for several parameters. (In fact, the incorrect T-S standard errors are always asymptotically smaller than the Cramér–Rao lower bound in the independence case.)

Our second example illustrates the case in which, due to sample size and other considerations, FIML methods are computationally impractical, and in which the assumption of independent random components across equations is not imposed by the theory. Following Topel (1984), we estimate the compensating wage differential that workers demand in return for accepting jobs that entail risk of future unemployment due to layoffs. Our data consist of 76,393 observations on prime-aged male labor force participants drawn from the March Current Population Surveys for the years 1977–1980. In the first stage, we estimate by maximum likelihood the determinants of the probabilities of two unemployment events, permanent and temporary layoffs, based on reported employment status at the survey date. To focus on the key aspects of the model, we may think of these as a single event, unemployment. Define $D_i = 1$ if the ith individual is unemployed, $D_i = 0$ otherwise. Let $u(\theta, X_{1i})$ be the probability that person i with observable characteristics X_{1i} is unemployed. Then the log-likelihood contribution for individual i is

$$\mathscr{L}_i = D_i \log u(\theta, X_{i1}) + (1 - D_i)\log(1 - u(\theta, X_{i1})). \quad (3)$$

The wage equation that incorporates the unemployment probability is

$$\ln \omega_i = X_{i2}\beta + u(\theta, X_{i1})\gamma_1 + u(\theta, X_{i1})R_i\gamma_2 + \varepsilon_i, \quad (4)$$

where ω_i is an individual's average weekly wage, X_{i2} is a vector of characteristics that affect productivity, and R_i is the proportion of lost labor earnings that would be replaced by unemployment insurance. Theory indicates that $\gamma_1 > 0$ (jobs that offer higher unemployment must offer higher wages), whereas $\gamma_2 < 0$.

Given some assumption about the functional form of $u(\theta, X_{i1})$ and standard regularity conditions, maximization of the likelihood defined by (3) yields a consistent estimator $\hat{\theta}$ of θ and a consistent estimate $\hat{V}(\hat{\theta})$ of the asymptotic covariance matrix of $\hat{\theta}$. In the actual problem studied by Topel (1984), the dimensionality of $\hat{V}(\hat{\theta})$ is large (72 × 72). This adds to the computational burden for any joint procedure. In addition, theory suggests no specific form for the joint distribution of the random components of (3) and (4), as is required by FIML. Neither, however, does theory imply (as in the first example) that the random components are orthogonal.

With these conditions, the T-S estimator replaces θ with $\hat{\theta}$, based on (3), in (4). The correct asymptotic covariance matrix for models of this type is stated in Theorem 2 in Section 4. Table 2 reports point estimates and corrected and uncorrected estimated standard errors for the T-S version of (4). For purposes of comparison, we report estimated standard errors under the assumption that the random components of (3) and (4) are independent ($Q_2 = 0$) and for the general case ($Q_2 \neq 0$), where Q_2 is a matrix defined in Section 4. Though the coefficient estimates are in accord with theory, our interest here is in the sensitivity of hypothesis tests to the standard error corrections. Thus note that for the imputed regressors [$u(\hat{\theta}, X)$ and $u(\hat{\theta}, X)R$] in the first two rows of the table, the estimated standard errors are about triple their uncorrected levels, indicating that standard procedures would vastly exaggerate the precision of the second-step estimates for these key variables. The estimated normal statistics for these variables fall from more than 11.0 to about 4.0. Thus in this case also the inferences that may be drawn from the estimates are extremely sensitive to the two-step procedure.

Two other features of these estimates are noteworthy. First, accounting for covariance in the random components ($Q_2 \neq 0$) of the two models had very little impact on the standard error adjustment. As in our first

Table 2. Estimated Equalizing Wage Differences for Unemployment

Variable	Estimate	Original standard error	Corrected standard error $Q_2 = 0$	Corrected standard error $Q_2 \neq 0$
$u(\hat{\theta}, X_{i1})$	2.53	.226	.612	.618
$u(\hat{\theta}, X_{i1}) \cdot R_i$	−2.58	.224	.671	.685
Years of schooling[a]	.070	$.73 \times 10^{-3}$	$.143 \times 10^{-2}$	$.146 \times 10^{-2}$
Years of potential labor market experience[a]	.034	$.45 \times 10^{-3}$	$.54 \times 10^{-3}$	$.54 \times 10^{-3}$
Experience^2 [b]	$-.53 \times 10^{-3}$	$.98 \times 10^{-5}$	$.10 \times 10^{-4}$	$.10 \times 10^{-4}$
Nonwhite[a]	−.181	.0055	.0142	.01415
Married[b]	.187	.0038	.0039	.0039
Resides in				
Central city[b]	.100	.0037	.0038	.0038
Non-SMSA[b]	.044	.0057	.0058	.0059

[a] Variable included in first-step model.
[b] Variable excluded from first-step model.
NOTE: Number of observations = 76,393; R^2 = .3245. Dependent variable is log(base e) of average weekly wage; $\hat{u}$ is imputed from auxiliary model. First-step model estimated by maximum likelihood. Sample consists of prime-aged (20–65 years old) males who report labor force participation (employment or unemployment) for 40 or more weeks in the previous calendar year. Other regressors include dummies for sample year and census region. Variable definitions are as follows: years of schooling = highest number of years of completed schooling; labor market experience = age-schooling − 6; nonwhite = 1 if individual is not white; married = 1 if individual is married, spouse present; central city = 1 if person resides in a central city; non-SMSA = 1 if person resides outside of a standard metropolitan statistical area.

example, all estimated standard errors in the model are adjusted upward. This is not necessary when the errors are correlated, as we demonstrate in Section 4. Second, the proportional adjustment in estimated standard errors is largest for the imputed regressors, somewhat smaller for variables that appear in both the first- and second-stage models, and negligible for regressors in X_2 that are not also elements of X_1. As we will show, this pattern of proportional adjustments is not a general result. In particular, when both the first- and second-step models are linear and when X_1 contains X_2, all estimated standard errors are adjusted by the same factor.

These results illustrate the potential importance for inference of correct procedures in T-S models. We now turn to a more general framework, on which these estimates are based.

3. THE BASIC MODEL

The two models we have discussed are special cases of a more general procedure, which we now consider in detail. In order to focus on main issues, we initially assume that the second-step model of interest is linear in both the exogenous variables and the unobservables. For each observation we let

$$y = x_2\beta + f(\theta, x_1)\gamma + u, \tag{5}$$

where x_2 is a $1 \times k$ vector of exogenous variables assumed to influence y, x_1 is a vector of variables from the first step that determines the unobservables, $f(\cdot)$, which is a $1 \times m$ vector of functions (possibly nonlinear) determined by the unknown parameters θ. For example, components of $f(\cdot)$ may represent expectations of future forcing variables in a dynamic model or estimated characteristics of jobs in a spatial model of wage determination. An explicit illustration of $f(\cdot)$ is given by the likelihood equation (3). Thus, as an example, $f(\theta, X_1)$ may be an imputed probability based on a first-step probit or logit model or simply the predicted values from a linear regression. By assumption, the vector θ is estimated independently in the first step of the T-S strategy. Finally, β and γ are conformable vectors of parameters that are to be estimated in the second step.

We assume that the random component of y satisfies

$$E(u \mid x_1, x_2) = 0 \qquad E(u^2 \mid x_1, x_2) = \sigma^2$$

and that these errors are independent across observations. More complicated error structures may be admitted without materially affecting our results. For convenience, we regard the observations on (x_1, x_2) as nonstochastic.

The only assumptions that we impose on the auxiliary model for the unobservables are that $f(\cdot)$ be twice continuously differentiable in θ for each x_1 and that the first step yields an estimator $\hat{\theta}$ of θ such that

$$\sqrt{n}(\hat{\theta} - \theta) \overset{A}{\sim} n(0, V(\hat{\theta})). \tag{6}$$

That is, $\hat{\theta}$ is consistent and $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normal with covariance matrix $V(\hat{\theta})$. The investigator is assumed to possess a consistent estimate $\hat{V}(\hat{\theta})$ of the covariance matrix of $\hat{\theta}$ from the first step.

We assume for the moment that (5) is estimated by least squares and seek the asymptotic distribution of $(\hat{\beta}, \hat{\gamma})$ from a regression of y on x_2 and the estimated values, $f(\hat{\theta}, x_1)$, from the first step. Denote by X_2 the $n \times k$ matrix of observations on x_2 and let F be the $n \times m$ matrix of predictions $f(\hat{\theta}, x_1)$. Denote the observation matrix for model (5) by $Z = (X_2, F)$. We maintain the assumption that this observation matrix satisfies the usual least squares condition

$$\lim_{n\to\infty} (1/n)Z'Z = Q_0, \tag{7}$$

where Q_0 is positive definite and symmetric. With these assumptions, writing (1) in terms of observables yields

$$y = x_2\beta + f(\hat{\theta}, x_1)\gamma + [f(\theta, x_1) - f(\hat{\theta}, x_1)]\gamma + u, \tag{8}$$

where the bracketed term is induced by the sampling distribution of $\hat{\theta}$. The least squares estimator $(Z'Z)^{-1}Z'Y$ is therefore

$$\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix} = \begin{bmatrix}\beta\\ \gamma\end{bmatrix} + (Z'Z)^{-1}Z'[F(\theta, x_1) - F(\hat{\theta}, x_1)]\gamma + (Z'Z)^{-1}Z'U. \tag{9}$$

Provided that the functions $f(\cdot)$ are continuously differentiable in θ for each x_1 and that the sample second moments of these derivatives are bounded, then the least squares estimates of β and γ will be consistent. (The proof is in the Appendix.) Multiplication of (9) by $\sqrt{n}$ yields

$$\sqrt{n}\begin{bmatrix}\hat{\beta} - \beta\\ \hat{\gamma} - \gamma\end{bmatrix} = (n^{-1}Z'Z)^{-1}n^{-1/2}Z'[F(\theta, x_1) - F(\hat{\theta}, x_1)]\gamma + (n^{-1}Z'Z)n^{-1/2}Z'U. \tag{10}$$

Under the regularity conditions stated in the forthcoming Theorem 1, the terms in (10) involving $F(\cdot)$ are asymptotically equivalent to their first-order approximations. (We outline the proof, which is a slight variant on the usual results for the asymptotic distribution of differentiable functions of random variables, in the appendix.) More precisely, expanding $F(\cdot)$ in a Taylor series about $\hat{\theta}$, we have

$$\operatorname*{plim}_{n\to\infty}\{n^{-1/2}Z'[F(\theta, x_1) - F(\hat{\theta}, x_1)]\gamma - [(n^{-1}Z'F^*)\sqrt{n}(\theta - \hat{\theta})]\} = 0, \tag{11}$$

where the elements in the matrix F^* are of the form

$$f^*_{ij} = \sum_{k=1}^{m} \gamma_k \frac{\partial f_k}{\partial \theta_j}(\hat{\theta}, x_{1i}). \tag{12}$$

[The proof of (11) is rather long and technically difficult. The proof of (11) and similar asymptotic equivalence results are available from the authors on request.] We maintain the assumption that the derivatives of F satisfy the following conditions, analogous to (7):

$$\lim_{n\to\infty} (1/n)Z'F^* = Q_1 \tag{13}$$

and

$$\left|\frac{\partial^2 f(\hat{\theta}, x_1)}{\partial\theta_i\partial\theta_j}\right|^2 \le M_2(x_1) \quad \text{for} \quad \|\hat{\theta} - \theta\| < \delta,$$

where $\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} M_2(x_{1i}) < \infty$. Substitution of (11) into (10) then yields

$$\sqrt{n}\begin{bmatrix}\hat{\beta} - \beta\\ \hat{\gamma} - \gamma\end{bmatrix} \stackrel{A}{=} (n^{-1}Z'Z)^{-1}n^{-1}Z'F^*(\sqrt{n}\,(\theta - \hat{\theta})) + (n^{-1}Z'Z)^{-1}n^{-1/2}Z'U, \tag{14}$$

where $\stackrel{A}{=}$ denotes asymptotic equivalence in probability, so the indicated quantities have identical limiting distributions.

To evaluate the asymptotic distribution of (14), we first consider the case in which the random components of the first- and second-step models are independent. As we indicated, this situation may be relevant when the first and second steps are estimated from different data or time periods, or if the investigator uses outside estimates of θ and $V(\hat{\theta})$ from another study. (In the first example presented in Section 2, this condition was implied by the structure of the model.) An asymptotic equivalence argument and the central limit theorem then imply that $n^{-1/2}Z'U \stackrel{A}{=} N(0, \sigma^2 Q_0)$, where Q_0 is defined as in (7). Under these conditions, we obtain the following theorem on the asymptotic distribution of the T-S estimator.

Theorem 1 (T-S Estimators With Independent Random Components). In the preceding model, if (a) $Z = (X_2, F(\theta, x_1))$ satisfies (7), (b) $f(\theta, x_1)$ is twice continuously differentiable in θ for each x_1 with the sample second moments of $\partial f/\partial\theta$ uniformly bounded in the sense of (13), and (c) $\hat{\theta}$ is a consistent estimator of θ, then the least squares estimator of (β', γ') is consistent. If, in addition, (d) the second derivatives of $f(\cdot)$ satisfy (13) and (e) the sequence of estimates $\hat{\theta}$ is independent of U with an asymptotic distribution satisfying (6), then the least squares estimator

$$\sqrt{n}[(Z'Z)^{-1}Z'Y - (\beta', \gamma')']$$

is asymptotically normal with mean vector 0 and covariance matrix

$$\Sigma = \sigma^2 Q_0^{-1} + Q_0^{-1}Q_1V(\hat{\theta})Q_1'Q_0^{-1}, \tag{15}$$

where Q_0 and Q_1 are described by (7) and (13). In addition, the quantities in (15) are consistently estimated by the following sample quantities:

$$\hat{\sigma}^2 = n^{-1}(Y - X_2\hat{\beta} - F(\hat{\theta}, x_1)\hat{\gamma})' \times (Y - X_2\hat{\beta} - F(\hat{\theta}, x_1)\hat{\gamma})$$

$$\hat{Q}_0 = n^{-1}Z'Z$$

$$Z = [X_2, F(\hat{\theta}, x_1)]$$

$$\hat{Q}_1 = n^{-1}Z'F^*$$

$$f^*_{ij} = \sum_{k=1}^{n} \hat{\gamma}_k \frac{\partial f_k}{\partial \theta_j}(\hat{\theta}, x_{1i}) \tag{16}$$

Proof. See Appendix.

Theorem 1 implies that the correct covariance matrix for the second-step estimators in the T-S procedure exceeds the commonly reported asymptotic covariance matrix, $\sigma^2 Q_0^{-1}$, by a positive-definite matrix. As a result, standard errors from a naive two-step procedure are understated. The latter matrix consists of the estimated covariance matrix from the first step pre- and post-multiplied by the estimated regression coefficients of the derivatives F^* on the explanatory variables X_2 and $f(\hat{\theta}, X_1)$. Thus, in practice, the estimated asymptotic covariance matrix is, using (16),

$$\hat{\Sigma} = \hat{\sigma}^2(Z'Z)^{-1} + (Z'Z)^{-1}Z'F^*\hat{V}(\hat{\theta})F^{*\prime}Z(Z'Z)^{-1}. \quad (15')$$

Equation (15′) shows that the information on the sampling distribution of $\hat{\theta}$, which is estimated in the first step, may be used to adjust the estimated covariances in the second-step equation. Thus the sampling distribution of $(\hat{\beta}'\hat{\gamma}')'$ depends crucially on the precision with which θ is estimated in the auxiliary model. Furthermore, (15′) illustrates that the increase in the estimated standard errors depends crucially on the correlation between the derivatives F^* and the explanatory variables Z and the precision of the first-step estimates. To the extent that the correlation is high or the first-step sampling error is significant, this adjustment will be important.

A special, though common, case of (15′) arises when the auxiliary model is linear with error ε. With only one imputed regressor, we have $F^* = \gamma_1 x_1$, $\hat{V}(\theta) = \hat{\sigma}^2(X_1'X_1)^{-1}$, so the second term in (15′) involves the projection of Z onto the space spanned by X_1. If all of the regressors in X_2 are also contained in X_1, then $X_1(X_1'X_1)^{-1}X_1'Z = Z$, so

$$\hat{\Sigma} = \hat{\sigma}^2(1 + (\hat{\gamma}^2\hat{\sigma}_\varepsilon^2/\hat{\sigma}^2))(Z'Z)^{-1}. \quad (17)$$

In this case, the correction inflates all standard errors in the second-step model by the same factor of proportionality. (With more than one imputed regressor, the adjustment factor becomes $1 + \hat{\sigma}^{-2}\sum_{i=1}^{m}\gamma_i^2\sigma_i^2$, assuming that the auxiliary equations are also independent.) Clearly, in this case the amount by which a naive two-step procedure exaggerates the precision of the second-step estimator depends on *both* the relative error variance in the two models and the magnitude of the coefficient on the estimated regressor. When these quantities are nonnegligible, the proportional bias in estimated standard errors will be important.

Most important, note that (15′) involves only easily calculated sample statistics in conjunction with standard output from most computer packages. Thus even in situations in which joint estimation of the first- and second-step models is infeasible for reasons of cost, application of (15′) yields asymptotically correct estimates of the standard errors and covariances for the commonly used T-S estimator under the independence assumption. Equation (15′) was used to derive the corrected standard errors for Barro's (1977) model of unemployment and money growth, previously illustrated (see Table 1).

4. THE GENERAL CASE: NONINDEPENDENT RANDOM COMPONENTS

When the second-stage model and the model for the unobservables are estimated from the same or contemporaneous data, the assumption of stochastically independent random components is less attractive. For example, in estimating models that involve unobserved expectations, a common T-S procedure is to estimate the model for the unobservables over the same time period as in (5). Contemporaneous covariance in the random components of the models may be quite important in this case. In the following discussion, we will assume that the first-stage model has been estimated by maximum likelihood, since this situation is common in the applied literature.

Dropping the independence assumption, we proceed as before through (14). Applying well-known results for maximum likelihood estimation, we have the asymptotic equivalence:

$$\sqrt{n}\,(\theta - \hat{\theta}) \stackrel{A}{=} -R_1(\theta)^{-1}\sum_{i=1}^{n} n^{-1/2}\hat{l}(x_{1i};\theta), \quad (18)$$

where $\hat{l}(\cdot)$ is the column vector of first derivatives of the log-likelihood with respect to θ and

$$R_1(\theta) = -E\,n^{-1}[\partial^2 l(X_1;\theta)/\partial\theta\partial\theta'] \quad (19)$$

is Fisher's information matrix. Substitution of (18) into (14) yields the asymptotic equivalence

$$\sqrt{n}\begin{bmatrix}\hat{\beta}-\beta\\ \hat{\gamma}-\gamma\end{bmatrix} \stackrel{A}{=} -(n^{-1}Z'Z)^{-1}n^{-1}Z'F^*R_1 \times\left(\sum_{i=1}^{n} n^{-1/2}\dot{l}(x_{1i};\theta)\right) + (n^{-1}Z'Z)^{-1}\left(\sum_{i=1}^{n} n^{-1/2}z_i'u_i\right). \quad (20)$$

To evaluate the asymptotic distribution of (20), we must calculate the joint asymptotic distribution of the two sums on the right side.

In addition, one can establish that $n^{-1/2}z'u$ is asymptotically equivalent to its value when $f(\theta, x_1)$ is evaluated at the true θ. (The proofs of this and other technical results are available from the authors on request.) We therefore make no distinction between these quantities in what follows. The two sums in (20) are simply averages of the vectors $z_i'u_i$ and $\dot{l}(x_{1i};\theta)$, so the central limit theorem may be applied. Since both vectors have 0 mean, the sums are asymptotically normally distributed with a covariance matrix equal to the expectation of their cross products. By the homoscedasticity assumption and the usual maximum likelihood results, we obtain

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \sum_{i=1}^{n} z_i' u_i \\ \\ \sum_{i=1}^{n} \dot{l}(\theta;\, x_{1i}) \end{pmatrix} \overset{A}{\sim} N(0,\, \Omega)$$

$$\Omega = \begin{pmatrix} \sigma^2 Q_0 & E[z_i' u_i \dot{l}'(\theta,\, X_{1i})] \\ E[\dot{l}(\theta;\, X_{1i}) u_i z_i] & R_1(\theta) \end{pmatrix} \quad (21)$$

Were the first-step model estimated by regression, the off-diagonal expectation in (21) would have an explicit representation in terms of the error covariance of the two equations. We denote this expectation by Q_2. By the law of large numbers, this quantity can be consistently estimated by using

$$\hat{Q}_2 = n^{-1} \sum_{i=1}^{n} Z_i' \hat{u}_i \dot{l}'(\hat{\theta};\, x_{1i}). \quad (22)$$

Here the $\hat{u}_i$ are the estimated residuals from the second-step regression. Using this fact, the asymptotic distribution of the second-step estimators is given by

$$\sqrt{n}\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} \overset{A}{=} N(0,\, \Sigma), \quad (23)$$

where

$$\Sigma = \sigma^2 Q_0^{-1} + Q_0^{-1}[Q_1 R_1^{-1}(\theta) Q_1' - Q_1 R_1^{-1}(\theta) Q_2' - Q_2 R_1^{-1}(\theta) Q_1'] Q_0^{-1}. \quad (24)$$

In practice, the quantities in (24) are consistently estimated by the sample quantities in expressions (16) and (22) and by the estimated asymptotic covariance matrix from the first step. We summarize these results as

Theorem 2 (Asymptotic Distribution of the T-S Estimator). Under the conditions of Theorem 1, but allowing the first-stage parameters to be estimated from the same sample via maximum likelihood, the least squares estimates of the second-step parameters are consistent and asymptotically normally distributed with the asymptotic covariance matrix given by (23). These covariances are consistently estimated by applying the sample moments given in (16) and (22).

Proof. See Appendix.

Two points about Theorem 2 are in order. First, note that the result may be easily generalized, via conformable derivations, to situations in which either model or both models are nonlinear. Our assumption that only the first-step model is estimated by maximum likelihood is merely an attempt to conform to popular models in the literature. A more general situation will be analyzed shortly. Second, note that the conceptual problem addressed in deriving the asymptotic covariance matrix of the T-S estimator is similar to two-stage least squares (TSLS): second-stage standard errors of structural parameter estimates must account for the fact that reduced-form parameters are estimated in the first stage. In fact, some tedious calculations show that (21) is formally equivalent to the asymptotic covariance matrix of TSLS in the special case in which both models are linear and when $x_2 \subset x_1$. As in (17), in this special case all standard errors are adjusted by the same factor of proportionality.

A commonly advocated alternative to the T-S estimator is instrumental variables (IV) estimation. The two-step procedure and the standard error corrections presented in this article contain IV as a special case. In addition, the T-S estimator is appropriate in many cases in which IV estimation is not or in which IV is simply infeasible. The T-S estimator has the added advantage of considering the model's structure in estimating the first stage or reduced form.

To illustrate these points, assume that a second-step dynamic model contains, in addition to variables dated period t, an unobserved expectation based on information dated $t - 1$ or earlier. Then, after substituting the observable actual values for the unobserved expectations, the current variables are no longer valid instruments for the second-stage equation. They will in general be correlated with the prediction error. If a sufficient number of instruments remain, this will lead to less efficient estimation of the second-step parameters, since the values of current variables would be replaced by predicted values based on the last period's information when estimating the second stage. If the loss of current variables reduces the number of available instruments below that needed for identification, then the IV approach would be infeasible. The T-S procedure avoids these problems by considering the model's structure when creating the predicted values for second-stage estimation. Hence, the two-step procedure uses current variables to predict the current variables in the second step while imposing the constraint on the estimated reduced form that these current variables do not affect the unobserved expectation. In this context, the basic advantage of the T-S estimator over IV estimation is that it exploits the model's structure and associated constraints in estimating the reduced form.

5. EXTENSIONS

5.1 Two-Step Maximum Likelihood

Our previous analysis may be extended to the situation in which both the auxiliary and second-step models are to be estimated by maximum likelihood. The derivations are similar to those in Section 4, so we merely outline the results here.

We assume that the marginal distributions of the two random vectors y_1 and y_2 can be written as $F_1(y_1;\, \theta_1)$ and $F_2(y_2;\, \theta_1,\, \theta_2)$, where the dependence of these distributions on the vectors x_1 and x_2 has been suppressed to save on notation. The parameter vectors θ_1 and θ_2 are to be estimated. Denoting the natural logarithms of $F_1(\cdot)$ and $F_2(\cdot)$ by $L_1(\cdot)$ and $L_2(\cdot)$, the two-step maximum likelihood estimator satisfies

$$\sum_{i=1}^{n} \frac{\partial L_1(y_{1i};\, \hat{\theta}_1)}{\partial \theta_1} = 0 \qquad (25)$$

$$\sum_{i=1}^{n} \frac{\partial L_2(y_{1i};\, \hat{\theta}_1,\, \hat{\theta}_2)}{\partial \theta_2} = 0. \qquad (26)$$

Under standard regularity conditions (see Dhrymes 1974), the first-step maximum likelihood $\hat{\theta}_1$ is consistent. Denoting the true parameter values by θ_1^* and θ_2^*, maximization of $n^{-1}\Sigma L_2(y_{2i},\, \hat{\theta}_1,\, \theta_2)$ with respect to θ_2 is asymptotically equivalent to maximization of $n^{-1}\Sigma L_2(y_{2i},\, \theta_1^*,\, \theta_2)$. Thus the second-step estimator of θ_2 is also consistent.

Expanding the sum (25) and (26) about $\theta^* = (\theta_1^*, \theta_2^*)$ yields

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial L_1(y_{1i}, \theta_1^*)}{\partial \theta_1}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\frac{\partial^2 L_1(\theta_1^* + \lambda(\hat{\theta} - \theta_1^*))}{\partial \theta \partial \theta'} d\lambda \right) \sqrt{n}(\hat{\theta}_1 - \theta_1^*), \qquad (27)$$

$$-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial L_2(y_{2i}, \theta_1^*, \theta_2^*)}{\partial \theta_2}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} \int_0^1 \frac{\partial^2 L_2(\theta^* + \lambda(\hat{\theta} - \theta^*))}{\partial \theta_2 \partial \theta_1'} d\lambda \right] \sqrt{n}(\hat{\theta}_1 - \theta_1^*)$$

$$+ \frac{1}{n}\left[\sum_{i=1}^{n} \int_0^1 \frac{\partial^2 L_2(\theta^0 + \lambda(\hat{\theta} - \theta^*))}{\partial \theta_2 \partial \theta_2'} d\lambda \right] \sqrt{n}(\hat{\theta}_2 - \theta_2^*). \qquad (28)$$

To proceed, we must calculate the asymptotic joint distribution of these two random vectors. Define

$$R_1(\theta_1) = E \frac{\partial L_1}{\partial \theta_1} \left(\frac{\partial L_1}{\partial \theta_1} \right)' = -E \frac{\partial_2 L_1}{\partial \theta_1 \partial \theta_1'}$$

$$R_2(\theta_2) = E \frac{\partial L_2}{\partial \theta_2} \left(\frac{\partial L_2}{\partial \theta_2} \right)' = -E \frac{\partial L_2^2}{\partial \theta_2 \partial \theta_2'}$$

$$R_3(\theta) = E \frac{\partial L_2}{\partial \theta_1} \left(\frac{\partial L_2}{\partial \theta_2} \right)' = -E \frac{\partial^2 L_2}{\partial \theta_1 \partial \theta_2'}$$

$$R_4(\theta) = E \frac{\partial L_1}{\partial \theta_1} \left(\frac{\partial L_2}{\partial \theta_2} \right)'. \qquad (29)$$

Then by the central limit theorem, the joint distribution of the first partials of the two-step likelihood is

$$\left[\begin{matrix} \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} \dfrac{\partial L_1(y_{1i},\, \theta_1^*)}{\partial \theta_1} \\ \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} \dfrac{\partial L_2(y_{2i},\, \theta_1^*,\, \theta_2^*)}{\partial \theta_2} \end{matrix} \quad \overset{A}{\sim} N(0, \Omega) \right], \qquad (30)$$

where

$$\Omega = \begin{bmatrix} R_1(\theta_1) & R_4(\theta) \\ R_4(\theta)' & R_2(\theta_2) \end{bmatrix}. \qquad (31)$$

Equation (27) implies the asymptotic equivalence that was used previously:

$$\sqrt{n}\,(\hat{\theta}_1 - \theta_1^*) \overset{A}{=} -R_1(\theta_1)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial L_1(y_{1i}, \theta_1^*)}{\partial \theta_1}. \qquad (32)$$

Substitution of (32) into (28) and application of the law of large numbers then yields

$$\sqrt{n}(\hat{\theta}_2 - \theta_2^*) \overset{A}{=} -R_2(\theta_2)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial L_2(y_{2i};\, \theta_1^*,\, \theta_2^*)}{\partial \theta_2}$$

$$+ R_2(\theta_2)^{-1} R_3(\theta)' R_1^{-1}(\theta_1) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial L_1(y_{1i};\, \theta_1^*)}{\partial \theta_1}. \qquad (33)$$

Equation (33) is the two-step maximum likelihood analog of (14) for the linear model. Since the joint distribution of the two sums on the right side is given by (30), we obtain in this case $\sqrt{n}\,(\hat{\theta}_2 - \theta_2) \overset{A}{\sim} N(0, \Sigma)$,

$$\Sigma = R_2^{-1} + R_2^{-1}[R_3' R_1^{-1} R_3 - R_4' R_1^{-1} R_3 - R_3' R_1^{-1} R_4] R_2^{-1}, \qquad (34)$$

where the matrices $R_1(\theta_1) \cdots R_4(\theta)$ are defined in (29). Again, if the models are estimated from different samples, we may impose the restriction $R_4(\theta) = 0$, so the second and third terms in parentheses vanish. Again, commonly reported asymptotic standard errors are underestimated for this case.

5.2 Different First-Step Models for Subsamples of the Data

As a final case, we consider a model in which separate first-step models apply to groups of observations, but the second-step model is estimated on the pooled sample. For example, in the wage model illustrated in Section 2 separate auxiliary models for the determinants of unemployment might be estimated within each of a large number of industries, so θ would be indexed by industry, whereas the parameters of the wage equation are common to all individuals. Similarly, in a time series study the possibility of structural change may lead to separate equations for subperiods of the data. This type of model is common in applied research (e.g., Smith and Welch 1978; Card in press) and a simple extension of our previous results applies.

We assume that a separate first-step model has been estimated on each of S subgroups of the data indexed by s ($s = 1, 2, \ldots, S$), yielding an estimate $\hat{\theta}_s$ of the first-step parameters. The parameter vector θ_s need not have identical dimensions, nor must the functional form of the first-step models be the same across groups. Assuming that the random components of the S auxiliary models are independent, we proceed as before through (14), except that we define

$$F^* = \begin{bmatrix} F_1^* & & & & 0 \\ & F_2^* & & & \\ & & \ddots & & \\ & & & & \\ 0 & & & & F_s^* \end{bmatrix}, \quad (\theta - \hat{\theta}) = \begin{bmatrix} \theta_1 - \hat{\theta}_1 \\ \vdots \\ \theta_S - \hat{\theta}_S \end{bmatrix},$$

$$R(\theta) = \begin{bmatrix} R_1(\theta_1) & & & \\ & \ddots & & \\ 0 & & \ddots & 0 \\ & & & \\ & & & R_S(\theta_S) \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_S \end{bmatrix}, \tag{35}$$

where F_s^*, $R_s(\theta)$, and Z_s are as defined in Section 3 for each subgroup of the data. Letting $k_s = n_s/n$ be the sample share of group s, Theorem 2 applies with

$$Q_1 R^{-1}(\theta) Q_1' = \sum_{s=1}^{s} k_s Q_{1s} R_s^{-1}(\theta_s) Q_{1s}' \tag{36}$$

and

$$Q_1 R^{-1}(\theta) Q_2' = \sum_{s=1}^{s} k_s Q_{1s} R_s^{-1}(\theta_s) Q_{2s}'. \tag{37}$$

Here, the matrices Q_{js} conform to the limiting values of Section 3, but they are calculated within each subgroup as $n_s \to \infty$. Thus in this case the adjustment to the second-step dispersion matrix of the parameter estimates depends on a sample-share weighted average of the dispersion matrices for the individual auxiliary models. Thus even in this case the asymptotically correct procedure for estimating the precision of the second-step estimator need not be computationally burdensome, unless the number of subgroups in the data becomes quite large.

6. CONCLUSION

This article has developed asymptotically correct procedures for inference and hypothesis testing in a broad class of two-step econometric models. Other two-step procedures, such as IV or TSLS, are special cases that may be inappropriate in many econometric applications. A basic finding is that the correct covariance matrix for the two-step procedure is easily computed from output that is available from most standard regression packages. In the examples that we have studied, these corrections have an appreciable impact on statistical inference. We do not regard these cases as unusual. Thus our results suggest that standard two-step results common in the literature should be interpreted with extreme caution and that the relatively small cost of computing the correct covariance matrix of the estimator seems warranted in most econometric applications.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Department of Labor, Office of the Assistant Secretary for Policy. Points of view or opinions expressed in this article do not necessarily represent the official position of the U.S. Department of Labor. We are indebted to John Abowd, Lars Hansen, Lee Lillard, and Arnold Zellner for helpful discussions. Remaining errors are our responsibility.

APPENDIX

Proof of Theorem 1. Theorem 1 can be proved by using (10) and showing how the characteristic functions converge. (Note: It is easiest to regard X_1 and X_2 as fixed constants in this case. We adopt this in what follows.) Let

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} = V_n;$$

then by using (10) we obtain

$$\begin{aligned} &E \exp(i\lambda' V_n) \\ &= \int_{\hat{\theta}_n} \int_u \exp[i\lambda' Q_0^{-1} Q_1 \sqrt{n}(\hat{\theta} - \theta) + i\lambda' Q_0^{-1}(1/\sqrt{n}) Z' u] \\ &\quad \cdot dF_u(u)\, dF_{\hat{\theta}N}(\hat{\theta}) \\ &= \int_{\hat{\theta}_n} \exp[i\lambda' Q_0^{-1} Q_1 \sqrt{n}(\hat{\theta} - \theta)]\, dF_{\hat{\theta}}(\hat{\theta}_n) \\ &\quad \cdot \int_u \exp[i\lambda' Q_0^{-1}(1/\sqrt{n}) Z' u]\, dF_u(u); \end{aligned} \tag{A.1}$$

but

$$\int_{\hat{\theta}_n} \exp[i\lambda' Q_0^{-1} Q_1 \sqrt{n}(\hat{\theta} - \theta)]\, dF_{\hat{\theta}_n} \to \exp[-\lambda Q_0^{-1} Q_1 V(\hat{\theta}) Q_1' Q_0^{-1} \lambda]. \tag{A.2}$$

(A.2) follows from the assumption that $\sqrt{n}(\hat{\theta} - \theta) \overset{A}{\sim} n(0, V(\theta))$. Likewise, since u is iid, the Lindberg–Feller central limit theorem (Feller 1971) implies that

$$\int_u \exp[i\lambda' Q^{-1}(1/\sqrt{n}) Z' u]\, dF(u) \to \exp(-\lambda' Q_0^{-1} \lambda \sigma^2). \tag{A.3}$$

Hence, we obtain equation (A.4) that yields the asymptotic distribution desired:

$$\begin{aligned} &E \exp(i\lambda' V_n) \\ &\quad \to \exp(-\lambda'[Q_0^{-1} Q_1 V(\hat{\theta}) Q_1' Q_0^{-1} + \sigma^2 Q_0^{-1}]\lambda) \\ &\quad \Rightarrow \sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} \overset{A}{\sim} n(0, Q_0^{-1} Q_1 V(\hat{\theta}) Q_1' Q_0^{-1} + \sigma^2 Q_0^{-1}). \end{aligned} \tag{A.4}$$

Proof of Theorem 2. Under the standard assumptions on the first-stage maximum likelihood model, we have

$$\sqrt{n}(\hat{\theta} - \theta) \overset{A}{=} R_1(\theta)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(\theta_j x_{1i}). \quad \text{(A.5)}$$

Next $\lim_{n\to\infty}((1/n)Z'F^*) = Q_1$ by assumption; by using this and (A.5), we obtain (A.6), which corresponds to (20) in the text.

$$\sqrt{n}\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} \overset{A}{=} Q_0^{-1} Q_1 R_1(\theta)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}(\theta_j x_{1i}) + Q_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' u_i. \quad \text{(A.6)}$$

By the central limit theorem, we have

$$\frac{1}{\sqrt{n}} \sum_{L=1}^{n} \begin{bmatrix} \dot{l}(\theta, x_{1i}) \\ z_i' u_i \end{bmatrix} \overset{A}{\sim} n\begin{bmatrix} 0, & \begin{matrix} R_1(\theta) & Q_2 \\ Q_2' & \sigma^2 Q_0 \end{matrix} \end{bmatrix},$$

$$Q_2 = E(z_i' u_i \dot{l}'(\theta, x_{1i})) \quad \text{(A.7)}$$

From (A.7) and (A.6) we obtain the desired result:

$$\sqrt{n}\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} \overset{A}{\sim} n(0, V)$$

$$V = Q_0^{-1}[Q_1 R_1(\theta)^{-1} Q_1 - Q_2 R(\theta)^{-1} Q_1' - Q_1 R_1(\theta)^{-1} Q_2' + \sigma^2 Q_0] Q_0^{-1}. \quad \text{(A.8)}$$

The asymptotic equivalence used earlier and similar derivations imply that we can estimate this covariance matrix by evaluating at $\hat{\theta}$, $\hat{\gamma}$, and $\hat{\sigma}^2$. Formal proofs of these equivalences are available from the authors on request.

[*Received April 1984. Revised February 1985.*]

REFERENCES

Abowd, J., and Ashenfelter, O. (1981), "Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Differentials," in *Studies in Labor Markets*, ed. S. Rosen, Chicago: University of Chicago Press for the National Bureau of Economic Research, p. 141.

Barro, R. J. (1977), "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review*, 67, 101–115.

——— (1978), "Unanticipated Money, Output, and the Price Level in the United States," *Journal of Political Economy*, 86, 549–580.

Blanchard, L. J. (1983), "The Production and Inventory Behavior of the American Automobile Industry," *Journal of Political Economy*, 91, 365–400.

Card, David (in press), "Indexation in Long Term Labor Contracts: A Theoretical and Empirical Analysis," *Journal of Political Economy.*

Crawford, R. G. (1978), "Expectations and Labor Market Adjustments," *Journal of Econometrics*, 11, 207–232.

Dhrymes, P. J. (1974), *Econometrics, Statistical Foundations and Applications*, New York: Springer-Verlag.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications* (Vol. 2, 2nd ed.), New York: John Wiley.

Flavin, M. A. (1982), "The Adjustment of Consumption to Changing Expectations About Future Income," *Journal of Political Economy*, 89, 974–1010.

Leiderman, L. (1980), "Macroeconometric Testing of Rational Expectations and Structural Neutrality Hypotheses for the United States," *Journal of Monetary Economics*, 69–82.

Lilien, David (1982), "Sectoral Shifts and Cyclical Unemployment," *Journal of Political Economy*, 90, 777–793.

Mishkin, Frederic (1983), *The Rational Expectations Approach to Macroeconomics*, Chicago: University of Chicago Press for the National Bureau of Economic Research.

Sargent, T. J. (1978), "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," *Journal of Political Economy*, 86, 1009–1044.

Smith, J., and Welch, F. (1978), "Local Labor Markets and Cyclic Components in the Demand for College Trained Manpower," *Annales del INSEE*, 30/31 (April–Sept.), 559–630.

Topel, R. (1982), "Inventories, Layoffs, and the Short Run Demand for Labor," *American Economic Review*, 72, 769–787.

——— (1984), "Equilibrium Earnings, Turnover, and Unemployment: New Evidence," *Journal of Labor Economics*, 4, 500–522.