4

Inventory Models: The Estimation of Euler Equations

Kenneth D. West

1 Introduction

After a period of dormancy in the 1960s and 1970s, empirical work on inventories has enjoyed a resurgence in the 1980s and 1990s. In this chapter, I discuss some of the econometric issues raised by this recent work, and survey results from some recent empirical papers. For reasons of comparative advantage, and to avoid overlap with other chapters in this volume, I focus on a rational expectations version of the linear quadratic inventory model of Holt et al. (1960).

My aim is to illustrate recent developments in time series econometrics by showing how such developments have or might be applied to this often used inventory model. Some of these developments, such as the optimal linear combination of a given vector of instruments in the presence of serial correlation, are relatively well known, and appear in standard regression packages such as RATS. Others are not as well known, and, as far as I know, do not appear in standard software packages. The intended reader is one who nonetheless is willing to consider use of these techniques, but finds it difficult or tedious to plow through theoretical papers. From a theoretical econometric point of view, the discussion is informal; the interested reader may consult the cited references for discussion of underlying technical considerations.

Because of space constraints, the discussion is by no means self-contained, in that some issues that are likely to be encountered in empirical work are not discussed. Prominent among these is the question of how to model trends (unit roots and all that). I simply take as given that the researcher has somehow decided whether or not a unit root is appropriate, without asking how the decision was made or whether the testing procedure (if any) used in making the decision should be taken into account when conducting subsequent inference. Other relevant econometric issues that are not discussed here include those raised by continuous time models (e.g. Mosser, 1988; Christiano and Eichenbaum, 1989) and by aggregation (e.g. Blinder, 1981; Lovell, 1993; Schuh, 1993). As well, economic models other than the linear quadratic model are given short shrift, with only passing discussion of the flexible accelerator model (Lovell, 1961), and not a single mention of, for example, models that put inventories in the production function (e.g. Christiano, 1988; Ramey, 1989). See Blinder and Maccini (1991a; 1991b) for fine reviews that discuss these models as well as a broader array of economic (as opposed to econometric) aspects of recent inventory research.

Section 2 presents the linear quadratic model. Sections 3 to 6 discuss instrumental variables estimation of a first order condition of this model, and section 7 provides the solution and estimation of a decision rule implied by the model. Section 8 compares the approach analyzed in sections 3 to 6 with that of section 7. Section 9 surveys some recent estimates of the model. Readers uninterested in the econometric discussion may proceed directly to section 9 after familiarizing themselves with the notation defined in section 2.

2 The Linear Quadratic Model

A number of papers have followed Holt et al. (1960) and used a model in which a representative firm maximizes the expected present discounted value of future cash flows, with a cost function that includes linear and quadratic costs of production and of holding inventories. In some papers, sales and revenue are exogenous, in which case an equivalent objective is to minimize the expected present discounted value of future costs.

To state the problem formally, let p_t be real price (say, ratio of output price to the wage), S_t real sales, Q_t real production, H_t real end of period inventories, C_t real period costs, b a discount factor, $0 \leqslant b < 1$, and E_t mathematical expectations conditional on information known at time t, assumed equivalent to linear projections. The objective function, then, is

$$\max \Pi_t \equiv \lim_{T \to \infty} E_t \sum_{j=0}^{\infty} b^j (p_{t+j} S_{t+j} - C_{t+j}) \tag{4.1}$$

subject to $Q_t = S_t + H_t - H_{t-1}$,

$$C_t = 0.5a_0 \Delta Q_t^2 + 0.5a_1 Q_t^2 + 0.5a_2(H_{t-1} - a_3 S_t)^2 + u_{1t} Q_t + u_{2t} H_t.$$

For the moment, a_1 and a_2 are assumed positive, a_0 and a_3 nonnegative. The terms in a_0 and a_1 capture increasing costs of changing production and of production. The terms in a_2 and a_3 capture inventory holding and backlog costs. Section 9 discusses the role the a_is play in determining inventory behavior.

The scalars u_{1t} and u_{2t} are unobservable cost shocks that have zero mean and may be serially correlated, possibly with unit autoregressive roots. In this stripped down model, they capture any stochastic variation in costs. Some richer models surveyed briefly below include observable cost variables such as wages, raw materials prices, and interest rates.[1] Constant, trend, and seasonal terms, which also typically are included in estimation, or are removed prior to estimation, are omitted for simplicity.

An optimizing firm will not be able to cut costs by increasing production by one unit this period, storing the unit in inventory, and producing one less unit next period, holding revenue p_tS_t unchanged throughout. Formally, differentiating costs with respect to H_t gives[2]

$$\begin{aligned} &E_t\{a_0(\Delta Q_t - 2b\Delta Q_{t+1} + b^2\Delta Q_{t+2}) + a_1(Q_t - bQ_{t+1}) \\ &\qquad + ba_2(H_t - a_3S_{t+1}) + u_t\} = 0, \qquad (4.2) \\ &u_t = u_{1t} - bE_tu_{1t+1} + u_{2t}. \end{aligned}$$

Note that a_0, a_1, and a_2 are identified only up to scale: doubling all of these leaves the first order condition unchanged, apart from rescaling the unobservable disturbance u_t. Thus from this first order condition one can only aim to estimate ratios of these parameters.[3]

There are four independent unknowns to be estimated: b, a_3 and the ratios of (i) two of a_0, a_1, and a_2 to (ii) some linear combination of a_0, a_1, and a_2. Estimation of the discount rate b is, however, problematical. Analytical arguments, simulations, and empirical experience in estimating this and related models (Blanchard, 1983; West, 1986b; Gregory et al., 1992) indicate that the data are unlikely to yield sharp inferences about the value of b. Almost all the relevant literature has therefore imposed rather than estimated a value for b, which yields the additional benefit that the remaining three parameters (a_3 and two of a_0, a_1 and a_2, the latter two identified only up to scale) may be estimated linearly. A reasonable value of b comes from noting that with, say, monthly data, values for b of about 0.995 to 0.998 imply annual rates of discount of about 2% to 6%. In practice, estimates of the remaining parameters tend to be insensitive to exact choice of b (Blanchard, 1983; West, 1986a). Through the remainder of the discussion, therefore, I assume that a value of b is imposed, and there are three parameters to be estimated.

Two approaches have been used in estimating and testing this model. A *limited information* approach works off the first order condition (4.2), which, it should be noted, was derived without making a parametric assumption about the demand curve (e.g. does it depend on the price of competing products?) or market structure (competitive, monopolist, etc.). Econometric issues raised in this approach are discussed in sections 3 to 6. Section 3 discusses choice of instruments, section 4 covariance matrix estimation, section 5 methods for testing, and section 6 implications of unit root nonstationarity.[4] In sections 3–5, I assume that S_t is I(0), possibly around a trend that is not explicitly discussed, and that u_t is I(0) as well.

Section 7 discusses a second, *full information* approach, which makes a parametric assumption about demand and then solves for the firm's equilibium decision rule. Section 8 compares the limited and full information approaches.

3 Limited Information Estimation: Instrumental Variables

3.1 Introduction

This approach transforms (4.2) into an estimable equation and then uses nstrumental variables to obtain parameter estimates and test statistics. As is standard in instrumental variables estimation, a choice of left hand side variable (a normalization) is required. Asymptotically, all normalizations are equivalent, provided the linear combination of a_0, a_1, and a_2 that multiplies he left hand side variable is nonzero.

For concreteness, put $-(\partial^2 \Pi_t/\partial H_t^2)H_t = [a_0(1 + 4b + b^2) + a_1(1 + b) + ba_2]H_t \equiv cH_t$ on the left hand side;[5] the Legendre-Clebsch condition (a dynamic analog of the usual second order necessary condition; Stengel, 1986, p. 213) states that $-\partial^2 \Pi_t/\partial H_t^2 > 0$ and thus that this particular linear combination is nonzero. Then (4.2) may be rewritten

$$
\begin{aligned}
H_t &= (a_0/c)X_{0t+2} + (a_1/c)X_{1t+1} + (ba_2a_3/c)S_{t+1} + v_{t+2} \qquad (4.3)\\
&\equiv X_t'\beta + v_{t+2},\\
X_{0t+2} &\equiv -(\Delta S_t - 2H_{t-1} + H_{t-2}) + 2b(\Delta S_{t+1} + H_{t+1} + H_{t-1})\\
&\quad - b^2(\Delta S_{t+2} + H_{t+2} - 2H_{t+1}),\\
X_{1t+1} &\equiv -S_t + H_{t-1} + b(S_{t+1} + H_{t+1}),\\
v_{t+2} &\equiv u_t/c + e_{t+2},\\
e_{t+2} &\equiv -(a_0/c)(X_{0t+2} - E_tX_{0t+2}) - (a_1/c)(X_{1t+1} - E_tX_{1t+1})\\
&\quad - (ba_2a_3/c)(S_{t+1} - E_tS_{t+1}),\\
X_t &\equiv (X_{0t+2}, X_{1t+1}, S_{t+1})',\\
\beta &\equiv (\beta_1, \beta_2, \beta_3)' \equiv (a_0/c, a_1/c, ba_2a_3/c)',\\
c &\equiv a_0(1 + 4b + b^2) + a_1(1 + b) + ba_2.
\end{aligned}
$$

From an estimate of β, one can recover estimates of a_0/c and a_1/c directly, and of a_2/c and a_3 using

$$a_2/c = [1 - \beta_1(1 + 4b + b^2) - \beta_2(1 + b)]/b, \qquad a_3 = \beta_3/(ba_2/c).$$

The idea is to use a vector of instruments that is uncorrelated with v_{t+2}, but correlated with X_t, taking account of serial correlation of v_{t+2}.

3.2 Optimal Linear Combination of Given Vector of Instruments

Suppose we are given a vector of instruments Z_t of finite dimension q; since three parameters are to be estimated, an order condition is that $q \geqslant 3$. Let Z_t satisfy $EZ_t v_{t+2} = 0$, $EX_t Z_t'$ of rank 3. In West (1986a), for example, Z_t consisted of lags of H_t and S_t; Kashyap and Wilcox (1993) included lags of stock prices as well. Which $(3 \times q)$ matrix selects the optimal linear combination of instruments? The answer depends on the serial correlation properties of $Z_t v_{t+2}$, the vector of cross-products of the instruments, and the unobserved disturbance. In particular, since, as illustrated below, $Z_t v_{t+2}$ is serially correlated, the conventional two stage least squares (2SLS) estimator is not the most efficient.[6]

Let $\hat{\Omega}$ be a consistent estimate of the $(q \times q)$ matrix

$$\begin{aligned}\hat{\Omega} &\equiv \sum_{j=-\infty}^{\infty} EZ_t Z_{t-j}' v_{t+2} v_{t+2-j} \\ &\equiv \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j'), \qquad \Gamma_j \equiv EZ_t Z_{t-j}' v_{t+2} v_{t+2-j}. \end{aligned} \tag{4.4}$$

Ω is sometimes called the "long run" covariance matrix of $Z_t v_{t+2}$; procedures to obtain $\hat{\Omega}$ are discussed in section 4.

Let Z be a $T \times q$ matrix whose tth row is Z_t', and similarly let $X = [X_t']$ be $T \times 3$, $H = [H_t]$ be $T \times 1$. Hansen (1982) shows that given the vector of instruments Z_t, the $3 \times q$ matrix that selects the optimal linear combination is $EX_t Z_t' \Omega^{-1}$, with finite sample counterpart $T^{-1} X'Z\hat{\Omega}^{-1}$. Once offsetting factors of T^{-1} and T are dropped to keep the algebra uncluttered, the resulting estimator is

$$\begin{gathered}\hat{\beta} = (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'H, \qquad \hat{\beta} - \beta \approx N(0, \hat{V}), \qquad \hat{V} \equiv T(X'Z\hat{\Omega}^{-1}Z'X)^{-1} \\ \text{asymptotic } V = \text{plim } T\hat{V} = \text{plim } T^2(X'Z\hat{\Omega}^{-1}Z'X)^{-1}. \end{gathered} \tag{4.5}$$

Thus this estimator differs from the usual 2SLS one in that $\hat{\Omega}^{-1}$ replaces $(Z'Z)^{-1}$, although if $q = 3$, so that the equation is exactly identified and $X'Z$ is square and invertible, (4.5) and the 2SLS estimators are identical. Whether or not $q = 3$, 2SLS is consistent. But if $q > 3$, 2SLS is inefficient (larger asymptotic variance-covariance matrix), and, because of the serial correlation in $Z_t v_{t+2}$, the usual 2SLS variance-covariance matrix is inappropriate for inference whether or not $q > 3$.

The rather forbidding formula in (4.4) simplifies under assumptions often made in practice. If $u_t = 0$, so that there are no unobservable cost disturbances (e.g. Kashyap and Wilcox, 1993), $v_{t+2} = e_{t+2}$, and the regression disturbance is a sum of expectational errors. Now, as is well known, under rational expectations the expectational errors $X_{1t+1} - E_t X_{1t+1}$ and $S_{t+1} - E_t S_{t+1}$ are serially uncorrelated. But $X_{0t+2} - E_t X_{0t+2}$ depends on period $t+1$ and period $t+2$ information, and $X_{0t+3} - E_{t+1} X_{0t+3}$ on period $t+2$ and period $t+3$ information. This implies an MA(1) structure to $X_{0t+2} - E_t X_{0t+2}$ and thus to v_{t+2} as well. Since $E_t v_{t+2} = 0$, the vector $Z_t v_{t+2}$ is also MA(1): for $j > 1$, $E(Z_t v_{t+2} Z'_{t-j} v_{t+2-j}) = E[Z_t Z'_{t-j} v_{t+2-j} E(v_{t+2} | Z_t Z'_{t-j} v_{t+2-j})] = E[Z_t Z'_{t-j} v_{t+2-j} 0] = 0$. So in this case, $\Omega = \Sigma_{j=-1}^{1} E Z_t Z'_{t-j} v_{t+2} v_{t+2-j} = \Gamma_0 + (\Gamma_1 + \Gamma'_1)$.

If the unobservable cost shock u_t is present, Q_t, H_t and, in general, S_t will depend on that shock. But considerable simplification of (4.4) may nonetheless occur. If u_{1t} and u_{2t} and thus u_t are white noise (West and Wilcox, 1993b), $X_{0t+2} - E_t X_{0t+2}$ will depend in part on u_{t+1} and u_{t+2}, implying that v_{t+2}, which depends on u_t as well, will be MA(2). The vector $Z_t v_{t+2}$ will be MA(2) as well, so $\Omega = \Sigma_{j=-2}^{2} E Z_t Z'_{t-j} v_{t+2} v_{t+2-j} = \Gamma_0 + (\Gamma_1 + \Gamma'_1) + (\Gamma_2 + \Gamma'_2)$.

More generally, if u_t has an autoregressive component, so too will v_{t+2}. If u_t follows a particular parametric process, such as an AR(1), one can estimate the AR(1) parameter simultaneously with the cost parameters (Kollintzas, 1992).

3.3 Comparison with Two Stage Least Squares

What are the efficiency gains from using (4.5) rather than two stage least squares (2SLS)? Here, the 2SLS estimator $\dot{\beta}$ is

$$\dot{\beta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'H, \tag{4.6}$$

$$\text{asymptotic } \dot{V} = \text{plim } T^2[X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}\Omega(Z'Z)^{-1} Z'X[X'Z(Z'Z)^{-1}Z'X]^{-1}.$$

Plainly, the answer to this question depends on the data generating processes (DGPs) for the Z and X variables, as well as those of the unobservable shocks.

To illustrate what the gains might be, I have worked them out in a simple case. Sales are forecast from an exogenous AR(2), the cost shock u_t is serially uncorrelated, and two lags each of sales and inventories are used as instruments:

$$\begin{aligned} S_t &= \phi_1 S_{t-1} + \phi_2 S_{t-2} + v_{2t}, \\ E u_t u_{t-j} &= E u_t v_{2t-j} = 0 \quad \text{for } j \neq 0, \\ Z_t &= (H_{t-1}, H_{t-2}, S_{t-1}, S_{t-2})'. \end{aligned} \tag{4.7}$$

Given ϕ_1, ϕ_2, a_0, a_1, a_2, a_3, b, and the variance-covariance matrix of (u_t, v_{2t}) one may then solve for the data generating process for inventories and sales, using methods outlined in section 7 on full information estimation.

I tried three data generating processes, each corresponding to a different set of cost parameters, but all using a common set of parameters for the sales processes and shocks. These sets of cost parameters are summarized in table 4.1. They are intended to correspond to parameters found in some studies of

Table 4.1 Cost parameters for data generating processes

Mnemonic	a_0	a_1	a_2	a_3
E	0.01	0.25	0.50	0.70
R	0.20	−1.00	2.00	0.35
W	0.10	0.40	0.01	2.00

(a) a_0, a_1, a_2 and a_3 are defined in (4.1).
(b) For all three data generating processes, the discount rate was set at $b = 0.995$. The sales process was assumed to follow an exogenous AR(2), $S_t = 0.7S_{t-1} + 0.25S_{t-2} + v_{2t}$, with $\mathrm{var}(v_{2t}) = 0.128033$, $\mathrm{corr}(u_t, v_{2t}) = -0.5$, $\mathrm{var}(u_t) = 0.5$, with u_t the cost shock as defined in (4.1).

two digit manufacturing industries in the United States: Eichenbaum (1989, p. 862, top panel: mnemonic E), Ramey (1991, p. 323: R), and West (1986a, p. 393, top panel: W).[7] (The problem (4.1) is well posed for the R DGP despite the negative sign on a_1: see Killintzas, 1989; Ramey, 1991.)

I set the variance-covariance matrix of (u_t, v_{2t}) so that for the W DGP the unconditional variance-covariance matrix of (H_t, S_t) was approximately proportional to that of US monthly nondurables manufacturing, 1967–90; H_t are finished goods inventories, with $\mathrm{var}(S_t) = 1$ (a harmless normalization). The exact parameters are given in the notes to table 4.1. It should be noted that the implied reduced forms for DGPs E and R (not given in the table) are implausible in that H_t displays little serial correlation; both Eichenbaum (1989) and Ramey (1991) implicitly accounted for the serial correlation that is empirically present in H_t by allowing for serially correlated cost shocks, which I omit for simplicity.

Table 4.2 compares some standard errors implied by the asymptotic variance-covariance matrices V (defined in (4.5)) and $\dot{V}$ (defined in (4.6)). Define

$$c_t \equiv \sum_{j=0}^{\infty} E_t C_{t+j}$$

as the expected present discounted value of costs. Then the quantity in column 2 of table 4.2, $(1 + b)a_0 + a_1 \equiv \partial^2 c_t/\partial Q_t^2$, is the slope of marginal production cost, which Ramey (1991) has argued is of central economic interest. To my surprise, the optimal estimator yields little efficiency gains, for any of the DGPs: column 4, for example, indicates that the asymptotic variance of $T^{1/2}(\hat{a}_3 - a_3)$ is at best $(0.998)^2$ times smaller for the optimal than for the 2SLS estimator. Similar ratios apply for $(1 + b)a_0 + a_1$ and a_3 (columns 2 and 3).

No doubt there are other data generating processes for which the gains from the optimal estimator are quite large. But even for the data generating processes assumed here, the small efficiency gains do not argue that it is a matter of indifference which estimator one uses, if one is interested in

Table 4.2 Comparison of asymptotic variance-covariance matrices resulting from optimal and 2SLS linear combinations of instruments

1 DGP	*2* Ratio of SEs on $[(1+b)a_0+a_1]/c$	*3* Ratio of SEs on a_2/c	*4* Ratio of SEs on a_3
E	0.990	0.982	0.976
R	0.999	0.999	0.998
W	0.992	0.962	1.000

(a) a_0, a_1, and a_2 are defined in (4.1); c is defined in (4.3).
(b) In the ratios referenced in columns 2–4, the numerator is the standard error computed from the variance-covariance matrix of the estimator (4,5), which optimally combines a given set of instruments; the denominator is the standard error from that of the two stage least squares estimator (4.6). Both estimators were assumed to use a (3×4) linear combination of a (4×1) instrument vector consisting of two lags each of H_t and S_t. The ratios must lie between 0 and 1, smaller numbers indicating a greater efficiency gain from using the optimal estimator.
(c) These asymptotic comparisons may not accurately predict the actual finite sample performance of the two estimators.

performing inference on the estimated parameters. If one estimates by 2SLS, the appropriate covariance matrix is given in (4.6). Suppose one instead uses the traditional (and, in the present example, incorrect) covariance matrix

$$\tilde{V} = Ev_{t+2}^2 \text{ plim } T[X'Z(Z'Z)^{-1}Z'X]^{-1}. \tag{4.8}$$

How accurate are hypothesis tests?

Table 4.3 considers this question (asymptotically) for tests of three simple hypothesis tests. Interpretation of the tests is given at the foot of the table. By

Table 4.3 Asymptotic sizes of nominal 0.05 tests, when an inconsistent estimator of the 2SLS covariance matrix is used

DGP	$[(1+b)a_0+a_1]/c$	a_2/c	a_3
E	0.107	0.143	0.249
R	0.053	0.051	0.011
W	0.100	0.533	0.500

(a) This table presents the asymptotic size of a nominal 0.05 test of the null that the indicated parameter equals its population value, as given in table 4.2. For example, in the row E, column a_2/c, the null is that $a_2/c = 0.5$. It is assumed that the covariance matrix given in (4.8) is used in calculating the relevant standard error.
(b) These asymptotic calculations may not accurately predict the actual finite sample performance of such tests.

and large there are dramatic distortions. The a_3 entry for DGP W, for example, indicates that a test that should reject only 5% of the time in fact will reject 50% of the time. The implication is that it is important to use a consistent estimate of the covariance matrix. Since such a matrix requires estimation of Ω (see (4.7)), use of 2SLS in the end will be no simpler than use of the optimal estimator (4.5).

How accurately does the asymptotic theory underlying tables 4.2 and 4.3 apply in the finite samples used in practice? The answer to this question is not well known. Preliminary simulation results in West and Wilcox (1993a; 1993b) indicate that while the asymptotic theory often provides a good guide to finite sample performance, the estimator usually displays slightly more variability than is predicted by the asymptotic theory, and sometimes displays some bias as well; on occasion, test statistics are very poorly sized. In empirical work, the limited applicability of asymptotic theory is perhaps suggested by the sensitivity of estimates of the model to choice of left hand side variable (Krane and Braun, 1991; Ramey, 1991; Kashyap and Wilcox, 1993), a sensitivity not displayed by the asymptotic theory.

3.4 Alternative Instrumental Variables Techniques

What are the implications for empirical work of such inaccuracy in asymptotic approximations? For inference, one might want to use simulation or bootstrap techniques, although such techniques, which typically require specification of a data generating process for all the variables in the system, seem more natural in systems estimated by full rather than limited information techniques. See West (1992) for an illustration of the use of bootstrap techniques in a full information environment.

Alternatively, for estimation as well as inference one might want to consider estimators that have better asymptotic or finite sample properties. Full information estimators are discussed below. To motivate alternative limited information estimators, begin by considering how one chooses an instrument vector, a decision that so far has been taken as given. Obvious candidates for instruments include lags of S_t and H_t (equivalently, lags of H_t and Q_t, given the inventory identity $Q_t = S_t + \Delta H_t$). This is, indeed, done in many studies, with $Z_t = [H_t, S_t, H_{t-1}, S_{t-1}, \ldots]'$ if cost shocks are assumed absent, $Z_t = [H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, \ldots]'$ if cost shocks are assumed present but serially uncorrelated. Hansen (1985) shows that given the serial correlation in $Z_t v_{t+2}$, an increased number of lags used (i.e. an increase in the dimension of Z_t) yields a strict increase in the asymptotic efficiency of the estimator. This applies even if, as in the example in section 7, there is a finite number of lags of H_t and S_t in the reduced form for $(H_t, S_t)'$. Since this result by itself gives little practical guidance on the number of lags to use, it is advisable to experiment with various lag lengths to see if results are sensitive to the exact number of lags used (e.g. Eichenbaum, 1989).

Alternatively, if one is willing to specify and estimate a finite parameter ARMA model for H_t, S_t, and v_{t+2}, one can apply the formulas in Hansen (1985) to obtain a 3×1 instrument vector Z_t^* (say) that is optimal in the space of instrument vectors that rely on lagged H_ts and S_ts. Suppose for concreteness that the environment is as in section 3.3: the cost shock u_t is iid, so that $v_{t+2} \sim$ MA(2) and $EH_{t-j}v_{t+2} = ES_{t-j}v_{t+2} = 0$ for $j \geqslant 1$. Then Z_t^* is a linear combination of past H_ts and S_ts, with nonzero weights on *all* lags: $H_1, S_1, \ldots, H_{t-1}$, and S_{t-1} are all used to construct Z_t^*. See West and Wilcox (1993a), who show how to make this estimator operational, and indicate that for some but not all plausible DGPs there are large asymptotic efficiency gains from using (a) Z_t^* rather than (b) section 3.2's conventional GMM estimator with Z_t the set of lags of H_t and S_t in the reduced form for $(H_t, S_t)'$.

Ramey (1991), however, notes that any lags of H_t will be correlated with the disturbance v_{t+2} if unobservable cost shocks have an autoregressive component, implicitly argues that we do not have *a priori* evidence about the parametric structure of such a component, and explicitly calls for using instruments that she describes as "truly exogenous." These include oil prices, military spending, and dummies for the political party of the president.

It is difficult to evaluate Ramey's suggested instruments. On the one hand, there is some Monte Carlo evidence that instruments that are only weakly correlated with the vector of right hand side variables may perform poorly in finite samples, even if they are uncorrelated with the disturbance (Nelson and Startz, 1990); on the other hand, it is well known that instruments that are correlated with the disturbance will perform poorly even in large samples, no matter how strongly correlated with the vector of right hand side variables. Monte Carlo evidence on what Shea (1993) calls the tradeoff between "exogeneity" and "relevance" would be very useful.

4 Limited Information Estimation: Covariance Matrix Estimation

As (4.5) indicates, one must compute $\hat{\Omega}$ for inference and (if the efficient estimator is used) estimation. To discuss how to do so, let $\dot{v}_t$ be the 2SLS residual, $\dot{v}_{t+2} = H_t - X_t'\dot{\beta}$, where $\dot{\beta}$ is defined in (4.6). Consider estimating Γ_j (defined in (4.4)) by

$$\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} Z_t Z_{t-j}' \dot{v}_{t+2} \dot{v}_{t+2-j} \quad \text{for } j \geqslant 0. \tag{4.9}$$

For concreteness, focus initially on the case where the cost shock u_t is absent, so that $Z_t v_{t+2} \sim MA(1)$ and $\Omega = \Gamma_0 + (\Gamma_1 + \Gamma_1')$. The simplest technique used to estimate $\hat{\Omega}$ is the obvious one, $\hat{\Omega} = \hat{\Gamma}_0 + (\hat{\Gamma}_1 + \hat{\Gamma}_1')$. This is, indeed, consistent. But it is not guaranteed to be positive definite, and, indeed, in practice is not always positive definite (e.g. Cumby and Huizanga, 1990). In such a case, the variance-covariance matrix $\hat{V}$ (defined in (4.5)) will also fail to be positive definite, and there will be some linear combinations of parameters whose estimated standard errors will be negative.

While $\hat{\Gamma}_0 + (\hat{\Gamma}_1 + \hat{\Gamma}_1')$ will be positive definite asymptotically, if the model is right, there is an evident need for an estimator that will be positive definite by construction. Early proposals include Cumby, Huizanga, and Obstfeld (1983) and Eichenbaum, Hansen, and Singleton (1988). Since Ω is proportional to the spectral density of $Z_t v_{t+2}$ at frequency zero (e.g. Granger and Newbold, 1977), recent research has built on the well developed literature on estimators of spectral densities.

One part of this literature suggests constructing $\hat{\Omega}$ as

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{j=1}^{m} k_j(\hat{\Gamma}_j + \hat{\Gamma}_j'), \tag{4.10}$$

for weights k_j chosen to insure that $\hat{\Omega}$ is positive definite, and a suitably chosen bound m. A simple choice of weights are what are called the "Bartlett" ones,

$$k_j = 1 - [j/(m+1)] \Rightarrow k_1 = m/(m+1),\ k_2 = (m-1)/(m+1), \ldots, k_m = 1/(m+1). \tag{4.11}$$

The formal asymptotic theory underlying use of the estimator (4.10) requires that as the sample size T becomes arbitrarily large, then so too does m, but in such a fashion that $m/T^{1/2} \to 0$ (Newey and West, 1987a). This theory is applicable not only when the vector $Z_t v_{t+2}$ is MA(1), as in the simple case used to motivate the present discussion, but for any ARMA process for $Z_t v_{t+2}$, even when the order of the ARMA process is not known *a priori*.

It should be emphasized that even in the simple case that $Z_t v_{t+2}$ is known to be MA(1), the theory requires that m increase with T: even though one knows *a priori* that the population value of Γ_j is zero for $j > 1$, for sufficiently large T one will want to be using estimates of the form $\hat{\Omega} = \hat{\Gamma}_0 + [m/(m+1)](\hat{\Gamma}_1 + \hat{\Gamma}_1') + \ldots + [1/(m+1)](\hat{\Gamma}_m + \hat{\Gamma}_m')$ for $m > 1$. The reason is that one wants $\hat{\Omega}$ to well approximate $\Gamma_0 + (\Gamma_1 + \Gamma_1')$ in a large sample; if m is fixed at 1 independent of T, then $\hat{\Omega}$ will instead approximate $\Gamma_0 + (1/2)(\Gamma_1 + \Gamma_1')$.

But just how large should m be for a given sized sample, in the simple MA(1) example or more generally? While it seems unlikely that a fully automatic rule for selecting m will be satisfactory in all attempts to estimate (4.3), Andrews (1991) and Andrews and Monahan (1992) have developed procedures that may be used to produce an initial choice of m that (i) is asymptotically optimal in a certain sense, and (ii) can be used as a starting point in subsequent experimentation.

I will illustrate this using Newey and West's (1993) extension of those procedures, since it is simpler to explain (in my totally unbiased opinion). Let [.] denote "integer part of." An asymptotically optimal choice of m satisfies

$$m = [\gamma T^{1/3}], \tag{4.12}$$

$$\gamma = 1.1447(\{s^{(1)}/s^{(0)}\}^2)^{1/3}, \qquad s^{(1)} = 2\sum_{j=1}^{\infty} j\sigma_j, \qquad s^{(0)} = \sigma_0 + 2\sum_{j=1}^{\infty} \sigma_j,$$

$$\sigma_j = w'\Gamma_j w, \qquad w = (1, 1, \ldots, 1, 1)',$$

where w is $(q \times 1)$.

Suppose first that cost shocks are absent, so that $Z_t v_{t+2} \sim \mathrm{MA}(1)$, $\sigma_j = \Gamma_j = 0$ for $j > 1$, $s^{(1)} = 2\sigma_1$ and $s^{(0)} = \sigma_0 + 2\sigma_1$. Then one could set

$$m = [\hat{\gamma} T^{1/3}], \tag{4.13}$$

$$\hat{\gamma} = 1.1447(\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^2)^{1/3}, \qquad \hat{s}^{(1)} = 2\hat{\sigma}_1, \qquad \hat{s}^{(0)} = \hat{\sigma}_0 + 2\hat{\sigma}_1, \qquad \hat{\sigma}_j = w'\hat{\Gamma}_j w.$$

An alternative procedure for selecting m, applicable both when cost shocks are absent and when cost shocks follow any stationary process, is to choose m by

$$m = [\hat{\gamma} T^{1/3}], \tag{4.14}$$

$$\hat{\gamma} = 1.1447(\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^2)^{1/3}, \qquad \hat{s}^{(1)} = 2\sum_{j=1}^{n} j\hat{\sigma}_j, \qquad \hat{s}^{(0)} = \hat{\sigma}_0 + 2\sum_{j=1}^{n} \hat{\sigma}_j,$$

$$n = [4(T/100)^{2/9}].$$

Thus, if $T = 100$, $n = 4$; if $T = 300$, $n = 5$.[8]

Andrews and Monahan (1992) emphasize the possible benefits of combining what is called "prewhitening" with a procedure such as that just described. To illustrate Newey and West's (1993) prewhitened estimator, let

$$\hat{g}_t = Z_t \hat{v}_{t+2}, \qquad \hat{A} = \sum_{t=2}^{T} \hat{g}_t \hat{g}_{t-1}' \left(\sum_{t=2}^{T} \hat{g}_{t-1} \hat{g}_{t-1}'\right)^{-1}, \qquad \tilde{g}_t \equiv \hat{g}_t - \hat{A}\hat{g}_{t-1}, \tag{4.15}$$

$$n = [4(T/100)^{2/9}],$$

$$\tilde{\sigma}_j = T^{-1} \sum_{t=j+1}^{T} \{(w'\tilde{g}_t)(w'\tilde{g}_{t-j}), \quad j = 0, \ldots, n,$$

$$\hat{s}^{(1)} = 2\sum_{j=1}^{n} j\tilde{\sigma}_j, \qquad \hat{s}^{(0)} = \tilde{\sigma}_0 + 2\sum_{j=1}^{n} \tilde{\sigma}_j, \qquad \hat{\gamma} = 1.1447(\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^2)^{1/3}.$$

Thus, $\hat{A}$ is the $(q \times q)$ matrix of VAR(1) regression coefficients obtained by regressing cross-products of instruments and residuals on their first lag, and $\tilde{g}_t$ is the resulting $(q \times 1)$ vector of period t residuals. The idea is to apply a procedure such as that just described to the VAR residuals $\tilde{g}_t$, and then use $\hat{A}$ to adjust the result. Specifically, one sets

$$\hat{\Omega} = (I - \hat{A})^{-1} \tilde{\Omega} (I - \hat{A})^{-1\prime}, \tag{4.16}$$

$$\tilde{\Omega} = \{\tilde{\Gamma}_0 + \sum_{j=1}^{m} \{1 - j/(m+1)\}(\tilde{\Gamma}_j + \tilde{\Gamma}_j')\}, \qquad \tilde{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} \tilde{g}_t \tilde{g}_{t-j}', \qquad j = 0, \ldots, m,$$

$$m = [\hat{\gamma} T^{1/3}].$$

When cost shocks are absent, (4.13) and (4.14) are asymptotically equivalent; whether the prewhitened estimator (4.16) is asymptotically preferable to (4.13) and (4.14) depends on the underlying data generating process. The argument for prewhitening is not so much that it yields great asymptotic gains as that it seems to work well in simulations.

But with or without cost shocks, (4.13), (4.14), and (4.16) might yield different values of m in a given application. This illustrates the general point that regardless of the process followed by cost shocks, there are a number of reasonable rules to choose m. For the data generating processes considered in Newey and West (1993), the rules (4.14) and (4.16) worked relatively well. But since asymptotic theory does not yield a single value of m for a given data set and sample size, it is advisable to do some experimentation with a range of values, computing at least some test statistics using different $\hat{V}$s, each $\hat{V}$ relying on a $\hat{\Omega}$ computed with a different m and/or n; the hope is that results will not be sensitive to the exact values of n chosen. Further theoretical and simulation evidence on alternative rules are of great interest.

How do the rules developed to date compare with longer established testing procedures? The Monte Carlo evidence in Andrews (1991), Andrews and Monahan (1992), and Newey and West (1993) indicates that their recommended procedures are preferable to more traditional ones. This evidence indicates as well, however, that when data are highly serially correlated, tests may suffer serious size distortions even in sample sizes larger than those typically used in inventory studies: nominal 0.05 tests may have actual sizes of 0.20 or higher.

5 Limited Information Estimation: Testing

Let $\hat{\beta}$ be defined as in (4.5). Given a hypothesis H_0: $g(\beta) = 0$, where $g(\beta)$ is $s \times 1$ and $\partial g/\partial \beta$ is of row rank $s \leqslant 3$, the usual Wald statistic is appropriate,

$$g(\hat{\beta})'[(\partial g/\partial \hat{\beta}) \hat{V} (\partial g/\partial \hat{\beta})']^{-1} g(\hat{\beta}) \overset{A}{\sim} \chi^2(s), \tag{4.17}$$

where $\partial g/\partial \hat{\beta}$ denotes the $(s \times 3)$ matrix $\partial g/\partial \beta$ evaluated at $\hat{\beta}$. If the null is simply H_0: $\beta_i = 0$, then (4.17) is of course just the square of the usual t-statistic.

Let $\hat{v}_{t+2} = H_t - X_t'\hat{\beta}$, $\hat{v} = [\hat{v}_{t+2}]$. If the number of instruments q exceeds the number of regressors 3, one can test the model by computing

$$T^{-1} \hat{v}' Z \hat{\Omega}^{-1} Z' \hat{v} \overset{A}{\sim} \chi^2(q-3). \tag{4.18}$$

This is sometimes referred to as Hansen's (1982) "J-test," or a test of instrument residual orthogonality. One interpretation of this is as a test of whether the coefficients of an arbitrarily chosen set of $q - 3$ instruments would all be zero if one added these instruments to the regression equation (4.3) (Newey and West, 1987b). In practice, it has been difficult to turn a rejection by (4.18) into a constructive suggestion about how the first order condition (4.3) or underlying model (4.1) should be modified, at least in my experience.

An easier to interpret, although perhaps less powerful, test was suggested by West (1986a), applied by Krane and Braun (1991) and Dimelis and Ghali (1992), and extended by Kollintzas (1992). It may be shown that the model (4.1) implies that certain weighted sums of the variances, auto- and cross-covariances of H_t, S_t, Q_t, and the cost shocks are nonnegative. If $u_t = 0$ as in West (1986a), for example, we have

$$a_0[\text{var}(\Delta S) - \text{var}(\Delta Q)] + a_1[\text{var}(S) - \text{var}(Q)] - a_2\text{var}(H) + 2a_2a_3\text{cov}(H, S_{+1}) \geqslant 0, \tag{4.19}$$

where "var" denotes variance and "cov" denotes covariance. To compute the left hand side of (4.19), one may use estimates of the a_is from the Euler equation and of the indicated second moments from the obvious sample counterparts. A standard error may be computed from the joint variance-covariance matrix of the estimated a_is and second moments. See West (1986a) for details.

The left hand side of (4.19) is the difference in average per period costs between the policy actually followed and that of an alternative feasible policy that leaves sales unchanged but sets $Q_t = S_t$ and H_t to zero. If $a_0 = a_3 = 0$, and a_1, $a_2 > 0$, for example, (4.19) reduces to $a_1[\text{var}(S) - \text{var}(Q)] - a_2\text{var}(H) \geqslant 0$: the average per period reduction in production costs allowed by inventory holdings had better be bigger than the average costs of holding inventories themselves – or why would a firm hold inventories? See West (1986a), Kollintzas (1992), and section 8 for further discussion and interpretation.

6 Limited Information Estimation: Unit Roots

6.1 H_t, S_t Cointegrated

Assume now that $S_t \sim I(1)$, but $u_t \sim I(0)$. Kashyap and Wilcox (1993) emphasize that H_t and S_t are then cointegrated. Specifically, with a little bit of algebra, the first order condition (4.2) can be rewritten as

$$\begin{gathered} E_t\{a_0[\Delta S_t + \Delta^2 H_t - 2b(\Delta S_{t+1} + \Delta^2 H_{t+1}) + b^2(\Delta S_{t+2} + \Delta^2 H_{t+2})] \\ - ba_1(\Delta H_{t+1} + \Delta S_{t+1}) + a_1 \Delta H_t + ba_2(H_t - \gamma S_t)\} = -u_t, \qquad (4.20) \\ \gamma \equiv a_3 - [a_1(1-b)/a_2]. \end{gathered}$$

Let {.} denote the expression in braces. It may be shown that $\{.\} \sim I(0)$, and, since $\Delta S_t \sim I(0)$ by assumption, $H_t - \gamma_t \sim I(0)$: H_t and S_t are cointegrated with cointegrating parameter γ.[9]

Consider first the case where $E\Delta H_t \neq 0 (\Rightarrow E\Delta S_t \neq 0)$, as is probably appropriate when one has a long time series from a growing industry. It follows from West (1988a) that the discussion in section 3 still applies: the optimal linear combination of instruments is as given in that discussion, and the resulting coefficient vector is asymptotically normal. The discussions in sections 4 and 5 are applicable as well: (4.18) is asymptotically chi-squared (West, 1988a) and certain variance bounds tests (not the one described in section 5) may be performed as well (West, 1988b; 1990). Andrews and Monahan (1992) indicate that data dependent procedures to select m (defined in (4.10)) still are appropriate.

It sometimes will be more reasonable to assume that $E\Delta H_t = 0$, however. There is little secular movement of the Depression era data in Kashyap and Wilcox (1993), for example. To my knowledge, no one has directly summarized the implications for estimation of equations such as (4.3) under such circumstances. Park and Phillips (1988) and Sims, Stock, and Watson (1990) emphasize that in the related contexts that they consider, the entire coefficient vector will not be asymptotically normal with a full rank variance-covariance matrix.

But as Kashyap and Wilcox (1993) illustrate, inference about many objects of interest may be done in a conventional fashion. Individual coefficients will be asymptotically normal, and inference about such coefficients may proceed as usual if procedures described in section 4 are used to estimate the variance-covariance matrix (Park and Phillips, 1988; West, 1988a; Sims, Stock, and Watson, 1990; Andrews and Monahan, 1992; Hansen, 1992); Kashyap and Wilcox (1993) conjecture that the equation (4.18) statistic will still be asymptotically chi-squared, which seems reasonable given that the statistic can be interpreted as testing the joint significance of $(q - 3)$ of the instruments.

For some other linear rational expectations models, Stock and West (1988) and West (1988a) present Monte Carlo evidence indicating that the asymptotic normal approximation is adequate in sample sizes typically encountered.

6.2 H_t, S_t Not Cointegrated

Now suppose that $u_t \sim I(1)$. Then H_t and S_t obviously may not be cointegrated. For inference to proceed along standard lines, one must difference equation (4.3). The discussion in sections 4 and 5 now applies, with some obvious modifications. Differences of H_t and S_t are now prominent candidates for instruments, for example. And if u_t is a pure random walk, the disturbance of the differenced equation is MA(2).

Lack of cointegration between H_t and S_t may at first blush seem surprising. But note that the model under consideration in fact rationalizes such an occurrence if cost shocks are I(1), as indeed is typically maintained in the literature on real business cycles. And standard tests applied to US data at the two digit SIC code and more aggregate levels generally do not reject the null of no cointegration (Granger and Lee, 1989; West, 1990; Rossana, 1992).

7 Full Information Estimation

7.1 Solution of the Model

For algebraic simplicity, assume that the firm views revenue $p_t S_t$ as exogenous, so that the objective function becomes one of cost minimization. (The assumption of exogenous revenue over an infinite horizon obviously is silly. But it makes discussion of the relevant econometric issues relatively straightforward. Below I comment briefly on some implications of sales being endogenous: see Eichenbaum, 1984; Blanchard and Melino, 1986; Dimelis and Kollintzas, 1989; and West, 1990 for completely worked out examples of solution and estimation when sales are endogenous.) Also for simplicity, assume that both the firm and the econometrician forecast future sales from a univariate autoregression in S_t,

$$\begin{aligned} &S_t = \phi_1 S_{t-1} + \ldots + \phi_p S_{t-p} + v_{2t}, \\ &\phi_p \neq 0, \qquad E v_{2t} S_{t-j} = E v_{2t} H_{t-j} = 0 \quad \text{for } j > 0, \\ &1 - \phi_1 z - \ldots - \phi_p z^p = 0 \Rightarrow |z| \geqslant 1, \end{aligned} \tag{4.21}$$

where $|z|$ denotes the modulus of a complex number z. Note that (4.21) allows S_t to have a unit autoregressive root. The innovation v_{2t} is assumed uncorrelated with lagged H_{t-j}, in accord with the assumption that sales are exogenous.

For the moment, assume $a_0 \neq 0$. Let L be the lag operator. Use the identity $Q_t = S_t + \Delta H_t$ to rewrite the Euler equation (4.2) as

$$E_t\{f(L)H_{t+2} = D_t\}, \tag{4.22}$$

$$\begin{aligned} f(L) \equiv\ & 1 - b^{-2}a_0^{-1}[ba_1 + 2a_0 b(1+b)]L + b^{-2}a_0^{-1}[a_0(1+4b+b^2) + a_1(1+b) + ba_2]L^2 \\ & - b^{-2}a_0^{-1}[a_1 + 2a_0(1+b)]L^3 + b^{-2}L^4, \\ D_t =\ & -b^{-2}(\Delta S_t - 2b\Delta S_{t+1} + b^2\Delta S_{t+2}) - b^{-2}a_0^{-1}a_1(S_t - bS_{t+1}) + b^{-1}a_0^{-1}a_2 a_3 S_{t+1} \\ & - b^{-2}a_0^{-1}u_t. \end{aligned}$$

Let λ_i be the roots of the fourth order lag polynomial $f(L)$, $|\lambda_1| \leqslant \ldots \leqslant |\lambda_4|$. It may be shown that $\lambda_4 = 1/(b\lambda_1)$ and $\lambda_3 = 1/(b\lambda_2)$, so that at $|\lambda_1|$, $|\lambda_2| < 1/b$. Suppose further that $|\lambda_1|$, $|\lambda_2| < 1$, and, for expositional convenience, that $\lambda_1 \neq \lambda_2$. (See Kollintzas, 1989 on the relationship between the cost parameters, the discount rate, and the modulus of these roots.) If we are to obtain a nonexplosive solution, we must solve the stable roots λ_1 and λ_2 backwards, and the unstable roots λ_3 and λ_4 forwards: indeed, a transversality condition forces the firm to do so (Kollintzas, 1989). One may verify that the following then satisfies the Euler equation (4.22):

$$H_t = (\lambda_1 + \lambda_2)H_{t-1} - \lambda_1\lambda_2 H_{t-2} + b^{-1}\lambda_1\lambda_2(\lambda_1 - \lambda_2)^{-1} \times \sum_{j=0}^{\infty} \{[(b\lambda_1)^{j+1} - (b\lambda_2)^{j+1}]E_t D_{t+j}\}. \tag{4.23}$$

Note that if λ_1 and λ_2 are complex, they are complex conjugates, so that $\lambda_1 + \lambda_2$ and $\lambda_1\lambda_2$ are real.

Suppose finally, for simplicity, that the unobservable shock u_t is serially uncorrelated. Then one can use techniques such as those in Blanchard (1983) to solve for the reduced form, which expresses H_t in terms of lagged H_ts and S_ts and u_t, and for the decision rule, which expresses H_t in terms of lagged H_ts, current and lagged S_t, and current u_t. Define the scalars ρ_1, ρ_2, w_1, w_2, w_3, and w_4 the $(l \times p)$ vector e' and the $(p \times p)$ matrices Φ and D as

$$\begin{aligned}
&\rho_1 = \lambda_1 + \lambda_2, \qquad \rho_2 = -\lambda_1\lambda_2, \\
&w_1 = b^2\rho_2, \qquad w_2 = -\rho_2[b^2 + 2b + b(a_1/a_0) + (ba_2a_3/a_0)], \\
&w_3 = \rho_2[2b + 1 + (a_1/a_0)], \qquad w_4 = -\rho_2, \\
&e' = (1\ 0\ \ldots\ 0), \\
&\Phi = \begin{matrix} (\phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p) \\ (1 & 0 & \ldots & 0 & 0) \\ \ldots \\ (0 & 0 & \ldots & 1 & 0), \end{matrix} \\
&D = [I - b\rho_1\Phi - b^2\rho_2\Phi^2]^{-1}.
\end{aligned} \tag{4.24}$$

Then the reduced form is

$$\begin{aligned}
H_t &= \rho_1 H_{t-1} + \rho_2 H_{t-2} + \pi_1 S_{t-1} + \ldots + \pi_p S_{t-p} + v_{1t}, \\
(\pi_1, \ldots, \pi_p) &= e'D(w_1\Phi^3 + w_2\Phi^2 + w_3\Phi + w_4 I), \\
v_{1t} &= (\pi_p/\phi_p)v_{2t} + v_{Ht} \qquad (p \geqslant 2), \\
v_{1t} &= [(\pi_1 + \rho_2)/\phi_1]v_{2t} + v_{Ht} \qquad (p = 1), \\
v_{Ht} &= (\rho_2/a_0)u_t.
\end{aligned} \tag{4.25}$$

The underlying decision rule is

$$\begin{aligned}
H_t &= \rho_1 H_{t-1} + \rho_2 H_{t-2} + \delta_1 S_t + \ldots + \delta_p S_{t-p+1} + v_{Ht} \qquad (p \geqslant 2), \\
H_t &= \rho_1 H_{t-1} + \rho_2 H_{t-2} + \delta_1 S_t + \delta_2 S_{t-1} + v_{Ht} \qquad (p = 1),
\end{aligned} \tag{4.26}$$

where for $p \geqslant 2$,

$$\delta_1 = \pi_p/\phi_p, \qquad \delta_i = \pi_{i-1} - \delta_1\phi_{i-1} \quad (i = 2, \ldots, p),$$

and for $p = 1$,

$$\delta_2 = -\rho_2, \qquad \delta_1 = (\pi_1 - \delta_2)/\phi_1.$$

7.2 Estimation and Inference

One aims to estimate the pair of equations (4.25) and (4.21) jointly. Given estimates of the coefficients of (4.25) and an imposed value of b, one can retrieve estimates of the underlying parameters of the cost function (the a_is) using (Blanchard, 1983)

$$a_1/a_0 = \rho_1(b - \rho_2^{-1}) - 2b - 2, \tag{4.27}$$

$$a_2/a_0 = -b^{-1}[\rho_2^{-1}(1 + b\rho_1^2) + b^2\rho_2 + (1 + b)(a_1/a_0) - (1 + 4b + b^2)].$$

Given estimates of the sales process (4.21) as well, estimates of a_3 can be disentangled from π_1 (or, for that matter, from any of the other π_is as well, if $p \geqslant 2$).

There are $p + 3$ parameters to be estimated: $\phi_1, \ldots, \phi_p$, a_3, and, relative to some linear combination of the a_is, the values of two of a_0, a_1, or a_2. There are $2p + 2$ right hand side variables in the bivariate system (4.21), (4.25). If (*a*) $p = 1$, the system is exactly identified. One can estimate by OLS. Inference may proceed in standard fashion even if S_t has a unit root, so that H_t and S_t are cointegrated (West, 1988a; Sims, Stock, and Watson, 1990), subject to caveats discussed in section 6.

If (*b*) $p > 1$, the system is overidentified. Consider first (b)(i) a stationary model ($S_t \sim I(0) \Rightarrow H_t \sim I(0)$). Estimation may proceed by maximum likelihood, imposing the nonlinear overidentifying restrictions (Blanchard, 1983). Since the variance-covariance matrix of the disturbances in (4.21), (4.25) is unrestricted, an asymptotically equivalent procedure is nonlinear three stage least squares, which may be computationally simpler (Amemiya, 1977).

To my knowledge, the formal asymptotic theory has not been completely worked out for restricted estimates of the model in the case that (b)(ii) $S_t \Rightarrow I(1)$ ($\Rightarrow H_t$, S_t cointegrated) and $p > 1$ so that the system is overidentified (the complication results from the nonlinear cross-equation restrictions). See Gregory, Pagan, and Smith (1992) for discussion of estimation in a related model.

For stationary models, Hansen and Sargent (1982) suggest another estimator that is applicable if there are variables observed by the firm but not the economist that help predict S_t (say, reports from sales representatives about deals likely to close the next period). The idea is to estimate simultaneously the first order condition (4.3), the time series process for predicting S_t (the analogue to (4.21)),[10] and the reduced form for H_t. (In the example above, in which there is no such private information, this appears to yield no gains relative to conventional full information estimation as just described.) As far as I know, this estimator has yet to be applied.

For the discussion of empirical work below, it is useful to note the decision rules implied by two other specifications. First, if we generalize the assumption maintained so far that u_t is serially uncorrelated in favor of the assumption that it follows an exogenous AR(1),

$$u_t = \theta u_{t-1} + \varepsilon_{ut}, \tag{4.28}$$

then the reduced form and decision rule are as above, but with

$$v_{Ht} = (\rho_2/a_0)(1 - b\theta\rho_1 - b^2\theta\rho_2)^{-1}u_t.$$

If $\theta = 1$, u_t is a random walk. If $S_t \sim I(1)$ as well, one can difference (4.21) and (4.26) and proceed as described above.

Second, if $a_0 = 0$, $a_1 \neq 0$, the reduced forms and decision rules are

$$H_t = \rho H_{t-1} + \pi_1 S_{t-1} + \ldots + \pi_p S_{t-p} + v_{1t}, \tag{4.29}$$

$$\rho \text{ is the smaller root of } ba_1x^2 - (a_1 + ba_1 + ba_2)x + a_1 = 0,$$

$$(\pi_1, \ldots, \pi_p) = e'(I - b\rho\Phi)^{-1}[b\rho(1 + a_1^{-1}a_2a_3)\Phi^2 - \rho\Phi],$$

$$v_{1t} = (\pi_p/\phi_p)v_{2t} + v_{Ht},$$

$$v_{Ht} = -(\rho/a_1)(1 - b\theta\rho)^{-1}u_t,$$

$$H_t = \rho H_{t-1} + \delta_1 S_t + \ldots + \delta_p S_{t-p+1} + v_{Ht} \tag{4.30}$$

$$\delta_1 = \pi_p/\phi_p, \qquad \delta_i = \pi_{i-1} - \delta_1\phi_{i-1} \quad (i = 2, \ldots, p).$$

8 Comparison of Full and Limited Information Estimation

The limited information techniques described in sections 3–6 are less efficient but more robust than are the techniques described in section 7. That they are more robust is illustrated by the following example. Suppose that sales are not exogenous but are determined by the intersection of a demand curve and supply. The demand curve might be

$$S_t = -(1/\alpha)p_t + \text{demand shock}, \tag{4.31}$$

where $\alpha > 0$ and p_t is defined in (4.1); a supply curve is obtained from a first order condition obtained by differentiating (4.1) with respect to S_t and/or p_t. The assumption maintained above that sales are exogenous is a special case of (4.31) resulting when $\alpha \to \infty$, so that S_t = demand shock (Kollintzas, 1989).

We have seen in (4.25) and (4.26) that when $\alpha = \infty$, the reduced form of the model is a bivariate vector autoregression in S_t and H_t, and it may be shown that this holds even when $\alpha < \infty$. But if $\alpha < \infty$, S_t will be Granger-caused by H_t. The intuition in, say, a competitive market is that decisions about a firm's sales and inventories will be influenced by a comparison of this period's and next period's expected price, with the firm putting more in inventories the higher it expects next period's price to be (*ceteris paribus*). In equilibrium, then, industry wide H_t will help predict next period's price and thus next period's sales as well.

As a result, when the present value on the right hand side of (4.23) is projected onto past H_ts and S_ts, the H_ts will get nonzero coefficients. This means that the coefficients on the lagged H_ts in the reduced form or decision rule will not be related to the underlying cost parameters in the fashion given in (4.27), because these coefficients will reflect in part the ability of the H_t to predict future S_ts.[11] Thus, use of (4.27) will result in inconsistent estimates, while the instrumental variables estimation described in previous sections will still be consistent.

To see whether this might be a substantial problem in practice, I solved for the population values of ρ_1 and ρ_2 for a given set of values of a_0/c, a_1/c, a_2/c, and a_3, and for a range of αs; for each value of α, I then computed the values of a_0/c, a_1/c, and a_2/c that are implied by (4.27). Since use of (4.27) is

Table 4.4 Asymptotic bias of full information estimator, when sales are endogenous

	1	2	3	4	5	6
				Value implied by (4.27) of:		
	α	ρ_1	ρ_2	$a_0/2$	a_1/c	a_2/c
1	0.05	0.44	−0.09	0.07	0.11	0.36
2	0.19	0.73	−0.23	0.14	−0.01	0.16
3	0.32	0.81	−0.27	0.16	−0.02	0.12
True values:				0.17	−0.04	0.07

(a) a_0, a_1, and a_2 are defined in (4.1); c is defined in (4.3); α is defined in (4.31).
(b) Columns 2 and 3 present the coefficients in the decision rule for H_t when α is as indicated, a_0/c, a_1/c, and a_2/c are as given in the "True values" row, and $a_3 = -0.04$. These are the approximate values of the a_is estimated for aggregate inventories in West (1990, table III, row 1), when the demand shock in equation (4.31) follows a random walk. West's (1990) estimated value for α is that given in row 2, with the values in rows 1 and 3 being West's upper and lower bounds of the 95% confidence interval for α.
(c) Columns 4–6 give the values of a_0/c, a_1/c, and a_2/c that are implied by ρ_1 and ρ_2 under the incorrect presumption that $\alpha = \infty$ and sales are exogenous.
(d) To facilitate reading the table, only two digits are given. The values of α and the a_is actually used in the calculation are the three digit values given in West (1990).
(e) These asymptotic calculations may not accurately predict the actual finite sample performance of the full information estimator.

appropriate for $\alpha = \infty$, the question is whether plausible values of α are large enough that use of (4.27) results in little bias even though it is technically inappropriate.

I chose a data generating process consistent with one of the sets of estimates in West (1990).[12] Table 4.4 lists the values used. The table indicates that in this example, there is a plausible range of α for which one might be seriously misled by assuming that sales are exogenous ($\alpha = \infty$) when in fact sales are endogenous ($\alpha < \infty$). Row 1 of the table indicates that if $\alpha = 0.05$ (a value that is plausible in the sense that it falls within the 95% confidence interval for α given by West, 1990), use of estimates of H_t's reduced form would yield a value of a_1/c that not only is positive but is larger than that of a_0/c, at least in an arbitrarily large sample; in truth, however, a_1/c is negative and smaller in absolute value than a_0/c. Other values of α (rows 2 and 3) yield smaller biases.

Of course, full information estimation is still viable when $\alpha < \infty$; one must simply estimate α along with the cost function parameters (see West, 1990 for specifics). But then such information will yield inconsistent estimates if, say, there are costs of adjustment so that lags of S_t appear on the right hand side of (4.31), or if prices of competing products appear in the demand curve. The point is that limited information estimation is robust to possible misspecification of the demand curve, whereas full information estimation is not.

On the other hand, the full information technique will be more efficient, if the specification of demand is correct. For the three DGPs considered in tables 4.2 and 4.3, table 4.5 compares the full information technique with the limited information one that optimally uses two lags each of H_t and S_t as instruments

Table 4.5 Comparison of asymptotic variance-covariance matrices of limited and full information estimators

1	*2*	*3*	*4*
DGP	*Ratio of SEs on* $[(1+b)a_0 + a_1]/c$	*Ratio of SEs on* a_2/c	*Ratio of SEs on* a_3
E	0.783	0.815	0.921
R	0.832	0.834	0.988
W	0.501	0.372	0.394

(a) In the ratios referenced in columns 2–4, the numerator is computed from the variance-covariance matrix of the full information estimator that estimates (4.21) and (4.25) jointly, and the denominator from that of the limited information estimator (4.5) that uses as instruments the set of variables appearing in the reduced form (4.21) and (4.25). The ratios must lie between 0 and 1, smaller numbers indicating a greater efficiency gain from using the full information estimator. Limited information use of lags of H_t and S_t beyond those in the reduced form would result in smaller efficiency gains.

(b) These asymptotic comparisons may not accurately predict the actual finite sample performance of the two estimators.

(i.e. estimates as in (4.5), with $Z_t = [H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}]'$). Table 4.5 indicates that the gains from using full information are modest for DGPs E and R, with standard errors typically falling by less than 20%. The gains for the W DGP are, however, dramatic. Recall that the W DGP is the only one that generates substantial serial correlation in H_t. These results are consistent with those of West (1986b), who, using a different linear rational expectations model, found that the gains for full information estimation were large only for highly serially correlated data. Recall that the efficiency gains would be smaller if the limited information estimator used more lags of H_t and S_t; preliminary calculations in West and Wilcox (1993a) indicate that the use of additional lags sometimes but not always results in an estimator nearly as efficient as full information.

In sum, there is a tradeoff familiar from the literature on the traditional linear simultaneous equations models. While the full information method can yield substantial efficiency gains, it requires specification of demand, and there is much less consensus on the form of demand than of the cost function. In contrast to the traditional literature, however, full information estimation is much more complex computationally, since it requires nonlinear estimation. Recent literature on the linear quadratic model has tended to emphasize use of limited information techniques.

9 Empirical Evidence

9.1 Decision Rules Implied by the Linear Quadratic Model

For the benefit of readers who skipped section 7, as well as those who tried but could not stay awake through that section, I begin by summarizing the decision rules implied by the model (4.1) when sales follow an exogenous AR(p), $p \geqslant 2$, and the cost shock u_t follows an AR(1) with parameter θ (possibly with $\theta = 0$, so that u_t is serially uncorrelated):

$$S_t = \phi_1 S_{t-1} + \ldots + \phi_p S_{t-p} + v_{2t}, \tag{4.21}$$

$$u_t = \theta u_{t-1} + \varepsilon_{ut}. \tag{4.28}$$

The equations are repeated from section 7, as are the following decision rules:

$$H_t = \rho_1 H_{t-1} + \rho_2 H_{t-2} + \delta_1 S_t + \ldots + \delta_p S_{t-p+1} + v_{Ht} \qquad (a_0 \neq 0), \tag{4.26}$$

$$H_t = \rho H_{t-1} + \delta_1 S_t + \ldots + \delta_p S_{t-p+1} + v_{Ht}, \qquad (a_0 = 0). \tag{4.30}$$

As detailed in section 7, the parameters ρ_1, ρ_2, and ρ are functions of b, a_0, a_1, and a_2; the δ_is are functions of b, the a_is, and the ϕ_is; the disturbance v_{Ht} is AR(1) with parameter $\theta(\theta = 0 \Rightarrow v_{Ht}$ is serially uncorrelated).

Note one implication of the cost shock being serially correlated (of $\theta \neq 0$). If one multiplies (4.30) by $(1 - \theta L)$ and rearranges, H_{t-2} appears on the right hand side (as does S_{t-p}, though that is not important): serially correlated cost shocks, as well as nonzero costs of adjustment ($a_0 \neq 0$, see (4.26)), put a second lag of H_t in the decision rule and reduced form.

9.2 Unrestricted Estimates of the Decision Rule

Since Lovell's (1961) pioneering research, empirical work in inventories has been dominated by unrestricted estimates of equations like (4.26) or (4.30) (that is, estimates that are not restricted to accord with estimates of an equation to forecast sales, such as (4.21)). For the sake of completeness, I briefly sketch the "flexible accelerator" model that Lovell and others have used to rationalize the equation estimated, and then note a couple of stylized facts about estimation results.

The model supposes that the representative firm balances costs of adjusting inventories against costs of having inventories deviate from their frictionless target level H_t^*:[13]

$$\min 0.5(H_t - H_t^*)^2 + 0.5w(H_t - H_{t-1})^2 + u_t H_t; \quad H_t^* = \alpha E_t S_{t+1}. \tag{4.32}$$

In (4.32), $w > 0$ is the weight of the second cost relative to the first, $\alpha > 0$ is a parameter, and u_t is an unobservable disturbance that follows an AR(1) with parameter θ, possibly with $\theta = 0$ (as in (4.28)). The first order condition is

$$H_t = \rho H_{t-1} + (1 - \rho)\alpha E_t S_{t+1} + v_{Ht}, \tag{4.33}$$

$$v_{Ht} \equiv -(1 - \rho)u_t; \quad 0 < \rho \equiv w/(1 + w) < 1.$$

If, as in (4.21), S_t is modeled as evolving according to an exogenous autoregression of order p, so that $E_t S_{t+1} = \phi_1 S_t + \ldots + \phi_p S_{t-p+1}$, (4.33) may be put in estimable form as

$$H_t = \rho H_{t-1} + \delta_1 S_t + \ldots + \delta_p S_{t-p+1} + v_{Ht}, \quad \delta_i = (1 - \rho)\alpha\phi_i. \tag{4.34}$$

It will be recognized that (4.34) is the same as (4.30), although the two models do predict different relationships between (1) the δ_is and the ϕ_is, and (2)ρ and underlying cost parameters.

Blinder and Maccini (1991a; 1991b) have discussed some stylized facts about unrestricted estimates of (4.34), two of which bear repeating here. First, even conditional on current and lagged S_t there is considerable serial correlation in H_t, in that estimates of ρ and/or θ (the serial correlation parameter of u_t) tend to be near one. Moreover, both ρ and θ tend to be significantly different from zero. Second, the estimate of δ_1 tends to be positive. This regression result reflects the business cycle fact that inventories move procyclically, tending to be accumulated during business cycle expansions as S_t and Q_t rise, and to be decumulated during recessions as S_t and Q_t fall. Given the inventory identity $Q_t = S_t + \Delta H_t$, the positive correlation between S_t and ΔH_t produces the well known result that Q_t is more variable than S_t.

The next two sections consider how such facts might be explained by the model (4.1), and briefly surveys the empirical evidence, focusing largely on papers that have explicitly used (4.1).

9.3 Explaining Extreme Serial Correlation

The extreme serial correlation of H_t that is typically observed can be rationalized in either of two ways. First, this fact will follow if u_t is highly serially correlated. Second, irrespective of the serial correlation of u_t, it will follow if ρ_1 and ρ_2 (see (4.26)) are such that the larger root of $x^2 - \rho_1 x - \rho_2$ is near unity, or, when $a_0 = 0$, if ρ (see (4.30)) is near unity.

The model will yield such a root, or yield $\rho \approx 1$ when $a_0 = 0$, when the marginal inventory holding cost a_2 is small relative to the slopes of marginal costs of production a_1 and of changing production a_0. For example, when $a_0 = 0$, $\rho \to 1$ as $a_2/a_1 \to 0$; more generally, the larger root of $x^2 - \rho_1 x - \rho_2$ approaches 1 as $a_2 \to 0$ for any fixed positive values of a_1 and a_0.

Recent empirical estimates of (4.1) have given some support to both the cost shock and cost parameter explanations. Using two digit US manufacturing data, and estimating a Euler equation such as (4.3), Eichenbaum (1989) and Ramey (1991) found a_is that by themselves implied little serial correlation, but very high serial correlation of an unobservable cost shock, while West (1986a) found a_is that implied high serial correlation (but did not test for serial correlated cost shocks). Blanchard's (1983) full information estimation, applied to automobile data, got results similar to West's.

Distinguishing between the two explanations may be difficult. Recall from the discussion in section 9.1 that if both ρ and θ are nonzero, when (4.30) is transformed to have a serially uncorrelated disturbance the resulting decision rule will have H_{t-2} on the right hand side. It will therefore look similar to the decision rule (4.26), which was derived assuming $a_0 \neq 0$ and a serially uncorrelated cost shock. What is involved, then, is distinguishing between costs of adjustment and serial correlation, which is not easily done (Blinder, 1986b; McManus, Nankervis, and Savin, 1992). Below I discuss these explanations further.

A final point to be made at present concerns the plausibility of explaining the serial correlation with a relatively flat curve describing marginal inventory holding costs. If a model such as (4.32) is used to interpret estimates of (4.30) or (4.34), $\rho \approx 1$ implies that the percentage of the gap between H_t and H_t^* that is closed each period is small (e.g. $\rho = 0.8 \Rightarrow 20\%$ closed). Following Carlson and Wehrs (1974) and Feldstein and Auerbach (1976), many find this implausible on the grounds that monthly and even quarterly changes in inventory stocks rarely amount to more than a few days production. But in the context of a model such as (4.1), a finding that $\rho \approx 1$, or that $x^2 - \rho_1 x - \rho_2$ has a large root, does not seem to me to be prima facie implausible, at least in the absence of any independent evidence about how fast marginal production costs increase relative to marginal inventory holding costs.

9.4 Explaining Procyclical Inventory Movements

The procyclical character of inventory movements may be rationalized by the model in at least three ways, which are not mutually exclusive.[14] Before discussing empirical evidence, I sketch the logic of the three explanations. Formal proofs using variance bounds inequalities such as (4.19) may be found in West (1986a; 1988b; 1990).

The simplest (and, peculiarly, sometimes overlooked) explanation is simply that this is a result of the accelerator term $a_2(H_{t-1} - a_3 S_t)^2$. This term captures a tradeoff between inventory holding costs on the one hand and stockout or backlog costs on the other. More inventories means higher holding costs but lower probability of stocking out, with the level of H_t that balances the two competing costs increasing in expected sales. In an extreme case in which there were no production costs ($a_0 = a_1 = u_t = 0$), the firm would simply set $H_t = a_3 E_t S_{t+1}$: the more customers expected to walk in the door next period, the larger the inventory stock. In this case, positive serial correlation in S_t will cause H_t and ΔH_t to track actual as well as expected sales, and inventories clearly will move procyclically. And even if nonzero $a_0(\,)$, $a_1(\,)$, and u_t terms induce countercyclical movements (see below), as long as the influence of such terms is small enough relative to that of the accelerator term, H_t will move procyclically.

To understand the other two explanations, it is useful to first consider a set of circumstances under which H_t would *not* move procyclically. Suppose now that $a_0 = a_3 = u_t = 0$, a_1, $a_2 > 0$. Then with increasing marginal costs of production ($a_1 > 0$), no accelerator motive ($a_3 = 0$), and costs nonstochastic ($u_t = 0$), firms use inventories to smooth production in the face of randomly fluctuating sales: they build up inventories when sales are low, draw them down when sales are high.

One of the two remaining explanations for procyclical inventory movements emphasizes the possible role of stochastic movements in costs. Now allow for $u_t \neq 0$, but, for clarity, continue to assume $a_0 = a_3 = 0$. Firms will intertemporally substitute production out of periods in which u_t is high into periods in which u_t is low, drawing down H_t when u_t is high, building them up when u_t is low.[15] This will produce a tendency for ΔH_t and Q_t to move in the same direction; if this tendency is strong enough relative to the one described in the preceding paragraph, inventories may move procyclically, and certainly will if movements in S_t are also driven by u_t (as is suggested by real business cycle models).

The third explanation is that marginal production cost slopes down. For simplicity, set $u_t = a_3 = 0$. Assume for the moment that $0 > \partial^2 c_t / \partial Q_t^2 = (1 + b)a_0 + a_1$. Then, in contrast to the previous paragraph but one, firms will use inventories not to smooth but to bunch production, producing high output in periods when sales are high to exploit the diminished costs that come from high output levels, and producing low output in periods when sales are low. As Ramey (1991) has emphasized, inventories will move procyclically. If one instead makes the milder assumption that $a_1 < 0$ but $(1 + b)a_0 + a_1 > 0$, so that marginal production cost slopes down only when one abstracts from costs of adjusting production, there will still be a tendency for inventories to move procyclically; whether they do or not depends on whether the motive to smooth ($(1 + b)a_0 + a_1 > 0$) or bunch ($a_1 < 0$) is stronger.

In sum, the procyclical movement of inventories suggests a substantial role for the $a_2(H_{t-1} - a_3S_t)^2$ term, or for cost shocks, or for downward sloping marginal costs, or indeed for more than one of these. The evidence on each of these is mixed. Blanchard (1983) and Ramey (1991) find estimates of a_3 that are positive and significant at conventional levels; West (1986a) and Krane and Braun (1991) find estimates that usually are positive but rarely are significant; Kashyap and Wilcox (1993) and West (1990) find estimates that are of mixed sign and usually are insignificant.[16] Perhaps an indirect indication of the importance of the stockout or backlog costs underlying $a_3 > 0$ is that measures of order backlogs often are significant in inventory regressions (e.g. Maccini and Rossana, 1984; Blinder, 1986a).

Consider now the possibility that marginal production costs slope down, in the sense that $(1 + b)a_0 + a_1 < 0$. Ramey (1991) vigorously argues that this is the case. Most others, including some who have used similar data (Eichenbaum,

Table 4.6 Statistical significance of cost variables

	Wage	*Materials prices*	*Energy prices*	*Interest rate*	*Unobservable shock*
1 Blinder (1986b)	?	?		n	?
2 Durlauf and Maccini (1992)	?	n	n		
3 Eichenbaum (1989)					y
4 Maccini and Rossana (1981)	y	?		n	?
5 Maccini and Rossana (1984)	n	y		n	y
6 Miron and Zeldes (1988)	n	?	n	n	
7 Ramey (1991)	n	?	n		y

(a) All the studies used two digit manufacturing data from the US. The exact data, sample period, specification, and estimation technique vary from paper to paper.
(b) A "y" entry indicates that the variable in a given column was significantly different from zero at the 5% level in at least three-fourths of the data sets in a given study; an "n" that it was significant in at most one-fourth of the data sets; and a "?" that it was significant in more than one-fourth but fewer than three-fourths of the data sets. A blank indicates that the variable was not examined.
(c) Line by line sources: (1) table 1 (pp. 360–1) (2) table 3, inst. set 4 (3) table 2 (p. 861) (4) table 1 (p. 20) (5) table 3 (p. 231) and discussion on p. 227 (6) table II (p. 892) (7) table 1 (p. 323).

1989; West, 1986a) have come to the opposite conclusion; a possible exception is Krane and Braun (1991, pp. 574–5), one of whose specifications yielded insignificant but negative slopes in about half their data sets. However, a number of authors have found the production cost a_1 insignificantly different from zero (Blanchard, 1983; West, 1990; Kashyap and Wilcox, 1993); in at least one study (West, 1990), the negative (but insignificant) point estimate of a_1 was large enough in absolute value to imply procyclical inventory movements.

Finally, with reference to unobservable cost shocks, there is a persistent tendency for unobservable cost shocks to be highly autocorrelated when one allows for such a possibility, as do Eichenbaum (1989), West (1990), and Ramey (1991) (but not, for example, Blanchard, 1983; or West, 1986b). One would hope that such shocks are crude proxies for observable measures of costs such as factor prices. Unfortunately, this seems not to be the case, since, in practice, such factor prices rarely are significant. See table 4.6, which summarizes some results from both flexible accelerator and linear quadratic models that have been applied to two digit manufacturing data from the US.

9.5 Summary

Highly serially correlated cost shocks rationalize both the considerable serial correlation in H_t and the procyclical nature of movements in H_t. But as Blinder and Maccini (1991b) note, one cannot be very confident that such disturbances in fact capture stochastic variation in costs rather than model misspecification, given that observable measures of costs do not seem to influence inventory movements very much.

Alternatively, both stylized facts fall out of demand driven models if the inventory holding cost a_0 is small relative to the production costs a_0 and/or a_1, and either (1) a_1 is slightly negative, or (2) the effects of the accelerator term $a_2(H_{t-1} - a_3S_t)^2$ are large. Note that it is the *product* a_2a_3 that determines the strength of the accelerator term (see (4.3), (4.19), and the definition of the w_2 in (4.24)). So there may be a tension between arguing that a_2 is small (to obtain extreme serial correlation) and that a_2a_3 is large (to obtain procyclical movements), although given arbitrarily small a_2 there is an a_3 sufficiently large that the accelerator will yield arbitrarily strong effects. In any case, existing estimates do not come to a tight consensus about the magnitudes of these parameters. Perhaps microeconomic evidence from a more disaggregate level, such as in Bresnahan and Ramey (1991), will help narrow the range of plausible a_is, as well as help sharpen our understanding of the role of stochastic variation in costs in determining inventory movements.

Notes

I thank an anonymous referee, participants in a seminar at the Federal Reserve Board of Governors, Louis Maccini, Scott Schuh, and, especially, David Wilcox for helpful comments and discussions, and the National Science Foundation, the Sloan Foundation, and the Graduate School at the University of Wisconsin for financial support.

1 Formally, one adds to the cost function terms of the form (1)$c_1' w_t Q_t$, where c_1 is a parameter vector and w_t is a vector of input prices, and/or (2) $c_2 r_t H_t$, where r_t is the *ex ante* real interest rate.

2 An implicit assumption made is that one can parameterize the problem so that $\partial(p_{t+j}S_{t+j})/\partial H_t = 0$ for all j. Reasons why this might not be possible: (1) a nonnegativity constraint on inventories sometimes prevents the firm from holding revenue constant when it produces one fewer unit (Abel, 1985; Kahn, 1987); (2) in an imperfectly competitive industry, this period's level of inventories affects future revenues, for strategic reasons (Rotemberg and Saloner, 1989).

3 If data on output price p_t are available, an additional first order condition allows one to identify the a_is relative to the units in which p_t is measured (presumably, constant dollars, for US data) and not just relative to one another. But for simplicity of exposition, and for conformity with much empirical work, I focus on the case when such data are not used.

4 Hall (1992), Ogaki (1992), and Pesaran (1987) provide general discussions of the issues raised in these sections.

5 This implicitly assumes that $\partial(p_{t+j}S_{t+j})/\partial H_t = 0$ for all j and that Q_t is not a choice variable. This is a harmless assumption in that, in most of the literature, it is a matter of convenience as to whether one makes (1) H_t and S_t, or (2) H_t and Q_t the two choice variables. But see note 2 for some conditions under which one might not be able to set up the problem this way.

6 Nor is GLS correction for serial correlation generally desirable. The instruments generally (although not always) include lags of H_t and/or S_t, in which case application of a standard GLS transformation to eliminate serial correlation may yield inconsistent parameter estimates (Hayashi and Sims, 1983).

7 Eichenbaum sets $a_0 = 0$ and reports parameters that he denotes λ and α. I mapped his estimates into the present notation using $a_2/a_1 = \lambda + (1/\lambda) - 1 - (1/b)$, with $b = 0.995$, $a_3 = 2(1/\alpha)/(a_2/a_1)$. For computational convenience, I then set a_0 to a small nonzero number rather than zero. Eichenbaum also assumes that C_t includes a term of the form (in my notation) $(H_t - a_3 S_t)^2$ rather than $(H_{t-1} - a_3 S_t)^2$; I slur over this minor difference.

8 Just as $\tilde{\Gamma}_0 + (\tilde{\Gamma}_1 + \tilde{\Gamma}_1')$ might not be positive definite, $\hat{s}^{(0)}$ may be negative (as, of course, might $\hat{s}^{(1)}$ or even the underlying population quantity $s^{(1)}$ (but not $s^{(0)}$). But the fact that the ratio $\hat{s}^{(1)}/\hat{s}^{(0)}$ is squared in (4.13) means that m will be nonegative, and the resulting $\hat{\Omega}$ will be positive definite.

9 That $\{.\} \sim I(0)$ follows since $E_t\{.\} = -u_t \sim I(0)$ by assumption and $\{.\} - E_t\{.\} \sim I(0)$ as well. Further, $\{.\} \sim I(0) \Rightarrow H_t \sim I(1)$: if $H_t \sim I(d)$ for some $d > 1$, then the order of integration of H_t would be greater than that of any other variable and $\{.\}$ could not be $I(0)$, while if $H_t \sim I(0)$, $\{.\} \sim I(1)$. Incidentally, Q_t and S_t are "multicointegrated," in the terminology of Granger and Lee (1989; 1991).

10 Under these circumstances, equation (4.21) will not describe the time series process for S_t. For in this case H_t will Granger-cause S_t relative to an information set consisting of lagged H_ts and S_ts, and lagged H_ts will appear on the right hand side of S_t's time series process.

11 Let $E_t Y_t \equiv E_t \Sigma_{j=0}^{\infty} \{[(b\lambda_1)^{j+1} - (b\lambda_2)^{j+1}] D_{t+j}\}$ denote the present value on the right hand side of (4.23). Note that the problem described in the text is not trivially

circumvented by simply projecting Y_t onto past S_ts, and absorbing the difference between this projection and $E_t Y_t$ in the error term: the error term would then be correlated with the lagged H_ts.

12 In this example, the model is exactly identified, so full information estimation simply involves estimating the unrestricted reduced form. See West (1990). Overidentification would result if the demand curve were e.g. $S_t = -(1/\alpha)p_t + \phi_1 S_{t-1} + \phi_2 S_{t-2}$ + demand shock, with $\phi_1, \phi_2 \neq 0$.

13 As a rule, flexible accelerator studies measure inventories and sales in logs rather than levels. I ignore distinctions between logs and levels because results for the linear quadratic model do not change much when one rids the data of exponential growth before estimation (West, 1988b; 1990).

14 I consider here the *un*conditional relationship between inventories on the one hand and sales and production on the other; I will be deliberately vague about whether the analysis relates to levels or differences since I believe that neither the stylized facts nor their interpretation turns on how stationarity is induced. See Blinder (1986a) for discussion of a conditional relationship, specifically, an analysis of what makes the regression coefficient δ_1 positive; see Krane (1993) for evidence on the similar pattern that applies in the deterministic seasonal relationship between inventories and sales/production.

15 There is a sense in which inventories will move procyclically even if u_t follows a random walk and there are no possibilities for intertemporal substitution; see West (1990).

16 One interpretation of the disparity in estimates is that stockout or backlog costs indeed are central to explaining why inventories move procyclically, but are poorly modeled by a simple quadratic term like $a_2(H_{t-1} - a_3 S_t)^2$ (Krane, 1991). See Kahn (1987; 1992) for excellent work modeling such costs in a more sophisticated way.

References

Abel, Andrew B. (1985), "Inventories, Stock-Outs and Production Smoothing," *Review of Economic Studies* 52, 283–93.

Amemiya, Takeshi (1977), "The Maximum Likelihood and the Nonlinear Three Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model," *Econometrica* 45, 955–68.

Andrews, Donald W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817–58.

Andrews, Donald W.K., and Christopher J. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 60, 953–66.

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

Blanchard, Olivier J., and Angelo Melino (1986), "The Cyclical Behavior of Prices and Quantities: The Case of the Automobile Market," *Journal of Monetary Economics* 17, 379–408.

Blinder, Alan S. (1981), "Retail Inventory Behavior and Business Fluctuations," *Brookings Papers on Economic Activity* 2, 443–505.

Blinder, Alan S. (1986a), "Can the Production Smoothing Model of Inventories Be Saved?", *Quarterly Journal of Economics* CI, 431–53.

Blinder, Alan S. (1986b), "More on the Speed of Adjustment in Inventory Models," *Journal of Money Credit and Banking* 18, 355–65.

Blinder, Alan S. and Louis J. Maccini (1991a), "The Resurgence of Inventory Research: What Have We Learned?", *Journal of Economic Surveys* 5, 291–328.

Blinder, Alan S. and Louis J. Maccini (1991b), "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives* 5, 73–96.

Bresnahan, Timothy F. and Valerie A. Ramey (1991), "The Nature of Output Fluctuations: Evidence from Plant Level Data," manuscript, University of California at San Diego.

Carlson, John A. and William Wehrs (1974), "Aggregate Inventory Behavior," in G. Horwich and P. Samuelson (eds) *Trade, Stability and Growth: Essays in Honor of Lloyd A. Metzler*, New York: Academic Press.

Christiano, Lawrence J. (1988), "Why Does Inventory Investment Fluctuate So Much?", *Journal of Monetary Economics* 21, 247–80.

Christiano, Lawrence J., and Martin S. Eichenbaum (1989), "Temporal Aggregation and the Stock Adjustment Model of Inventories," 70–109 in T. Kollintzas (ed.) *The Rational Expectations Inventory Model*, New York: Springer Verlag.

Cumby, Robert E. and John Huizanga (1990), "Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions," NBER Technical Working Paper no. 92.

Cumby, Robert E., John Huizanga and Maurice Obstfeld (1983), "Two Step, Two Stage Least Squares Estimation in Models with Rational Expectations," *Journal of Econometrics* 21, 333–55.

Dimelis, Sophia P., and Moheb A. Ghali (1992), "Classical and Variance Bounds Tests of Production and Inventory Behavior," manuscript, Athens University of Economics and Business.

Dimelis, Sophia P., and Tryphon Kollintzas (1989), "A Linear Rational Expectations Equilibrium Model for the American Petroleum Industry," 110–98 in T. Kollintzas (ed.) *The Rational Expectations Inventory Model*, New York: Springer Verlag.

Durlauf, Steven N. and Louis J. Maccini (1992), "Measuring Noise in Inventory Models," manuscript, The Johns Hopkins University.

Eichenbaum, Martin S. (1984), "Rational Expectations and the Smoothing Properties of Finished Goods," *Journal of Monetary Economics* 14, 71–96.

Eichenbaum, Martin S. (1989), "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," *American Economic Review* 79, 853–64.

Eichenbaum, Martin S., Lars Peter Hansen and Kenneth J. Singleton (1988), "A Time Series Analysis of A Representative Agent Model of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 51–78.

Feldstein, Martin S. and Alan J. Auerbach (1976), "Inventory Behavior in Durable Goods Manufacturing: The Target Adjustment Model," *Brookings Papers on Economic Activity* 351–96.

Granger, Clive W.J. and T.H. Lee (1989), "Investigation of Production, Sales and Inventory Relationships Using Multicointegration and Non-symmetric Error Correction Models," *Journal of Applied Econometrics* 4, S145–S159.

Granger, Clive W.J. and T.H. Lee (1991), "Multicointegration," 179–90 in R. F. Engle and C.W.J. Granger (eds) *Long Run Economic Relationships: Readings in Cointegration*, Oxford: Oxford University Press.

Granger, Clive W.J. and Paul Newbold (1977), *Forecasting Economic Time Series*, New York: Academic Press.

Gregory, Allan W., Adrian R. Pagan and Gregor W. Smith (1992), "Estimating Linear Quadratic Models with Integrated Processes," manuscript, Queens University.

Hall, Alistair (1992), "Some Aspects of Generalized Method of Moments Estimation," manuscript, North Carolina State University.

Hansen, Bruce (1992), "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica* 60, 967–72.

Hansen, Lars Peter (1982), "Asymptotic Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029–54.

Hansen, Lars Peter (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics* 30, 203–38.

Hansen, Lars Peter, and Thomas J. Sargent (1982), "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models,: *Journal of Monetary Economics* 9, 263–96.

Hayashi, Fumio, and Christopher A. Sims (1983), "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments," *Econometrica* 51, 783–98.

Holt, Charles C., Franco Modigliani, John F. Muth and Herbert A. Simon (1960), *Planning Production, Inventories and Work Force*, Englewood Cliffs, NJ: Prentice-Hall.

Kahn, James A. (1987), "Inventories and the Volatility of Production," *American Economic Review* 77, 667–80.

Kahn, James A. (1992), "Why is Production More Volatile than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory Holding," *Quarterly Journal of Economics* CVII, 481–510.

Kashyap, Anil K. and David W. Wilcox (1993), "Production and Inventory Control at the General Motors Corporation during the 1920s and 1930s," *American Economic Review* 83, 383–401.

Kollintzas, Tryphon A. (1989), "The Linear Rational Expectations Inventory Model: An Introduction," 1–33 in T. Kollintzas (ed.) *The Rational Expectations Inventory Model*, New York: Springer Verlag.

Kollintzas, Tryphon A. (1992), "A Generalized Variance Bounds Test, with an Application to the Holt et al. Inventory Model," manuscript, Athens University of Economics and Business.

Krane, Spencer D. (1991), "The Distinction between Inventory Holding and Stockout Costs: Implications for Target Inventories, Asymmetric Adjustment, and the Effect of Aggregation on Production Smoothing," manuscript, Federal Reserve Board of Governors.

Krane, Spencer D. (1993), "Induced Seasonality and Production Smoothing Models of Inventory Behavior," forthcoming, *Journal of Econometrics*.

Krane, Spencer D. and Steven N. Braun (1991), "Production Smoothing Evidence from Physical Product Data," *Journal of Political Economy* 99, 558–81.

Lovell, Michael C. (1961), "Manufacturers' Inventories, Sales Expectations, and the Acceleration Principle," *Econometrica* 29, 293–314.

Lovell, Michael C. (1993), "Simulating the Inventory Cycle," *Journal of Economic Behavior and Organization* 21, 147–79.

Maccini, Louis J. and Robert J. Rossana (1981), "Investment in Finished Goods Inventories: An Analysis of Adjustment Speeds," *American Economic Review* 71 (May), 17–22.

Maccini, Louis J. and Robert J. Rossana (1984), "Joint Production, Quasi-Fixed Factors of Production, and Investment in Finished Goods Inventories," *Journal of Money, Credit and Banking* XVI, 218–36.

McManus, Douglas A., John C. Nankervis and N.E. Savin (1992), "Multiple Optima and Asymptotic Approximations in the Partial Adjustment Model," University of Iowa Working Paper 92–10.

Miron, Jeffrey A. and Stephen P. Zeldes (1988), "Seasonality, Cost Shocks and the Production Smoothing Model of Inventories," *Econometrica* 56, 877–908.

Mosser, Patricia C. (1988), "Empirical Tests of the (S, s) Model for Merchant Wholesalers," 261–84 in A. Chikan and M. Lovell (eds) *The Economics of Inventory Management*, Amsterdam: Elsevier Science Publishers.

Nelson, Charles R. and Richard A. Startz (1990), "The Distribution of the Instrumental Variables Estimator and its *t*-ratio When the Instrument is a Poor One," *Journal of Business* 63, 125–40.

Newey, Whitney K., and Kenneth D. West (1987a), "A Simple, Positive Semidefinite. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 703–8.

Newey, Whitney K., and Kenneth D. West (1987b), "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review* 28, 777–86.

Newey, Whitney K., and Kenneth D. West (1993), "Automatic Lag Selection in Covariance Matrix Estimation," manuscript, University of Wisconsin.

Ogaki, Masao (1992), "An Introduction to Generalized Method of Moments," University of Rochester Center for Economic Research Working Paper no. 314.

Park, Joon Y. and Peter C.B. Phillips, (1988), "Statistical Inference in Regressions with Integrated Processes: I," *Econometric Theory* 4, 468–97.

Pesaran, M. Hashem, (1987), *The Limits to Rational Expectations*, New York: Basil Blackwell.

Ramey, Valerie A. (1989), "Inventories as Factors of Production and Economic Fluctuations," *American Economic Review* 79, 338–54.

Ramey, Valerie A. (1991), "Nonconvex Costs and the Behavior of Inventories," *Journal of Political Economy* 99, 306–34.

Rossana, Robert J. (1992), "The Long-Run Implications of the Production Smoothing Model of Inventories: An Empirical Test," manuscript, Wayne State University.

Rotemberg, Julio J., and Garth Saloner (1989), "The Cyclical Behavior of Strategic Inventories," *Quarterly Journal of Economics* CIV, 73–98.

Schuh, Scott (1993), "Linear-Quadratic Inventory Models and Aggregation," manuscript, Federal Reserve Board of Governors.

Shea, John (1993), "The Input–Output Approach to Demand-Shift Instrumental Variables Estimation," *Journal of Business and Economic Statistics*, 11, 145–55.

Sims, Christopher A., James H. Stock and Mark Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58, 113–44.

Stengel, Robert F. (1986), *Stochastic Optimal Control: Theory and Application*, New York: Wiley.

Stock, James H. and Kenneth D. West (1988), "Integrated Regressors and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics* 21, 85–95.

West, Kenneth D. (1986a), "A Variance Bounds Test of the Linear Quadratic Inventory Model," *Journal of Political Economy* 94, 374–401.

West, Kenneth D. (1986b), "Full Versus Limited Information Estimation of a Rational Expectations Model: Some Numerical Comparisons," *Journal of Econometrics* 33, 367–86.

West, Kenneth D. (1988b), "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica* 56, 1397–418.

West, Kenneth D. (1988b), "Order Backlogs and Production Smoothing," 305–18 in Attila Chikan and Michael Lovell (eds) *The Economics of Inventory Management*, Amsterdam: Elsevier.

West, Kenneth D. (1990), "The Sources of Fluctuations in Aggregate Inventories and GNP," *Quarterly Journal of Economics* CV, 939–72.

West, Kenneth D. (1992), "Sources of Cycles in Japan, 1975–1987," *Journal of the Japanese and International Economies* 6, 71–98.

West, Kenneth D., and David W. Wilcox (1993a), "Finite Sample Behavior of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," forthcoming, *Journal of Business and Economic Statistics.*

West, Kenneth D., and David W. Wilcox, (1993b), "Some Evidence on the Finite Sample Behavior of an Instrumental Variables Estimator of the Linear Quadratic Inventory Model," in R. Fiorito (ed.) *Inventory Cycles and Monetary Policy*, Berlin: Springer Verlag, 253–82.