This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher:

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

Publication Date: October 1973

Chapter Title: On The Determination of Systematic Parameter Variation in The Linear Regression Model

Chapter Author: David A. Belsley

Chapter URL: http://www.nber.org/chapters/c9939

Chapter pages in book: (p. 484 - 491)

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

ON THE DETERMINATION OF SYSTEMATIC PARAMETER VARIATION IN THE LINEAR REGRESSION MODEL

BY DAVID A. BELSLEY*

This paper examines the general problem of time varying parameters in the linear regression model. Systematic, non-stochastic variation of β, linearly dependent upon "outside" variates, is highlighted. A moving-window regression technique is examined as a means of determining relevant outside variates. Computationally efficient algorithms are given for the technique. The procedure is seen to be biased, but not badly so for outside variates that move slowly over time.

INTRODUCTION

A model of increasing interest to econometricians is of the form

(1) $$y(t) = x'(t)\beta(t) + \varepsilon(t) \qquad t = 1 \dots T$$

where

$x(t)$ is a K-vector of variates

$\beta(t)$ is a K-vector of time-variant parameters

and

$\varepsilon(t)$ is distributed with zero mean and $V(\varepsilon) = \sigma_\varepsilon^2\Sigma$.

Let us further assume that

(2) $$\beta(t) = \Gamma z(t) + u(t)$$

where

Γ is a $K \times R$ matrix of constants

$z(t)$ is an R-vector of variates (which might include some x's)

$u(t)$ is distributed with zero mean. $V(u) = \sigma_u^2\Omega$ independently of $\varepsilon(t)$.

Various cases of this model have been treated elsewhere. Brown–Durbin [2] consider tests on the residuals of such a model in pyramidal regressions against $H_0: \beta(t) = \beta\ \forall t$. Rao [10] considers the case $\Gamma z(t) = \beta\ \forall t$, Σ and Ω unknown, and shows the OLS estimator is both a minimum-variance estimator of β and a minimum-variance predictor of $\beta(t)$—although with different variances. Theil [11] working with the same model with $K = 1, \Sigma = I$, Ω unknown, produces estimators of all parameters with an ingenious multi-step regression technique. Quandt has dealt with the cases (i) $R = 2$, var $u(t) \equiv 0$,

$$z(t) = \begin{cases} (1 \quad 0) & 1 \leq t \leq t_0 \\ (1 \quad 1) & t_0 \leq t \leq T \end{cases}$$

* Research supported by National Science Foundation Grant GJ-1154x to the National Bureau of Economic Research, Inc. Research Report W0005. This report has not undergone the full critical review accorded the National Bureau's studies, including review by the Board of Directors.

in [7, 8] and (ii) $R = 2$, $\operatorname{var} u(t) \equiv 0$,

$$z(t) = \begin{cases} (1 \quad 0) \text{ with prob } \lambda \\ (1 \quad 1) \text{ with prob } (1-\lambda) \end{cases}$$

in [9]. Variants of the above occur in Farley and Hinch [5], Burnett and Guthrie [3], Duncan and Jones [4]. Holland [6] has considered maximum likelihood estimates in the case $K = R = 2$, $\Sigma = I_T$, $\Omega = I_R$, and has suggested an iterative procedure for producing a solution. This approach has, no doubt, similar asymptotic properties to Theil's method. In what follows we shall always assume $\Sigma = \sigma_\varepsilon^2 I$ and that $Eu(t)u'(s) = 0$, $s \neq t$.

If the z's Are Known

When the z's are known, few real problems are encountered. (1) and (2) give

(3)
$$\begin{aligned} y(t) &= x'(t)\Gamma z(t) + x'(t)u(t) + \varepsilon(t) \\ &= [x'(t) \otimes z'(t)]\Lambda + \eta(t) \equiv w'(t)\Lambda + \eta(t) \end{aligned}$$

where

$$\eta(t) = x'(t)u(t) + \varepsilon(t)$$

$$E\eta(t) = 0, \qquad \operatorname{var}(\eta(t)) = \sigma_\varepsilon^2 + \sigma_u^2 x'(t)\Omega x(t)$$

$$\operatorname{cov}(\eta(t), \eta(s)) = 0 \qquad t \neq s,$$

and

$$\Lambda = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_R \end{bmatrix}, \qquad \gamma_i = i\text{th column of } \Gamma.$$

Two cases arise: when $V(u(t)) \equiv 0$ and not.

In the case $V(u(t)) \equiv 0$, the $u(t) = 0$ with Prob 1 and (3) is

(4)
$$y(t) = w'(t)\Lambda + \varepsilon(t)$$

an equation amenable to OLS, producing

(5)
$$\hat{\Lambda} = (W'W)^{-1}W'Y,$$

where the cap letters are the obvious summary matrices of the data.

When $V(u(t)) \neq 0$ it is clear that the disturbance term $\eta(t)$ in (3) is heteroskedastic and OLS is unbiased but inefficient. If Ω is known, but σ_ε^2 and σ_u^2 are not, a generalization of Theil's two-step OLS–GLS technique [11] seems possible—although it would appear to involve a great deal of computation. When Ω is not known, there seems to be no solution in the literature, but a generalization of Holland's maximum-likelihood approach seems indicated.

For many econometric models, the assumption that $V(u(t)) \equiv 0$ may not be too bad, at least as an approximation. This argues simply that the β coefficients are changing over time, perhaps with great complexity, but systematically with respect to outside variates. Such a generalization of the standard linear model is worthwhile in itself. In the next section, then we assume the β's vary systematically and without random component. As we have seen, this case is easily handled when the z's are known.

When the z's Are Not Known

When the z's are not known, things become more complicated as a practical matter. Clearly, alternative z matrices could be subjected to usual equation-testing techniques by direct substitution into (4). While this technique has much to say for it for small K and R, (5) involves the computation of an inverse of a $KR \times KR$ matrix, and even moderate K and R produce costly and time-consuming computation.

For moderate K and R, then, a more efficient search technique is required. An intuitively appealing approach that is often suggested regresses the Y's on the x's, ignoring parameter variation, and then compares the resulting residuals with potentially proper z variates, either graphically or with subsidiary regressions, to see if any sharp, systematic relationship can be discerned. This approach has much to offer in the event that the "missing z variates" are additive (i.e., the true equation is of the form $y(t) = x'(t)\beta + z'(t)\gamma + \varepsilon(t)$)* but can be misleading in a model such as (1). Indeed, one can see that

$$
\begin{aligned}
(6) \qquad e_t &= y(t) - W'(t)\hat{\Lambda} \\
&= y(t) - x'(t)\hat{\Gamma} z(t) \\
&= x'(t)(\Gamma - \hat{\Gamma})z(t) + \varepsilon(t),
\end{aligned}
$$

and hence, the e_t generally depend in a complicated and non-linear way on the z's.

A two-step approach is suggested from the following. We suppose $V(u(t)) \equiv 0$ and (2) becomes

$$
(7) \qquad \beta(t) = \Gamma z(t).
$$

If one had an independent unbiased estimator $b(t)$ of $\beta(t)$ $(t = 1 \ldots T)$ we could write

$$
\begin{aligned}
(8) \qquad b(t) &= \beta(t) + v(t) \\
Ev(t) &= 0 \\
Evv' &\equiv \sigma_v^2 \Sigma_t \\
Ev(t)v'(s) &= 0 \qquad s \neq t.
\end{aligned}
$$

And (7) becomes

$$
(9) \qquad b(t) = \Gamma z(t) + v(t)
$$

a block-equation in Γ amenable to OLS block regression. Such a model is readily applicable to testing alternative z's and, thus, affords an efficient search technique

* See Belsley note on Additive Misspecification [1].

when the appropriate z's are unknown. The problem, of course, is in determining the estimators $b(t)$ of $\beta(t)$. At the moment, I know of no means of producing unbiased, time-independent estimators; but OLS itself can produce estimators whose time dependence wears off completely in at most K periods and whose bias will be small when the z's move "slowly" from period to period, i.e., when Δz is small relative to z.

Consider the following procedure:

Using the first $K + \tau$ observations only, regress Y on X to obtain $b(1)$ where $b(1)$ is associated with some period in the interval $[1, K + \tau]$ – this is to be discussed further below. Then add the next period, drop the first and obtain $b(2)$, etc. This moving regression window can actually be made computationally quite efficient—involving inversions of only 2×2 matrices at each step—using the iterative algorithms shown in Appendix A giving

$$(10)\quad b(T) = b(T-1) + (X'_T X_T)^{-1}\{x(T)[y(T) - x'(T)b(T-1)] - x(T-K-\tau-1)[y(T-K-\tau-1) - x'(T-K-\tau-1)b(T-1)]\}.$$

and

$$(11)\quad (X'_T X_T)^{-1} = (X'_{T-1}X_{T-1})^{-1}B[D^{-1} + B'(X'_{T-1}X_{T-1})^{-1}B]^{-1}B'(X'_{T-1}X_{T-1})^{-1}$$

where

$$B = [x(T) \qquad x(T-K-\tau-1)] \text{ a } K \times 2 \text{ matrix}$$

and

$$D = D^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The above results in a $T - K - \tau + 1$ time series on the $b(t)$. There are several problems with these series which limit their use in a regression on the z's, but in some common cases these problems may not cause much harm. Let us look at the $b(t)$ in greater detail. (See Appendix A for definitions of symbols.)

$$(12)\quad b(T^*) = (X'_T X_T)^{-1} X'_T Y_T$$

where we let T^* be in the interval $[T - K - \tau \ldots T]$.

Then

$$(13)\quad \begin{aligned} \beta(t) &= \beta(T^*) + \Gamma(z(t) - z(T^*)) \\ &\equiv \beta(T^*) + \Gamma\Delta z(t) \\ &\simeq \beta(T^*) + \Gamma\frac{\partial z(T^*)}{\partial t}(t - T^*). \end{aligned}$$

Hence

$$(14)\quad \begin{aligned} y(t) &= x'(t)[\beta(T^*) + \Gamma\Delta z(t)] + \varepsilon(t) \\ &= x'(t)\beta(T^*) + x'(t)\Gamma\Delta z(t) + \varepsilon(t) \\ &= x'(t)\beta(T^*) + [x'(t) \otimes \Delta z'(t)]\Lambda + \varepsilon(t) \\ &\equiv x'(t)\beta(T^*) + n'(t)\Lambda + \varepsilon(t) \end{aligned}$$

and

(15) $$Y_t = X_t\beta(T^*) + N_t\Lambda + \varepsilon_t.$$

Putting (15) into (12) for $t = T$ gives

(16) $$b(T^*) = \beta(T^*) + (X'_TX_T)^{-1}X'_TN_T\Lambda + (X'_TX_T)^{-1}X'_T\varepsilon_T$$

and

(17) $$Eb(T^*) = \beta(T^*) + (X'_TX_T)^{-1}X'_TN_T\Lambda.$$

$b(T^*)$ is thus biased, but we note the bias depends on

$$N_T = [x'(t) \otimes \Delta z'(t)] \qquad t = T \ldots T - K - \tau$$
$$\simeq \left[x'(t) \otimes \frac{\partial z'(T^*)}{\partial t}(t - T^*)\right].$$

The elements of this matrix will be small (and the bias in $b(T^*)$ likewise) as $\Delta z'(t)$ is small (on a relative basis). The approximation in the last term reminds us $\Delta z'(t)$ will be small if z moves "slowly" over time and if the periods included in the regression window are close to T^*, the base period of the Tth regression (see 12)). This last statement argues strongly that τ should be picked small—indeed $\tau = 0$ is reasonable—and T^* should be chosen in the middle of the period. The former argues the method may cause trouble with abrupt, strong movements in the z's.

From (16) we see that

(18) $$V(b(T^*)) = \sigma_\varepsilon^2(X'_TX_T)^{-1}$$

and

$$\text{cov}\,(b(T^*)b(T^* - n)) = \sigma_\varepsilon^2(X'_TX_T)^{-1}X'_TM_nX_{T-n}(X'_{T-n}X_{T-n})^{-1},$$

where M_n is a matrix with 1's in the n'th super diagonal and zeros elsewhere, and hence $M_n = 0$ for $n > K + \tau$.

Therefore, $b(T^*)$ and $b(T^* - n)$ are uncorrelated after $K + \tau$ periods but are correlated with the surrounding $K + \tau$ values.

Let $B = [b(T^*)]$ a $(T - K - \tau) \times K$ matrix of the iteratively calculated $b(T^*)$'s. Using B in (9) with OLS produces

(19) $$\hat{\Gamma} = (Z'Z)^{-1}Z'B, \qquad \text{a block regression.}$$

Clearly this regression is inefficient because of the autocorrelation mentioned above. Also the error term does not have zero mean. But, for slowly moving z, and τ small, this latter problem is minimized. The inefficiency due to autocorrelation may be tolerable in many instances, for our purpose in using (19) and its associated tests is principally to identify the appropriate z's. Once this is done, we can return to the z-known case above for efficient estimates of the Γ's.

The advantage of (19) is that each successive test on the z's can be obtained with only an $R \times R$ inversion during its regression on B—which need be calculated only once on the basis of the X data alone. Further, because of (10) and (11), the B matrix may be determined with only one $K \times K$ regression inversion and a sequence of $T - K - r$ 2×2 inversions. This technique is, for moderate K and R, computationally much more efficient. It also allows the investigator to see directly and test directly the influence the possible z's have on the β's.

New Directions

The above procedure looks promising. Its formal properties should be further investigated. In addition, other means of estimating the $\beta(t)$'s should be considered (or devised). Clearly an unbiased estimator would help, and one less sensitive to abrupt changes in z would make Quandt-like jumps (as opposed to quantum jumps) more easily analyzed.

Another like of attack is to generalize the maximum-likelihood approach à la Holland. Unfortunately, these estimators do not allow for a sequence of steps first using the X data, and then the z data. Searching for excluded z's is thereby made less direct and the computation will be great for each trial z matrix.

Monte Carlo experiments with the proposed procedure are warranted to obtain experience with its behavior under likely situations. Such experiments will be the subject of a future report.

Boston College, and
National Bureau of Economic Research

Appendix A
Recursive Inversion Algorithm: Drop One–Add One

Consider a recursive algorithm for generating the inverse Grammian matrix when one observation is added and another deleted. We assume the next observation is added and the oldest is deleted.

Let

$$X_R = \begin{bmatrix} x'(R-r) \\ \vdots \\ x'(R) \end{bmatrix}$$

where $r + 1$ is the number of periods upon which any regression is based.

$$X_{R-1} = \begin{bmatrix} x'(R-1-r) \\ \vdots \\ x'(R-1) \end{bmatrix}. \qquad \text{Define } X \equiv \begin{bmatrix} x'(R-r) \\ \vdots \\ x'(R-1) \end{bmatrix}$$

so

$$X_R = \begin{bmatrix} X \\ x'(R) \end{bmatrix} \quad \text{and} \quad X_{R-1} = \begin{bmatrix} x'(R-r-1) \\ X \end{bmatrix}.$$

Then

$$X'_R X_R = X'X + x(R)x'(R)$$

$$= X'X + x(R-r-1)x'(R-r-1) + x(R)x'(R)$$

$$- x(R-r-1)x'(R-r-1)$$

$$= X'_{R-1}X_{R-1} + [x(R) \quad x(R-r-1)]D\begin{bmatrix} x'(R) \\ x'(R-r-1) \end{bmatrix}$$

$$\equiv A + \mathrm{BDB}' \quad \text{where} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = D^{-1}.$$

Using Rao [p. 29, Ex. 2.9]

$$(X'_R X_R)^{-1} = A^{-1} - A^{-1}B[D^{-1} + B'A^{-1}B]^{-1}B'A^{-1}$$

where $A \equiv (X'_{R-1}X_{R-1})^{-1}$ — assumed already computed — and $B = [x'(R) \quad x(R-r-1)]$, a $K \times 2$ matrix. Hence $D^{-1} + B'A^{-1}B$ is 2×2 and this is all that requires inversion to gain $(X'_R X_R)^{-1}$.

Appendix B
Iterative Procedure for b_R: Add One–Drop One

Let X_R be x data for the $r + 1$ periods ending in period R. Likewise for Y_R. Let X be the x data for the r periods $R - 1 \ldots R - r$, so $X_R = \begin{bmatrix} X \\ x'(R) \end{bmatrix}$ and $X_{R-1} = \begin{bmatrix} x'(R) - r - 1) \\ X \end{bmatrix}$. Likewise for Y.

Let b_R be the OLS estimate based on the $r + 1$ periods through R, i.e.,

$$b_R = (X'_R X_R)^{-1}X'_R Y_R$$

$$= (X'_R X_R)^{-1}[X' \; x(R)]\begin{bmatrix} Y \\ y(R) \end{bmatrix}$$

$$= (X'_R X_R)^{-1}[X'Y + x(R)y(R)]$$

$$= (X'_R X_R)^{-1}[X'Y + x(R-r-1)y(R-r-1) + x(R)y(R)$$

$$- x(R-r-1)y(R-r-1)]$$

$$= (X'_R X_R)^{-1}(X'_{R-1}Y_{R-1} + BDW)$$

where

$$B = [x(R) \quad x(R-r-1)],$$

$$D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \text{and} \quad W = \begin{bmatrix} y(R) \\ y(R-r-1) \end{bmatrix}.$$

Thus

$$b_R = (X'_R X_R)^{-1}[(X'_{R-1}X_{R-1})b_{R-1} + BDW].$$

but

$$X'_{R-1}X_{R-1} = X'_R X_R - BDB'.$$

Thus

$$\begin{aligned} b_R &= (X'_R X_R)^{-1}[X'_R X_R b_{R-1} - BDB' b_{R-1} + BDW] \\ &= b_{R-1} + (X'_R X_R)^{-1} BD[W - B' b_{R-1}] \\ &= b_{R-1} + (X'_R X_R)^{-1}\{x(R)[y(R) - x'(R) b_{R-1}] \\ &\quad - x(R-r-1)[y(R-r-1) - x'(R-r-1) b_{R-1}]\}. \end{aligned}$$

Bibliography

[1] Belsley, D. A., "Using Regression Residuals for Detecting Additive Misspecification," unpublished NBER mimeo, Nov. 1972.

[2] Brown, R. L. and J. Durbin, "Methods of Investigating Whether a Regression Relationship is Constant Over Time," *Selected Statistical Papers I*, European Meetings, 1968.

[3] Burnett, Thomas D. and Donald Guthrie, "Estimation of Stationary Stochastic Regression Parameters," *JASA*, Vol. 65, No. 332, Dec. 1970.

[4] Duncan, David B. and Richard Jones, "Multiple Regression with Stationary Errors," *JASA*, December, 1968.

[5] Farley, John U. and Melvin J. Hirsch, "A Test for a Shifting Slope Coefficient in a Linear Model," *JASA*, Vol. 65, No. 331, Sept. 1970.

[6] Holland, Paul, "A Note on an Iterative Maximum-Likelihood Estimation of Shifting Regression Parameters," unpublished NBER mimeo, July, 1972.

[7] Quandt, Richard E., "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *JASA*, March 1957, pp. 873–880.

[8] ———, "Test of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes," *JASA*, March 1960, pp. 324–330.

[9] ———, "A Further Approach to the Estimation of Switching Regressions," Princeton Econometric Research Memorandum #122, March 1971.

[10] Rao, C. R., "The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves," *Biometrica*, 52, 3 and 4, p. 447, 1965.

[11] Thiel, H., *Principals of Econometrics*, Section 12.4, Wiley, 1971.