Paper presented at the econometric Society European Meeting 1989 in Munich, Germany.

Variance Estimation in a Random Coefficients Model*

by

Ekkehart Schlicht

Darmstadt Institute of Technology

Schloss, 6100 Darmstadt

West Germany

(c) November 1988

revised March 1989

ABSTRACT

Consider the regression model

$$y_t = a_t' x_t + u_t \qquad t = 1,2,\ldots, T; \; u_t \sim \mathcal{N}(0,\sigma^2)$$

with $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^n$ observations, $a_t \in \mathbb{R}^n$ coefficients to be estimated and $u_t \in \mathbb{R}$ normal disturbances for the time periods $t=1,2,\ldots,T$. The coefficients are assumed to be generated by a random walk with normal disturbances $v^t \in \mathbb{R}^n$

$$a_t = a_{t-1} + v^t \qquad t=1,2,\ldots,T; \; v^t \sim \mathcal{N}(o,\Sigma)$$

The variance-covariance matrix Σ is assumed diagonal

$$\Sigma = \begin{bmatrix} \sigma_1{}^2 & & & 0 \\ & \sigma_1{}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix} , \quad \sigma_i^2 > 0, \; i=1,2,\ldots,n$$

Thus the variances in the model are σ^2 and Σ or $(\sigma^2, \sigma_1^2, \ldots \sigma_n^2)$. This paper develops a method for estimating these variances by means of certain "expected statistics estimators". These estimators are compared to maximum likelihood estimators.

A Windows program that implements the estimation procedure is available at
https://ideas.repec.org/c/lmu/muenso/684.html.
A console version with source code can be found here:
https://ideas.repec.org/c/lmu/muenso/719.html

Variance Estimation in a Random Coefficients Model*

by

Ekkehart Schlicht
Darmstadt Institute of Technology
Schloss, 6100 Darmstadt
West Germany

revised March 1989

Comments welcome

The research reported in this paper has been financially supported by the Deutsche Forschungsgemeinschaft. I thank also Daniela Diekmann, Theo Dijkstra, Walter Krämer and Ralf Pauly for valuable comments and suggestions.

Typesetting by Christine Woerlein

Introduction

Consider the regression model

$$y_t = a_t' x_t + u_t \qquad t= 1,2,\ldots, T;\ u_t \sim N(0,\sigma^2)$$

with $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^n$ observations, $a_t \in \mathbb{R}^n$ coefficients to be estimated and $u_t \in \mathbb{R}$ normal disturbances for the time periods $t=1,2,\ldots,T$. The coefficients are assumed to be generated by a random walk with normal disturbances $v^t \in \mathbb{R}^n$

$$a_t = a_{t-1} + v^t \qquad t=1,2,\ldots,T;\ v^t \sim N(0,\Sigma)$$

The variance-covariance matrix Σ is assumed diagonal

$$\Sigma = \begin{bmatrix} \sigma_1{}^2 & & & 0 \\ & \sigma_1{}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}, \quad \sigma_i^2 > 0,\ i=1,2,\ldots,n$$

Thus the variances in the model are σ^2 and Σ or $(\sigma^2, \sigma_1{}^2, \ldots \sigma_n^2)$. The estimation problem is the following: Given the observations $(x_1, x_2, \ldots, x_T)$ and $(y_1, y_2, \ldots, y_T)$, how to estimate the time path of the coefficients $(a_1, a_2, \ldots a_T)$ and the variances σ^2 and Σ?

The main difficulty here is to obtain estimates for the variances. Once the variances are determined it is relatively easy to give estimates for the coefficients, either by recursive Kalman filtering or, still easier, by the method described in Schlicht (1985, 52-56).

One possibility would be to estimate the variances by the maximum likelihood method. The purpose of this paper is to propose a variance estimator which compares favorably to the maximum likelihood estimator in several respects:

- it is asymptotically equivalent to the maximum likelihood estimator;
- it is computationally much easier to implement;
- it has a direct intuitive interpretation also in small samples;
- and it seems to work better in small samples.

The plan of the paper is as follows: Part 1 gives some notation and preliminary results. Part 2 introduces the "expected statistics" estimators and compares them with maximum likelihood estimators. The appendix gives a numerical illustration.

1. The Model

1.1 Notation

Define

$$y := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad u := \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}, \quad a := \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{bmatrix}, \quad v := \begin{bmatrix} v^2 \\ v^3 \\ \vdots \\ v^T \end{bmatrix}$$

order $\quad T\times 1 \quad T\times 1 \quad Tn\times 1 \quad (T-1)n\times 1$

$$X := \begin{bmatrix} x_1' & & & 0 \\ & x_2' & & \\ & & \ddots & \\ 0 & & & x_T' \end{bmatrix}, \quad P := \begin{bmatrix} -I & I & & & 0 \\ & -I & I & & \\ & & \ddots & \ddots & \\ 0 & & & -I & I \end{bmatrix}$$

order $\quad T \times Tn \quad (T-1)\, n \times Tn$

and write (1), (2) as

$$y = Xa + u, \quad u \sim N(0, \sigma^2 I)$$

$$Pa = v, \quad v \sim N(0, S), \quad S := I \otimes \Sigma.$$

Define further

$$Q: = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

which permits us to write

$$P = Q \otimes I$$

Denote further by $e_i \in \mathbb{R}^n$ the n-th column of an nxn identity matrix and define

$$P_i := Q \otimes e_i'$$

which permits us to write

$$v_i := P_i\ a$$

where $v_i = (v_i^2, v_i^3, \ldots v_i^T)'$ denotes the time path of the change in the i-th coefficient.

1.2 A Likelihood Function

Consider now the time averages of the coefficients

$$\bar{a} := \frac{1}{T} \sum_{t=1}^{T} a_t$$

By using the Tn x n matrix

$$Z := \frac{1}{T}\begin{bmatrix} I \\ I \\ \vdots \\ I \end{bmatrix}$$

can be expressed also als

(14) $$Z'a = \bar{a}$$

We note

(15) $$PZ = 0,\ Z'Z = I,\ P'(PP')^{-1}P + ZZ' = I$$

Define the Tn xTn matrix

(16) $$\tilde{P} = \begin{bmatrix} P \\ Z' \end{bmatrix}$$

Eqs. (7) and (14) can be combined now to

$$\tilde{P}\ a = \begin{bmatrix} v \\ \bar{a} \end{bmatrix}$$

Since $\tilde{P}^{-1} = (P'(PP')^{-1}, Z)$, this can be solved for a:

(18) $$a = P'(PP')^{-1}\ v + Z\bar{a}$$

Inserting this into (6) yields

$$y = XZ\bar{a} + w, \quad w := XP'(PP')^{-1}v + u \tag{19}$$

We note that

$$XZ = \frac{1}{T} \cdot \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix}$$

Thus (19) stands for a standard GLS regression in the time-averages $\bar{a}$ of the coefficients, and it is reasonable to assume that XZ has full rank:

$$r(XZ) = n \tag{21}$$

The disturbances w in (19) are normally distributed

$$w \sim N(0,V), \quad V := XP'(PP')^{-1}S(PP')^{-1}PX' + \sigma^2 I$$

likelihood function associated with (19) is therefore

$$L\,(\bar{a},\sigma^2,\sigma_1{}^2,...\sigma_n^2) := \log\det V + (y' - \bar{a}'Z'X')\,V^{-1}(XZ\bar{a}-y) \tag{23}$$

Minimization with respect to $\bar{a}$ yields the Aitken estimate

$$\hat{\bar{a}} = (Z'X'V^{-1}XZ)^{-1}\,Z'X'V^{-1}y \tag{24}$$

We may thus view $\hat{\bar{a}}$ as a function of the variances and the observations and insert it into (23) in order to obtain a concentrated likelihood function

$$(25) \qquad L^*(\sigma^2,\sigma_1{}^2,\ldots,\sigma_n^2) := L(\hat{\bar{a}},\sigma^2,\sigma_1{}^2,\ldots,\sigma_n^2) + \text{constants}$$

which could be used, in principle, to determine the variances This can, however, be simplified considerably.

1.3 Estimates for the Coefficients

For given $\bar{a}$, y, and X, the system (18), (22) defines the conditional normal distribution of a with mode and expectation equal to

$$(26) \qquad Z\bar{a} + P'(PP')^{-1}S(PP')^{-1}PX'V^{-1}(y-XZ\bar{a})$$

We replace the parameter $\bar{a}$ by its estimate $\hat{\bar{a}}$ and take the resulting expression as our estimate for the coefficients $\hat{a}$

$$(27) \qquad \hat{a} := Z\hat{\bar{a}} + P'(PP')^{-1}S(PP')^{-1}PX'V^{-1}(y-XZ\hat{\bar{a}})$$

This estimate can be represented also in a different way.

<u>Proposition 1</u> (Schlicht 1985, 55-56) The estimate $\hat{a}$ in (27) satisfies

$$M\,\hat{a} = X'y$$

where

$$M = (X'X + \sigma^2 P'S^{-1}P)$$

is nonsingular.

<u>Proof.</u> Eq. (28) is proved by evaluating the left-hand side explicitly, which leads to the result $X'y$.

In order to prove nonsingularity of M, consider its rank:

$$r(M) = r\left\{(X', \sigma P'S^{-\frac{1}{2}})\begin{bmatrix} X \\ \sigma S^{-\frac{1}{2}}P \end{bmatrix}\right\}$$

$$= r(X', P')$$

If (X', P') were not of full rank, there would exist vectors $c_t \in \mathbb{R}^n$, $t = 1,2,\ldots,T$, not all of them zero, such that

(31) $$X'c_1 = P'\begin{bmatrix} c_2 \\ c_3 \\ \vdots \\ c_T \end{bmatrix}$$

is satisfied. If (31) is premultiplied by Z' from (13), this leads to $Z'X'c_1 = 0$ which implies, together with (21), $c_1=0$. Since P' is of full rank $(T-1)\cdot n$, this implies also that $c_2, c_3, \ldots, c_T$ are zero. This proves the proposition.

In view of Prop. 1, the estimate $\hat{a}$ can be given a direct desriptive characterization: It minimizes the weighted sum of squares

$$\frac{1}{\sigma^2} u'u + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} v_i'v_i$$

This minimization is, for given variances, equivalent with the minimization of the expression

$$\begin{aligned} Q &:= u'u + \sigma^2 v'S^{-1}v \\ &= (y'-a'X')(Xa-y) + \sigma^2 a'P'S^{-1}Pa \end{aligned}$$

Eq. (28) is just the first-order condition for a minimum of Q with respect to a.

1.4 Another Representation of Likelihood

We may define the estimated disturbances associated with the estimated coefficients in a natural way:

$$\hat{u} := y-X\hat{a}, \quad \hat{v} := P\hat{a}, \quad \hat{v}_i := P_i\hat{a}, \quad i = 1,2,\ldots,n$$
$$\hat{w} := XP'(PP')^{-1}\hat{v} + \hat{u}.$$

All these are functions of the variances (and the observations). We may insert them into (32) and obtain the estimated sum of squares as a function of the variances:

$$\hat{Q} := \hat{u}'\hat{u} + \sigma^2 \hat{v}'S^{-1}\hat{v}$$

Position 2 (Schlicht 1985,55). The concentrated likelihood function L^*, as defined in Eq. (23), is equivalently given by

$$(35) \qquad L^*(\sigma^2,\Sigma) = \log \det V + \frac{1}{\sigma^2} \cdot \hat{Q}$$

Proof. The first terms in (23) and (35) are identical. We must prove that the second term in (23) is equal to $\hat{Q}/\sigma^2$. From (19), (24), and (33) we find for this term

$$(36 \qquad \hat{w}'V^{-1}\hat{w} = \frac{1}{\sigma^4} \hat{u}'V \hat{u}$$

Using the definition of V and the relation $X'\hat{u} = \sigma^2 P'S^{-1}\hat{v}$, which can be derived from (28), (29), and (33), this reduces to

$$\hat{w}'V^{-1}\hat{w} = \frac{1}{\sigma^2} \hat{u}'\hat{u} + \hat{v}'S^{-1}\hat{v} = \frac{1}{\sigma^2} \hat{Q}$$

which completes the proof

1.5 Notes on Computation of the Maximum Likelihood Estimates

The representation (35) of the likelihood function makes it possible to actually do maximum likelihood estimation since a inversion of V is avoided. The determinant of V can be determined practically since each element of V can be expressed by a simple formula (Schlicht 1985, 57-78). The sum of squares $\hat{Q}$ is also rather easy to compute since it requires, basically, to solve the system (28) for $\hat{a}$. The matrix M is a very simple symmetric band matrix of band width (n-1). The system can be solved accurately and efficiently by a Cholesky decomposition. When actually doing these computations, I encountered repeatedly the problem, however, that the likelihood function was rather badly behaving for short time series. An example is provided in the appendix. Further, the intuitive understanding of the estimation procedure seemed hard to me to obtain. This led to the development of another kind of estimator, which will be described in the following part of the paper.

2. Variance Estimation

2.1 The Heuristic Argument

The estimated coefficients $\hat{a}$ along with the estimated disturbances are random variables. Their distribution is determined by the true variances along with the observations. We may write for instance

$$(38) \qquad \hat{a} = M^{-1}X'y = M^{-1}X'(Xa+u)$$

by using (28) and (6). This gives $\hat{a}$ in terms of the true coefficients a and the true disturbances. Since

$$(39) \qquad X'(Xa+u) = X'Xa + X'u + \sigma^2P'S^{-1}P - \sigma^2P'S^{-1}P$$

and $v = Pa$ from (7) Eq. (38) can be re-written as

$$(40) \qquad \hat{a} = a + M^{-1}(X'u - \sigma^2P'S^{-1}v).$$

Premultiplication of (40) with P_i yields

$$(41) \qquad \hat{v}_i = v_i + P_iM^{-1}(X'u - \sigma^2P'S^{-1}v), \qquad i = 1,2,\ldots,n$$

Similarily, $\hat{u} = y - X\hat{a} = X(a-\hat{a}) + u$ can be formed and

(42) $$\hat{u} = u - XM^{-1}(X'u - \sigma^2 P'S^{-1}v)$$

is obtained.

Thus $\hat{u}$ and $\hat{v}_1$ $\hat{v}_2, \ldots \hat{v}_n$ are linear functions of the normal random variables u and v, and we may calculate the expectation of the squared errors:

(43) $$E(\hat{u}'\hat{u}) = \sigma^2(T - \mathrm{tr}XM^{-1}X')$$

(44) $$E(\hat{v}_i'\hat{v}_i) = \sigma_i^2(T-1) - \sigma^2 \; \mathrm{tr}\; P_i M^{-1} P_i' \qquad i = 1,2,\ldots n$$

deriving (43) and (44) we note that

(45) $$X'X + \sigma^2 P'S^{-1}P \equiv X'X + \sum_{i=1}^{n} \frac{\sigma^2}{\sigma_i^2} P_i'P_i$$

and that $E(\xi'\xi) = E(\mathrm{tr}(\xi\xi'))$ for any random vector ξ.)

The expectations (43) and (44) are functions of the variances and the observations:

$$f_o(\sigma,\Sigma) := \sigma^2 - \frac{\sigma^2}{T}\; \mathrm{tr}\; XM^{-1}X' = E(\frac{1}{T}\, \hat{u}'\hat{u})$$

(47) $$f_i(\sigma,\Sigma) := \sigma_i^2 - \frac{\sigma^2}{T-1}\cdot \mathrm{tr}\; P_i M^{-1} P_i' = E(\frac{1}{T-1}\, \hat{v}_i'\hat{v}_i)$$
$$i = 1,2,\ldots,T.$$

On the other hand, the estimated errors $\hat{v}_i$ and $\hat{u}$ are functions of the variances and the observations, too, and the corresponding "empirical variances" can be written as functions of the theoretical variances again:

$$(48) \qquad m_o(\sigma^2,\Sigma):=\frac{1}{T}\,y'(I-M^{-1}X')(I-XM^{-1})y=\frac{1}{T}\,\hat{u}'\hat{u}$$

$$(49) \qquad m_i(\sigma^2,\Sigma):=\frac{1}{T-1}y'M^{-1}P_i{}'P_iM^{-1}y=\frac{1}{T-1}\,\hat{v}_i{}'\hat{v}_i \qquad i=1,2,\ldots,n$$

The proposed estimation procedure is to select variances $\hat{\sigma}^2$ and $\hat{\Sigma}$ such that the "empirical variances" (48), (49) are just equal to the corresponding expectations (46) and (47):

$$(50) \qquad m_i(\hat{\sigma}^2,\ \hat{\Sigma})=f_i(\hat{\sigma}^2,\ \hat{\Sigma}), \qquad i=0,1,2,\ldots,n$$

We call these estimators "expected statistics estimators". The intuition underlying these estimators is straightforward: We select the variances such that some observed statistics - i.e. the values of the moments (48) and (49) - are just equal to their expectations under the assumption that the postulated variances are the true variances.

Before we proceed to analyze our variance estimators further, a small digression on the underlying estimation principle might be in place.

Some Remarks on the Method of Expected Statistics.

The method of expected statistics is obviously a simple generalization of the well-known method of moments where theoretical moments are equated to their empirical counterparts. It leads actually to very familiar results in many cases, as the following two examples might indicate.

<u>1. The Parameters of a Normal Distribution</u>. Consider a random draw $(x_1, x_2, \ldots, x_n)$ from a normal population with unknown mean μ and unknown variance σ^2. In order to employ the method of expected statistics, we need two statistics. Take the mean $\bar{x}$ and the variance s^2

(51) $$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$s^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i \quad \bar{x})^2$$

Since x_i is normally distributed, $\bar{x}$ and s^2 are random variables with the expectations

$$E(\bar{x}) = \mu$$

and

(54) $$E(s^2) = (1 - \frac{1}{n}) \cdot \sigma^2$$

Equating (51) with (53) and (52) with (54) gives the estimators for μ and σ^2:

(55) $$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

(56) $$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

which are just the usual unbiased moment estimators.

2. Parameter Estimation in the Classical Regression Model.

Consider, as a further example, the classical regression problem

(57) $$y = Y\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

with $\alpha \in \mathbb{R}^n$, $\varepsilon \in \mathbb{R}^T$, $y \in \mathbb{R}^T$ and Y a real Txn matrix. Observations are Y and y, and the parameters β and σ^2 are to be estimated.

We may calculate the expectation of the empirical cross-correlations $Y'y$:

(58) $$E(Y'y) = E(Y'Y\beta + Y'\varepsilon) = Y'Y\beta$$

This is equated to the observed vector $Y'y$ and yields the least squares estimate

$$\hat{\beta} = (Y'Y)^{-1}Y'y \tag{59}$$

We may further calculate the expected variance of the estimated error $\hat{\varepsilon} = y-Y\hat{\beta} = (I-Y(Y'Y)^{-1}Y')\varepsilon$

which is

$$E(\hat{\varepsilon}'\hat{\varepsilon}) = \sigma^2(T-n) \tag{60}$$

Equating this expectation with the calculated value of $\hat{u}'\hat{u}$ yields the usual best quadradic unbiased estimator

$$\hat{\sigma}^2 = \frac{1}{T-n} \hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{T-n} y'(I-Y(Y'Y)^{-1}Y')y \tag{61}$$

In a similar but less straightforward fashion we may also obtain the GLS estimators via expected statistics, and we could interpret the Aitken-estimator (24) for $\bar{a}$ along these lines.

2.3 Another Characterization

Consider the function

(62)

$$K(\sigma^2,\sigma_1{}^2,..\sigma_n^2):= \log \det M + \frac{1}{\sigma^2}\hat{Q}-T(n-1)\ \log \sigma^2+(T-1)\Sigma \log \sigma_i^2$$

which we wish to minimize. We note (using the "envelope theorem" and representation (45)) that

(63) $$\frac{\partial}{\partial\sigma^2} \log \det M = \sum_i \frac{1}{\sigma_i^2}\ \operatorname{tr} P_i M^{-1} P_i{}'$$

$$\frac{\partial}{\partial\sigma_i^2} \log \det M = - \frac{\sigma^2}{\sigma_i^4}\ \operatorname{tr} P_i M^{-1} P_i{}' \qquad i= 1,2,\ldots..,n$$

$$\frac{\partial}{\partial\sigma^2}\hat{Q} = \sum_i \frac{1}{\sigma_i^2}\hat{v}_i{}'\hat{v}_i$$

$$\frac{\partial}{\partial\sigma_i^2}\hat{Q} = - \frac{\sigma^2}{\sigma_i^4}\hat{v}_i{}'\hat{v}_i, \qquad i= 1,2,\ldots..,n$$

$$\operatorname{tr} M^{-1}M = \operatorname{tr} XM^{-1}X + \Sigma \frac{\sigma^2}{\sigma_i^2} P_i\ M^{-1}P_i{}' = Tn$$

Necessary conditions for a minimum of (62) are:

$$\frac{\partial K}{\partial \sigma^2} = \sum_i \frac{1}{\hat{\sigma}_i^2} \operatorname{tr} P_i M^{-1} P_i' - T(n-1) \frac{1}{\hat{\sigma}^2}$$

$$- \frac{1}{\hat{\sigma}^4} \hat{Q} + \frac{1}{\hat{\sigma}^2} \sum_i \frac{1}{\hat{\sigma}^2} \hat{v}_i' \hat{v}_i = 0$$

$$(65) \qquad \frac{\partial K}{\partial \sigma_i^2} = \frac{\sigma^2}{\hat{\sigma}_i^4} \operatorname{tr} P_i M^{-1} P_i' - \frac{1}{\hat{\sigma}_i^4} \hat{v}_i' \hat{v}_i + (T-1) \frac{1}{\hat{\sigma}_i^2} = 0$$

$$i = 1,2,\ldots,n$$

The first term in (64) is equal to $(Tn - \operatorname{tr}XM^{-1}X')/\hat{\sigma}^2$, and the last two terms add to $\hat{u}'\hat{u}/\hat{\sigma}^4$. Thus we may write instead

$$(66) \qquad \hat{\sigma}^2\left(1 - \frac{1}{T} \operatorname{tr}XM^{-1}X'\right) = \frac{1}{T} \hat{u}'\hat{u}$$

$$(67) \qquad \hat{\sigma}_i = \frac{1}{T-1} \hat{v}_i'\hat{v}_i - \frac{\hat{\sigma}}{T-1} \cdot \operatorname{tr} P_i M^{-1} P_i' \qquad i = 1,2,\ldots,n$$

Comparing these equations with our estimation equations (46) - (50) we see that they are equivalent. In case K has a unique minimum we might characterize our variance estimators therefore also as minimizers of K.

Asymptotic Equivalence With Maximum Likelihood Estimators

In this section it will be shown that the "statistics criterion" K, as defined in (62) is asymptotically equivalent to the "Likelihood criterion" L^* as given in (25) or (35). We show that

$$(68)\qquad \frac{\log \det M - T(n-1)\log \sigma^2 + (T-1)\sum_{i=1}^{n} \log \sigma_i^2 + \hat{Q}/\sigma^2}{\log \det V + \hat{Q}/\sigma^2}$$

approaches unity if T goes to infinity

Consider the Tn x Tn matrixes

$$\tilde{P} = \begin{bmatrix} P \\ Z' \end{bmatrix} \qquad \tilde{S} = I \otimes \Sigma$$

Note that $\tilde{P}^{-1} = \tilde{P}'(\tilde{P}\tilde{P}')^{-1}$ and consider

$$(70)\qquad \tilde{V} = X\tilde{P}^{-1}\,\tilde{S}\,\tilde{P}'^{-1}X' + \sigma^2 I$$

which is obtained by substituting P and S in the definition (22) of V by $\tilde{P}$ and $\tilde{S}$. Since

$$(\tilde{P}\tilde{P}')^{-1} = \begin{bmatrix} (PP')^{-1} & 0 \\ 0 & I \end{bmatrix}$$

we find

(72) $$\tilde{P}^{-1}\,\tilde{S}\,\tilde{P}'^{-1} - P'(PP')S(PP')^{-1}P = Z\Sigma Z'$$

which tends to zero with increasing T. This implies that

$$\frac{\det\,(\tilde{V})}{\det\,(V)} \rightarrow 1 \text{ for } T \rightarrow \infty$$

and we may approximate det V by det $\tilde{V}$ for large T.

Consider now the matrix

(74) $$\tilde{M} := X'X + \sigma^2\,\tilde{P}'\,\tilde{S}^{-1}\,\tilde{P}$$

which is obtained by substituting $P'S^{-1}P$ by $\tilde{P}'\tilde{S}^{-1}\tilde{P}$

in the definition (29) of M.

We note that

$$\tilde{M} - M = Z\Sigma^{-1}Z'$$

which approches zero for large T, and we may approximate M by $\tilde{M}$ large T.

We are going to consider now how $\tilde{V}$ and $\tilde{M}$ are interrelated. Define the matrix

(76) $$A := (X\ \tilde{P}^{-1}\ \tilde{S}^{1/2})$$

We note that

(77) $$\tilde{V} = AA' + \sigma^2 I$$

and

(78) $$\tilde{M} = \tilde{P}' \cdot \tilde{S}^{-1/2}(A'A + \sigma^2 I)\ \tilde{S}^{-1/2}\ \tilde{P}$$

Denote the T eigenvalues of AA' by $\mu_1, \mu_2, \ldots, \mu_T$. These are also eigenvalues of A'A, but A'A has in addition Tn−T zero eigenvalues. The eigenvalues of $AA' + \sigma^2 I$ are $\lambda_i = \mu_i + \sigma^2$, $i=1,2,\ldots,T$. These are also eigenvalues of $A'A + \sigma^2 I$, but this matrix has, in addition, the eigenvalue σ^2 with multiplicity Tn−T. Since the determinant of a matrix is equal to the product of its eigenvalues, we obtain

(79) $$\det (A'A + \sigma^2 I) = (\sigma^2)^{Tn-T} \cdot \det (AA' + \sigma^2 I)$$

and, together with (77) and (78),

(80) $$\det \tilde{M} = (\sigma^2)^{Tn-T} \cdot \det \tilde{S}^{-1} \cdot \det \tilde{P}\tilde{P}' \cdot \det \tilde{V}$$

Since

$$(81) \qquad \det \tilde{P} = \frac{1}{T} \cdot \det \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$= \frac{1}{T} \det \begin{bmatrix} -1 & & & & 0 \\ & -1 & & & \\ & & \ddots & & \\ & & & -1 & \\ 1 & 2 & \cdots & \cdots & T \end{bmatrix} = (-1)^{T-1}$$

we find $\det \tilde{P}\tilde{P}' = 1$. We note further that $\det \tilde{S} = \left(\prod_{i=1}^{n} \sigma_i^2\right)^T$ take logarithms in (80), rearrange terms, and obtain

$$(82) \qquad \frac{\log \det \tilde{M} + T \cdot \sum_{i=1}^{n} \log \sigma_i^2 - T\,(n-1) \log \sigma^2 + \hat{Q}/\sigma^2}{\log \det \tilde{V} + \hat{Q}/\sigma^2} = 1$$

Compare this with (68). For large T we can approximate M by $\tilde{M}$, V by $\tilde{V}$ and T-1 by T. This establishes the asymptotic equivalence between maximum likelihood estimators and the expected moments estimators proposed here.

2.5 Computation

In this section, we drop the circumflexes and denote our estimates simply by σ^2, σ_i^2, etc. Multiply Eq. (64) by σ^2 and Eqs (65) by σ_i^2. If we add the resulting equations, we obtain

$$\sigma^2 = Q/(T-n)$$

is inserted into (62) and we obtain the concentrated loss function which involves only the variance ratios

(84) $$\rho_i := \sigma_i^2/\sigma^2 \qquad i = 1,2,\ldots,n$$

Note that Q and M are functions of these variance ratios, rather than of the variances themselves:

$$M = M(\rho), \; Q = Q(\rho)$$

Disregarding constants, the resulting loss function can be written as

(86) $$H(\rho) = \log\det M(\rho) + (T-n)\log Q(\rho) + (T-1)\Sigma\log \rho_i$$

We shall refer to this function as the "statistics criterion" henceforth.

The estimation equations (46) - (50) may be expressed in terms of the variance ratios as

$$\rho_i = g_i(\rho) \qquad \text{with}$$

$$(87)\quad g_i(\rho) := \frac{1}{T-1}\left(v_i{}'v_i \cdot \frac{(T-n)}{Q} + \mathrm{tr}\, P_i M^{-1} P_i{}' \right), \qquad i=1,2,\ldots\ldots,n$$

where $v_i{}'v_i$, Q and M are functions of ρ.

In order to calculate tr $P_i M^{-1} P_i'$ we use the decompostition $M = BB'$ which has been used for solving the normal equation, and we note that tr $P_i M^{-1} P_i'$ is equal to the sum of all squared elements of $B^{-1} P_i'$. We need not store B^{-1} (which is not banded) in order to do this calculation, it is only necessary to compute two colums of B^{-1} at a time. In this way, we determine $g_i(\rho)$ and update the weights according to

$$(88)\qquad \rho_i' = \rho_i{}^2 / \, g_i(\rho), \qquad i=1,2,\ldots\ldots,n$$

This process has been found to converge in many examples. (I have not found a single case where (88) dit not converge). It has not been possible up to now to establish general concavity of the statistics criterion, however.

2.6 Comparison With the Maximum Likelihood Estimator

The likelihood (35) may be expressed in terms of the variance ratios by using

$$W = \frac{1}{\sigma^2} V \tag{89}$$

which is a function only of the variance ratios. This leads to

$$L^* = \log \det W + \frac{1}{\sigma^2} \hat{Q} + T \cdot \log \sigma^2 \tag{90}$$

which may be compared with (86).

Minimization with respect to σ^2 leads to $\sigma^2 = Q/T$ which may be inserted into (90). We disregard constants and write the resulting likelihood function as

$$L^{**}(\rho) = \log \det W(\rho) + T \cdot \log Q(\rho) \tag{91}$$

This is the "likelihood criterion" which may be compared with the statistics criterion (86). In order to minimize this function, we may calculate the deviatives with respect to ρ_i and put them to zero. The resultung conditions (given in Schlicht 1985;58) are numerically rather complicated, however, and much less tractable than (87). They involve an inversion of a full (rather than banded) T x T matrix. If T is large, this is practically infeasible, but then the expected statistics estimators, which are much easier to compute, are equivalent, and the estimators proposed here seem better. If T is small, however, we typically encounter convergency problems. It has been observed, as a rule, that the function L^{**} has no reasonable minima if T is small, whereas the minimization of (86) give at least a definite result. The example given in the appendix illustrates that.

3. Concluding Comments

The proposed variance estimator seems to be a useful alternative to maximum likelihood estimators. Many questions are still open - uniqueness and consistency in particular.
The asymptotic equivalence of the proposed estimator and the maximum likelihood estimator in conjunction with computational manageability and (arguably) better performance in small samples might render it even the superior alternative.

Let me conclude with a quite general remark regarding the estimation of the time-path of the coefficients in (1) - (3): We cannot recover the coefficients a from the observations on X and y since there are much more coefficients than data points. We can, however obtain sensible guesses about the state of the economy, and these are our estimates $\hat{a}$ as given in (27). They denote the expected mean of the distribution of a which remains a random variable with non-zero variance even if we enlarge the time horizon and the sample size to infinity.

If we generate data and coefficients according to (1) and (2) on a computer, we may compute estimates for the variance ratios $\hat{\rho}$ and compare the estimated time-path of the coefficients $\hat{a}(\hat{\rho})$ with the estimation $\hat{a}(\rho)$ we would get if we had used the true variance ratios ρ for computing $\hat{a}$, but it does not make very much sense to compare $\hat{a}(\hat{\rho})$ with the true time-path of the coefficients a, since they deviate randomly from their expectation. In Monte-Carlo studies we should take not the true coefficients, but rather $\hat{a}(\rho)$ as the benchmark.

APPENDIX

Assume $n = 2$, $T= 100$, $\sigma^2 = .1$, $\sigma_1^2 = .1$ and $\sigma_2^2 = .01$, $a_{11} = 1$ $a_{12} = 2$ and generate coefficients according to (2). Let e_t denote a random variable uniformely distributed over the interval $[.5, 1.5]$ and generate observations $x_{1,t} = 1$ and $x_{2,t} = e_t$ for all $t=1,\ldots 100$. Generate a time series of y_t according to (1). A possible outcome is summarized in Table 1.

From x and y we may compute the likelihood criterion (93) and the statistics criterion (88) for alternative variance ratios. This is done in Table 2.

We note that the true variance ratios are $\rho_1 = 1$ and $\rho_2 = .1$, and that the minimum both of the likelihood and of the statistics criterion is fairly close to this (We may further compute the variances $\frac{}{T-1} \sum_{t=1}^{T} (a_{it} - a_{i-1})^2$ from the data and compute their ratios. These "empirical variances" and the corresponding "empirical variances ratios" are also given in the tables).

If we use only T = 25 rather than T = 100, we obtain table 3. We see that the two criteria suggest different results

We find in particular that the minimization of the likelihood criterion leads to rather unreasonable corner solutions. It is my impression that this is a quite general phenomena in small samples, which is even more pronouced when we deal with more than two explanatory variables. The "expected statistics" estimators, on the other hand, do not seem to tend to corner solutions.

Figures 1 and 2 illustrate, finally, the decomposition. Fig. 1 depicts the time path of the true coefficients (light) and the time path of the optimal estimates $\hat{a}(\rho)$ (heavily drawn curve). Figure 2 depicts the time path of the optimal estimates $\hat{a}(\rho)$ and with the estimated time-path of the coefficients $\hat{a}(\hat{\rho})$, computed $\hat{\rho}_1 = 7.2948$ and $\hat{\rho}_2 = 1.4684$ (light). We see that the estimated variance ratios are greater than the true values, and the resulting time-paths exhibit slightly more variability than $\hat{a}(\rho)$. The paths $\hat{a}(\rho)$ and $\hat{a}(\hat{\rho})$ are qualitatively very similar. We observe also a rather close connection between the true coefficients a and their expectations $\hat{a}(\hat{\rho})$ and $\hat{a}(\rho)$.

As an aside we note further that the averages of the true coefficients are (4.7953, 1.6742). The estimated averages are $\hat{\lambda} = (5.1580, 1.3803)$. Estimating λ by OLS yields (6.2160, .3210) which differs significantly from the true averages. Thus the assumption of time-invariant coefficients, although not unreasonable in the example, leads to a considerable underestimation of the influence of the exogeneous variable x_2.

APPENDIX B

Expected Statistics Estimators: A Definition

by Ekkehart Schlicht,

Technische Hochschule, Schloss, 6100 Darmstadt

September 1989

The expected statistics estimators introduced in the text can be defined as follows:

Consider the model given by the density function

$$f(y|x,\theta)$$

where

y endogenous observables

x exogenous observables

θ exogenous non-observables, parameters

A statistic is a function

$$s(y,x,\theta)$$

Define the expected statistic as

$$S(x,\theta) = E\{s(y,x,\theta)|x,\theta\}$$

A solution $\hat{\theta}$ of

$$s(y,x,\hat{\theta}) = S(x,\hat{\theta})$$

is termed <u>expected statistics estimator.</u>

The set of solutions to this equation is determind by the model, the statistics selected, and the obervations.

If this estimation principle has been proposed somewhere, please let me know!

References

Ekkehart Schlicht (1985)

Isolation and Aggregation in Economics,

Berlin-Heidelberg-New-York-Tokyo: Springer

t	y_t	x_{2t}	a_{1t}	a_{2t}
1	3.4777	.8568	1.0000	2.0000
2	3.2192	1.3142	.4395	2.0318
3	3.1478	1.3762	.6005	2.0684
4	3.6158	1.3756	1.0356	2.0179
5	4.5672	1.4389	1.4116	1.9939
6	4.7632	1.1618	1.8959	2.0405
7	4.5122	1.4462	1.6556	1.9873
8	4.9027	1.2127	1.7221	2.0672
9	3.7182	1.0832	1.5019	2.0011
10	3.3405	.5779	1.3245	2.1217
11	3.0737	.7879	1.2587	2.0227
12	3.7493	1.4371	1.2494	1.9329
13	3.4204	.7780	1.3726	2.0514
14	3.3395	.7790	1.2386	2.0791
15	3.2518	1.0459	1.4285	2.1699
16	2.8992	.5290	1.5666	2.3230
17	5.1609	1.4879	1.5812	2.2965
18	4.1873	1.0257	2.0418	2.3285
19	5.2289	1.4063	1.9897	2.3588
20	5.0190	1.2098	2.2039	2.3516
21	5.3432	1.3103	2.2618	2.2680
22	3.7682	.5639	2.8404	2.2436
23	5.2956	1.1062	3.1117	2.2492
24	4.6552	.7366	3.2603	2.3524
25	5.7304	.7848	3.3934	2.3166
26	6.2307	1.0120	3.5689	2.2995
27	5.1997	.5016	3.4223	2.1741
28	6.0287	1.4527	3.2869	2.1286
29	5.2403	.9914	3.2875	2.1221
30	5.8486	1.3618	3.1521	1.9428
31	6.0498	1.2297	3.2437	1.7550
32	5.8823	1.1094	3.8664	1.7224
33	5.4459	.8381	4.1722	1.6731
34	5.7773	1.4476	3.8352	1.5946
35	5.2862	.9910	3.9690	1.5227
36	6.1220	1.2049	4.3198	1.5681
37	5.8631	1.3865	3.9785	1.5798
38	5.8245	1.4898	3.4866	1.5534
39	5.6376	.6666	4.2037	1.6472
40	5.8816	.6655	4.7851	1.6964
41	6.3972	.9208	5.0016	1.7348
42	6.9732	1.0285	5.5799	1.5956
43	7.3886	1.1056	6.0284	1.4407
44	7.2552	.6656	6.4158	1.3668
45	8.3801	1.2278	6.5616	1.3880
46	9.0300	1.3585	6.8640	1.4871
47	8.1361	.8197	6.6288	1.5646
48	8.0489	.7605	6.7422	1.6707
49	7.9433	.5438	6.6634	1.5273
50	8.9568	1.3155	6.8313	1.5170

Table 1 (cont.)

t	y_t	x_{2t}	a_{1t}	a_{2t}
51	7.7385	.9156	6.6807	1.4629
52	7.3690	.8551	6.3126	1.3509
53	7.0655	.7406	6.0766	1.5048
54	7.9063	1.2100	6.2481	1.3929
55	8.3738	1.3415	6.3434	1.4009
56	7.2681	.7547	6.1577	1.1500
57	7.2228	.8431	6.3542	1.1528
58	7.9992	1.3133	6.5360	1.1990
59	7.6035	1.1219	6.2761	1.1540
60	6.9535	.8827	5.9564	1.0318
61	7.2546	.9956	6.1276	1.0470
62	7.2167	.7956	6.3520	1.0873
63	7.4084	1.1131	6.1554	1.1786
64	7.6927	1.4207	6.1223	1.2878
65	7.2911	.6858	6.6461	1.3539
66	8.1202	.6104	6.7691	1.4368
67	8.3668	.8745	6.7186	1.6480
68	8.5673	.9961	6.8942	1.6926
69	8.5418	1.2697	6.7626	1.6588
70	9.1700	1.4169	6.6472	1.5814
71	7.5813	.8489	6.7627	1.5028
72	8.4218	1.4253	6.7701	1.5288
73	7.2846	.5194	6.3379	1.3987
74	7.9252	.7129	6.7391	1.2307
75	8.3108	.7689	6.8765	1.3663
76	8.7761	1.3523	6.9093	1.5491
77	9.4917	1.4482	6.9746	1.6582
78	8.8959	.9283	6.9600	1.6850
79	7.0910	.5201	6.7436	1.4821
80	9.0190	1.1843	6.5211	1.5490
81	7.8375	.9134	6.3191	1.4573
82	7.3081	.6581	6.4404	1.4587
83	7.7055	.7668	6.6858	1.4252
84	7.6565	1.1270	6.0856	1.4073
85	6.9751	.6991	6.1550	1.5420
86	6.8473	.6034	6.2800	1.4578
87	8.2602	1.1392	6.3202	1.2982
88	7.8593	.7662	6.5825	1.4841
89	7.7420	.7498	6.6738	1.4990
90	7.2886	.8759	6.1741	1.6880
91	7.9529	1.0237	6.5046	1.5893
92	8.6688	1.3162	6.5615	1.4438
93	8.1529	1.2911	6.3616	1.3520
94	7.7060	.6071	6.3456	1.4279
95	7.8110	1.3270	5.7255	1.5697
96	7.2386	1.4959	5.0272	1.4605
97	5.9060	.6488	4.9153	1.4719
98	5.2760	.6627	5.0139	1.4690
99	6.8830	1.3133	4.8343	1.6209
100	7.0699	1.0491	5.5069	1.6300

(cont.) Table 1

Likelihood Criterion

r(1) \ r(2)	.0001	.001	.01	.1	1	10	100	1000	10000
.0001	25445.80	17266.34	7560.63	3730.09	1749.44	459.07	125.94	271.72	491.03
.001	17512.53	13809.20	7115.05	3631.17	1712.28	454.40	125.84	271.72	491.03
.01	6581.34	6287.34	4955.40	3005.32	1451.73	414.15	124.91	271.73	491.05
.1	2284.54	2270.31	2162.16	1679.48	835.74	249.28	117.36	271.84	491.22
1	613.26	612.12	602.51	542.51	317.18	75.89	93.90	272.21	492.19
10	24.62	24.59	24.35	22.41	10.49	0.00	85.94	273.73	494.34
100	64.52	64.53	64.58	65.00	67.53	79.74	137.97	289.21	498.74
1000	269.87	269.87	269.90	270.14	271.41	275.31	291.18	358.15	516.70
10000	495.52	495.52	495.54	495.75	496.82	499.05	503.25	519.52	587.38

Note: add 411.96 to obtain true values.

Statistics Criterion

r(1) \ r(2)	.0001	.001	.01	.1	1	10	100	1000	10000
.0001	62395.79	41087.50	27517.30	22617.76	20425.52	19136.94	18813.58	18960.40	19179.66
.001	41325.46	24407.56	13788.16	9221.65	7085.48	5825.38	5505.05	5651.74	5870.97
.01	26552.59	12974.13	7632.61	4574.19	2790.85	1742.06	1457.32	1604.30	1823.47
.1	21196.47	7883.15	3742.48	2147.95	1072.37	467.97	337.21	490.83	709.89
1	19307.64	6001.27	1948.23	769.85	329.27	87.91	113.71	292.12	511.59
10	18703.32	5394.03	1343.87	218.73	0.00	33.27	164.59	363.77	585.56
100	18743.78	5433.24	1380.45	255.83	56.71	148.08	328.66	540.22	763.98
1000	18948.90	5638.19	1584.90	459.17	259.43	352.46	537.96	752.86	977.27
10000	19174.41	5863.67	1810.33	684.43	484.32	577.03	762.33	976.94	1200.96

Note: add 825.93 to obtain true values.

EXAMPLE1:4, T= 100
Theoretical variances s(0)= .1 , s(1)= .1 , s(2)= .01
Variance ratios r(1)= 1 , r(2)= .1
Empirical variances s(0)= 9.50110145991E-2 , s(1)= 8.77194654477E-2 , s(2)= 9.59396827311E-3
Variance ratios r(1)= .923255749008 , r(2)= .100977432076
88/10/14

Table 2

Likelihood Criterion

r(1) \ r(2)	.0001	.001	.01	.1	1	10	100	1000	10000
.0001	318.68	307.25	237.97	106.15	12.42	0.00	40.10	93.51	148.58
.001	309.06	298.61	233.90	105.61	12.40	.00	40.10	93.51	148.58
.01	245.13	239.88	202.29	100.57	12.15	.02	40.11	93.52	148.59
.1	114.38	113.69	107.33	70.96	9.98	.18	40.21	93.59	148.65
1	24.94	24.86	24.10	18.17	.60	1.59	41.02	94.12	149.15
10	2.65	2.65	2.63	2.50	2.34	11.06	45.09	96.03	150.78
100	39.19	39.19	39.20	39.29	40.04	43.97	60.97	100.96	153.29
1000	92.18	92.18	92.19	92.26	92.77	94.66	99.64	117.70	158.35
10000	147.21	147.21	147.22	147.28	147.77	149.37	151.88	157.01	175.19

Note: add 70.64 to obtain true values.

Statistics Criterion

r(1) \ r(2)	.0001	.001	.01	.1		10	100	1000	10000
.0001	9361.15	6129.97	5089.31	4692.92	4546.88	4531.19	4571.45	4624.70	4679.70
.001	6130.03	2899.14	1861.64	1467.24	1320.94	1304.95	1345.12	1398.36	1453.35
.01	5093.12	1865.48	853.49	483.71	340.56	324.02	363.93	417.13	472.12
	4697.75	1472.13	488.02	184.36	69.07	54.44	93.95	147.05	202.03
	4555.92	1329.61	347.97	72.35	4.36	5.07	45.32	98.20	153.14
10	4530.79	1304.07	321.78	49.42	0.00	18.94	62.30	114.93	169.75
100	4567.56	1340.73	358.14	85.33	37.11	59.88	105.12	157.73	212.43
1000	4620.42	1393.58	410.95	138.02	89.53	112.09	157.17	209.45	264.00
10000	4675.37	1448.53	465.90	192.95	144.43	166.88	211.81	263.86	318.20

Note: add 174.69 to obtain true values.

EXAMPLE1:4, T= 25
Theoretical variances s(0)= .1 , s(1)= .1 , s(2)= .01
Variance ratios r(1)= 1 , r(2)= .1
Empirical variances s(0)= .143950380656 , s(1)= 7.71959837362E-2 , s(2)= 5.10016245454E-3
Variance ratios r(1)= .536268006967 , r(2)= 3.54300032505E-2
88/10/18

Table 3

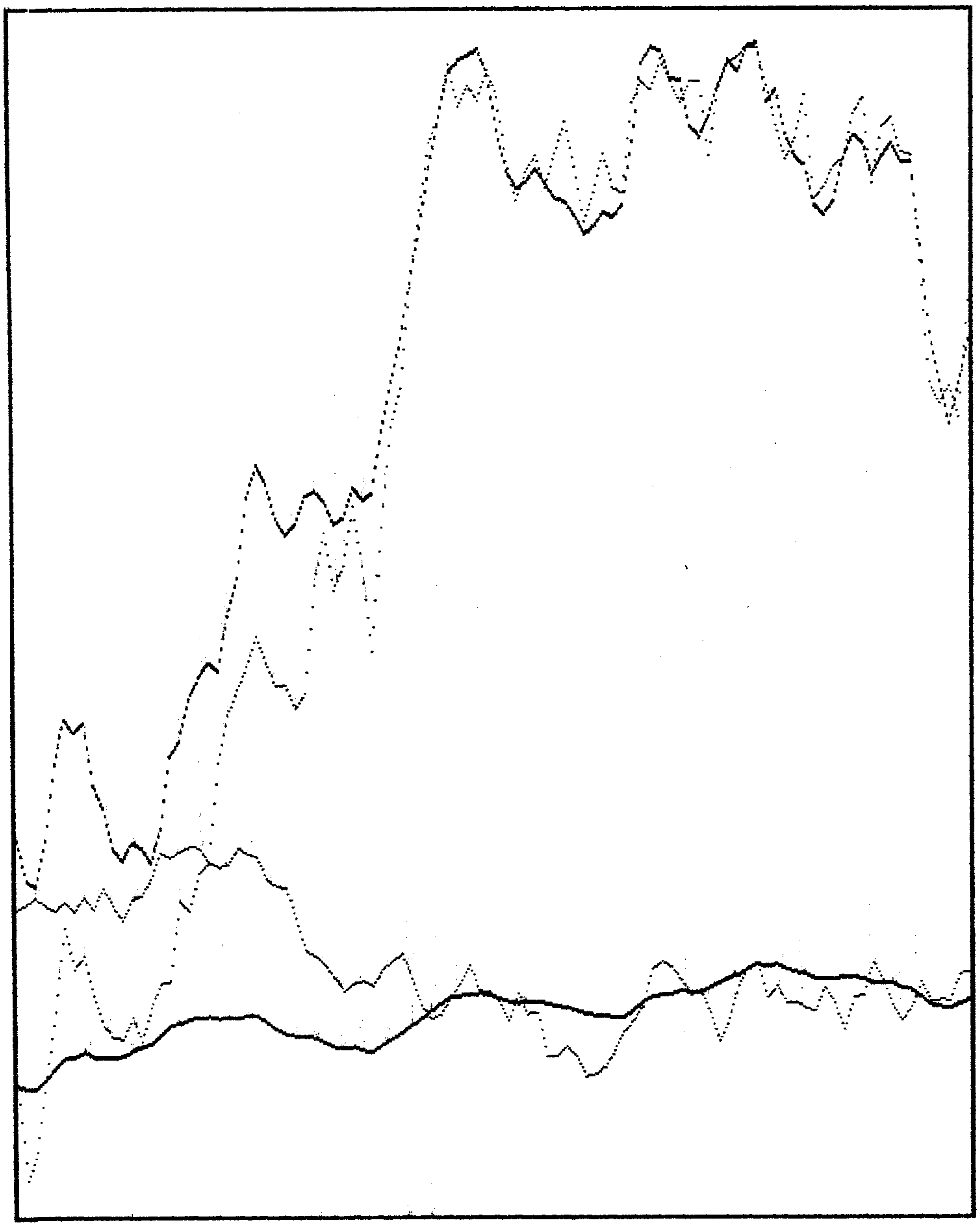

Figure 1

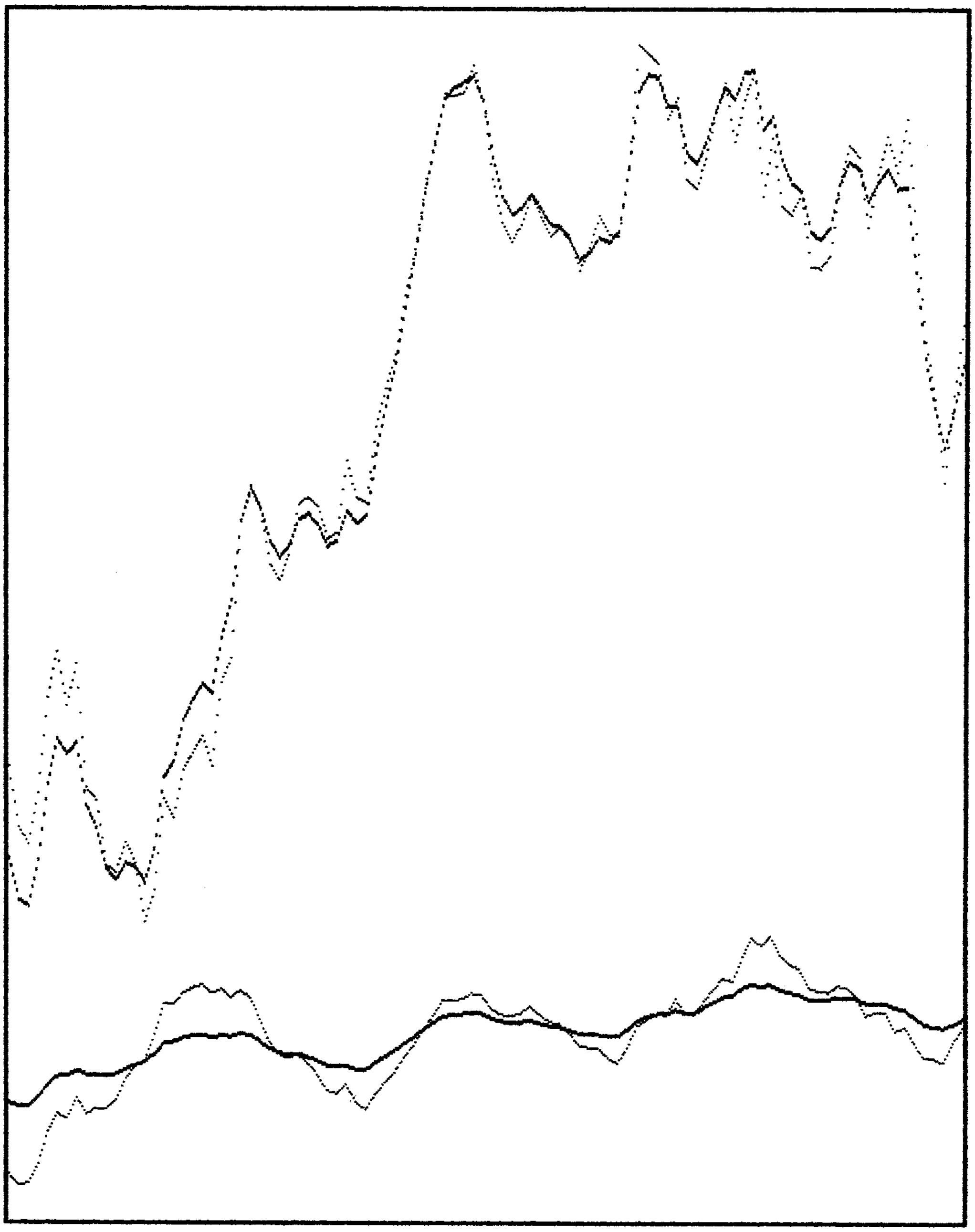

Figure 2