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 inventory theory 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
Inventory Models: The Estimation of Euler Equations 189

include those raised by continuous time models e.g. Moser, 1988; Cristiano 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 Macchi (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, \( h \) a discount factor, \( 0 < h < 1 \), and \( E \) mathematical expectations conditional on information known at time \( t \). assumed equivalent to linear projections. The objective function, then, is

\[
\max_{Q_t} \lim_{t \to \infty} E_t \sum_{j=0}^{\infty} h^j (p_{t+j} S_{t+j} - C_{t+j})
\]

subject to

\[
Q_t = S_t + H_{t-1} - H_t
\]

\[
C_t = 0.5a_0 Q_t^2 + 0.5a_1 Q_t^3 + 0.5a_2 (H_{t-1} - a_3 S_t)^2 + u_{0t} Q_t + u_{0t} H_t.
\]

For the moment, \( a_0 \) and \( a_1 \) are assumed positive, \( a_2 \) 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 of the \( a_2 \) play in determining inventory behavior.

The scalars \( u_{0t} \) and \( u_{0t} \) 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 \(pS_t\) unchanged throughout. Formally, differentiating costs with respect to \(H_t\) gives\(^1\)

\[
E_t(a_t dQ_t - 2b_q dQ_{t-1} + b^2 dQ_{t-2}) + \alpha_1 Q_t - b Q_{t-1} + a_t S_t + a_1 = 0.
\]

(4.2)

\[u_t = u_{t-1} - bE_{t-1}u_{t-1} + u_0.\]

Note that \(a_1\), \(a_1\), 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.\(^1\)

These are four independent unknowns to be estimated: \(b\), \(a_1\), and the ratios of \((i)\) two of \(a_2\), \(a_3\), and \(a_2\) to \((ii)\) same linear combination of \(a_2\), \(a_3\), and \(a_4\). 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 \(a_2\) and two of \(a_3\), \(a_4\), and \(a_5\), the latter two identified only up to scale) may be estimated directly. A reasonable value of \(b\) comes from noting that with, say, monthly data, values for \(b\) of about 0.985 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 with a parametric assumption about the demand curve (e.g. does it depend on the price of competing products?) or market structure (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.\(^1\) In sections 3–5, I assume that \(S_t\) is I(1), 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 equilib-
3 Limited Information Estimation: Instrumental Variables

3.1 Introduction

This approach transforms (4.2) into an estimable equation and then uses instrumental 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 \(a_n, a_n, a_n\) that multiplies the left hand side variable is non-zero.

For concreteness, put \(-\partial^2/\partial H_i\partial H_j\)(H) = \[a_n(1 + b)+b\gamma]+a_n(1 + b)+ba_n\] H for the left hand side. The Legendre-Clebsch condition (a dynamic analog of the usual second order necessary condition; Stengel, 1986, p. 213) states that \(-\partial^2/\partial H_i\partial H_j\) > 0 and thus that this particular linear combination is nonzero. Then (4.2) may be rewritten

\[
H_i = (a/c)X_{t+1} + (a/c)X_{t+1} + (b/a/c)S_{t+1} + u_{t+1} \quad (4.3)
\]

\[
= X_t^\prime \beta + u_{t+1},
\]

\[
X_{t+1} = (aS_t - 2H_{t+1} + H_{t+1}) + 2b(aS_{t+1} + H_{t+1} + H_{t+1}) - b(aS_{t+1} + H_{t+1} + H_{t+1}).
\]

\[
X_{t+1} = -S_{t+1} + H_{t+1} + b(S_{t+1} + H_{t+1}).
\]

\[
u_{t+1} = \frac{a/c}{e_{t+1}},
\]

\[
e_{t+1} = -(a/c)(X_{t+1} - E_{X_{t+1}}) - (a/c)(X_{t+1} - E_{X_{t+1}}) - (b/a/c)S_{t+1} - E_{S_{t+1}}.
\]

\[
X_t = (X_{t+1} - X_{t+1}, S_{t+1})^\prime,
\]

\[
\beta = (\beta_1, \beta_2, \beta_3)^\prime = (a/c, a/c, ba/c)^\prime.
\]

\[
e = a_n(1 + b) + a_n(1 + b) + ba_n.
\]

\[
From an estimate of \(\beta\), one can recover estimates of \(a/c\) and \(a/c\) directly, and of \(a/c\) and \(a/c\) using \(a/c = (1 - \beta_2(1 + b) + \beta_2(1 + b))/\beta_2(1 + b)/a_2 = \beta_2(ba)/c\).
\]

The idea is to use a vector of instruments that is uncorrelated with \(v_{t+1}\), but correlated with \(X_t\), taking account of serial correlation of \(v_{t+1}\).
3.2 Optimal Linear Combination of Given Vector of Instruments

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

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

$$\hat{\theta} = \sum_{j=1}^{q} EZ_i Z_i' \beta_j v_{i,t},$$

$$\beta_j = E \theta Z_i' \beta_j v_{i,t},$$

(4.4)

$q$ is sometimes called the "long run" covariance matrix of $Z_i v_{i,t}$; procedures to obtain $\hat{\theta}$ are discussed in section 4.

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

$$\hat{\beta} = (X' Z_i \hat{\beta} Z' Z' Y)' X' Z_i \hat{\beta} Z' E_{T},$$

$$\hat{\beta} = \hat{\beta} - N(0, \hat{V}),$$

$$\hat{V} = \hat{V} (X' Z_i \hat{\beta} Z' Y)'^{-1}$$

(4.5)

Thus this estimator differs from the usual 2SLS one in that $\hat{\beta}$ 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 $Z_i v_{i,t}$, 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 $n = 0$, so that there are no unobservable cost disturbances (e.g. Kashyap and Wilcox, 1993), $v_{i,t} = \epsilon_{i,t}$, and the regression disturbance is a sum of expectation errors. Now, as we well known, under rational expectations the expectation errors $X_{it-1} - E_t X_{it}$ and $S_{i,t-1} - E_t S_{i,t}$ are
Inventory Models: The Estimation of Sales Equations

seriously uncorrelated. But \( X_{t-2} - E X_{t-1} \) depends on period \( t + 1 \) and period \( t + 2 \) information, and \( X_{t-2} - E X_{t-1} \), on period \( t + 2 \) and period \( t + 3 \) information. This implies an MA(1) structure to \( X_{t-2} - E X_{t-1} \) and thus to \( v_{t-2} \) as well. Since \( E v_{t-2} = 0 \), the vector \( Z_{t-2} \) is also MA(1), for \( j > 1 \), \( E(\Sigma_{t-2} Z_{t-j}, v_{t-j+1}) = E(\Sigma_{t-2} Z_{t-j}, v_{t-j+1}) = E(\Sigma_{t-2} Z_{t-j}, v_{t-j+1}) = 0 \). So in this case, \( B = \Sigma_{t-2} E Z_{t-j}, v_{t-j+1} = \tau + f(t) + f(t) \).

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 \( x_t \) and \( u_t \) are white noise (West and Wilcox, 1972b), \( X_{t-2} - E X_{t-1} \) will depend in part on \( u_{t-2} \) and \( u_{t-1} \), implying that \( v_{t-2} \), which depends on \( u_t \) as well, will be MA(2). The vector \( Z_{t-1} \) will be MA(2) as well, so \( Z_t = \Sigma_{t-1} E Z_{t-j}, v_{t-j+1} = \tau + f(t) + f(t) + f(t) \).

More generally, if \( u_t \) has an autoregressive component, so too will \( v_{t-1} \). 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 (Kollatzas, 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 \( \hat{\beta} \) is

\[
\hat{\beta} = (X'Z'X)^{-1}X'Z'Y
\]

asymptotically \( \hat{\beta} = \text{plim} T [X'Z'X]^{-1}X'Z'Y \approx (X'Z'X)^{-1} \). But

\[
X'X = \text{plim} T [X'Z'X]^{-1}X'Z'Y \approx (X'Z'X)^{-1}
\]

Plainly, the solution to this question depends on the data generating processes (DGP) 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 instraments:

\[
Z_t = \phi_1 \delta_{t-1} + \phi_2 \delta_{t-2} + \tau_{t-2},
\]

\[
E(u_{t-j}) = E(u_{t-2}) = 0 \text{ for } j \neq 0
\]

\[
Z_t = (H_{t-1}, H_{t-2}, \delta_{t-1}, \delta_{t-2})'.
\]

Given \( \phi, \theta, \omega_{t_0}, \omega_{t_1}, \omega_{t_2}, \omega_{t_3} \), and the variance-covariance matrix of \( (u_t, v_t) \) 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

<table>
<thead>
<tr>
<th>Mnemonic</th>
<th>$a_2$</th>
<th>$a_1$</th>
<th>$a_0$</th>
<th>$a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>0.01</td>
<td>0.25</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>R</td>
<td>0.20</td>
<td>-1.00</td>
<td>2.00</td>
<td>0.33</td>
</tr>
<tr>
<td>W</td>
<td>0.50</td>
<td>0.40</td>
<td>0.01</td>
<td>2.00</td>
</tr>
</tbody>
</table>

(1) $a_2$, $a_1$, $a_0$, and $a_3$ are defined in (4.1).
(2) For all three data generating processes, the discount rate was set at $\delta = 0.953$. The sales process was assumed to follow an exogenous $AR(2)$, $\delta = 0.75$, $\gamma = 0.25$, $\gamma = 0.50$, with $\text{var}(\delta) = 0.1250$, $\text{var}(\gamma) = 0.5$, $\text{cor}(\delta, \gamma) = -0.5$, $\text{var}(\gamma) > 0.5$, with $\delta$ the cost shock as defined in (4.1).

Two digit manufacturing industries in the United States: Evensenbaum (1989, p. 862; top panel: mnemonic E), Ramey (1991, p. 323: R), and west (1986, p. 393, top panel: W). (The problem (4.1) is well posed for the R DGP despite the negative sign on $a_1$, see Killianzas, 1989; Ramey, 1991.)

I set the variance-covariance matrix of $(\delta, \gamma)$ so that for the W DGP the unconditional variance-covariance matrix of $(\delta, \gamma)$ was approximately proportional to that of US monthly nondurables manufacturing, 1967-90, $H_d$ are finished goods inventories, with $\text{var}(\delta) = 1$ (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 $H$ displays little serial correlation, both Eichenbaum (1989) and Ramey (1991) explicitly accounted for the serial correlation that is empirically present in $H$, 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 $\tilde{V}$ (defined in (4.5)) and $\tilde{V}$ (defined in (4.6)). Define

$$c_{i,t} = \sum_{t'=1}^T \tilde{e}_{i,t'}$$

as the expected present discounted value of costs. Then the quantity in column 2 of table 4.2, $(1 + \beta_0 + a_1, \beta_0 a_1)$, 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 $\rho^{\tilde{V}}(a_1 - a_3)$ is at best 0.998 times smaller for the optimal than for the 2SLS estimator. Similar ratios apply for $(1 + h)a_0 + a_1$ 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, $\tilde{V}$ one is interested in.
Table 4.2 Comparison of asymptotic variance-covariance matrices resulting from optimal and 2SLS linear combinations of instruments

<table>
<thead>
<tr>
<th></th>
<th>DGP</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ratio of SEs on ( [(1 + b)a_0 + a_1]c )</td>
<td>Ratio of SEs on ( a_2/c )</td>
<td>Ratio of SEs on ( a_1 )</td>
</tr>
<tr>
<td>E</td>
<td>0.990</td>
<td>0.982</td>
<td>0.976</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>0.992</td>
<td>0.962</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

(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 \((3 \times 4)\) linear combination of a \((4 \times 1)\) instrument vector consisting of two lags each of \( X \) and \( S \). 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

\[
\tilde{V} = En^{-1/2} \text{plim} \left( X'Z(Z'Z)^{-1}Z'X \right)^{-1}.
\]  

(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

<table>
<thead>
<tr>
<th></th>
<th>DGP</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ratio of SEs on ( [(1 + b)a_0 + a_1]c )</td>
<td>( a_2/c )</td>
<td>( a_1 )</td>
</tr>
<tr>
<td>E</td>
<td>0.107</td>
<td>0.143</td>
<td>0.249</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.053</td>
<td>0.051</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>0.100</td>
<td>0.533</td>
<td>0.500</td>
<td></td>
</tr>
</tbody>
</table>

(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 $\alpha_i$ 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 $\Omega$ (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 variables (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 technique 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 $S_i$ and $H_i$ (equivalently, lags of $H_i$ and $Q_i$ given the inventory identity $Q_i = S_i + dH_i$). This is, indeed, done in many studies, with $Z_i = [H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, \ldots, \text{if cost shocks are assumed absent}]$ or $Z_i = [H_{t-1}, S_{t-1}, H_{t-2}, S_{t-2}, \ldots, \text{if cost shocks are assumed present but serially uncorrelated}]$. Hansen (1982) shows that given the serial correlation in $Z_{t+1}$, an increased number of lags used (i.e., an increase in the dimension of $Z_i$) yields a strict increase in the asymptotic efficiency of the estimator. This applies even if, as in the example in section 1, there is a finite number of lags of $H_i$ and $S_i$ in the reduced form for $(H_i, S_i)$. 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 $\varepsilon_t$, one can apply the formulas in Hansen (1985) to obtain a $3 \times 1$ instruments vector $Z_t$ (say) that is optimal in the space of instrument vectors that rely on lagged $H_t$ and $S_t$. Suppose for concreteness that the environment is as in section 3.3: the cost shock $\nu_t$ is iid, so that $\nu_{t-1} = \nu_{t-2} = \cdots = \nu_{t-k} = 0$ for $j \geq l$. Then $Z_t$ is a linear combination of past $H_t$ and $S_t$, with nonzero weights on all lags: $H_{t-1}, \ldots, H_{t-l},$ and $S_{t-1}, \ldots, S_{t-l}$ 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 lag of $H_t$ will be correlated with the disturbance $\varepsilon_{t-2}$ 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 he 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 $\tilde{\varepsilon}$ for inference and (if the efficient estimator is used) estimation. To discuss how to do so, let $\tilde{\varepsilon}_t$ be the 2SLS residual, $\tilde{\varepsilon}_t = H_t - X_t \hat{\beta}$, where $\hat{\beta}$ is defined in (4.6). Consider estimating $\gamma_J$ (defined in (4.4)) by

$$\tilde{\gamma}_J = T^{-1} \sum_{t=J}^T Z_t \tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-1}, \tilde{\varepsilon}_{t-2}, \ldots, \tilde{\varepsilon}_{t-J},$$

for $j \geq 0$. (4.9)

For concreteness, focus initially on the case where the cost shock $\nu_t$ is absent, so that $Z_t, \varepsilon_t = MA(1)$ and $\tilde{\varepsilon}_t = \tilde{\varepsilon}_t - (t - T) / T$. The simplest technique used to estimate $\gamma_J$ is the obvious one, $\tilde{\gamma}_J = \tilde{\gamma}_J + (t - T) / T$. 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 Huizinga, 1990). In such a case, the variance-covariance matrix $V$ (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 seed for an estimator that will be positive definite by construction. Early proposals include Cumby, Huizinga, and Obstfeld (1983) and Eichenbaum, Hansen, and Singleton (1988). Since $D$ is proportional to the spectral density of $Z_{1t}$ 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 $D$ as

$$\hat{D} = \hat{\gamma}_0 + \sum_{j=1}^{m} k_j (\hat{\gamma}_1 + \hat{\gamma}_1')$$

(4.15)

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

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

(4.11)

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

It should be emphasized that even in the simple case that $Z_{1t}$ 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 $\gamma_j$ is zero for $j > 1$, for sufficiently large $T$ one will want to be using estimates of the form $\hat{D} = \hat{\gamma}_0 + \ln(m+1)\left[ (\hat{\gamma}_1 + \hat{\gamma}_1') + \cdots + \left(1/((m-1)(m+1))\right)(\hat{\gamma}_m + \hat{\gamma}_m') \right]$ for $m > 1$. The reason is that one wants $D$ to well approximate $\gamma_0 + (\gamma_1 + \gamma_1')$ in a large sample; if $m$ is fixed at 1 independent of $T$ then $D$ 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
Inventory Models: The Estimation of Euler Equations 199

\[ m = [z_1] \]

\[ y = 1.1447(\{z_{1(1)}^{(1)}\})^{1/3}, \quad z_{11}^{(1)} = 2 \sum_j j \sigma_j, \quad z_{11}^{(1)} = \sigma_1 + 2 \sum_{j=1}^n \sigma_j, \]

\[ j \sigma_j = w' f_j w, \quad w = (1, 1, \ldots, 1, 1)' \]

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

Suppose first that cost shocks are absent, so that \( z_{1(1)}^{(1)} - MA(1), \sigma_j = f_j = 0 \) for \( j > 1 \), \( z_{11}^{(1)} = 2 \sigma_1 \) and \( z_{11}^{(1)} = \sigma_1 + 2 \sigma_1 \). Then one could set

\[ m = [z_1]^{1/3} \]

\[ \hat{y} = 1.1447(\{z_{1(1)}^{(1)}\})^{1/3}, \quad \hat{z}_{11}^{(1)} = 2 \sigma_1, \quad \hat{z}_{11}^{(1)} = \sigma_1 + 2 \sigma_1, \]

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

\[ m = [z_1]^{1/3} \]

\[ \hat{y} = 1.1447(\{z_{1(1)}^{(1)}\})^{1/3}, \quad \hat{z}_{11}^{(1)} = 2 \sum_{j=1}^n j \hat{\sigma}_j, \quad \hat{z}_{11}^{(1)} = \hat{\sigma}_1 + 2 \sum_{j=1}^n \hat{\sigma}_j, \]

\[ n = [4(T/100)]^{1/3} \]

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

\[ \hat{A} = z_{1(1)+} A \quad \hat{z}_{1(1)+} = \hat{z}_{1(1)+} \quad \hat{z}_{1(1)+} = \hat{z}_{1(1)+} \]

\[ n = 4(T/100)^{1/3} \]

\[ \hat{\sigma}_j = T^{-1} \sum_{i=j}^{T} (w' \hat{\varepsilon}_i) (w' \hat{\varepsilon}_i), \quad j = 0, \ldots, n, \]

\[ \hat{z}_{11}^{(1)} = \sum_{j=1}^n j \hat{\sigma}_j, \quad \hat{z}_{11}^{(1)} = \hat{\sigma}_1 + 2 \sum_{j=1}^n \hat{\sigma}_j, \]

\[ \hat{y} = 1.1447(\{z_{1(1)}^{(1)}\})^{1/3} \]

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 \( \hat{\varepsilon}_i \) 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 \( \hat{\varepsilon}_i \) and then use \( \hat{A} \) to adjust the result. Specifically, one sets

\[ \hat{A} = (I - \hat{A})^{-1} \hat{\varepsilon}_i (I - \hat{A}) \]

(4.16)
\[ \hat{\Omega} = \sum_{j=1} \left( 1 - \frac{j}{m+1} \right) \tilde{\Omega}_j, \quad \tilde{\Omega}_j = T^{-1} \sum_{i=1}^{\tilde{v}} \tilde{\epsilon}_j \tilde{\epsilon}_j' \]  
\[ m = \left[ \frac{\gamma T}{n} \right] \]

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 by 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 \( V_k \) and relying on a \( V \) computed with a different \( m \) and/or \( n \); the hope is that results will not be sensitive to the exact values of \( m \) 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 \( \tilde{\beta} \) 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 \( s \leq 3 \), the usual Wald statistic is appropriate,

\[ g(\hat{\beta})' \left( \partial g / \partial \beta \right) \tilde{V} \left( \partial g / \partial \beta \right)' g(\hat{\beta}) \triangleq \hat{\chi}^2(\nu) \]

(4.17)

where \( g(\hat{\beta})' \left( \partial g / \partial \beta \right) \tilde{V} \left( \partial g / \partial \beta \right)' g(\hat{\beta}) \) is of rank \( \nu \) and \( \hat{\chi}^2(\nu) \) is the usual \( \nu \)-statistic.

Let \( \hat{\tilde{\Omega}} = \sum_{j=1} \left( 1 - \frac{j}{m+1} \right) \tilde{\Omega}_j \), \( \hat{\tilde{v}} = [\hat{\tilde{v}}_1, \ldots, \hat{\tilde{v}}_s] \). If the number of regressors \( q \) exceeds the number of regressors \( 3 \), one can test the model by computing

\[ T^{-1/2} Z \hat{\tilde{\Omega}}^{-1/2} \tilde{v} \sim \chi^2(q - 3). \]

(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 \( H, S, \bar{Q} \), and the cost shocks are nonnegative. If \( u_t = 0 \) as in
West (1986a), for example, we have

\[ a_0 \var{\Delta S_1} + \sigma_0 \var{\Delta Q_1} + \sigma_0 \var{\Delta Q_2} + \sigma_0 \var{\Delta H_1} + 2 \sigma_0 \cov{\Delta S_1, \Delta H_1} > 0, \]

(4.19)

where "\( \var{\cdot} \)" denotes variance and "\( \cov{\cdot, \cdot} \)" denotes covariance. To compute the
left-hand side of (4.19), one may use estimates of the \( \alpha \)s 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 \( \alpha \)s 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 \( Q = S \) and \( H = 0 \). If \( \alpha_1 = \alpha_2 = 0 \), and
\( \alpha_0, \alpha_3 > 0 \), for example, (4.19) reduces to \( \sigma_0 \var{\Delta S_1} - \sigma_0 \var{\Delta Q_1} - \sigma_0 \var{\Delta H_1} = 0 

(4.19)

for average per period reduction in production costs allowed by inventory
holdings 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 \( S_t \equiv \{0\} \), but \( u_t \sim \{0\} \). 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:

\[ E_t[\alpha_0 \var{\Delta S_1} + \alpha_1 \var{\Delta H_1} - 2\alpha_2 \var{\Delta S_{1,1}} + \alpha_3 \var{\Delta H_{1,1}} + \beta^2 \var{\Delta S_{1,1} + \alpha_3 \var{\Delta H_{1,1}}}]
\]

\[ - b_s(\Delta H_{1,1} + \Delta S_{1,1}) + \alpha_1 \Delta H_t + b_s(\Delta S_t - \gamma S_t) = - \nu_t, \]

(4.20)

\[ \gamma = \alpha_0(1 - b)/a. \]
Let \( \{h_t\} \) denote the expression in braces. It may be shown that \( \{h_t\} \) and, since \( \Delta s_t \sim i.i.d. \), \( H \), \( S \), and \( \gamma \) are cointegrated with cointegrating parameter \( \gamma \).

Consider first the case where \( E \Delta H_t = 0 \) (as \( E s_t = 0 \)) 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 (4.19)) 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 Depreciation 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 stay 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; Hagger, 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 \( H_t \), \( S \), Not Cointegrated

Now suppose that \( w_t \sim \Delta f(t) \). 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 \( w_t \) is a pure random walk, the disturbance of the difference 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 (Stranger and Lee, 1989; West, 1990; Roissasa, 1992).

7 Full Information Estimation

7.1 Solution of the Model

For algebraic simplicity, assume that the firm views revenue $p_S$ 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 $S_t$,

$$S_t = \phi_1 S_{t-1} + \ldots + \phi_p S_{t-p} + \epsilon_t,$$

$$\phi_p \neq 0, \quad E(\epsilon_{t+1} | \mathcal{F}_t) = 0 \quad \text{for} \ j > 0, \quad 1 - \phi_1 - \ldots - \phi_p = 0 \Rightarrow |\gamma| = 1,$$

(4.21)

where $|\cdot|$ denotes the modulus of a complex number $\gamma$. Note that (4.23) allows $S_t$ to have a unit autoregressive root. The innovation $\epsilon_t$ is assumed uncorrelated with lagged $H_{t-j}$, in accord with the assumption that sales are exogenous.

For the moment, assume $a_{xx} \neq 0$. Let $L$ be the lag operator. Use the identity $Q_t = S_t + aH_{t-j}$ to rewrite the Euler equation (4.2) as

$$E(\epsilon_t | \mathcal{F}_t) = D_1,$$

(4.22)

$$f(L) = 1 - b^{-1} a_{xx} [a_{uu} + 2ab(1 + b) + b^{-1} a_{xx} a_{uu} (1 + 4b + b^2) + a_{x} (1 + b) + b_{x}] L^{-1},$$

$$- b^{-2} a_{xx} [a_{u} + 2ab(1 + b)L^{-1} + b^{-1} x,,$$

$$D_2 = - b^{-1} (a_{x} S_{t-j} + b_{x} S_{t-j} - b^{-1} a_{xx} a_{uu} (S_{t-j} - b_{x}) + b^{-2} a_{xx} a_{uu} S_{t-j} - b^{-1} a_{xx} a_{uu},$$

(4.23)

Let $\lambda_1$ be the roots of the fourth order lag polynomial $f(L)$, $|\lambda_1| = \ldots = |\lambda_4|$. It may be shown that $\lambda_1 = 1/(b_{xx})$ and $\lambda_2 = 1/(b_{uu})$, so that at $|\lambda_1, |\lambda_2| < 1/b$. Suppose further that $|\lambda_3, |\lambda_4| < 1$, and, for expositional convenience, that $\lambda_3 = \lambda_4$. (See Koivunen, 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 $\lambda_1$ and $\lambda_2$ backwards, and
the unstable roots $\lambda_1$ and $\lambda_2$ forwards: indeed, a transversality condition forces the firm to do so (Kollintzas, 1989). One may verify that the following transcendental equation holds:

$$H_i = (k_i + \lambda_1)H_{i-1} - \lambda_1k_iH_{i-2} + h^{-1}\lambda_1\lambda_2(\lambda_i - \lambda_2)^{-1} \times \sum_{j=0}^{\infty} [((b\lambda_i)^{*-1} - (b\lambda_1)^{*-1})E_iD_{i,j}].$$

Note that if $\lambda_1$ and $\lambda_2$ 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 $w$ is serially uncorrelated. Then one can use techniques such as those in Blanchard (1983) to solve for the reduced form, which expresses $H_i$ in terms of lagged $H_{i-1}$ and $S_{i-1}$ and $u_i$, and for the decision rule, which expresses $h_i$ in terms of lagged $H_{i-1}$, current and lagged $S_i$, and current $u_i$. Define the scalars $p_1$, $p_2$, $v_1$, $w_1$, $v_1$ the $(i \times p)$ vector $v'$ and the $(p \times p)$ matrix $\Phi$ and $D$ as:

$$p_1 = \lambda_1 + 2\beta, \quad p_2 = -\lambda_1\lambda_2,$$

$$w_1 = b'd_0, \quad w_2 = -p_2[b^2 + 2b + b(a_i/a_0) + (b\lambda_1a_i/a_0)],$$

$$w_3 = p_2[2b + 1 + (a_i/a_0)], \quad w_4 = -p_2,$$

$$v' = (1 \ 0 \cdots 0),$$

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{i-1} & \phi_i \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \end{bmatrix}$$

$$D = [I - b\phi_1\phi_0 - b^2p_1\phi_1]^{-1}.$$  

Then the reduced form is:

$$H_i = p_1H_{i-1} + p_2H_{i-2} + \phi_1S_{i-1} + \cdots + \phi_{i-1}S_{1-1} + \psi u_i,$$

$$v_{i1} = (\phi_1\phi_0)v_{i1} + v_{i2}, \quad (p \geq 2),$$

$$v_{i2} = (\phi_i\phi_{i-1})v_{i2} + v_{i3}, \quad (p \geq 1),$$

$$v_{i3} = (\phi_{i-1}\phi_{i-2})v_{i3} + v_{i4}, \quad (p \geq 2).$$

The underlying decision rule is:

$$H_i = p_1H_{i-1} + p_2H_{i-2} + \phi_1S_{i-1} + \cdots + \phi_{i-1}S_{1-1} + V_{im} \quad (p \geq 2),$$

$$H_i = p_1H_{i-1} + p_2H_{i-2} + \phi_1S_{i-1} + \phi_{i-1}S_{1-1} + v_{i1}, \quad (p \geq 1).$$
where for \( p \geq 2 \),
\[
\delta_i = \pi_i \phi_i, \quad \delta_i = \pi_{i-1} - \delta_i \phi_{i-1} \quad (i = 2, \ldots, p),
\]
and for \( p = 1 \),
\[
\delta_1 = -\rho_1, \quad \delta_1 = (\pi_1 - \delta_1) \phi_1.
\]

### 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 \( \beta \), one can retrieve estimates of the underlying parameters of the cost function (the \( \alpha_i \)) using (Blanchard, 1983)
\[
a_{i}/a_{i-1} = \rho_i (b_i - b_i^2) - 2b - 2, \quad (4.27)
a_{i}/a_{i-1} = -b_i \{ [p_i](1 + b_i b_i) + b_i b_i + (1 + b_i) (a_i/a_{i-1}) - (1 + 4b + b_i^2) \}.
\]

Given estimates of the sales process (4.21) as well, estimates of \( a_i \) can be disentangled from \( \pi_i \) (or, for that matter, from any of the other \( \pi_i \) as well, if \( p \geq 2 \)).

There are \( p + 3 \) parameters to be estimated: \( \pi_1, \ldots, \pi_p, \pi_0, \) and, relative to some linear combination of the \( a_i \)'s, the values of two of \( \alpha_0, \alpha_0, \) or \( \alpha_0 \). 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 \( S \) has a unit root, so that \( H_i \) and \( S \) are cointegrated (West, 1988a; Sims, Stock, and Watson, 1990), subject to exceptions discussed in section 6.

If \( (b) \ p > 1 \), the system is overidentified. Consider first (b(i)) a stationary model \( (S \sim R(0)) \Rightarrow (H_i \sim R(0)) \). Estimation may proceed by maximum likelihood, imposing the nonlinear overdetermining 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 (Amenyia, 1977).

To my knowledge, the formal asymptotic theory has not been completely worked out for restricted estimates of the model in the case that \( (b(ii)) S_i \Rightarrow R(1) \Rightarrow H_i \Rightarrow S_i \) 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 \( S_i \) (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 \( S_i \) (the analogue to (4.21)),(13) and the reduced form for \( H_i \). (In the example above, in which there is no such private information, this appears to yield no gains
then the reduced form and decision rule are as above, but with
\[ u_{i} = \theta_{i}v_{i-1} + u_{i}, \]  
(4.28)
and
\[ \mu = (\sigma_{u}/a_{0})(1 - b_{0}p_{0} - b_{1}p_{1})^{-1}u_{i}. \]
If \( \theta = 1 \), \( u_{i} \) is a random walk. If \( S_{i} \sim (F) \) 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
\[ H_{i} = \rho H_{i-1} + \lambda_{i}S_{i-1} + \ldots + \lambda_{p}S_{i-p} + \nu_{i}, \]  
(4.29)
\( \rho \) is the smaller root of \( ba_{1}x^{2} - (a_{1} + ba_{0})x + a_{0} = 0, \)
\[ (\sigma_{u}, \ldots, \sigma_{p}) = \left( \psi_{i}^{1} - b_{0}p_{0}^{-1}b_{1}p_{1}^{1} + a_{1}^{1}u_{i}a_{0}^{-1}p_{0}^{1} - \rho \phi_{0} \right), \]
\[ \nu_{i} = (\sigma_{i}^{1}p_{0})u_{i} + \nu_{i}, \]
\[ \nu_{i} = (\sigma_{i}^{1}p_{0})u_{i} + \nu_{i}, \]
and
\[ H_{i} = \rho H_{i-1} + \delta_{i}S_{i} + \ldots + \delta_{p}S_{i-p} + \nu_{i} \]  
(4.30)
\( \delta_{i} = \sigma_{i}^{1}p_{0}, \quad \delta_{i} = \sigma_{i-1}^{1} - \delta_{p} \) \( (i = 2, \ldots, p). \)

8 Comparison of Full and Limited Information Estimation

The limited information techniques described in sections 3-6 are less efficient but more robust than 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 interaction of a demand curve and supply. The demand curve might be
\[ S_{i} = - (1/\alpha)p_{i} + \text{demand shock} \]  
(4.31)
where \( \alpha > 0 \) and \( p_{i} \) is defined in (4.1); a supply curve is obtained from a first order condition obtained by differentiating (4.1) with respect to \( S_{i} \) and/or \( p_{i}. \) The assumption maintained above that sales are exogenous is a special case of (4.31) resulting when \( \alpha = \infty, \) so that \( S_{i} = \text{demand shock} \) (Kollmorgen, 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 \( S_{i} \) and \( H_{i} \), and it may be
shown that this holds even when \( a < \infty \). But if \( a < \infty \), \( S_t \) will be Granger-caused by \( H_t \). The intuition is, 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 \( H_{t-1} \) and \( S_{t-1} \), the \( H_{t-1} \) will get nonzero coefficients. This means that the coefficients on the lagged \( H_{t-1} \) 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 \( S_t \). 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 \( \rho \) and \( \rho_1 \) for a given set of values of \( a_{u/c}, a_{c}/a_{u/c}, \) and \( a_{c} \), and for a range of \( \alpha \); for each value of \( \alpha \), I then computed the values of \( a_{u/c}, a_{c}/a_{u/c}, \) and \( a_{c} \) that are implied by (4.27). Since use of (4.27) is

<table>
<thead>
<tr>
<th>Table 4.4</th>
<th>Asymptotic bias of full information estimator, when sales are endogenous</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>1 0.05</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.44</td>
</tr>
<tr>
<td>( \rho_1 )</td>
<td>-0.09</td>
</tr>
<tr>
<td>( a_{u}/a_{c} )</td>
<td>0.07</td>
</tr>
<tr>
<td>( a_{c}/a_{u/c} )</td>
<td>0.11</td>
</tr>
<tr>
<td>( a_{c} )</td>
<td>0.16</td>
</tr>
</tbody>
</table>

(a) \( a_{o}, a_{1}, \) and \( a_{2} \) are defined in (4.1); \( c \) is defined in (4.3); \( a \) is defined in (4.31).
(b) Columns 2 and 3 present the coefficients in the decision rule for \( H_t \) when \( a \) is as indicated, \( a_{u/c}, a_{c}/a_{u/c}, \) and \( a_{c} \) are as given in the "True values" row, and \( a_{u}/a_{c} = -0.04 \).
(c) These are the approximate values of the \( a \) 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 \( a \) 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 \( a \).
(d) Columns 4-6 give the values of \( a_{u/c}, a_{c}/a_{u/c}, \) and \( a_{c} \) that are implied by \( \rho \) and \( \rho_1 \) under the incorrect presumption that \( a = \infty \) and sales are exogenous.
(e) To facilitate reading the table, only two digits are given. The values of \( a \) and the \( a \) actually used in the calculation are the three digit values given in West (1990).
(f) These asymptotic calculations may not accurately predict the actual finite sample performance of the full information estimator.
appropriate for \( y = \sigma \), the question is whether plausible values of \( \sigma \) 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). Table 4.4 lists the values used. The table indicates that in this example, there is a plausible range of \( a \) for which one might be seriously misled by assuming that sales are exogenous \((a = \gamma)\) when in fact sales are endogenous \((a < \gamma)\). Row 1 of the table indicates that if \( a = 0.65 \) (a value that is plausible in the sense that it falls within the 95% confidence interval for \( a \) given by West, 1990), use of estimates of \( R \)'s reduced form would yield a value of \( \psi / \sigma \) that is not only positive but is larger than that of \( a / \sigma \), at least in an arbitrarily large sample; in truth, however, \( a / \sigma \) is negative and smaller in absolute value than \( \psi / \sigma \). Other values of \( a \) (rows 2 and 3) yield smaller biases.

Of course, full information estimation is still visible when \( a < \gamma \); one must simply estimate \( a \) 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 \( S \) 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

<table>
<thead>
<tr>
<th>GNP</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ratio of SEs on ([1 + b \sigma_t + a] / \sigma)</td>
<td>Ratio of SEs on (a / \sigma)</td>
<td>Ratio of SEs on (a / \sigma)</td>
</tr>
<tr>
<td>E</td>
<td>0.783</td>
<td>0.815</td>
<td>0.921</td>
</tr>
<tr>
<td>R</td>
<td>0.832</td>
<td>0.834</td>
<td>0.988</td>
</tr>
<tr>
<td>W</td>
<td>0.901</td>
<td>0.372</td>
<td>0.294</td>
</tr>
</tbody>
</table>

(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 (2.1) 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 (4.2) and (4.25). The ratios must be 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.
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 \geq 2, and the cost shock \( \omega \) follows an AR(1) with parameter \( \theta \) (possibly with \( \theta = 0 \), so that \( \omega_t \) is serially uncorrelated):

\[
\begin{align*}
S_t &= \phi_1 S_{t-1} + \ldots + \phi_p S_{t-p} + \epsilon_t, \\
S_t &= \theta S_{t-1} + \epsilon_t.
\end{align*}
\]  

(4.21)  

(4.28)

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

\[
\begin{align*}
H_t &= \rho H_{t-1} + \rho H_{t-1} + \delta_t S_{t-1} + \ldots + \delta_t S_{t-p+1} + \epsilon_m \quad (\rho_1 \neq 0), \\
H_t &= \rho H_{t-1} + \delta_t S_{t-1} + \ldots + \delta_t S_{t-p+1} + \epsilon_m \quad (\rho_1 = 0).
\end{align*}
\]  

(4.26)  

(4.30)

As detailed in section 7, the parameters \( \rho_1, \rho_2, \) and \( \rho \) are functions of \( \lambda, \omega_t, \) and \( \delta_t \); the \( \delta_t \) are functions of \( \lambda \), the \( \omega_t \), and the \( \epsilon_m \); the disturbance \( \epsilon_m \) is AR(1) with estimator \( \theta = 0 \Rightarrow \epsilon_m \) is serially uncorrelated.

Note one implication of the cost shock being serially correlated (of \( \theta \neq 0 \)).

If one multiplies (4.30) by \( (1-\theta) \) and rearranges, \( H_{t-1} \) appear on the right.
hand side (as does \( S_i \), though that is not important): serially correlated cost shocks, as well as nonzero costs of adjustment (\( \alpha \neq 0 \), 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 \( H_t^* \):

\[
\min 0.5(H_t - H_t^*)^2 + 0.5\omega(H_t - H_{t-1})^2 + \nu, H_t; \quad H_t^* = \alpha E_t S_{i+1}.
\]  
(4.32)

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

\[
H_t = \rho H_{t-1} + (1 - \rho) \alpha E_t S_{i+1} + \nu_t, \quad \nu_t = (1 - \rho) \omega \nu_{t-1} \quad 0 < \rho < \frac{\omega}{\omega + \rho} < 1.
\]  
(4.33)

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

\[
H_t = \rho H_{t-1} + \delta_1 S_i + \ldots + \delta_p S_{i-p+1} + \nu_t, \quad \delta_i = (1 - \rho) \alpha \phi_i.
\]  
(4.34)

It will be recognized that (4.34) is the same as (4.30), although the two models do predict different relationships between \( H_t \) and the \( S_i \)'s, and (2)p and underlying cost parameters.

Blinder and Mascini (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 \( S_i \), there is considerable serial correlation in \( H_t \), in that estimates of \( \rho \) and \( \delta \) (the serial correlation parameter of \( u_t \)) tend to be near one. Moreover, both \( \rho \) and \( \delta \) tend to be significantly different from zero. Second, the estimate of \( \delta \) 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 \( S_i \) and \( Q_t \) rise, and to be decumulated during recessions as \( S_i \) and \( Q_t \) fall. Given the inventory identity \( Q_t = S_t + \Delta H_t \), the positive correlation between \( S_t \) and \( \Delta H_t \) produces the well-known result that \( Q_t \) is more variable than \( S_t \).
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 $a_t$ is highly serially correlated. Second, irrespective of the serial correlation of $a_t$, it will follow if $\rho$ and $\theta$ (see (4.26)) are such that the larger root of $x^2 - \rho x - \theta$ is near unity, or, when $a_t = 0$, if $\rho$ (see (4.30)) is near unity.

The model will yield such a root, or yield $\rho = 1$ when $a_t = 0$, when the marginal inventory holding cost $a_t$ is small relative to the slope of marginal cost of production $a_t$ and of changing production $a_t$. For example, when $a_t = 0$, $\rho \to 1$ as $a_t / a_t \to 0$. More generally, the larger root of $x^2 - \rho x - \theta$ approaches $1$ as $a_t \to 0$ for any fixed positive values of $a_t$ and $a_t$.

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 2 Euler equation such as (4.3), Eichenbaum (1987) and Ramsey (1991) found $a_t$ that by themselves implied a finite serial correlation, but very high serial correlation of an unobservably cost shock, while West (1986) found $a_t$ that implied high serial correlation but 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 $\rho$ and $\theta$ are nonzero, when (4.30) is transformed to have a serially uncorrelated disturbance the resulting decision rule will have $H_t$ on the right hand side. It will therefore look similar to the decision rule (4.26), which was derived assuming $a_t = 0$ and a serially uncorrelated cost shock. What is involved, then, is distinguishing between costs of adjustment and serial correlation, which is not easy done (Blinder 1986b; Malmquist, Nenkevich, and Savin, 1992). Below I discuss these explanations further.

A final point to be made at this point 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), $\rho = 1$ implies that the percentage of the gap between $H_t$ and $H_{t-1}$ 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 $\rho = 1$, or that $x^2 - \rho x - \theta$ 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 Procylical Inventory Movements

The procylical character of inventory movements may be rationalized by the model in at least three ways, which are not mutually exclusive. 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 \( a_t (H_{t-1} - \alpha_t S_{t-1}) \). This term captures a tradeoff between inventory holding costs on the one hand and stockout or backlogging 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 \( (a_0 = \alpha_0 = \mu_0 = 0) \), the firm would simply set \( H_t = \alpha_t E(S_{t+1}) \); 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 \( \Delta H_t \) to track actual as well as expected sales, and inventories clearly will move procyclusically. And even if nonzero \( a_t(\lambda) \), \( \alpha_t(\lambda) \), and \( \mu_t \) 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 procyclusically.

To understand the other two explanations, it is useful to first consider a set of circumstances under which \( H_t \) would not move procyclusically. Suppose now that \( a_t = \alpha_t = \mu_t = 0 \), \( \alpha_t > 0 \). Then with increasing marginal costs of production \( (\alpha_t > 0) \), no accelerator motive \( (a_t = 0) \), and costs nonstochastic \( (\mu_t = 0) \), 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 procyclusic inventory movements emphasizes the possible role of stochastic movements in costs. Now allow for \( u_t \neq 0 \), but, for clarity, continue to assume \( a_t = \alpha_t = 0 \). 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. This will produce a tendency for \( \Delta H_t \) and \( \Delta Q_t \) to move in the same direction; if this tendency is strong enough relative to the one described in the preceding paragraph, inventories may move procyclusically, 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 \( u_t = \alpha_t = 0 \). Assume for the moment that \( 0 < a_t < \alpha_t E(\Delta Q_t) = (1 + b)\alpha_t + 1 \). 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 procyclusically. If \( a_t \) instead makes the weaker assumption that \( a_t < 0 \) but \( (1 + b)\alpha_t + 1 \) > 0, so that
marginal production cost slopes down only when one abstracts from costs of adjusting production, there will be a tendency for inventories to move procyclically; whether they do or not depends on whether the motive to smooth $(1 + b)a_t + a_i > 0$ or bunch effects $(a_i < 0)$ is stronger.

In sum, the procyclical movement of inventories suggests a substantial role for the $a_i(1 + b)a_t + a_i$ 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 $a_i$ that are positive and significant at conventional levels; West (1980a) and Krane and Braun (1991) find estimates that usually are positive but rarely are significant; Kashyap and Wilcox (1993) and Wu (1990) find estimates that are of mixed sign, and usually are insignificant. Perhaps an indirect indication of the importance of the stockout or backlog costs underlying $a_i > 0$ is that measures of order backlogs often are significant in inventory regressions (e.g., Maccini and Rossana, 1984; Blander, 1986a).

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

| Table 4.5 Statistical significance of cost variables |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Wage prices     | Energy prices   | Interest rate   | Unobservable shock |
| 1 Blander (1986a) | ?               | ?               | n               | ?               |
| 2 Daggett and Maccini (1992) | ?               | n               | n               | ?               |
| 3 Eichenbaum (1989) | y               | ?               | n               | ?               |
| 4 Maccini and Rossana (1981) | y               | y               | n               | ?               |
| 5 Maccini and Rossana (1984) | n               | y               | n               | ?               |
| 6 Miron and Zeldes (1988) | ?               | ?               | 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 "?" 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 set; an "n" that is was significant at most one-fourth of the data set; and a "?" that is was significant in more than one-fourth but fewer than three-fourths of the data set. 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. 257 (6) table II (p. 892) (7) table 1 (p. 323).
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 \( a_t \), 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 \( a_t \) 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 \( a_t \) is small relative to the production costs \( a_t \) and/or \( a_t \), and either (1) \( a_t \) is slightly negative, or (2) the effects of the accelerator term \( a_t(\dot{H}_t - a_tS_t) \) are large.

### Notes

I thank an anonymous referee, participants in a seminar at the Federal Reserve Board of Governors, Louis Maccini, Scott Schub, and, especially, David Wilcox for helpful
Inventory Models: The Estimation of Euler Equations 215

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(t)Q, where c(t) is a parameter vector and w(t) is a vector of input prices, and/or (2) τf(t)H, where τf is the ex-ante real interest rate.

2 An implicit assumption made is that one can parameterize the problem so that \( \partial (p_j, s_j) / \partial H_j = 0 \) 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 rates, for strategic reasons (Rotemberg and Saloner, 1989).

3 If data on output price \( p_j \) are available, an additional first order condition allows one to identify the \( a_i \)s relative to the units in which \( p_j \) 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 \( \partial (p_j, s_j) / \partial H_j = 0 \) for all \( j \) and that \( Q_j \) 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 Eisenbaum sets \( a_0 = 0 \) and reports parameters that he denotes \( \lambda \) and \( a_1 \). I mapped his instances into the present notation using \( \lambda(a_1) = \lambda + (1/2a_1 - 1 - (1/b)), \) with \( b = 0.995, a_1 = 2(1/a_0)/(a_0/\alpha_1) \). For computational convenience, I then set \( a_0 \) to a small nonzero number rather than zero. Eisenbaum also assumes that \( C \) includes a term of the form (in my notation) \( (H_t - a_0 S_t)^2 \) rather than \( (H_t - a_0 S_t)^2 \); I shall over this minor difference.

8 Just as \( f_1 + (f_1 + f_2) \) might not be positive definite, \( s_i \) may be negative (as, of course, might \( Z_i \) or even the underlying population quantity \( Z_i \) but not \( Z_i \)). But the fact that the ratio \( s_i / s_j \) is squared in (4.13) means that \( s \) will be nonnegative, and the resulting \( \sigma \) will be positive definite.

9 That \( \lambda > 0 \) follows since \( E_i (\lambda - a_0 - b_0) \) by assumption and \( \lambda - E_i (\lambda - a_0 - b_0) \) as well. Further, \( \lambda - b_0 \geq H_t - a_0 S_t \) if \( H_t - a_0 S_t > b_0 \) for some \( d > 1 \), then the order of integration of \( H_t \) would be greater than that of any other variable and \( \lambda \) could not be \( b_0 \), while \( \lambda - b_0 \leq a_0 S_t \). Incidentally, \( \lambda \) and \( S_t \) are “multicointegrated,” in the terminology of Granger and Lee (1988, 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 lags \( H_t \) and \( S_t \), and lagged \( H_t \) will appear on the right hand side of \( S_t \)’s time series process.

11 Let \( E_Y Y = E_X X (X'X)^{-1} - (X'X)^{-1} X' \) 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 $S_t$, and absorbing the difference between this projection and $E_tY_t$ in the error term: the error term would then be correlated with the lagged $H_t$.

12 In this example, the model is exactly identified, so full information estimation simply involves estimating the unrestricted reduced form: See West (1980). Over-identification would result if the demand curve were e.g. $S_t = (1 + dp_r + \rho S_{t-1} + \phi_t)$, demand shock, with $\phi_t = 0$.

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 replaces the data of exponential growth before estimation (West, 1988b; 1990).

14 I consider here the unconditional 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 turn on how stationarity is induced. See Blinder (1984a) for discussion of a conditional relationship, specifically, an analysis of what makes the regression coefficient $\alpha$ positive; see Knaus (1991) for evidence on the similar pattern that applies in the deterministisch seasonal relationship between inventories and sales-production. There is a sense in which inventories will move procyclically even if $\alpha$ follows a random walk and there are no possibilities for intertemporal substitution; see West (1990).

15 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 $\alpha(Z_{t-1} - \alpha S_t^2$ (Knaus, 1991). See Kahn (1987; 1992) for excellent work modeling such costs in a more sophisticated way.

References


