Volume Title: Annals of Economic and Social Measurement, Volume 2, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm73-4

Publication Date: October 1973

Chapter Title: Time-Varying Regression Coefficients: A Mixed Estimation Approach and Operational Limitations of the General Markov Structure

Chapter Author: J. Philip Cooper

Chapter URL: http://www.nber.org/chapters/c9942

Chapter pages in book: (p. 525 - 530)

Annals of Economic and Social Measurement, 2/4, 1973

TIME-VARYING REGRESSION COEFFICIENTS: A MIXED ESTIMATION APPROACH AND OPERATIONAL LIMITATIONS OF THE GENERAL MARKOV STRUCTURE

BY J. PHILLIP COOPER*

This paper applies the Mixed estimation approach to the problem of time-varying regression coefficients. A general Markov form is imposed on the time-structure of the coefficients to facilitate estimation. The limitations of the approach are briefly considered.

In this paper, the Mixed estimation approach of Theil and Goldberger (1961) and Theil (1963) is applied to the problem of time-varying regression coefficients. Stochastic prior information for regression coefficients is developed by imposing a general Markov form on the time structure of the coefficients. Given the covariance matrix of this prior information, estimators of the variance of the regression disturbances, the initial values of the Markov form and the individual regression coefficients follow easily. The specification of the time structure of the coefficients remains a very difficult problem and it is discussed only briefly here.

Assume the following regression model:

$$y_t = X_t\beta_t + \varepsilon_t \qquad t = 1, 2, \ldots, N \tag{1}$$

where

y_t is a single observation on the dependent variable,
X_t is a $1 \times k$ row vector of exogenous regressors,
β_t is a $k \times 1$ vector of unknown regression coefficients,
ε_t is a disturbance term such that $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_t\varepsilon_{t+j}) = 0$ for $j \neq 0$.

For the complete sample of N observations, we write

$$\underset{(N \times 1)}{y} = \underset{(N \times Nk)}{X}\;\underset{(Nk \times 1)}{\beta} + \underset{(N \times 1)}{\varepsilon} \tag{2}$$

where

$$y = (y_1, y_2, \ldots, y_N)'$$

$$\beta = (\beta'_1, \beta'_2, \ldots, \beta'_N)'$$

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$$

and

$$X = \begin{pmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_N \end{pmatrix}.$$

* Assistant Professor of Business Economics at the Graduate School of Business, University of Chicago. The author gratefully acknowledges the helpful discussions he had with Stanley Fischer of the University of Chicago and Alexander H. Sarris of N.B.E.R.

Accepted for publication June 1973

Consider now the following general structure for the time-varying coefficients:

$$\underset{(k\times 1)}{\beta_{t+1}} = \underset{(k\times k)}{T}\underset{(k\times 1)}{\beta_t} + \underset{(k\times 1)}{\xi} + \underset{(k\times 1)}{u_{t+1}} \qquad t = 0, 1, \ldots, N-1 \tag{3}$$

where

$$E(u_j) = 0 \qquad j = 1, \ldots, N$$

$$E(u_i u_j') = H\delta_{ij} \qquad i, j = 1, \ldots, N$$

$$E(u_j \varepsilon_i) = 0 \qquad i, j = 1, \ldots, N$$

T, ξ, β_0 H given. (δ_{ij} is the Kronecker delta).

Equation (3) can be rewritten by substituting recursively as

$$\beta_t = T^t \beta_0 + \left(\sum_{i=0}^{t-1} T^i\right)\xi + \sum_{i=0}^{t-1} T^i u_{t-i} \qquad t = 1, 2, \ldots, N. \tag{4}$$

In stacked matrix form, this is

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{pmatrix} = \begin{pmatrix} T \\ T^2 \\ T^3 \\ \vdots \\ T^N \end{pmatrix}\beta_0 + \begin{pmatrix} I \\ I + T \\ I + T + T^2 \\ \vdots \\ I + T + T^2 + \ldots + T^{N-1} \end{pmatrix}\xi \tag{5}$$

$$+ \begin{pmatrix} I & 0 & 0 & \ldots & 0 \\ T & I & 0 & \ldots & 0 \\ T^2 & T & I & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ T^{N-1} & T^{N-2} & T^{N-3} & \ldots & I \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{pmatrix}$$

and with further compacting

$$\beta = T_1\beta_0 + T_2\xi + T_3 u \tag{6}$$

where the additional definitions necessary are obvious from the correspondence between (5) and (6).

Now, the structure can be cast in the Mixed estimation format by rewriting (6) as

$$(T_1\beta_0 + T_2\xi) = \beta - T_3 u. \tag{7}$$

Comparing (7) with the standard form of the Mixed model, i.e.,

$$r = R\beta + v,$$

we see:

$$R \equiv I$$

$$r \equiv T_1\beta_0 + T_2\xi$$

$$v \equiv -T_3 u$$

and

$$E(v) = 0$$

$$E(vv') = T_3(I_N \otimes H)T_3'.$$

Letting $E(vv') \equiv V$, mechanical application of the Mixed estimation formulae leads to

$$\hat{\beta} = \left(\frac{1}{\sigma_\varepsilon^2}X'X + V^{-1}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}X'y + V^{-1}r\right) \tag{8}$$

$$\operatorname{cov}(\hat{\beta}) = \left(\frac{1}{\sigma_\varepsilon^2}X'X + V^{-1}\right)^{-1}.$$

This is, of course, the same computational result arrived at by Sarris (1973, eqs. 56, 57) under the Bayesian approach.[1]

The Mixed estimator is not valid if we interpret the elements of β as random. The form (7), however, indicates that we might regard β as a constant but unknown parameter vector about which we have stochastic prior information. In particular, β is equal to the given value $(T_1\beta_0 + T_2\xi)$, but up to a random error $(-T_3u)$: "given" should be read as "given, at least so far as the current sample is concerned." The similarity of the results has at least one obvious benefit—computer programs for the Mixed estimator already exist.

In the typical application of the Mixed estimator, then, r and V are supplied as prior information but σ_ε^2 is unknown: to evaluate (8), σ_ε^2 is replaced by a consistent estimate providing the resulting approximate Mixed estimator of β with at least desirable large sample properties. In the present context, therefore, much information is required: values must be supplied for T, ξ, β_0, and H. Furthermore, we cannot hope to use the usual approach to estimating σ_ε^2 from the regression model (2) since the $Nk \times Nk$ matrix $X'X$ is not of full rank and is thus singular.

All is not yet lost, however. Assume now that just T, ξ, and H are given. Substitute (6) into (2) and arrange terms to get

$$(y - XT_2\xi) = (XT_1)\beta_0 + (XT_3u + \varepsilon) \tag{9}$$

or

$$y^* = X^*\beta_0 + \varepsilon^* \tag{9'}$$

and

$$E(\varepsilon^*) = 0$$

$$E(\varepsilon^*\varepsilon^{*\prime}) = \sigma_\varepsilon^2\left[I + \frac{1}{\sigma_\varepsilon^2}XT_3(I_N \otimes H)T_3'X'\right].$$

For any value of σ_ε^2, β_0 can be estimated by a simple application of Generalized

[1] The occasional, approximate correspondence between Bayesian and Mixed estimators has received some limited attention: see, for example, Theil (1971, Section 12.9) and Zellner (1971, Section 4.2).

Least Squares to (9'):[2] the appropriate value to use for σ_ε^2 can be found through a standard search procedure.

To illustrate, let $Q \equiv XT_3(I_N \otimes H)T_3'X'$ be written in the form

$$Q = G'DG \tag{10}$$

where G is an orthogonal matrix that diagonalizes Q, and D is an $N \times N$ diagonal matrix. Then,

$$\begin{aligned} E(\varepsilon^*\varepsilon^{*\prime}) &= \sigma_\varepsilon^2\left[I + \frac{1}{\sigma_\varepsilon^2}G'DG\right] \\ &= \sigma_\varepsilon^2 G'\left[I + \frac{1}{\sigma_\varepsilon^2}D\right]G \\ &= \sigma_\varepsilon^2 G'D^*G \end{aligned} \tag{11}$$

where D^* is a diagonal matrix whose elements are $(1/\sigma_\varepsilon^2)$ times the corresponding elements of D plus unity. (We start with some arbitrary value for $\sigma_\varepsilon^{\ 2}$.) Let $P = D^{*-1/2}G$ be used to transform (9') to

$$(Py^*) = (PX^*)\beta_0 + (P\varepsilon^*) \tag{12}$$

or

$$y^{**} = X^{**}\beta_0 + \varepsilon^{**} \tag{12'}$$

and

$$\begin{aligned} E(\varepsilon^{**}) &= 0 \\ E(\varepsilon^{**}\varepsilon^{**\prime}) &= D^{*-1/2}G(\sigma_\varepsilon^2 G'D^*G)G'D^{*-1/2} \\ &= \sigma_\varepsilon^2 I. \end{aligned}$$

With these properties, we estimate β_0, conditional upon the chosen value for σ_ε^2, as

$$\hat{\beta}_0 = (X^{**\prime}X^{**})^{-1}X^{**\prime}y^{**} \tag{13}$$

and can form a new value to use for σ_ε^2 by

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{N-k}\right)(y^{**} - X^{**}\hat{\beta}_0)'(y^{**} - X^{**}\hat{\beta}_0). \tag{14}$$

Actually, it can be shown that if we add the assumption of normality so that

$$\begin{aligned} \varepsilon &\sim \mathcal{N}(0, \sigma_\varepsilon^2 I) \\ u &\sim \mathcal{N}(0, [I_N \otimes H]) \end{aligned}$$

and use N instead of $N - k$ as the deflator for $\hat{\sigma}_\varepsilon^2$, then (13) and (14) give the maximum likelihood estimators of β_0 and σ_ε^2, respectively. Iterative solution of (13) and (14) leads to a local maximum of the likelihood function.

In summary, given T, ξ and H, we can use an iterative technique, based simply on ordinary least squares of transformed data, to estimate σ_ε^2 and β_0 and then

[2] Estimation of β_0 from the current sample violates the spirit of the mixed estimator; it is to be regarded as an asymptotically valid approximation in which β_0 is replaced by a consistent estimate in the same manner as $\hat{\sigma}_\varepsilon^2$ replaces σ_ε^2.

employ the estimates in calculating the approximate Mixed estimator of β. This is essentially the point to which Sarris (1973) has brought us, though, through a different and more lengthy route.[3] Obviously, the same asymptotic properties apply that he claims. The question remains as to what to do about T, ξ and H.

The problem of identification of so many more parameters ($\frac{3}{2}k[k + 1]$) is formidable in the typical sample size. It is reasonable to assume that the investigator would be willing to impose some structure on the movement of the time varying parameters. One special case is that of $T = 0$, in which β_0 drops out of the problem. We could then estimate ξ, the "mean" of the time varying coefficients, in the same manner as that by which β_0 was handled. That is, we might rewrite (9) as

$$y = (XT_2)\xi + (Xu + \varepsilon) \tag{15}$$

and use iterative Generalized Least Squares analogous to (13) and (14) for $\hat{\xi}$ and $\hat{\sigma}_\varepsilon^2$: applying (8) we would then calculate (for nonsingular H)

$$\hat{\beta} = \left(\frac{1}{\hat{\sigma}_\varepsilon^2}X'X + [I_N \otimes H^{-1}]\right)^{-1}\left(\frac{1}{\hat{\sigma}_\varepsilon^2}X'y + [I_N \otimes H^{-1}]T_2\hat{\xi}\right). \tag{16}$$

Another special case is the random walk with zero drift:

$$\beta_{t+1} = \beta_t + u_{t+1} \qquad t = 0, 1, \ldots, N - 1;$$

that is,

$$T = I \quad \text{and} \quad \xi = 0.$$

Both these cases still leave us the specification of H.[4] If we were to assume further that H is diagonal, then it might be possible to obtain estimates of the k diagonal elements. Such an assumption was employed by Hildreth and Houck (1968) in treating the first of the special cases given above.[5] Estimates of the diagonal elements of H follow from an auxiliary regression in which the dependent variable is the vector of squared residuals of an ordinary least squares regression of y on (XT_2).[6]

[3] The iterative technique differs somewhat from the one Sarris describes in his Section 4. He specifies $E(u_i u_j') = \sigma_u^2 H\,\delta_{ij}$ with σ_u^2 unknown and defines $\theta = \sigma_u^2/\sigma_\varepsilon^2$ as the parameter on which he iterates. σ_u^2 is not identifiable, though: starting with $(1/\lambda)H$ instead of H simply leads to an estimate of σ_u^2 that is λ times larger than previously.

[4] This is a good point at which to quote the assessment by Rosenberg (1972) of the problem remaining. "... the complex covariance structure of the parameters must either be specified *a priori* or else estimated from the data: the former seems unlikely, and the latter is virtually impossible." In the presentation here in the text, (6) and (7) show the covariance (of the stochastic prior information) to be $T_3(I_N \otimes H)T_3'$. An attempt has, thus, been made to split the problem into components—one due to T and the other due to H—because it was felt that an investigator in a particular case might have more to say about one of these than the other; for example, he might be willing to say something *a priori* about the numerical values in T, but be unable to do more than restrict the general form of H to, say, diagonality. Logically, of course, the Mixed estimation approach as generally presented, is founded upon our being able to make these specifications of T and H (and also β_0) *before* we face the current sample.

[5] Hildreth and Houck's model excluded the additive disturbance vector ε. The difference can easily be resolved if one of the regressors is identically unity—i.e., the model includes an intercept. For a detailed discussion of this point in a somewhat different context, see Hsiao (1972).

[6] Other variants are available to ensure the nonnegativity of the k diagonal elements.

The second case mentioned above *cannot* be handled in a similar manner, however. Rewriting (9) for this case, we obtain

(17) $$y = (XT_1)\beta_0 + (XT_3u + \varepsilon).$$

In (15) the compound disturbance term has a diagonal covariance matrix, a feature not shared by the compound disturbance term of (17). Hildreth and Houck (1968) observed that "the extension of procedures of [their] paper to [the non-diagonal case] is straightforward in principle but complicated in some details."[7] It appears that the computational problems associated with H unknown and $T \neq 0$ require further investigation before the generality of the structure of (3) becomes very useful in the absence of a good deal of prior information.

University of Chicago

References

Hildreth, C. and Houck, J. P., "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association* (June, 1968), 584–595.

Hsiao, C., "Some Estimation Methods for a Random Coefficients Model," Technical Report No. 77, the Economics Series, Institute for Mathematical Studies in the Social Sciences (November, 1972).

Rosenberg, B., "The Estimation of Stationary Stochastic Regression Parameters Reexamined," *Journal of the American Statistical Association* (September, 1972), 650–654.

Sarris, A. H., "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients," NBER Computer Research Center, Research Report W0013 (May, 1973); forthcoming in *Annals of Economic and Social Measurement*.

Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association* (1963), 401–414.

———. *Principles of Econometrics* (New York: John Wiley and Sons, Inc., 1971).

——— and Goldberger, A. S., "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review* (1961), 65–78.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics* (New York: John Wiley and Sons, Inc., 1971).

[7] Page 585, footnote 1. Hsiao (1972) echoes this sentiment in almost identical language.