Journal of Econometrics 33 (1986) 367-385. North-Holland

FULL- VERSUS LIMITED-INFORMATION ESTIMATION OF A RATIONAL-EXPECTATIONS MODEL
Some Numerical Comparisons

Kenneth D. **WEST***

Princeton University, Princeton, NJ 08544, USA

Received March 1985, final version received May 1986

This paper compares numerically the asymptotic distributions of parameter estimates and test statistics associated with two estimation techniques: (a) a limited-information one, which uses instrumental variables to estimate a single equation [Hansen and Singleton (1982)], and (b) a full-information one, which uses a procedure asymptotically equivalent to maximum likelihood to simultaneously estimate multiple equations [Hansen and Sargent (1980)]. The paper compares the two with respect to both (1) asymptotic efficiency under the null hypothesis of no misspecification, and (2) asymptotic bias and power in the presence of certain local alternatives. It is found that (1) full-information standard errors are only moderately smaller than limited-information standard errors, and (2) when the model is misspecified, full-information tests tend to be more powerful, and its parameter estimates tend to be more biased. This suggests that at least in the model considered here, the gains from the use of the less robust and computationally more complex full-information technique are not particularly large.

1. Introduction

A variety of techniques have been proposed to estimate rational-expectations models, with the development of new techniques still continuing.' The number of empirical rational-expectations studies applying one or more of these techniques also is growing, as is evident from the pages of practically any major journal. There does not, however, appear to be any intermediate work comparing numerically the distributions of different estimators and test statistics under various circumstances. It is therefore not clear how much better in practice any 'best' estimation technique is likely to be, with respect to either efficiency of parameter estimates or power of tests of overidentifying restrictions.

*I thank Whitney Newey, an anonymous referee and various seminar audiences for helpful comments; John Londregan for research assistance; and the Princeton University Committee for Research in the Humanities and Social Sciences and the National Science Foundation for partial financial support. This paper was revised while I was a National Fellow at the Hoover Institution.

'A partial list of recent contributions includes Hansen and Sargent (1982), Hansen and Singleton (1982), Hayashi and Sims (1983), Chow (1983) and Fair and Taylor (1983).

This paper is a preliminary step in filling this intermediate gap. For a representative rational-expectations model, it compares numerically some asymptotic properties of what are probably the two most commonly used methods of estimation of rational-expectations models. One method works directly with the first-order conditions, or Euler equations, of the models, and may be applied to both linear and non-linear models [Hansen and Singleton (1982)]. The other method works with the solution to the first-order conditions, and may be used only when those conditions are linear in the model's variables [Chow (1975), Hansen and Sargent (1980)].

The major feature distinguishing the two methods qualitatively appears to be well known [Hansen and Sargent (1982), Wallis (1980)]. It is the same feature that distinguishes single- and multiple-equation estimators in the traditional linear simultaneous-equations model: the method that works with the first-order conditions is a limited-information (LI) technique, that which works with the solution to those conditions a full-information (FI) technique. The FI technique is valid only under certain auxiliary assumptions not required for the LI technique to be valid. When these assumptions hold, however, FI parameter estimates are more efficient.

While this basic LI-FI distinction appears to be well known, less well understood is just how much more efficient is the FI method. And this is a matter of considerable interest to any researcher considering estimation of a linear model. The FI method not only requires additional assumptions but, in contrast to three-stage versus two-stage least squares in the traditional linear simultaneous-equations model, also is substantially more difficult to apply. Even for models linear in their variables, it is highly non-linear in the parameters and in practice has tended to entail numerical problems [Sargent (1978, pp. 477, 483–484), Blanchard (1983, pp. 386–387)]. The LI method, by contrast, is often linear in the parameters for models linear in their variables [West (1984, 1985a, 1986)]. Unless the payoff in terms of efficiency is substantial, the standard practice of using FI techniques to estimate linear models[2] may not be warranred.

This paper studies that payoff, for a representative linear rational-expectations model, borrowed from Hansen and Sargent (1980). It compares some asymptotic properties of the FI and LI estimates. It first derives exact closed-form expressions for the asymptotic distributions of parameter estimates and of certain test statistics. For LI estimation it assumes that a procedure asymptotically equivalent to the Hansen and Singleton (1982) instrumental-variables technique is used, for LI tests a test asymptotically equivalent to the Hansen (1982) test of residual-instrument orthogonality. For

[2] The earlier practice of using LI techniques [e.g., McCallum (1976)] appears to be less common in recent work. Some recent examples of LI estimation of linear rational-expectations models occurs in my own work [West (1984, 1985a, 1986)].

FI estimation it assumes a procedure asymptotically equivalent to maximum likelihood assuming normality is used, for FI tests a test asymptotically equivalent to a likelihood-ratio test of the model's cross-equation restrictions.

The paper writes these asymptotic distributions as functions of the 'deep' structural parameters representing technology and stochastic characteristics of forcing variables. It then calculates the numerical values of these expressions, for a range of plausible values of the structural parameters. Specifically, the distributions it considers are those needed to compare full and limited estimation with respect to: (1) the asymptotic efficiency of parameter estimates under the null hypothesis of no misspecification, and (2) the asymptotic bias and power against two local alternatives: (a) a certain adjustment cost has been omitted from the model, and (b) a certain cost of deviating from a target level has been omitted from the model.[3]

It is found that (1) FI standard errors are only moderately smaller than LI standard errors, infrequently more than 20 percent smaller, and (2) FI tests are slightly more powerful against the adjustment-cost alternative, notably more powerful against the target-level-cost alternative. But FI parameter estimates are usually somewhat more biased as well, often considerably more so.

It is of course not clear whether these results will prove robust in this model to choice of other parameters or alternatives, let alone to consideration of other linear models, of non-linear models or of small-sample properties. That the results reported here have some general applicability is perhaps indicated by the similarity of this paper's model to a number of models actually estimated [e.g., Sargent (1978), Kennan (1979), Rotemberg (1983), Blanchard (1983), West (1986)]. In any case, it appears that at least for the model and parameters used here, the asymptotic gains from full-information estimation are not enormous.

The plan of the paper is as follows. Section 2 briefly sets out the model and derives the first-order condition and full-information solution. Section 3 first outlines the LI and FI estimation and test procedures and then describes the two local alternatives described above. Section 4 presents some numerical values of the formulas derived in section 3 for various values of the structural parameters. Section 5 has conclusions. An appendix outlines the derivation of the necessary formulas, with full details in the appendix to the working paper version of this paper [West (1985b)].

2. The model

The starting point for the analysis that follows is the dynamic labor-demand model of Hansen and Sargent (1980). A firm chooses its labor force n_t to

[3] The working paper version of this paper [West (1985b)] considered a third alternative as well, that expectations were static. This was removed from the present paper to save space.

minimize the expected present discounted value of costs:

$$\lim_{s\to\infty} \mathrm{E}_0 \sum_{t=0}^{s} b^t \left[0.5a_0 n_t^2 + 0.5a_1(\Delta n_t)^2 + n_t w_t + n_t u_t\right] \tag{1}$$

where

E_t = mathematical expectations, conditional on information known at time t,

b = fixed real discount rate, $0 < b < 1$,

a_0, a_1 = positive parameters,

n_t = labor demanded,

w_t = real wage,

u_t = white-noise technology shock.

Deterministic terms in (1) are suppressed for notational simplicity.

The structural parameters of economic interest are b, a_0 and a, [see Sargent (1979) for an economic discussion of the model]; when efficiency of parameter estimates is analyzed below, the parameter vector of interest will be $\alpha = (b, a_0, a_1)$.

A first-order condition, or Euler equation, for this model may be found by differentiating the objective function (1) with respect to the choice variable n_t:

$$\mathrm{E}_t\left[-ba_1 n_{t+1} + (a_1 + ba_1 + a_0)n_t - a_1 n_{t-1} + w_t + u_t\right] = 0. \tag{2}$$

This may be written in estimable form:

$$\begin{aligned} n_{t+1} &= (1 + b^{-1} + b^{-1}a_1^{-1}a_0)n_t - b^{-1}n_{t-1} + b^{-1}a_1^{-1}w_t + b^{-1}a_1^{-1}u_t \\ &\quad + (n_{t+1} - \mathrm{E}_t n_{t+1}) \\ &\equiv \gamma_0 n_t + \gamma_1 n_{t-1} + \gamma_2 w_t + v_{t+1} \\ &\equiv X_{0t}'\gamma + v_{t+1}. \end{aligned} \tag{3}$$

$X_{0t} = (n_t, n_{t-1}, w_t)'$, $\gamma = (\gamma_0, \gamma_1, \gamma_2)'$, and v_{t+1} is an MA (1) error: The expectational error $n_{t+1} - \mathrm{E}_t n_{t+1}$ clearly depends in part on u_{t+1}. Thus v_{t+1} depends on both u_t and u_{t+1}, and of course on the innovations in other forcing variables reflected in $n_{t+1} - \mathrm{E}_t n_{t+1}$, and so is MA (1).

The limited-information instrumental-variables technique described in section 3 estimates $\hat{\gamma}$ and then recovers $\hat{\alpha} = (\hat{b}, \hat{a}_0, \hat{a}_1)$ from $\hat{\gamma}$. It tests the model by checking whether the instruments are orthogonal to the estimated residuals. Details are presented in the next section.

With two additional assumptions, one can use standard dynamic programming techniques [Chow (1975)], or the algebraic techniques of Hansen and

Sargent (1980), to solve (1). The two assumptions are: (1) the choice variable n_t does not Granger-cause the forcing variable w_t relative to the firm's information set, and (2) a transversality condition rules out explosive behavior in n_t. The Hansen-Sargent (1980) technique for solving (1) is then as follows [the dynamic programming technique produces the identical solution, see Chow (1980)]. Write the polynomial in n_{t+1} as

$$1-(1+b^{-1}+b^{-1}a_1^{-1}a_0)L+b^{-1}L^2\equiv(1-\lambda L)(1-b^{-1}\lambda^{-1}L). \tag{4}$$

It may be shown that $0<\lambda<1$, so $1<b^{-1}\lambda^{-1}$ [Hansen and Sargent (1980, p. 96)]. Solve the stable root A backwards, the unstable root $b^{-1}\lambda^{-1}$ forwards. This gives the dynamic programming solution

$$n_t=\lambda n_{t-1}-\lambda a_1^{-1}\sum_{j=0}^{\infty}(b\lambda)^j\mathrm{E}_t w_{t+j}+e_{1t}, \tag{5}$$

where e_{1t} is white noise, $e_{1t}=-\lambda a_1^{-1}u_t$. Eq. (5) may be estimated when a third additional assumption is made, concerning the process used by agents to forecast w,. Here we follow the bulk of the literature [e.g., Sargent (1978)] and assume that w, is forecast from a univariate AR(p), $p\geq 2$,

$$\Phi(L)w_t=(1-\phi_1L-\phi_2L^2-\cdots-\phi_pL^p)w_t=e_{2t} \tag{6}$$

$$\Rightarrow w_t=\sum_{i=1}^{p}\phi_i w_{t-i}+e_{2t}\equiv X_{2t}'\phi+e_{2t}.$$

In this case, (5) becomes

$$\begin{aligned} n_t&=\lambda n_{t-1}+\sum_{i=0}^{p-1}\delta_{i+1}w_{t-i}+e_{1t}\\ &\equiv\delta_0 n_{t-1}+\sum_{i=0}^{p-1}\delta_{i+1}w_{t-i}+e_{1t}\\ &\equiv X_{1t}'\delta+e_{1t}, \end{aligned} \tag{7}$$

with $X_{1t}=(n_{t-1},w_t,\ldots,w_{t-p+1})'$, $\delta=(\lambda,\delta_1,\ldots.\delta_p)'=(\delta_0,\delta_1,\ldots,\delta_p)'$, and [Hansen and Sargent (1980, p. 99)]

$$\begin{aligned} &\delta_j=-\lambda a_1^{-1}\Phi(b\lambda)^{-1}\tilde{\delta}_j, \qquad j\geq 1,\\ &\tilde{\delta}_1\equiv 1,\\ &\tilde{\delta}_j\equiv\sum_{k-j}^{p}(b\lambda)^{k-j+1}\phi_k, \qquad j\geq 2. \end{aligned} \tag{8}$$

A full-information technique for retrieving the model's structural parameters estimates (7) jointly with (6), subject to the restrictions written out in (8); it then recovers $\hat{\alpha}$ from δ and ϕ. To test the model, it checks whether the restrictions written out in (8) hold. In practice this usually requires estimates of δ and ϕ not subject to the restrictions (8), although in principle one could use a Lagrange multiplier test on the restricted estimates.

3. Asymptotic distributions

3.1. *Efficiency of parameter estimates*

A variety of techniques have been suggested to estimate eqs. (3), (6) and (7) (see footnote 1). But in practice, two estimation techniques are most commonly used. The most common LI estimation technique is an optimal instrumental-variables estimator [used in, e.g., Hansen and Singleton (1982)]. The most common FI estimation technique is maximum likelihood assuming normal disturbances [used in, e.g., Sargent (1978)]. Our asymptotic results therefore assume that techniques asymptotically equivalent to these are used. See Hansen and Sargent (1982) for a theoretical comparison of the asymptotic properties of these two techniques, including an explanation of why the FI procedure is more efficient.

An outline of the two techniques now follows. Algebraic and econometric details are in the appendix to the working paper version of this paper [West (1985b)].

The $(p + 1) \times 1$ vector of instruments Z_t that may be used in the LI technique is easily determined by inspection of (6) and (7): one lag of labor demand n_{t-1} and p lags of the real wage w_t. Stack these into a $T \times (p+1)$ vector Z, where T is the sample size. Define X_0 as a similarly stacked matrix of X_{0t}, n as a column vector of n_{t+1}, and $\hat{S}$ as T times an estimate of $\mathrm{E}\sum_{j=-1}^{1} Z_t(\mathrm{E}v_t v_{t-j})Z'_{t-j}$. The two-step LI estimator suggested by Hansen and Singleton first gets $\hat{S}$, obtaining $\mathrm{E}v_t v_{t-j}$ from the moments of two-stage least-squares residuals and $\hat{S}$ from the obvious moments of instrument–residual cross-products.[4] In a second step, this estimator sets $\hat{\gamma} = (X_0'Z\hat{S}^{-1}Z'X_0)^{-1}X_0'Z\hat{S}^{-1}X_0'n$. [An alternative, asymptotically equivalent technique is described by Hayashi and Sims (1983).] The standard formula for the asymptotic covariance matrix of $\hat{\gamma}$, and thus of $\hat{\alpha}$, may be calculated numerically, given own- and cross-moments of n_t, w_t and disturbances. Given the parameters b, a_n, and a_w, and the moments of v_t (equivalently, e_{1t}), the required n_t and w_t moments may be recovered from (6) and (7), in a fashion

[4] For simplicity we assume that the researcher makes no correction for conditional heteroscedasticity.

analogous to that used to obtain moments from the Yule-Walker equations [Anderson (1971, pp. 181–182)].[5]

The LI test suggested by Hansen and Singleton (1982) sees whether the instruments are orthogonal to the disturbances:

$$\left(T^{-1/2}\sum_{t=1}^{T} Z_t\hat{v}_{t+1}\right)'(T^{-1}\hat{S})^{-1}\left(T^{-1/2}\sum_{t=1}^{T} Z_t\hat{v}_{t+1}\right), \tag{9}$$

where $\hat{v}_{t+1}$ are residuals [actual n_{t+1} minus n_{t+1} fitted from an estimate of eq. (3)]. This statistic converges to a central $\chi^2(p-2)$ random variable under the null, to a non-central $\chi^2(p-2)$ random variable under the local alternatives described below.

The FI technique, again, was assumed to be asymptotically equivalent to maximum likelihood. Maximum-likelihood estimation entails minimizing the log of the determinant of the two by two variance-covariance matrix of the residuals to (6) and (7), subject to the cross-equation constraint written out in (8) [Sargent (1978)]. (An alternative, asymptotically equivalent technique, is non-linear three-stage least squares.[6]) Again, the standard formula for the asymptotic covariance matrix of the parameter vector and thus of $\hat{\alpha}$ may be calculated, given values of b, a_1, a_2, and of own- and cross-moments of the disturbances e_{1t} and e_{2t}. An FI test of the model is obtained by estimating an unrestricted version of (6) and (7) and performing a likelihood ratio test. Wald and Lagrange multiplier tests of course are asymptotically equivalent.

3.2. *Asymptotic bias and power of tests*

We consider asymptotic bias and power under two local alternatives: (1) presence of second-order costs of adjustment, and (2) the presence of a cost of having labor demand deviate from an exogenous target level.

It will be convenient to formulate each alternative as adding an extra component to the residuals of the regression equations (3) and (7), although each such component will depend at least in part on observable random variables. Each component goes to zero in the usual way as the sample size T goes to infinity. That is, under each alternative, the residual in the Euler equation is $v_{t+1}+T^{-1/2}c_{1t}$, and the residual to the full-information equation (7) is $e_{1t}+T^{-1/2}c_{2t}$, for suitable defined c_{1t} and c_{2t}.

[5]Note that under the conditions used to derive (5) from (3), if w_t is covariance-stationary [roots of $\phi(L)$ within the unit circle], then n_t alone and n_t and w_t jointly are covariance-stationary [see Hansen and Sargent (1980)].

[6]Non-linear three-stage least squares is asymptotically equivalent because the variance–covariance matrix of (6) and (7) is unrestricted and the model is linear in the variables [Amemiya (1977, p. 365)].

The asymptotic bias of the LI estimator is calculated as the mean of the asymptotic distribution of the estimator. The asymptotic bias of the non-linear FI estimator is calculated as in Kiefer and Skoog (1984) as the derivative of this bias with respect to c_{2t}. See the appendix for details.

The two alternatives to be considered are:

(I) Costs of adjustment of order two. This alternative has been chosen because there is often little theoretical guidance as to the number and precise form of adjustment costs to be included in a cost-of-adjustment model [Sargent (1978, p. 479), Taylor (1980, p. 30)]. Misspecification of adjustment costs therefore seems possible in empirical studies using models like (1).

Consider, then, the alternative that a per period cost of the form $0.5T^{-1/2}a_2(\Delta^2 n_t)^2$ is present in the objective function, $\Delta^2 n_t \equiv (n_1 - n_{t-1}) - (n_{t-1} - n_{t-2})$. Under this alternative, the Euler equation for a sample of size T is

$$n_{t+1} = X\&y + v_{t+1} + T^{-1/2}a_2 b^{-1}a_1^{-1}\mathrm{E}_t\left(b^2\Delta^2 n_{t+2} - 2b\Delta^2 n_{t+1} + \Delta^2 n_t\right). \tag{10}$$

The corresponding FI solution is'

$$\begin{aligned} n_t = {} & (\lambda_{1T} + \lambda_{2T})n_{t-1} - \lambda_{1T}\lambda_{2T}n_{t-2} \\ & - \lambda_{1T}\lambda_{2T}(b\lambda_{1T} - b\lambda_{2T})^{-1}\left(T^{-1/2}a_2\right)^{-1} \\ & \times \left(b\lambda_{1T}\sum_{j=0}^{\infty}(b\lambda_{1T})^j \mathrm{E}_t w_{t+j} - b\lambda_{2T}\sum_{j=0}^{\infty}(b\lambda_{2T})^j \mathrm{E}_t w_{t+j}\right) + e_{1t}. \end{aligned} \tag{11}$$

'Strictly speaking, the error terms in (10) and (11) should be subscripted by the sample size T. By using the algebra outlined in the appendix, it is easy to show that the error term in the FI solution is

$$e_{1tT} = -\lambda_{1T}\lambda_{2T}\left(T^{-1/2}a_2\right)^{-1}u_t.$$

The error in the LI solution (10) may be constructed from the above equation since $v_{t+1} = u_t + (n_{t+1} - \mathrm{E}_t n_{t+1})$. This dependence of the error terms on sample size is ignored in the text for the following reason. Estimates of the moments of the residuals in both (10) and (11) converge in probability to the moments of the error terms under the null. That is, as the sample size $T \to \infty$, $T^{-1}\sum e_{1tT}^2 \to \mathrm{E}e_{1t}^2$. Similar relations hold for the residual in the Euler equation (10), for not only the $\mathrm{E}v_t^2$ but for $\mathrm{E}v_t v_{t+1}$ as well. It therefore follows that the ***asymptotic*** distribution of the estimators under the alternative is the same whether or not we take care to note the dependence of the residuals on the sample size T. We have chosen to ignore this dependence, since it is notationally simpler.

λ_{1T} and λ_{2T} are the roots to a certain fourth-order polynomial whose coefficients are functions of b, a_1, a, and $T^{-1/2}a_2$. As $T \to \infty$, $\lambda_{2T} \to 0$ and $\lambda_{1T} \to \lambda$. See the appendix. The regression equation for a sample of size T is thus

$$n_t = X_{1t}'\delta + e_{1t} + T^{-1/2}\tilde{\psi}_t, \tag{12}$$

where

$$\begin{aligned} T^{-1/2}\tilde{\psi}_t = {} & (\lambda_{1T} + \lambda_{2T} - \lambda) n_{t-1} \\ & - \lambda_{1T}\lambda_{2T} n_{t-2} + \lambda a_1^{-1} \sum_{j=0}^{\infty} (b\lambda)^j \mathrm{E}_t w_{t+j} \\ & - \lambda_{1T}\lambda_{2T}(b\lambda_{1T} - b\lambda_{2T})^{-1}(T^{-1/2}a_2)^{-1} \\ & \times \left(\lambda_{1T} \sum_{j=0}^{\infty} (b\lambda_{1T})^j \mathrm{E}_t w_{t+j} - b\lambda_{2T} \sum_{j=0}^{\infty} (b\lambda_{2T})^j \mathrm{E}_t w_{t+j} \right). \end{aligned}$$

(2) *Cost of deviating from a target level.* This alternative has been chosen for two reasons. The first is that there is often little theoretical guidance as to whether a cost of deviating from a target level should be included in a model with costs of adjustment. [Compare, for example, the labor-demand models in Kennan (1979) and Sargent (1978), or compare the two inventory models in West (1986).] Omission of a target-level cost that in fact is present therefore seems possible in empirical studies using models like (1). The second reason for considering this target-level alternative is that such a cost rationalizes serial correlation in the disturbances to eqs. (3) and (7), at least in the plausible case where the target level itself is serially correlated. Such serial correlation is sometimes found in cost-of-adjustment models [e.g., Sargent (1978)] but is not always tested for. Misspecification of the stochastic properties of the disturbances to (3) and (7) therefore seems possible in empirical studies using models like (1).

Consider, then, the alternative that a per period cost of the form $0.5T^{-1/2}a_3(n_t - s_t)^2$ has been omitted from the model. a_3 is a strictly positive parameter. The firm is thus penalized in proportion to how far its labor demand is from a target level s_t. We assume that s_t is covariance-stationary, and expand the firm's information set to include lags of s_t. We also assume that s_t as well as w, are univariate AR(p) with respect to this information set.[a] Thus, (6) continues to describe the w, process and (17) describes the s_t

[a]The assumption that the univariate autoregressive representations of s_t and w_t are of the same order is not restrictive. Our notation allows suitable ρ_i or ϕ_i to be zero if the order of the two processes is different.

process,

$$\rho(L)s_t = (1 - \rho_1 L - \cdots - \rho_p L^p)s_t = e_{3t}. \tag{13}$$

$\mathrm{E}e_{3t}e_{2t}$ or $\mathrm{E}e_{3t}e_{1t}$ are non-zero. (Otherwise the presence of the target level cost would not lead to inconsistent parameter estimates.)

The Euler equation for a sample of size T is then

$$n_{t+1} = X_{0t}'\gamma + v_{t+1} + T^{-1/2}a_3 b^{-1}a_1^{-1}(n_t - s_t). \tag{14}$$

The corresponding full-information solution is[9]

$$n_t = \lambda_T n_{t-1} - \lambda_T a_1^{-1}\sum_{j=0}^{\infty}(b\lambda_T)^j \mathrm{E}_t w_{t+j}$$

$$+ T^{-1/2}a_3\lambda_T a_1^{-1}\sum_{j=0}^{\infty}(b\lambda_T)^j \mathrm{E}_t s_{t+j} + e_{1t}. \tag{15}$$

λ_T is the smaller of two roots to a certain quadratic equation. As $T \to \infty$, $\lambda_T \to \lambda$. See the appendix. The regression equation for a sample of size T is thus

$$n_t = X_{1t}'\delta + e_{1t} + T^{-1/2}\tilde{\kappa}_t,$$

where

$$T^{-1/2}\tilde{\kappa}_t = (\lambda_T - \lambda)n_{t-1} + \lambda a_1^{-1}\sum_{j=0}^{\infty}(b\lambda)^j \mathrm{E}_t w_{t+j}$$

$$-\lambda_T a_1^{-1}\sum_{j=0}^{\infty}(b\lambda_T)^j \mathrm{E}_t w_{t+j}$$

$$+ T^{-1/2}a_3\lambda_T a_1^{-1}\sum_{j=0}^{\infty}(b\lambda_T)^j \mathrm{E}_t s_{t+j}. \tag{16}$$

Against both this and the adjustment-cost alternative, both the LI and FI tests have power, and both techniques estimate all parameters inconsistently. For any given set of structural parameters, either technique may be less biased. The following section determines which technique dominates in these and other respects, for a variety of parameter values.

[9]Again, the error term in (15), strictly speaking, should be subscripted by the sample size T. See footnote 7.

Table 1

Parameter values.

b	0.99
a_0	0.1; 5; 1; 2; 10; 50
a_1	1; 2; 10; 100; 250; 500
$\sigma_1^2, \sigma_{12}, \sigma_2^2$	1, 0.5, 1
	10, 2, 1
	10, −3, 1
	1, 2, 10
	1, −3, 10
$\phi_1, \phi_2, \phi_3, \phi_4$	1.64, -0.842, 0.152, -0.0084
	1.90, −1.39, 0.299, -0.0178
	1.80, -1.26, 0.258, -0.0445
	0.9554, 0.0033, 0.0754, -0.1849

4. Numerical results

4.1. *Efficiency of* parameter estimates

The formulas derived in the appendix were used to calculate the asymptotic LI and FI variances of b, a_0 and a_1 for a variety of different values for the structural parameters b, a_0, a_1, $\mathrm{E}e_{1t}^2 = \sigma_1^2, \mathrm{E}e_{1t}e_{2t} = \sigma_{12}, \mathrm{E}e_{2t}^2 = \sigma_2^2$ and $\phi_1, \ldots, \phi_p$. These values are listed in table 1: one value of b, six of a_0, six of a_1, five of $\sigma_1^2, \sigma_{12}, \sigma_2^2$, and four of $\phi_1, \ldots, \phi_p$. The total number of covariance matrices calculated for each estimation technique was thus 720 = 6 × 6 × 5 × 4. The logic underlying the choices of structural parameter values is as follows.

The value of b was chosen to accord with economic theory. The annual discount rate corresponding to $b = 0.99$ is about 4 percent, assuming data are quarterly. Only one value of b was used because an earlier version of this paper tried various values of b and found that results were virtually identical for the various values.

The values of a_0 and a_1 were chosen to accord with results of some empirical linear rational-expectations studies of labor and of other variables. These suggested a wide range of possible values for the ratio of costs of adjustment a_0 to quadratic costs a_1. Sargent (1978, pp. 483, 485-487) found 100: 1 and 1000: 1, Blanchard's (1983, pp. 386, 387-388) ratios varied from 100: 1 to 1 : 4, West's (1986, tables III and IV) from 100: 1 to 1 : 30. The values chosen thus span the range of values empirically found, and mimic the tendency for the cost of adjustment a_0 to be larger than the quadratic cost a_1.

The order of the AR was set to four, which accords with some previous studies [Sargent (1978), Blanchard (1983)]. The first three polynomials have roots chosen so that the forcing variable displays hump-shaped responses to shocks. The first of the three has four real roots, and factors as $(1 - 1.34L +$

Table 2

Ratios of FI to LI standard errors

Parameter	Min	Ratio Percentile 10	25	50	75	90	Max
h	0.26	0.59	0.73	0.88	0.96	0.99	1.00
a_0	0.08	0.46	0.64	0.83	0.93	0.97	1.00
a_1	0.10	0.43	0.62	0.83	0.97	0.99	1.00

$0.42L^2)(1-0.2L)(1-0.1L)$ [Blanchard's (1981) autoregression for aggregate GNP is the source for the second-order polynomial]. The second of the four polynomials has two real and two complex roots, and factors as $(1-[0.8+0.5i]L)(1-[0.8-0.5iL])(1-0.2L)(1-0.X)$. The third has four complex roots, and factors as $(1-[0.8+0.5i]L)(1-[0.8-0.5i]L)(I-[0.1+0.2i]L)(1-[0.1-0.2i]L)$. The fourth and final polynomial was taken directly from Sargent's empirical study of aggregate U.S. labor demand (1978, p. 481).

The values of σ_1^2, σ_{12} and σ_2^2 were assigned somewhat arbitrarily. They allow either e,, or e_{2t} to have larger variance, and allow a positive or negative covariance between the two shocks.

Table 2 has a summary comparison of the asymptotic LI and FI standard errors in estimation of b, a, and a,. As may be seen, the improvement in efficiency from using FI is moderate. Standard errors (s.e.s.) tend to drop by less than 20 percent when estimated by FI rather than LI. The improvement can be dramatic – in at least one case, the standard error dropped by over 90 percent. But about three-fourths of the FI s.e.s. were at best 40 percent smaller than the corresponding LI s.e.s. A quarter of them were at best 7 percent smaller.

To get a feeling for the implications of such a reduction in standard errors, consider the following. A 20 percent reduction in the standard error of a given estimate corresponds to a 25 percent increase in its z-statistic (ratio of estimate to standard error, distributed as a standard normal random variable). This will turn a z-statistic significant at the 90 percent level into one significant at the 96 percent level. It will turn a z-statistic significant at the 95 percent level into one significant at the 98 percent level. This improvement does not seem enormous.

There was surprisingly little variation in the ratios of LI/FI s.e.s. across parameter values. The distributions of these ratios were calculated holding each parameter value in turn constant. That is, distributions were studied for 21 = 6 + 6 + 5 + 4 overlapping subsets of the parameters. In general, these distributions were very similar to the aggregate distribution reported in table

2. For example, the median ratio of s.e.s. was in general within 0.10 of the corresponding overall figures reported in table 2. The only exceptions were for the extreme values of a,. FI performed relatively well when a, = 0.1, relatively poorly when a, = 50. When a, = 0.1, the median values of the ratios of FI to LI s.e.s. were 0.77 for b, 0.60 for a, and 0.61 for a_1. The corresponding figures for a, = 50 were 0.97, 0.96 and 0.92.

4.2. *Asymptotic bias and power of tests*

Calculations of bias and power were performed for all 720 sets of structural parameters using only one value of the parameters basic to each alternative. We set $a_2 = 1$ in the adjustment-cost misspecification alternative; $a_3 = 1$, $\rho_1 = 0.95$, $\rho_2 = \rho_3 = \rho_4 = 0$, $\mathrm{E}e_{1t}e_{3t} = \mathrm{E}e_{2t}e_{3t} = 0.1$, $\mathrm{E}e_{3t}^2 = 1$ in the target-level misspecification alternative. After calculating the LI and FI bias in the three parameters b, a_0 and a, for a given one of the 720 sets, we computed the ratios of the absolute value of the LI to FI bias in each of the three. It is a summary of these ratios that we present."

The power of the LI test (9) was compared to the power of the FI likelihood-ratio test for a given alternative and a given one of the 720 sets by first computing the ratio of the two relevant non-centrality parameters. This ratio was then used to calculate the probability that the LI test would reject the null given that (1) the researcher has set the size of the test to 0.05 (i.e., has decided to reject the null only if the probability of seeing the statistic under the null is less than 0.05), and (2) the probability that the FI test would reject the null was (a) 50 percent or (b) 95 percent.

The two LI probabilities of rejecting the null were calculated as follows. Since the test statistic has two degrees of freedom, assumption (1) in the preceding paragraph implies that the investigator rejects the null only if the test statistic is above 5.99. With a non-centrality parameter of 4.95, the probability is 0.50 that a non-central chi-square variable with two degrees of freedom is greater than 5.99; with a non-centrality parameter of 15.45 the probability is 0.95. Suppose, for example, that the ratio of the LI to FI non-centrality parameter is 0.3. The probability that the LI test would reject under the previous paragraph's assumption 2(a) is then 0.17, since 0.17 is the

"The LI bias in γ and the FI bias in δ and ϕ are linear in a_2 and a_3. But since a is a non-linear function of γ (LI) and δ and ϕ (FI), the bias in a is a non-linear function of a_2 and a_3. Ideally one might therefore want to try a range of a_2 and a, in comparing the relative LI and FI bias. This was not done for two reasons. The first is that the values chosen, $a_2 = 1$ and $a_3 = 1$, suffice to yield inference about the effects of small misspecitications, which would seem to be the misspecifications of greatest interest. They yield such an inference because with the sample size T used in practice being around 100, $T^{-1/2}$ is about 0.1. So in sample sizes typically observed, each omitted cost parameter has a value of about 0.1 and so is small relative to most of the table 1 values of included costs. The second reason a range of a_2 and a_3 were not tried is that such experimentation would be quite burdensome computationally.

Table 3
Adjustment-cost misspecification alternative: Ratio of absolute LI bias to absolute FI bias.

Parameter	Min	Ratio Percentile 10	25	50	75	90	Max
b	0.005	0.13	0.41	1.02	2.41	6.53	867
a_0	0.003	0.07	0.25	0.89	2.31	7.49	666
a_1	0.002	0.18	0.61	1.10	2.20	6.86	967

Median probability of LI rejecting null conditional on a given FI probability of rejection

	Given FI probability 0.50	0.95
LI probability	0.35	0.83

probability that a &i-squared variable with two degrees of freedom and a non-centrality parameter of 1.49 (= 4.95 x 0.3) will be greater than 5.99. Similarly, the probability that the LI test would reject under assumption 2(b) is 0.47. We calculated such probabilities of LI rejection for each alternative for all 720 sets of parameters, and report the median of the resulting values.

Information on comparative LI and FI performance in the presence of adjustment-cost misspecitication is reported in table 3. The LI and FI biases are of roughly comparable magnitude for all three parameters. The median bias ratio was about one for all three parameters, see the column labelled '50 in the upper half of table 3. There was, however, a considerable spread in each of the three ratios of biases. For each parameter, each technique was in certain cases at least a couple of orders of magnitude less biased than the other technique.

The distributions of the bias ratios were calculated for the twenty-one overlapping subsets of the parameters defined above. The results were similar to those reported in the upper half of table 3. There was in each subset quite a spread of ratios of absoiute bias, for each of the three parameters. The median ratio of biases was usually about one. The relative LI bias in all three parameters tended to be least when a, was big. Otherwise no pattern was evident.

The median probabilities of LI rejection under this alternative are given in the bottom half of table 3. The median probability is 0.35 when the probability of FI rejection is 0.50. The median probability is 0.83 when the FI probability is 0.95. The LI test procedure therefore does not appear to be much less powerful than the FI test procedure in the presence of this alternative. There was no particular pattern to the median LI probabilities for the twenty-one overlapping sets of parameters (not reported in table 3).

Table 4

Target-level-cost misspecification alternative: Ratio of absolute LI bias to absolute FI bias.

Parameter	Min	Ratio Percentile 10	25	50	75	90	Max
b	0.0006	0.01	0.09	0.19	0.48	1.18	72
a_0	0.002	0.10	0.30	0.86	1.86	5.02	331
a_1	0.003	0.05	0.15	0.37	1.05	2.60	314

Median probability of LI rejecting null conditional on a given FI probability of rejection

	Given FI probability 0.50	0.95
LI probability	0.07	0.12

Information on the comparative LI and FI performance when a target-level cost has been omitted from the objective function is reported in table 4. LI is notably less biased than FI in its estimates of b and a,, having a smaller bias in about three fourths of the cases. See the '75' column in the upper half of table 4. LI is also somewhat less biased in its estimation of a,. As was the case for the adjustment-cost alternative, there was quite a spread in the bias ratio. Calculations were also performed for the twenty-one overlapping subsets (not reported in table 4). Unlike the adjustment-cost alternative, the relative LI bias in b tended to be least when a, was small. Otherwise no pattern was evident.

Finally, the median probability of LI rejection given a 0.50 or 0.95 probability of FI rejection is reported in the lower half of table 4. The median LI probability is 0.07 when the FI probability is 0.50; the median value is 0.12 when the FI probability is 0.95. The LI test therefore appears to be substantially less powerful than the FI test in this case. Once again, there was no particular pattern to the median LI probabilities for the twenty-one overlapping sets of parameters (not reported in table 4).

5. **Conclusions**

In the model and parameters considered here, limited-information estimation of a single first-order condition compares asymptotically to full-information estimation of a pair of equations subject to cross-equation constraints as follows.

Under the null hypothesis of no misspecification, FI estimation is in general only slightly more efficient than LI estimation. FI standard errors typically are no more than 20 percent smaller than LI standard errors. There are, however,

some cases in which the gains from FI estimation were much larger. Such cases tend to occur when the quadratic cost a, is small. This suggests the following strategy for estimation of a linear rational-expectations model with costs of adjustment. Begin with simple limited-information estimation of the relevant Euler equation. Then undertake the substantially more complex full-information estimation only if the LI estimate of quadratic costs is small – say, less than one.

Since rational-expectations models have a marked tendency to reject tests of overidentifying restrictions [e.g., Sargent (1978)], it is also important to decide on an estimation strategy when the investigator suspects that the null hypothesis of no misspecification is false. Here the results of this paper argue for FI estimation, although such an inference is not clearcut. FI tests tend to be considerably more powerful than LI tests. But FI parameter estimates are also somewhat more biased. The case for using FI estimation does not seem to be especially strong for any particular values of quadratic costs a_0, or, for that matter, of any other parameter.

We close by stating that a desirable direction for further research is to use Monte Carlo techniques to establish the small-sample properties of the LI and FI estimators. This paper's results should help limit the burdensome calculations required by such research. These results indicate, for example, that it is probably not necessary to try a large variety of stochastic processes for forcing variables, since an asymptotic comparison of the two estimators does not appear to depend on the properties of the process. It does, however, appear necessary to consider a range of technology parameters and, if studying the properties of the estimators when the model is misspecified, a variety of alternatives. As detailed above, the relative performance of the two estimators depends quite sensitively on these.

Appendix

This outlines the derivation of the formulas necessary to calculate the distribution of the LI and FI estimates. Details may be found in the appendix to the working paper version of this paper [West (1985b)].

The basic source for the distribution of the LI estimator under the null was Hansen (1982), with the heterocedasticity correction suppressed. The basic source for the FI estimator under the null was the non-linear three-stage least-squares formulas in Amemiya (1977).

The LI bias under an alternative was calculated using $\hat{\gamma} = y + (X_0'Z\hat{S}^{-1}Z'X_0)^{-1}X_0'Z\hat{S}^{-1}Z'T^{-1/2}c_1$, c_1 a stacked vector of c_{1t}, with the bias calculated as the plim of $T^{1/2}$ times the second term. The FI bias was obtained from the non-linear first-order condition need to estimate the $(p+3) \times 1$ parameter vector $\equiv \hat{\theta}$. This first-order condition was implicitly differenti-

ated with respect to c_{2t} and $\partial\hat{\theta}/\partial c_{2t}$ was calculated. As in Kiefer and Skoog (1984), the bias in the FI estimate was approximated as $(\partial\hat{\theta}/\partial c_{2t})c_{2t}$.

The non-centrality parameter for each estimate was calculated as a suitable quadratic form in its bias. Hansen (1982) and Rao (1973) were the basic references consulted.

Application of the above was routine for LI. Some work was required for FI under the alternatives. A brief description follows for each of these two.

Consider the alternative that a cost of adjustment of order two was left out of the model. Let $a_{2T}=T^{-1/2}a_2$. If we differentiate (1) with the term $a_{2T}(\Delta^2 n_{t+2})^2$ in the objective function, and divide the result by $b^2 a_{2T}$, we obtain

$$\begin{aligned}
&\mathrm{E}_t\{1-b^{-2}a_{2T}^{-1}[(2b^2+2b)a_{2T}+a_1 b]L \\
&\quad +b^{-2}a_{2T}^{-1}[(b^2+4b+1)a_{2T}+(1+b)a_1+a_0]L^2 \\
&\quad -b^{-2}a_{2T}^{-1}[(2b+2)a_{2T}+a_1]L^3+b^{-2}L^4\}n_{t+2} \\
&=-b^{-2}a_{2T}^{-1}(w_t+u_t). \qquad \text{(A.1)}
\end{aligned}$$

Factor the lag polynomial in n_{t+2} as

$$\begin{aligned}
&(1-\lambda_{1T}L)(1-\lambda_{2T}L)(1-b^{-1}\lambda_{1T}^{-1}L)(1-b^{-1}\lambda_{2T}^{-1}L) \\
&=1-(\lambda_{1T}+\lambda_{2T}+b^{-1}\lambda_{1T}^{-1}+b^{-1}\lambda_{2T}^{-1})L \\
&\quad +(b^{-2}\lambda_{1T}^{-1}\lambda_{2T}^{-1}+2b^{-1}+\lambda_{2T}b^{-1}\lambda_{1T}^{-1}+\lambda_{1T}b^{-1}\lambda_{2T}^{-1}+\lambda_{1T}\lambda_{2T})L^2 \\
&\quad -(b^{-1}\lambda_{1T}+b^{-1}\lambda_{2T}+b^{-2}\lambda_{1T}^{-1}+b^{-2}\lambda_{2T}^{-1})L^3+b^{-2}L^4. \qquad \text{(A.2)}
\end{aligned}$$

The roots to the polynomial come in pairs $\lambda_{jT}, b^{-1}\lambda_{jT}^{-1}, |\lambda_{jT}|<1$, $\lim \lambda_{1T}=\lambda$, $\lim \lambda_{2T}=0$. (The limits are taken as the sample size $T\to\infty$.) Eq. (11) in the text comes from solving the unstable roots forwards, the stable roots backwards.

As is evident from eq. (12), calculation of the asymptotic distribution of the FI estimate requires calculation of functions like $\lim T^{1/2}(\lambda_{1T}+\lambda_{2T}-\lambda)$. This can be done by equating the coefficients on the polynomial in (A.l) and (A.2) and then taking limits. For example, $\lim T^{1/2}\lambda_{2T}$ may be calculated by

equating the coefficients on L to obtain

$$2 + 2b^{-1} + a_1 b^{-1} a_{2T}^{-1} = b^{-1}\lambda_{1T}^{-1} + b^{-1}a_{1T} + \lambda_{1T} + \lambda_{2T}$$

$$\Rightarrow \lambda_{2T} a_{2T}^{-1} = a_1^{-1} + a_1^{-1}\lambda_{2T} b\left[b^{-1}\lambda_{1T}^{-1} + \lambda_{1T} + \lambda_{2T} - 2 - 2b^{-1}\right]$$

$$\Rightarrow \lim \lambda_{2T} a_{2T}^{-1} = a_1^{-1}$$

$$\Rightarrow \lim T^{1/2}\lambda_{2T} = a_2/a_1,$$

since

$$\lim \lambda_{2T} = 0, \quad \lim \lambda_{1T} = \lambda, \quad \lim T^{1/2} a_{2T} = a_2. \tag{A.3}$$

Other required limits may be calculated analogously.

The target cost alternative involves similar algebra. Let $a_{3T} = T^{-1/2} a_3$. The Euler equation for a sample of size T is

$$\mathrm{E}_t\left\{\left[1 - \left(1 + b^{-1} + b^{-1}a_1^{-1}a_0 + b^{-1}a_1^{-1}a_{3T}\right)L + b^{-1}L^2\right] n_{t+1}\right.$$
$$\left. = b^{-1}a_1^{-1}\left(w_t - a_{3T}s_t + u_t\right)\right\}. \tag{A.4}$$

The lag polynomial in n_{t+1} may be factored as

$$(1 - \lambda_T L)\left(1 - b^{-1}\lambda_T^{-1} L\right). \tag{A.5}$$

We have $0 < \lambda_T < 1$ [Hansen and Sargent (1980)].

The full-information solution for a sample of size T is thus eq. (15). Equating the coefficients on the two lag polynomials in (A.4) and (A.5) and taking limits again yields the formulas needed to calculate the asymptotic distribution of the FI estimates.

References

Amemiya, Takeshi, 1977, The maximum likelihood and the nonlinear three squares estimator in the general nonlinear simultaneous equations model, Econometrica 45. 955-968.

Anderson, T.W., 1971, The statistical analysis of time series (Wiley, New York).

Blanchard, Olivier, 1981, What is left of the multiplier accelerator?, American Economic Review 71, 150–154.

Blanchard, Olivier, 1983, The production and inventory behavior of the american automobile industry, Journal of Political Economy 91, 365–400.

Chow, Gregory, 1975, Analysis and control of dynamic economic systems (Wiley, New York).

Chow, Gregory, 1980, Estimation of rational expectations models, Journal of Economic Dynamics and Control 2, 241-255.

Chow, Gregory, 1983, Econometrics (McGraw-Hill, New York).

Fair. Ray C. and John B. Taylor, 1983, Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models, Econometrica 51, 1169-1185.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029–1054.

Hansen, Lars Peter and Thomas Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, Journal of Economic Dynamics and Control 2, 7-46; reprinted in: Lucas and Sargent, eds. (1981) 91-136.

Hansen, Lars Peter and Thomas Sargent, 1982, Instrumental variables procedures for estimating *linear rational* expectations models, Journal of Monetary Economics 9, 263-296.

Hansen, Lars Peter and Kenneth Singleton, 1982, Generalized instrumental variable estimation of nonlinear rational expectations models, Econometrica 50. 1269-1286.

Hayashi, Fumio and Christopher Sims, 1983, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, Econometrica 51. 783-79X.

Kiefer, Nicholas M. and Garry Skoog, 1984, Local asymptotic specification error analysis, Econometrica 52, 873-886.

Lucas, Robert E., Jr. and Thomas J. Sargent, eds., 1981, Rational expectations and econometric practice (University of Minnesota Press, Minneapolis, MN).

Maddala, G.S., 1977, Econometrics (McGraw-Hill, New York).

McCallum, Bennett, 1976, Rational expectations and the natural rate hypothesis: Some consistent estimates, Econometrica 54, 43-53.

Rao, CR., 1973, Linear statistical inference and its applications (Wiley, New York).

Rotemberg, Julio, 1982, Sticky prices in the United States, Journal of Political Economy 86, 673–700.

Sargent, Thomas J., 1978, Estimation of dynamic labor demand schedules under rational expectations, in: Lucas and Sargent, eds. (1981) 463-500.

Sargent, Thomas J., 1979, Macroeconomic theory (Academic Press, New York).

Taylor, John B., 1980, Recent developments in the theory of stabilization policy, Center of the Study of American Business working paper 53.

Wallis, Kenneth, 1980, Econometric implications of the rational expectations hypothesis, in: Lucas and Sargent, eds. (1981) 329-354.

West, Kenneth D., 1984, Speculative bubbles and stock price volatility, Princeton University financial research memorandum 54.

West, Kenneth D., 1985a, A specification test for speculative bubbles, Princeton University Woodrow Wilson School discussion paper 97.

West, Kenneth D., 1985b, Full versus limited information estimation of a rational expectations model: Some numerical comparisons, Princeton University Woodrow Wilson School discussion paper 86.

West, Kermeth D., 1986, A variance bounds test of the linear quadratic inventory model, Journal of Political Economy 94, 374-401.