ED
ESTIMATING DISTRIBUTION LAGS IN SHORT PANELS
WITH AN APPLICATION TO THE SPECIFICATION OF
DEPRECIATION PATTERNS AND CAPITAL STOCK CONSTRUCTS

Ariel Pakes

Zvi Griliches

Estimating Distribution Lags in Short Panels with an
Application to the Specification of Depreciation
Patterns and Capital Stock Constructs

ABSTRACT

In this paper, we investigate the problem of estimating distributed lags in short panels. Estimates of the parameter of distributed lag relationships based on single time-series of observations have been usually rather imprecise. The promise of panel data is in the N repetitions of the time-series that it contains which should allow one to estimate the identified lag parameters with greater precision. On the other hand, panels tend to track their observations only over a relatively short time interval. Thus, some assumptions will have to be made on the contribution of the unobserved presample x's to the current values of y before any lag parameters can be identified from such data. In this paper we suggest two such assumptions; both of which are, at least in part, testable, and outline appropriate estimation techniques. The first places reasonable restrictions on the relationship between the presample and insample x's, while the second imposes conventional functional form constraints on the lag coefficients. The paper concludes with an example which investigates empirically how to construct a "capital stock" for profit or rate of return regressions.

Ariel Pakes
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

(617) 868-3921

Zvi Griliches
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

(617) 495-2181

ESTIMATING DISTRIBUTED LAGS IN SHORT PANELS WITH AN APPLICATION TO
THE SPECIFICATION OF DEPRECIATION PATTERNS AND
CAPITAL STOCK CONSTRUCTS

I. INTRODUCTION

The problem we deal with in this paper arises, as is often the
case in econometrics, because we do not have all the data that we
would like to have. We often expect that changes in our independent
variables have relatively long lasting effects, but have only fairly
short time series to uncover them. Consider, for example, the simple
model

$$y_t = w_0 x_t + w_1 x_{t-1} + \ldots + w_m x_{t-m} + u_t$$

where we have suppressed the constant term for simplicity of exposi-
tion. Here $m + 1$ is the total length of the lag structure and the
problem we wish to consider arises when m is large relative to T,
or to $T + \theta$, where T is the total number of observations on y,
and θ is the available number of lagged observations on x. A
typical example might be $T = 20$, $m = 6$, and $\theta = 2$. In this case,
one would have to give up 4 observations on y to get the whole lag
"in", and estimate 8 parameters (including the constant term) on the
basis of 16 observations. The resulting estimates are unlikely to be
very precise. One way of gaining precision is to impose additional
structure on the w's, choosing the form of the lag a priori -- poly-
nomial, geometric, etc. Alternatively, one may try to increase the
sample in some other dimension, adding more individuals, states or
commodities. The sample becomes then a time-series panel with $N \times T$
observations.

A significant number of panel data sets are now available. Though N in these data sets is often quite large, it is not yet clear whether they will yield "better" estimates of distributed lag coefficients. Part of the problem lies in the fact that panel data sets usually track their observations only over a rather short time interval. Thus, if in the example introduced above $T + \theta$ were less than seven while $m > 7$, we could not estimate any of the coefficients consistently (as N grows large) without bringing in additional information. The problem is, of course, that in short panels the contribution of the unobserved presample x's to the current values of y (or the truncation remainders) are likely to be particularly important. Since the value of the truncation remainder is determined by the lag coefficients and the presample x's one has to constrain either the lag structure, or the stochastic process generating these x's, in order to identify any of the coefficients of interest. The first alternative has been discussed in the context of a single time series by Klein (1958), Dhrymes (1971), Maddala and Rao (1971), Pesaren (1973), and others. In Section III, we discuss how their solutions can be adapted to the panel data situation. Because there are usually many more degrees of freedom available in panel data, we concentrate on solutions which do not restrict the lag coefficients a priori. Section II considers, therefore, the problem of estimating a sequence of unrestricted lag coefficients when one is willing to impose some structure on the stochastic processes generating the x's. The models we deal with allow for the possibility of individual-specific time-invariant unobservable factors which impact both on y given past x's and on the distribution of the x's themselves (they

allow for heterogeneity and correlated effects). We consider explicitly the identification of lag coefficients when the regression function of recent x's on all past values of x and an individual specific constant term depends only on this constant term and a small number of the immediately preceeding x's . Examples of stochastic processes which statisfy such assumptions are mixture processes consisting of individual-specific components plus an integrated autoregressive deviate. The parameters defining this process may differ from year to year (they need not be homogeneous over time). It is shown that given this assumption one can identify the leading lag coefficients without making any assumptions on the shape of the rest of the lag distribution and test for alternative simplifications (for, e.g., that there is no individual-specific component, or that the lag coefficients, at least after a few free lags, have an autoregressive structure). These assumptions do not allow, however, for the identification of the tail of the lag structure (at least not without further restrictions). One cannot provide answers to questions involving the entire sequence of the lag coefficients (such as their sum) without imposing constraints on the lag structure itself (Sims, 1974). On the other hand, the early w's are often of significant interest in themselves, especially since unconstrained estimates of their values may provide an empirical basis for restricting the tail of the lag structure in further work. This brings us then to Section III where we discuss the estimation of distributed lag coefficients from panel data when one is willing to restrict the shape of the lag structure a priori.

An advantage of using the approach discussed here is that, in general, it can be checked against the observed behavior of the data. The tests we consider consist of a set of constraints on the second order population moments of the vector $r_i' = (y_i', x_i')$; where y_i is the T element vector of observations on y for individual i , and x_i is the $T + \theta$ vector of observations on x . In most cases these constraints are restrictions on the matrix of coefficients obtained from the regression of each of the T y_{it} 's on all $(T + \theta)$ of the x_{it} 's . In each of our models this matrix, labelled Π by Chamberlain (1978), consists of a combination of the lag coefficients of interest, and of the regression functions of the unobserved components (the truncation remainders, and the individual-specific sources of heterogeneity), on x_i .

We conclude with an empirical example. It consists of regressing the operating profits of firms on a distributed lag of their past investment expenditures. The example, therefore, investigates empirically how to construct a "capital stock" for profit or rate of return regressions. We find that the lag coefficients rise monotonically over the first three periods and then remains fairly constant over the next four or five. This differs significantly from lag structures based on straight line or declining balance depreciation schemes (in both of these the lag coefficients decline monotonically in τ). When we compare the fit of the equations estimated from our models to the fit of a system of regressions functions based on the usual capital stock constructs (net or gross), our equations do quite a bit better. On the other hand, when we compare the restricted Π matrix

implied by our model to an unrestricted Π matrix, we find that we can, using traditional testing procedures, reject the null hypothesis embodied in our assumptions. It is doubtful whether the relationship between profits and past investments is truly stable over different firms and time periods.

II. DISTRIBUTED LAGS FROM PANEL DATA WITH PRIOR STRUCTURE ON THE x-PROCESS

The distributed lag model which we assume is given by A1.

A1:
$$y_{it} = \alpha_i + \sum_{\tau=0}^{\infty} w_\tau x_{i\,t-\tau} + u_{it}$$

where, if $\underset{\sim}{x}_i{}' = [x_{i\,-\theta}, \ldots, x_{iT}]$ and $\underset{\sim}{u}_i{}' = [u_{it}, \ldots, u_{iT}]$, it is assumed that the N observations on the vector $(\underset{\sim}{x}_i{}', \underset{\sim}{u}_i{}'\,\alpha_i)$ are independent draws from a common distribution function in which $E^*[\underset{\sim}{u}_i \mid \underset{\sim}{x}_i, D_i] = 0$, where D_i denotes the indicator function which takes the value of one for individual i and zero elsewhere, and $\Sigma_{xx'} = E[\underset{\sim}{x}_i, \underset{\sim}{x}_i{}']$ is positive definite. The sequence $\{w_\tau\}_{\tau=0}^{\infty}$ is a set of unknown finite parameters.

Here (and in the discussion below) $E^*(z_1 \mid z_2)$ is the operator which provides the minimum mean square error linear predictor (or the regression function) of z_1 given z_2. Some preliminary comments on A1 are in order. For ease of exposition we have omitted time-specific constant terms from it and from all equations to be introduced below. Since we will be concerned with the limiting distribution of the parameter estimates in dimension N (holding T fixed), time-specific

constant terms, if required, can be added without creating any addi-
tional estimation problems. Al does assume that the distributed lag
coefficients do not differ either over time or between individuals
(though random differences in these coefficients as defined for
example by Swamy, 1970, are covered by Al). Note, however, that the
model does allow for individual-specific constant terms (the α_i).
This permits one to deal explicitly with the problem of unobserved
individual-specific components which are constant over time and affect
both the conditional distribution of y_i given the x sequence
corresponding to individual i and the marginal distribution of $x_{\sim i}$.
Given the α_i , we are also assuming that u_{it} is not correlated
with either the x's observed prior to, or those observed after,
period t [i.e., $E^*(u_{\sim i} \mid x_{\sim i}, D_i) = 0$] . Since the models presented
here will, in general, be overidentified, one may, in any particular
example, wish to relax the assumptions of constancy of the lag para-
meters and $E^*(u_{\sim i} \mid x_{\sim i}, D_i) = 0$. These assumptions do, however, allow
for a relatively straightforward exposition of our results concerning
the estimation of lag structures in short panels.

It is convenient to rewrite Al as

$$y_{it} = \alpha_i + \sum_{\tau=0}^{t+\theta} w_\tau x_{i\,t-\tau} + b_{it} + u_{it} \tag{1}$$

where $b_{it} = \sum_{\tau=\theta+1}^{\infty} w_{t+\tau} x_{i\,-\tau}$, for i = 1, ..., N and t = 1, ..., T .

In (1), b_{it} is the truncation remainder (or the contribution of
presample x to the current value of y) for individual i in

period t . Let the matrix of coefficients obtained from the regression function of y_i on x_i be Π , defined by $E^*(y_i \mid x_i) = \Pi x_i$. We will say that a lag coefficient is identified if it can be calculated from the elements of Π .[1] Given Al, equation (1) makes it clear that each row of Π will contain a combination of the lag coefficients of interest and of the coefficients from the regression functions of the two unobserved components, α_i and b_{it} , on x_i . Our problem is, therefore, to separate out the lag coefficients from the coefficients defining these two regression functions.

The distinction between α_i and b_{it} is that α_i is constant over time while b_{it} is not. The estimation of models containing the α_i has been discussed extensively in the past (see Mundlak, 1978, and the literature cited there) and therefore the particular difficulties we shall encounter in identifying (at least some of) the lag coefficients from Π in our models arise from the existence of the truncation remainders (the b_{it}). Since $b_{it} = \sum_{\tau=\theta+1}^{\infty} w_{t+\tau} x_{i-\tau}$, in order to identify any one of the lag coefficients one has to restrict either the sequence of lag coefficients (the w_τ), or the relationship between presample x's and the observed data. This section deals with the second alternative.

---

1. This will be shown to be equivalent to stating that a lag coefficient is identified if it can be calculated from the second order population moments of $r_i' = (x_i', y_i')$.

II.1 The Identification Problem When There Is No Heterogeneity

The phrase 'no heterogeneity' will be used to denote a situation in which it is appropriate to assume that

A2: $\alpha_i = 0,$ for $i = 1, \ldots, N$.

Since the basic ideas underlying our results are easier to grasp when we maintain A2, and since the addition of the α_i does not change the identification status of the lag coefficients in the models discussed here, we begin with this special case.

Given both A1 and A2 the Π matrix (which, in this special case, we denote by Π^{nh}) has the form

$$\Pi^{nh} = W + B$$

where   (2)

$$
W = \begin{bmatrix}
0 & 0 & \cdots & & 0 & w_0 & w_1 & \cdots & w_\theta \\
0 & 0 & \cdots & 0. & w_0 & w_1 & w_2 & \cdots & w_{\theta+1} \\
\cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & & \cdot & \cdot & \cdot & & \cdot \\
0 & w_0 & w_1 & \cdots & & w_{T-3} & w_{T-2} & w_{T-1} & \cdots & w_{T+\theta-2} \\
w_0 & w_1 & w_2 & \cdots & & w_{T-2} & w_{T-1} & w_T & \cdots & w_{T+\theta-1}
\end{bmatrix}
$$

and B is defined by

$$E^*[b_i \mid x_i] = B x_i .$$

Clearly, in order to be able to identify any of the lag coefficients from Π^{nh} , B must have a structure which depends on only a few unknown parameters. B will be restricted if it is reasonable to

assume that the stochastic process generating the x sequences restricts the regression of presample on insample x . The particular special case that we assume is given by

A3: $\qquad E^*(x_{i \ -\theta-q} \mid \underset{\sim}{x}_i) = \sum_{j=1}^{p} \rho_j^{(q)} x_{i \ j-\theta-1}$

for $q \geq 1$; that is in the regression of the unseen past x's on $\underset{\sim}{x}_i$ only the first (the closest) p coefficients are non-zero.

A sufficient condition for A3 is that the regression of the $T + \theta - p$ most recent observations of x on their entire past history depends only on their preceding p values.[2] This would hold, for example, if the recent x's were a realization from a pth order autoregressive process. It would also be true for more general processes such as an integrated autoregression (where the order of the integration plus the order of the autoregression sum to p), with possibly changing coefficients over time.[3]

Since each element of $\underset{\sim}{b}_1$ is just a different linear combination of the same presample x's , given A3, the B matrix in (2) will have the form

$$B = [\ \overbrace{0}^{T+\theta-p} \ \vdots \ \overbrace{B_p}^{p} \ ] \ . \qquad (3)$$

2. Mathematically the condition is that for $q \geq 1$,

$E^*[(x_{iT}, \ \cdots, \ x_{i \ - \ \theta+p})' \mid (x_{i \ - \ \theta+p-1}, \ \cdots, \ x_{i \ -\theta} \ , \ x_{i \ -\theta-q})] =$
$E^*[(x_{iT}, \ \cdots, \ x_{i \ -\theta+p})' \mid (x_{i \ - \ \theta+p-1}, \ \cdots, \ x_{i \ -\theta})]$

3. See Anderson (1978) and the literature cited there for details on the use of such processes in multiple time series.

That is, B may be partitioned into a T by $T + \theta - p$ matrix of zeros and a T by p matrix of free coefficients (B_p). In this case then a consistent estimate of Π^{nh} yields consistent estimates of the leading $T + \theta - p$ lag coefficients without imposing any restrictions on the lag structure.

It will prove helpful to provide a more explicit solution for Π^{nh}. Noting that given A3

$$E^*(b_{it} \mid \underset{\sim}{x_i}) = \sum_{j=1}^{p} \phi_{t\ j-\theta-1}\ x_{i\ j-\theta-1} \tag{4a}$$

where $\phi_{t\ j-\theta-1} = \sum_{q=1}^{\infty} w_{t+\theta+q}\ \rho_j^{(q)}$ and using, for simplicity, the assumption that $p = \theta + 1$, the system of regression equations defined by $E^*[\underset{\sim}{y_i} \mid \underset{\sim}{x_i}]$ has the form

$$E^*[y_{i1} \mid \underset{\sim}{x_i}) = \sum_{j=1}^{P} \beta_{1\ j-\theta-1},x_{i\ j-\theta-1}$$

$$E^*(y_{i2} \mid \underset{\sim}{x_i}) = w_0 x_2 + \sum_{j=1}^{p} \beta_{2\ j-\theta-1},x_{i\ j-\theta-1}$$

$$\tag{4b}$$

$$E^*(y_{i3} \mid \underset{\sim}{x_i}) = w_0 x_3 + w_1 x_2 + \sum_{j=1}^{p} \beta_{3\ j-\theta-1},x_{i\ j-\theta-1}$$

$$\vdots$$

$$E^*(y_{iT} \mid \underset{\sim}{x_i}) = w_0 x_T + \cdots + w_{T+\theta-p} x_2 + \sum_{j=1}^{P} \beta_{T\ j-\theta-1},x_{i\ j-\theta-1}$$

where $\beta_{t\ j-\theta-1} = w_{t+\theta+1-j} + \phi_{t\ j-\theta-1}$, for $t = 1, \ldots, T$, and $j = 1, \ldots, p$.

(4) clarifies how one identifies the leading lag coefficients from Π under assumptions A1 to A3. Each year's value of y is regressed on current and all previous insample values of x. The last p values of x in each regression equation have non-zero partial correlations with the truncation remainders (the b_{it}). Hence their coefficients do not identify the parameters of the lag distribution. The $t + \theta - p$ leading coefficients in each equation are, in fact, estimates of the leading lag coefficients. As t increases, one gradually uncovers the lag structure. Of course, if $T + \theta$ is small relative to p, we will not be able to build up much information on the tail of the lag distribution. This simply reflects the fact that short panels, by their very nature, do not contain unconstrained information on that tail. However, even when $T + \theta$ is small, the initial consistent estimates of the first few lag coefficients privided by (4) may contain enough information on the lag structure to allow one to restrict it to be a member of a family of distributions which depend on a small number of parameters and concentrate on estimating those parameters thereafter. This gets slightly more complicated when there is also heterogeneity, a topic which we turn to now.

II.2 Identification With Heterogeneity

Recall that the reason for dealing explicitly with the α_i was to allow for the possibility of unobserved individual-specific factors which cause differences in both y_i given all the individual's past

x's , and in x_i .[4/] It follows that if there is heterogeneity, A3

is not likely to provide a good approximation to the regression func-

tion of presample on insample x . That is, if y conditional on all

past x's has an individual-specific component, and the diverse unob-

served factors which lead to that component have an independent effect

on insample x's , then these same factors are likely to have an inde-

pendent effect on presample x's . The simplest way to incorporate

this possibility into our framework is to allow the regression func-

tion of presample on insample x to depend directly on the α_i .

That is, we would drop A2 of the last subsection (thus allowing for

heterogeneity) and replace A3 with the assumption

$$A4: \quad E^*[x_{i,-\theta-q} \mid x_i, D_i] = c_q \alpha_i + \sum_{j=1}^{p} \rho_j^{*(q)} x_{i,-\theta+j-1} , \quad \text{for } q \geqslant 1 .$$

Note that A4 does constrain the individual-specific differences in the

x process to be proportional to the α_i . Since this assumption may

be too restrictive for many examples we develop below the generaliza-

---

4. If the α_i were not correlated with x_i , i.e., if

   $E^*(\alpha_i \mid x_i) = 0$, then, given A1, the α_i would become a set of

   'random' (or uncorrelated) effects in the sense of Balestra and

   Nerlove (1966). Since the discussion of II.1 did not restrict

   $E[u_i u_i' \mid x_i]$, in this case we can simply define $\tilde{u}_i = u_i + \alpha_i$,

   substitute $\tilde{u}_i$ for u_i in subsection II.1, and apply the results

   obtained there.

tions required when we replace the α_i in A4 with a separate unobserved individual specific component, say m_i .

Let $\tilde{x}_{it} = (x_{it} - a_t m_i)$ for $i = 1, \ldots, N$, all t , and some sequence $\{a_t\}$. Then if $\tilde{x}_{it}$ is substituted for x_{it} in the discussion of the sufficient conditions for A2 (see p.9) those conditions become sufficient for A4. Familiar examples of processes which satisfy this condition are processes in which, for

$$t = - \theta + p, \ldots, T ,$$

$$x_{it} = \gamma^t \alpha_i + \sum_{\tau=1}^{p} \kappa_\tau^t x_{i\ t-\tau} + \xi_{it} , \quad \text{where}$$

$E^*[\xi_{it} \mid D_i, x_{i\ t-1}, x_{i\ t-2}, \ldots] = 0$. With $\gamma^t = \gamma$ and $\kappa_\tau^t = \kappa_\tau$ for $t = -\theta + p, \ldots, T$, this is a mixture process consisting of an individual-specific component and an integrated autoregressive deviate whose coefficients are stable over time (MaCurdy, 1982, gives several examples of the use of such processes).

Using A4 to find $E^*(b_i \mid D_i, \underset{\sim}{x}_i)$ and substituting the result into A1 yields

___

5. A simple example where the restriction is relevant is provided by Mundlak (1961). In that article, the problem is to estimate a production function from panel data on the inputs (x) and output (y) of a group of firms operating in a competitive market. In this case the α_i represent unobserved interfirm productivity differences that are constant over time and known to management at the time input decisions are made.

$$E^*(y_{it} \mid \underset{\sim}{x}_i, D_i) \tag{5}$$

$$= c_t \alpha_i + \sum_{\tau=0}^{t+\theta-p} w_\tau x_{i\,t-\tau} + \sum_{j=1}^{P} \beta^*_{t\,j-\theta-1} x_{i\,j-\theta-1}$$

where $c_t = [1 + \sum_{q=1}^{\infty} w_{t+\theta+q}\, c_q]$, and $\beta^*_{t\,j-\theta-1} = w_{t+\theta+1-j} + \sum_{q=1}^{\infty} w_{t+\theta+q}\, \rho^{*(q)}_j$,

for $t = 1, \ldots, T$, and $j = 1, \ldots, p$. Note that in equation (5) the α_i have different coefficients in the different years of the sample. This follows from the fact that the truncation remainders are different linear combinations of the presample x's. If $w_\tau = 0$ for $\tau > T + \theta$, that is, if the length of the lag is less than or equal to $T + \theta$, then there is no truncation remainder in period T, and $c_T = 1$. In the general case, however, equation (5) has a single unobserved factor with free 'factor loadings' (the c_t). Since the contribution of the α_i in year t can only be identified relative to its contribution in some other year, a normalization of the factor loadings is required and one which is consistent with the possibility that the length of the lag is less than or equal to $T + \theta$ is $c_T = 1$. To show which lag coefficients are identified in this case we rewrite (5) as

$$y_{it} - \bar{c}_t y_{i1} = \sum_{\tau=0}^{t-1} w_\tau x_{i\,t-\tau} + v_{it} ,$$

for $i = 1, \ldots, N$ and $t = 2, \ldots, T$, where $\bar{c}_t = c_t/c_1$, and $v_{it} = y_{it} - c_t y_{i1} - E^*(y_{it} - c_t y_{it-1} \mid \underset{\sim}{x}_i, D_i)$, so that $E^*(\underset{\sim}{v}_i \mid \underset{\sim}{x}_i, D_i) = 0$ by construction, and we have assumed for simplicity that $\theta + 1 = p$. This is a standard simultaneous equation system and will identify the leading $T + \theta - p$ lag coefficients (if

$\theta + 1 = p$, $T + \theta - p = T - 1$) if there is an equation (year) in which at least one x is excluded from the determinants of $y_{it} - \bar{c}_t y_{i1}$ while still being correlated with y_{i1}. If $T \geq 3$, then x_{i3}, x_{i4}, $\ldots$, x_{iT} are excluded from the equation determining $y_{i2} - \bar{c}_2 y_{i1}$, and provided they have a non-zero partial correlation with the α_i, they will be correlated with y_{i1}.

To find the Π matrix corresponding to (5) (that is, the matrix under assumptions A1 and A4 which we denote by Π^{sh}), we use the fact that $E^*(\underset{\sim}{y_i} \mid \underset{\sim}{x_i}) = E^* \{ (E^*(\underset{\sim}{y_i} \mid \underset{\sim}{x_i}, D_i) \mid \underset{\sim}{x_i} \}$, to derive

$$\Pi^{sh} = W + B + \underset{\sim}{c} \underset{\sim}{\alpha}' \tag{6}$$

where W and B are defined in (2) and (3), $\underset{\sim}{c}' = [1, c_{T-1}, \ldots, c_1]$, and α_i denotes the vector of coefficients resulting from the regression of α_i on $\underset{\sim}{x_i}$ [i.e., $E^*(\alpha_i \mid \underset{\sim}{x_i}) = \sum_{r=-\theta}^{T} \alpha_r x_{ir}$]. Recall that $\Pi^{nh} = W + B$ so that the difference in the Π matrices that results from allowing for the α_i (heterogeneity) is just $\underset{\sim}{c} \underset{\sim}{\alpha}'$, the matrix of coefficients obtained from the regression of $\underset{\sim}{c} \alpha_i$ on $\underset{\sim}{x_i}$.

If one were to allow for two different sources of heterogeneity, one in the distributed lag model (the α_i) and one in the process generating the x's (the m_i), then to identify the leading lag coefficients we would require $T \geq 4$, and the Π matrix (which in this case we denote by Π^{th}) would be given by

$$\Pi^{th} = W + B + \underset{\sim}{c}^* \underset{\sim}{m}' + \underset{\sim}{\ell} \underset{\sim}{\alpha}' \tag{7}$$

where W , B , and $\underset{\sim}{\alpha}$ are defined as in (6), $\underset{\sim}{c}^{*\prime} = [1, c_1^*, ..., c_T^*]$, $\underset{\sim}{m}$ is the vector of coefficients obtained from the regression of

m_i on $\underset{\sim}{x}_i$ [i.e. $E^*(m_i \mid \underset{\sim}{x}_i) = \sum_{r=-\theta}^{T} m_r x_{ir}$] , and $\underset{\sim}{\ell}$ here, and in the discussion below, is a vector of ones. Note that when two sources of heterogeneity are allowed for it is m_i which has a variable coefficient over time, and if the length of the lag is less than $T + \theta$, its coefficient in year T is zero. To allow for this possibility the normalization, $c_1^* = 1$ is used (instead of $c_T = 1$ as in Π^{sh}).[6]/

II.3 Some Further Considerations: Estimation and Testing

Having discussed the identification of the lag parameters under alternative assumptions, we turn now to the consideration of some issues associated with the choice of estimators and test statistics for such models.

In the context of panel data one is interested in the limiting properties of estimators as N (the number of individuals) grows large, and these depend on how the sample moments behave as the sample grows in this dimension. Chamberlain (1982) provides an explicit discussion of the properties of alternative estimators of Π and of relevant test statistics under the assumption that the N obser-

---

6. With two sources of heterogeneity there is an additional normalization to be imposed on one of the elements in $\underset{\sim}{\alpha}$ or in $\underset{\sim}{m}$.

vations on the vector $\underset{\sim}{r}_i' = (\underset{\sim}{x}_1', \underset{\sim}{y}_i')$ are independent draws from

some multivariate distribution with finite fourth order moments. Note

that neither this assumption, nor any of the assumptions made prior to

it, imply that $E\{(\underset{\sim}{y}_i - \Pi\underset{\sim}{x}_1')(\underset{\sim}{y}_i - \Pi\underset{\sim}{x}_i)' \mid \underset{\sim}{x}_i\} = \Omega$ (independent of

i) . Thus random coefficients models, and models in which there is a

complicated non-linear relationship between the α_i (or m_i) and

$\underset{\sim}{x}_i$ [i.e. $E^*(\alpha_i \mid \underset{\sim}{x}_i) \neq E(\alpha_i \mid \underset{\sim}{x}_i)$] do not necessarily violate any of

our assumptions (for a discussion of related points and alternative

stochastic assumptions, see, in addition, White, 1980, 1980a, 1980b,

and 1982). With these assumptions maximum likelihood (based on a nor-

mal density for $\underset{\sim}{r}_i$) or an asymptotically equivalent estimator of Π

is (strongly) consistent and asymptotically normal but need not be

efficient, nor need its covariance matrix be given by the traditional

information matrix (see also MaCurdy, 1981). The point to note here

is that unless further assumptions on the covariance matrix of distur-

bances from the regression of $\underset{\sim}{y}_i$ on $\underset{\sim}{x}_i$ are relevant, one must

take care in formulating standard errors and test statistics from maxi-

mum likelihood estimation procedures (Chamberlain, 1982, provides both

the appropriate formulae and more efficient estimators, under these

assumptions). Of course, if $E\{(\underset{\sim}{y}_i - \Pi\underset{\sim}{x}_i)(\underset{\sim}{y}_i - \Pi\underset{\sim}{x}_i) \mid \underset{\sim}{x}_i\} = \Omega$

then maximum likelihood is efficient, and variances and test

statistics are formed in the usual manner.[7/]

Looking at the structure of the Π matrices, it is clear that for $t \geq 4$, Π^{th} can be obtained from a set of restrictions on Π , Π^{sh} from restrictions on Π^{th} , and Π^{th} from restrictions on Π^{sh} . There is, therefore, a nested sequence of restrictions that can be used to decide between alternative models. In particular, if there is no heterogeneity, then the elements of $\Pi = [\pi_{tr}]$ ought to be close to zero for $r > t \geq p - \theta$ (that is all leading x's for years where $t \geq p - \theta$ should have insignificant coefficients); and its elements should have a "stationary structure" (should depend only on $t - r$) for $t \geq r > p - \theta$. The test for a model with no heterogeneity involves, therefore, only equality and zero restrictions.

---

7. Note that our discussion is based on estimating the Π matrix rather than on estimating the form of the model containing either or both of the sequences $\{\alpha_i\}_{i=1}^{N}$ and $\{m_i\}_{i=1}^{N}$ (the system in equation (5) for example). Since they affect the distribution of only T (a small number) of sample points, one cannot estimate them consistently as $N \to \infty$ and an attempt to do so would lead, in general, to inconsistent estimates of the other parameters of interest. That is, in our problem $\{\alpha_i\}_{i=1}^{N}$ and $\{m_i\}_{i=1}^{N}$ are sets of incidental parameters in the sense of Neyman and Scott (1948). By going directly to the Π matrix we are, in fact, constructing the density of $y_{\sim i}$ conditionals on $x_{\sim i}$, and the structural parameters of the model (i.e. the parameters governing the distributions of each $r_{\sim i}$), but marginal to the model's incidental parameters; see Kiefer and Wolfowitz (1956).

This assumes that p is known at the outset. If this is not the case then one can build a nested testing sequence for p by starting with p sufficiently large and testing the constraints on Π arising from lower values of p. In all cases the test of the null hypothesis that $p = p^* - 1$ against the alternative that $p = p^*$ consists of testing whether the left row of the submatrix B_p^* (see equation (3)) is a vector of zeroes. Provided $p^* \geq \theta + 1$, the null hypothesis imposes $T - 1$ additional constraints on the Π matrix.[8/]

Note also that if we denote the matrix of second order population moments

$$\Sigma_{rr'} = \begin{bmatrix} \Sigma_{xx'} & \Sigma_{xy'} \\ & \Sigma_{yy'} \end{bmatrix} \quad ,$$

___

8. The choice of rejection criteria for such a sequence of tests is discussed in Geweke and Meese (1979). Their discussion is likely to be particularly relevant when p can take on a wide range of values (so that a long sequence of tests may be relevant). In our case, one is likely to have fairly narrow _a priori_ bounds on p. The assumptions in the text constrain only Π. If, however, one were willing to assume that the leads which receive zero coefficients in the projection of past x's on future x's , (or on future x's and D_i) are independent of time, then one could test A2 (A4) directly by considering the regression of early observed x's on later ones (or on the later ones and D_i , and the projection of D_i onto $\underset{\sim}{x}_i$). The procedure should prove helpful in suggesting a relevant range for p.

then the only constraints our assumptions place on $\Sigma_{rr'}$, are con-
straints on $\Pi' = [\Sigma_{xx'}]^{-1} \Sigma_{xy'}$; a matrix which also contains all
the parameters of interest. Assumption A2 or A4 does impose an addi-
additional set of constraints on moments from the joint distribution
of presample and insample x , but the effect of either assumption on
$\Sigma_{rr'}$, is limited to its effect on Π . However, if one can assume
that the process generating $x_{\sim i}$ is one of the simpler stochastic pro-
cesses that satisfies A2 or A4), then the constraints are in terms of
the elements of both Π and $\Sigma_{xx'}$. In this case more efficient esti-
mates of the lag coefficients and more powerful test statistics can be
obtained by imposing the constraints on both of these matrices. For
example, if $x_{it} = m_i + z_{it}$, where z_{it} is a pth order auto-
regressive deviate with partial correlations ρ_1 , ..., ρ_p and
variance σ^2 for i = 1, ..., N and all t (this is a simple mixture
process satisfying A4), then $\Sigma_{xx'} = E[x_{\sim i} \, x'_{\sim i}] = \sigma_m^2 \, \ell \ell' + \Sigma^*$, where
$\sigma_m^2 = E \, m_j^2$, and Σ^* is the covariance matrix for $T + \theta$ consecutive
observations on a pth order autoregressive deviate (formulae for the
elements of Σ^* in terms of the parameters ρ_1 , ..., ρ_p and σ^2 are
provided in Galbraith and Galbraith, 1974). Clearly then the $(T + \theta) \cdot$
$(T + \theta + 1) \, / \, 2$ elements of $\Sigma_{xx'}$ are a function of only $p + 2$

parameters. Moreover, these same parameters determine some of the elements of Πth (or Π^{sh}).[9/]

Finally, though our discussion has been concerned with models which contain only a single regressor it is not difficult to extend our results to cases where there are more than one of them. In this case there would be a vector of x's for each individual in every

---

9. To see this note that in equation (7), $\underset{\sim}{m} = \Sigma_{xx'}^{-1} \underset{\sim}{\ell} \, q_m^2$. It is straightforward to verify that $\Sigma_{xx'}^{-1} = \Sigma^{*-1} - \theta_1 \, \Sigma^{*-1} \, \underset{\sim}{\ell}\underset{\sim}{\ell}' \, \Sigma^{*-1}$ where $\theta_1 = \sigma_m^2 / (1 + \underset{\sim}{\ell}' \, \Sigma^{*-1} \, \underset{\sim}{\ell}\sigma_m^2)$. Thus $\underset{\sim}{m} = \Sigma^{*-1} \underset{\sim}{\ell} \, \theta_1$, which is a function of the same parameters as $\Sigma_{xx'}$ (in fact $m_r = m$ for $T - p > r > p - \theta$, while $m_{-\theta+j} = m_{T-j}$ for $0 \leq j \leq p$). Finally the c_q and $\rho_j^{(q)}$ coefficients in A4 will be determined from the backcasting function for this process. In particular the $\rho_j^{(q)}$ will follow the recursion formulae for backcasting coefficients from autoregressive processes $(\rho^{(q+1)} = \rho_j \rho_1^{(q)} + \rho_{j+1}^{(q)}$, where it is understood that $\rho_j^{(.)} = 0$ for $j > p$ and $\rho_j^{(1)} = \rho_j$ for $j = 1, \ldots, p$; see Box and Jenkins, 1970), and $c_q = 1 - \sum_{j=1}^{P} \rho^{(q)}$. Given the definition of c [equation (5)], and of B and W [equation (3)], these formulae provide an additional set of non-linear constraints connecting the elements of Π to those of $\Sigma_{xx'}$.

period, say, x_{it} , and one requires assumptions on the regression

of presample x_{it} on insample x_{it} . The assumptions which are

analogous to A2 and A4 are, respectively,

$$E^*[x_{i \, -\theta-q} \mid x_{i \, -\theta} , \cdots , x_{iT}] = E^*[x_{i \, -\theta-q} \mid x_{i \, -\theta} , \cdots , x_{i \, -\theta+p}] \quad \text{and}$$

$$E^*[x_{i \, -\theta-q} \mid x_{i \, -\theta} , \cdots , x_{iT}, D_i] = E^*[x_{i \, -\theta-q} \mid x_{i \, -\theta} , \cdots , x_{i \, -\theta-p}, D_i]$$

for $q \geq 1$. These would be satisfied, for example, by a time inhomo-

geneous p^{th} order vector autoregression, and by a mixture process

consisting of individual specific components (one for each regressor)

and a pth order vector autoregression.[10/] In either case the Π

matrix would allow for the coefficients of the earliest p values of

each regressor in every year to be unconstrained regardless of how many

of the regressors had a distributed lag effect on y .

III. Distributed Lags from Panel Data with Prior Structure on the Lag Coefficients

As noted earlier in order to identify any of the lag coefficients

from panel data one must impose some structure on the relationship

between the truncation remainders and the observed data. The last sec-

tion discussed the possibility of restricting the relationship between

presample and insample x ; in this section we consider restricting the

lag coefficients themselves. The particular example we focus on is a

sequence of lag coefficients which, after a few free lags, has an auto-

---

10. Kiefer, 1979, discusses the problem of determining whether a

smaller number of individual-specific components would suffice.

regressive structure [such structures have been widely used the past; see, e.g., Dhrymes (1971), Griliches (1967) and Nerlove (1972)]. This restriction is formalized in the assumption

A5:
$$
w_\tau = \begin{cases} w_\tau & \text{for } \tau \le k_1 \\ \sum\limits_{q=1}^{Q} \delta_q w_{\tau-q} & \text{otherwise} \end{cases}
$$

where the roots of the characteristic equation $1 - \sum\limits_{q=1}^{Q} \delta_q z^q = 0$, say $\lambda_1^{-1} \ldots, \lambda_q^{-1}$ lie outside the unit circle. For simplicity we assume that $k_1 = \theta$, and that $\lambda_1, \ldots, \lambda_Q$ are all real and distinct.

If A5 is substituted into A1 the set of truncation remainders from the resulting distributed lag model follow an exact Q^{th} order auto-regression. They can, therefore, be written as $b_{it} = \sum\limits_{j=1}^{Q} \lambda_j^t b_{ij}$ where $b_{ij} = w_{\theta-j+1} \sum\limits_{\tau=1}^{\infty} \lambda_j^\tau x_{i-\tau-\theta}$; that is, $b_{i1}, \ldots, b_{iQ}$ are a set of Q unobserved initial conditions. Under A5, then, the distributed lag model A1 becomes a system of T regressions with $Q + 1$ freely correlated unobserved factors (α_i , $b_{i1}, \ldots, b_{iQ}$), Q of them having factor loadings which decay geometrically over time. Since the identification properties of a model in which there are $Q + 1$ unob-served factors are a straightforward generalization of a model with two of them, we deal first with the familiar case of a (modified) geometric or Koyck lag in which $Q = 1$, and then point out the modi-fications required for $Q > 1$.

Setting $Q = 1$ into A5, and substituting the result into A1, we have:

$$y_{it} = \alpha_i + \sum_{\tau=0}^{\theta} w_\tau x_{i\ t-\tau} + w_\theta \sum_{\tau=\theta+1}^{t+\theta} \delta^{\tau-\theta} x_{i\ t-\tau}$$

$$+ \delta^{t-1} b_i + u_{it} \qquad\qquad (8)$$

where $b_i = w_\theta \sum_{\tau=1}^{\infty} \delta^\tau x_{i\ -\tau-\theta}$, that is, b_i is the truncation

remainder in period one (and the initial condition when $Q = 1$) for

$i = 1, \ldots, N$ and $t = 1, \ldots, T$.

Given the discussion of the last section it is not difficult to
see why the lag coefficients in (8) are identified from the Π
matrix. Recall that to identify the lag parameters one requires a set
of restrictions on the matrix $B = [\beta_{tr}]$, where $E^*[b_{\sim i} \mid x_{\sim i}] = Bx_{\sim i}$.
Given our assumptions $b_{it} = \delta b_{i\ t-1}$, so that $E^*[b_{it} \mid x_{\sim i}] =$
$\delta E^*[b_{i\ t-1} \mid x_{\sim i}]$ which implies that $\beta_{tr} = \delta \beta_{t-1\ r}$ for
$r = 1, \ldots, T + \theta$ and $t = 2, \ldots, T$.$\underline{11/}$

11. Note the difference here between the estimates of (modified)
    geometric lag structures from panel data and those from a single
    time series. In the latter case the relevant limiting dimension
    is T , and as T gets larger the dependence of consistent esti-
    mators of the lag coefficients on the value chosen for b_i
    generally disappears (Dhrymes, 1971). When T is short this
    result no longer applies. In particular with large N as in
    panel data we have a set of truncation remainders and the proper-
    ties of their distribution (i.e., $E^*(b_i \mid x_i)$) do affect the pro-
    perties of alternative estimates of the lag coefficients (see
    also Pesaren, 1973).

It follows that the Π matrix (which in this case we denote by Π^g)
has the form:

$$\Pi^g = W^* + \underset{\sim}{\delta}\underset{\sim}{\beta}' + \underset{\sim}{i}\underset{\sim}{\alpha}' \tag{9}$$

where $\underset{\sim}{i}$ and $\underset{\sim}{\alpha}$ are defined as in (7) , $\underset{\sim}{\delta} = [1, \delta', \ldots, \delta^T]$, $\underset{\sim}{\beta}$
is the vector of coefficients from the regression of b_i on $\underset{\sim}{x}_i$

$[\text{i.e.,} \quad E^*(b_i \mid \underset{\sim}{x}_i) = \sum_{r=-\theta}^{T} \beta_r x_{ir}]$ and

$$W^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & w_0 & w_1 & w_\theta \\ 0 & 0 & \cdots & 0 & w_0 & w_1 & w_2 \cdots w_\theta \delta \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & \cdot \\ w_0 & w_1 \cdots w_\theta & w_\theta \delta \cdots & w_\theta \delta^{T-\theta-2} & w_\theta \delta^{T-\theta-1} & \cdots w_\theta \delta^{T-1} \end{bmatrix}$$

Π^g will identify the lag coefficients provided $T \geq 3$.$\underline{12/}$ In the
more general case with $Q > 1$ one would replace $\underset{\sim}{\delta}\underset{\sim}{\beta}'$ in (9) by

---

12. To see this note that (8) implies that

$$(y_{it} - y_{i\,t-1}) - \delta(y_{i\,t-1} - y_{i\,t-2}) = w_0 x_{it} + [w_1 - w_0(1-\delta)] x_{i\,t-1}$$

$$+ \sum_{\tau=2}^{\theta} [w_\tau - (1-\delta)w_{\tau-1} + w_{\tau-2}] x_{i\,t-\tau} + v_{it}$$

where $E^*[\underset{\sim}{v}_i \mid \underset{\sim}{x}_i] = 0$ for $i = 1, \ldots, N$ and $t = 2 , \ldots, T$.
In this form we have a standard simultaneous equations system.
Provided $T \geq 3$, $x_{i3}, \ldots, x_{iT}$ will serve as instruments for the
equation determining $y_{i2} - y_{i1}$.

$$\sum_{j=1}^{Q} \underset{\sim}{\delta}_j \underset{\sim}{\beta}_j{}'$$ and the Π matrix would identify the lag coefficients if

$T \geq Q + 2$.

One further point is worthy of note. In this section we have not restricted the relationship between presample and insample x in any way, and as a result the Π matrix resulting from assumptions A1 and A5 are not, in general, nested inside the Π matrices of Section II. If, however, in addition to A5 we assume A2 or A4 of the last section, then the Π matrices from the models in this section can be simplified further (the additional restrictions are straightforward special cases of the restrictions developed in Section II) and the models presented here will be nested to those presented in the last section. Thus, if either A2 or A4 were relevant, and the initial consistent estimates of a free lag distribution obtained from the techniques developed in the last section indicated that one of the simple lag structures satisfying A5 seemed plausible, one could test and then estimate this lag structure in a second round.

IV. AN EXAMPLE: ESTIMATING THE SHAPE OF THE DEPRECIATION FUNCTION

A common empirical construct is the "stock of capital" based on some variant of the perpetual inventory method. It is, in fact, a distributed lag function of past investment expenditure with the depreciation formula implying a set of lag weights (w's). These weights are, usually, assumed a priori based on mortality assumptions (straight line, geometric, etc.) and some scattered evidence on the length of life of selected assets. Alternatively, one accepts the

depreciation assumptions used by firms and reported in their annual statements, even though they may be heavily affected by tax and other accounting considerations that have little to do with the economic measure of depreciation one is after. In general, there is no obvious unique right way to construct "capital"; it all depends on the purpose for which such a variable is to be constructed (Cf Griliches 1963, and the articles in Usher, ed. 1980).

One of the more common uses for capital is to compute and compare "rates of return" across projects, firms, or industries. Such measures assume implicitly that there exists a stable relationship between earnings (gross or net profits) and past investments, and that firms or industries differ only by a factor of proportionality in the yield on their past investments, with the time shape of these yields being the same across firms and implicit in the assumed depreciation formula. In this section we investigate this question empirically, using panel data and the methods developed in Section II to estimate an unconstrained distrubuted lag relationship between firm profits and their past investments.

The data come from the Compustat tape (1980). Profits are for the years 1964-72 inclusive (thus $T = 9$) and are defined to equal the operating income of firms deflated by the implicit G.N.P. deflator and an index of the average gross rate of return (1972 = 1.0) taken from Feldstein and Summers (1977, table 1, column 3, page 216). The observed investment data are for the years 1961-71 (thus, assuming that this year's investment does not effect this years profits, we have two presample observations on investment). Investment is defined to equal the reported gross investment of firms deflated by the impli-

cit G.N.P. investment deflator. Firms were chosen from the tape if
there was no evidence of a major merger, acquisition, or sale of a
part of the firm during the sample period; and if all the required
variables for the firm were available. This yielded a sample (N) of
258 firms.[13]/

All results reported here are based on two simplifications.
First, after making an adjustment for heteroskedasticity, we shall
assume that the conditional (on x_i) covariance matrix of distur-
bances from the regression of y_i on x_i does not vary over obser-
vations. This implies that maximum likelihood estimation procedures
provide consistent estimates of all parameters and their standard
errors, and maximum likelihood is used throughout. The adjustment
made for heteroskedasticity was to divide all figures for each firm by
the square root of the average rate of investment over the 1961-71
period as a whole. The second simplification to be used here concerns
the relationship between presample and insample x . Our results
focus on the model which allows for heterogeneity and, therefore,
assumption A4 will be used to describe this relationship. It will be
assumed that p in A4 equals three. This is based on a preliminary
examination of the observed x sequences (see note 8) which indicated
that a p of 2 or 3 would do.

_____

13. It also, no doubt, resulted in some selection bias. The esti-
    mates presented here do, however, allow for firm effects or
    heterogeneity, and this should take care of most of the sample
    selectivity problem.

The Compustat tape for the years of interest contains more information on the history of the firms x's than just the reported investment series. It reports also the firm's gross (k^g) and net (k^n) capital stocks at the beginning of each year. These numbers are essentially linear combinations of the firm's past investment expenditures (including investments in years that we do not have in our data). Net capital makes adjustments for reported depreciation expenses while k^g does not.[14/] The availability of k^n allows us to make comparisons between the results from the models based on the techniques developed in this paper and results that would be obtained if one assumed A1 but constrained the lag coeffiecients such that

$$k^n_{it} = \sum_{\tau=0}^{\infty} w_\tau x_{i\, t-\tau} \; ,$$ (a model which assumed that firms report the

relevant measure of capital stock). To build the comparison model, note that under the latter assumptions $E^*[y_{it} \mid k^n_{\sim i}, D_i] = \alpha_i + a_1 k^n_{it}$ and, $E^*[\alpha_i \mid k^n_{\sim i}] = a_2 k^n_{i.}$, so that

---

14. Though there are some additional adjustments, the basic recursion formulae for k^g and k^n are: $k^g_{it} = k^g_{i\, t-1} + x_{i\, t-1} - dis_{t-1}$ where dis_{t-1} = discards in period t-1 , and; $k^n_{it} = k^n_{i\, t-1} + x_{i\, t-1} - dis_{t-1} - dep_{t-1}$, where dep_{t-1} is the firms reported depreciation in period t-1 . Depreciation is that part of past investments written off in the given year. The precise formulae for the fraction of each year's investment that is written off in subsequent years differ among firms, the most frequently used being a straight line depreciation scheme.

$$y_{it} = a_1 k^n_{it} + a_2 k^n_{i.} + v_{it} \qquad (10)$$

where $E*[v_i \mid k^n_i] = 0$, and $k^n_{i.}$ is each firm's mean k^n_{it} over the

sample period, for $i = 1, \ldots, N$, $t = 1, \ldots, T$.

Here we shall also assume that $E[v_i v'_i \mid k^n_i] = \Omega_k$, and use

maximum likelihood estimation techniques. Note that the model in (10)

is not nested to the model in Section II which allows for a single

source of heterogeneity. In particular though k^n_{it} imposes the par-

ticular lag structure implicit in the Compustat calculations, it also

uses information on (presample) x's not used in the models of the

last section.

We shall also make use of the information contained in k^g . Note

that if $k^g_{i \, -\theta} = \sum_{\tau=1}^{\infty} f_\tau x_{i \, -\theta-\tau}$, and the model with a single source

of heterogeneity is used then that model (A4) implies that

$$k^g_{i \, -\theta} = \sum_{\tau=1}^{p} \beta_{k\tau} x_{i \, -\theta+\tau-1} + c_k [\sum_{r=-\theta}^{T} \alpha_r x_{ir}] + \varepsilon_{ki} \qquad (11)$$

where $E*[\varepsilon_{ki} \mid x_i] = 0$, for $i = 1, \ldots, N$. If we assume that

$E[\varepsilon^2_{ki} \mid x_i] = \sigma^2_k$, and allow ε_{ki} to be freely corrected with

$\varepsilon_i = y_i - \Pi x_i$, then equation (11) can be added to the system in (6),

to produce a system with $T + 1$ equations, which, since it uses the

information on the presample x's contained in k^g , should provide

more precise parameter estimates than one could derive from using (6)

alone.

Table 1 summarizes the parameter estimates. Estimates of the
unconstrained Π matrix indicated that there was heterogeneity (see
p.18). Column (1) lists the estimated lag coefficients based on
allowing for a single source of heterogeneity (the model of assump-
tions A1 and A4, or the Π matrix in (6)) while column 2 adds
equation (11) to this system. The algorithm used to compute the maxi-
mum likelihood estimates of column (1) did not actually meet its con-
vergence criteria. It seemed that the data could not distinguish
effectively between alternative combinations of the vectors $\underset{\sim}{\alpha}$ and $\underset{\sim}{c}$
and the last few lag coefficients when using system (6); but conver-
gence was rapid once we added equation (11). The resulting estimates,
listed in column (2), yield w_1 through w_8 which are all positive,
the first seven being estimated with a fair amount of precision. As
one should expect from the structure of the Π matrix, the standard
errors of the lag coefficients rise monotonically in τ . While it is
clear that the lag coefficients increase until $\tau = 3$; w_3 , w_4 , w_5 ,
w_6 , w_7 and w_8 are not very different from each other (though there
is weak evidence of a slight fall in the values of the last one,
perhaps two, coefficients). The observed value of the χ^2_4 test sta-
tistic of the null hypothesis that $w_3 = w_4 = w_5 = w_6 = w_7 = w$ was
2.05 which is below the expected of a χ^2_4 deviate, and imposing this
restriction produced the estimates of column (3). A test of the addi-
tional constraint that $w_8 = w$, produced an observed χ^2_1 test sta-
tistic of 1.91 (still not significant at usual levels of significance),

Table 1

The Relationship of Profits to Past Investment Expenditures for U.S. Manufacturing Firms:
Parameter Estimates Allowing for Heterogeneity*

Parameter (Standard Error)	Without k^g_{-2} (System 6)	With k^g_{-2} (System 6+11)	Column 2 with $w_3=\ldots=w_7=w$	Column 2 with $w_3=\ldots=w_8=w$	Comparison Model (System 10)	3 Years Investment $+ k^n_{i,t-4}$	3 Years Investment $+ k^g_{i,t-4}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
w_1	0.067 0.028	0.068 0.027	0.073 0.026	0.061 0.026		0.073 0.022	0.057 0.021
w_2	0.115 0.033	0.112 0.032	0.118 0.030	0.097 0.030		0.104 0.022	0.077 0.022
w_3	0.224 0.041	0.222 0.040	0.224 0.032	0.191 0.032		0.141 0.024	0.120 0.024
w_4	0.172 0.046	0.208 0.046	.	.			
w_5	0.072 0.049	0.198 0.050	.	.			
w_6	0.096 0.062	0.277 0.057	.	.			
w_7	-0.122 0.094	0.202 0.076	0.224 0.032	.			
w_8	-0.259 0.133	0.087 0.103	0.063 0.097	0.191 0.032			
Coefficient of:							
$k^n_{i,t}$					0.095 0.012		
$k^n_{i,t-4}$						0.103 0.011	
$k^g_{i,t-4}$							0.045 0.006

Table 1 (continued)

* The dependent variable is gross operating income deflated by the implicit GNP deflator and an index of the overall rate of return in manufacturing (1972 = 1.0). The w_τ refer to the coefficients of gross investment expenditures in period $t-\tau$ deflated by the implicit GNP producer durable investment deflator. k^n_{it} and k^g_{it} are the Compustat measures of net and gross capital at the beginning of the year deflated by the deflator provided by Mairesse and Siu, 1981. k^g_{-2} refers to undeflated gross capital in 1961 as reported by Compustat. All variables are divided by the square root of the firm's mean investment expenditures over the 1961-71 period. Dummy variables for the nine time periods are included in all equations. $N = 258$ and $T = 9$.

and imposing this constraint resulted in the estimates of column
(4).[15/] The estimated lag coefficients, then, rise over the first

three periods and remain fairly constant over the next four or five.

Note that this pattern is distinctly different from the pattern

implied by either straight line or declining balance depreciation.

Both of these imply coefficients which decline monotonically in τ ,

and in the second case the decline is greatest in the earlier periods.

Column (5) of Table 1 provides the coefficient of k_{it}^n obtained

from estimating the comparison model (equation 10) while Table 2 pro-

vides some measures of goodness of fit. All measures of goodness of

fit are based on the trace of the estimated covariance matrices of

disturbances from the alternative models, and the traces themselves

are presented in row 1 of the table. Rows 2 and 3 normalize, that is

divide, row 1 by alternative numbers. In row 3 we divide by $\sum_{t=1}^{T} s_{yt}^2$,

where s_{yt}^2 is the sample variance of profits in year t , so that one

minus the numbers in this row provide an indication of the fraction of

the variance in profits over the nine year period accounted for by the

model estimated. In row 2 we divide by the trace from a model where

the coefficients in the regression of $\underset{\sim}{y}_i$ on $\underset{\sim}{x}_i$ are left

unconstrained; that is the model which allows for a free Π matrix.

Comparing the first two columns of the table we see that the trace for

the comparison model (which uses A1 and k_{it}^n as defined by Compustat)

is about 75 percent larger that the trace of the model underlying

---

15. A test of the constraint that $w_\tau = \delta w_{\tau-1}$ for $\tau \geq 8$ was

   strongly rejected by the data.

Table 2

Measures of Goodness of Fit for Table 1

Measure	Column 2 (1)	Column 5 (2)	Column 6 (3)	Column 7 (4)	Free Π Matrix (5)
1. Trace $\hat{\Omega}$ [1]	298.7	515.8	342.5	347.1	253.6
2. (Trace $\hat{\Omega}$)/253.6	1.18	2.04	1.35	1.37	1.0
3. (Trace $\hat{\Omega}$)/1208.4 [2]	0.25	0.43	0.28	0.29	0.21

[1] $\hat{\Omega}$ = estimated covariance matrix of disturbances from the system of profit regressions.

[2] $1208.4 = \sum_{t=1}^{T} s^2_{yt}$, where s^2_{yt} is the sample variance in y_t.

the estimates of column (2) in Table 1 (which use A1 and A4 of Section II). The latter figure is still, however, 18 percent larger than the trace of a model which allows for a free Π matrix (the trace of the comparison model is 104 percent larger than the trace from the model with the free matrix). The model with a free Π accounts for about 79 percent of the total variance in profits, the model based on A1 and A4 accounts for about 75 percent of this variance, while the comparison model accounts for about 57 percent of it.

Coming back to the lag structure, recall that the estimates of column (2) suggested that it rose over the first three periods and was fairly flat over the next four or five, with perhaps a slight decline in the last few coefficients estimated. A simple way to approximate such a lag structure might, therefore, be to allow for three free lag coefficients and then substitute either $k^g_{i\ t-4}$ or $k^n_{i\ t-4}$ for

$\sum_{\tau=4}^{\infty} w_\tau x_{i\ t-\tau}$ (the choice would depend on the rate of decline in the coefficients which were not estimated). Column (6) (which uses $k^n_{i\ t-4}$) and column (7) (which uses $k^g_{i\ t-4}$) of Table 1 present estimates of these models (the system estimated is constructed in a manner analogous to the construction of the system for the comparison model, see the discussion of system 10). The estimated coefficients in both these columns also rise over the initial three year period. Turning to columns (3) and (4) of Table 2, it is clear that the trace of the estimated coveriance matrices of disturbances for the systems which allow for these three free initial coefficients are much smaller than the trace of the covariance matrix for the comparison model

(which does not). Their values are only 66 percent of the values of
the trace for the comparison model, but they are still over 15 percent
larger than the trace of the covariance matrix which uses the esti-
mates of column (2).

So far we have compared the estimates from the model based on a
single source of heterogeneity and a set of free lag coefficients to a
model which uses capital stock figures from the Compustat tape, and,
in addition, shown that the model's estimates seem to be consistent
with the information from auxiliary calculations. Though the model
does well in these respects, it does not really provide a completely
adequate summary of the data. This becomes evident when we formally
test the restrictions on the Π matrix embodied in the system con-
sisting of equations (6) and (11). The observed value of the $\chi^2_{55}/55$
test statistic for the restrictions was 5.64 which indicates rejection
of the null hypothesis at traditional levels of significance.[16] Thus
though the model underlying the estimates of column (2) seems to pro-
vide a better summary of the relationship between past investment and
current profits than conventional capital measures, none of these
models provide a really adequate account of the interaction between
profits and past investment expenditures. This model failure
reflects the basic implausibilities in the original maintained hypo-
thesis: (a) that the relationship between investment and profits is
essentially the same for different firms and industries and, (b) that

---

16. Though in judging this test statistic one should keep in mind the
rather large size of our sample.

the time shape of the relationship (the depreciation pattern) is independent of calander time, that the contribution of a particular vintage of investment is independent of the circumstances and factor prices that prevailed at the time it was purchased and from subsequent changes in them. Nevertheless, what is clear is that the usual depreciation schemes which assume that the contribution of past investments declines rapidly and immediately with age are wrong. If anything there may be an "appreciation" in the early years as investments are completed, shaken down, and adjusted to.

References

* We are grateful to Gary Chamberlain, Lung-fei Lee, Christopher Sims,
   and to the participants in the Harvard-MIT econometric workshop,
   and in seminars at the Hebrew University and the NBER, for helpful
   comments. Gregory Clark and Bronwyn Hall provided capable
   research assitance and valuable suggestions. This work has been
   supported by NSF Grants PRA-13740 and SOC79-042479, and by BSF
   Grant 1881/79. All errors are, of course, solely the authors
   responsibility.

Anderson, T.W., 1978, Repeated measurements on autoregressive pro-
   cesses, Journal of the American Statistical Association, Vol. 73,
   371-378.

Balestra, P. and M. Nerlove, 1966, Pooling cross section and time
   series data in the estimation of a dynamic model: The demand for
   natural gas, Econometrica 34, 585-612.

Box, G.E.P and G.M. Jenkins, 1970, Time series analysis: Forecasting
   and control, (Holden-Day, San Francisco).

Chamberlain, G., 1978, Analysis of coveriance with qualitative data,
   Harvard Institute of Economic Research Discussion Paper #678.

Chamberlain, G., 1982, Multivariate regression models for panel data,
   Journal of Econometrics 5-46.

Compustat II, 1980, Standard and Poor's Compustat Services, Inc.,
   Englewood, CO.

Dhrymes, P.J., 1971, Distribution lags: Problems of estimation and
   formulation, (Holden-Day, San Francisco).

Feldstein, M., and L. Summers, 1977, Is the rate of profit falling?,
   Brookings Papers on Economic Activity, 211-228.

Galbraith, R.F., and J.I. Galbraith, 1974, On the inverses of some
   patterned matrices arising in the theory of stationary time
   series, Journal of Applied Probability 11, 63-71.

Geweke, J., and R. Meese, 1979, Estimating regression models of finite
   but unknown order, Social System Research Institute, The
   University of Wisconsin, Madison, Discussion Paper #7925.

Griliches, Z., 1963, Capital stock in investment function: Some
   problems of concept and measurement, in: C. Christ et. al.,
   Measurement in economics: Essays in memory of Yehuda Grundfeld,
   (Stanford University Press, Stanford), 115-137.

Griliches, Z., 1967, Distributed lags: A survey, Econometrica 35,
   16-49.

References (continued)

Kiefer, N.M., 1979, Regression systems with fixed effects with a factor structure, May, Unpublished paper.

Kiefer, J. and J. Wolfowitz (1956), Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, Annals of Mathemetical Statistics, Dec., Vol. 27, No. 4, pp. 887-906.

Klein, L.R., 1958, The estimation of distributed lags, Econometrica, Vol. 26, No. 4, pp. 553-565.

MaCurdy, T.E., 1981, Asymptotic properties of quasi-maximum likelihood estimates and test statistics, Technical Working Paper 14, National Bureau of Economic Research, Cambridge, MA.

MaCurdy, T.E., 1982, The use of time series processes to model the error structure of earnings in a longitudinal data analysis, Journal of Econometrics 18, 83-114.

Maddala, G.S. and A.S. Rao, 1971, Maximum likelihood estimation of Solow's and Jorgenson's distributed lag models, Review of Economics and Statistics 53, pp. 80-88.

Mairesse, J. and A. Siu, 1981, The R&D and Productivity data master file, unpublished manuscript, National Bureau of Economic Research, Cambridge, MA

Mundlak, Y., 1961, Empirical production function free of management bias, Journal of Farm Economics 43, 44-56.

Mundlak, Y., 1978, On the pooling of time series and cross section data. Econometrica 46, 69-85.

Nerlove, Marc ,1972, Lags in economic behavior, Econometrica 40, 221-251.

Neyman, Jerzy and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica, 1-32.

Pesaran, M.H., 1973, The small sample problem of truncation remainders in the estimation of distributed lag models with autocorrelated errors, International Economic Review 14, 120-131.

Sims, C, 1974, Distributed lags, in: Intriligator and Kendrick (eds.) Frontiers of Quantitative Economics, Vol II, (North-Holland, Amsterdam and New York).

Swamy, P.A.V.B., 1970, Efficient inference in a random coefficient regression model, Econometrica 38, 83-114.

References (continued)

Usher, D., (editor), 1980, The measurement of capital, Studies in Income and Wealth, 45, National Bureau of Economic Research (University of Chicago Press, Chicago and London).

White, H., 1980, Using least squares to approximate unknown regression functions, International Economic Review 21, 149-170.

White, H., 1980a, Nonlinear regression on cross-section data, Econometrica 48, 721-746.

White, H., 1980b, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817-838.

White, H., 1982, Instrumental variable regression with independent observations, Econometrica 50, 483-500.