Volume Title: Errors in Provisional Estimates of Gross National Product

Volume Author/Editor: Rosanne Cole

Volume Publisher: UMI

Volume ISBN: 0-870-14207-0

Volume URL: http://www.nber.org/books/cole70-1

Publication Date: 1970

Chapter Title: An Error Model APPENDIX

Chapter Author: Rosanne Cole

Chapter URL: http://www.nber.org/chapters/c3292

Chapter pages in book: (p. 97 - 103)

APPENDIX

An Error Model

A simple error model is developed here in order to illustrate the types of error that GNP estimates may contain and some of the properties of the errors. First considered is the nature of the errors that would be associated with the use of a related series to interpolate between two benchmark estimates; next, the errors that are introduced when the related series consists of preliminary data and is used to extrapolate the last benchmark; and finally, the change in errors when a new benchmark is introduced.

Let ψ denote a variable whose value is known only in benchmark periods *0* and *N;* and χ, a related variable for which there is a continuous series. Suppose an estimate of the value of ψ in each of the periods between *0* and *N* is obtained by first interpolating arithmetically and then using deviations in χ from its simple arithmetic trend to estimate the corresponding deviations in ψ.[1] More specifically, suppose that the trend values of ψ and χ in period *i*, denoted ψ_i^T and χ_i^T, respectively, are estimated as

$$(1)\quad \psi_i^T = \psi_0 + \frac{i}{N}(\psi_N - \psi_0) \quad \text{and}$$

$$\chi_i^T = \chi_0 + \frac{i}{N}(\chi_N - \chi_0),\ i = 0, \cdots, N.$$

[1] This procedure, which is the one considered by Friedman ("The Interpolation of Times Series"), may or may not be the one actually followed. The specific method used to interpolate and extrapolate the benchmarks is described only as "by means of related series" in the detailed descriptions of how GNP estimates are constructed (*National Income* and *U.S. Income and Output*). However, it is sufficient for the present purpose, which is to illustrate the types of error that the estimates may contain, rather than to estimate the magnitude of the errors or to replicate GNP estimates. Other methods, as Friedman has shown, can be reduced to special cases of his more general one.

If the deviations from these trend values are defined,

(2) $$u_i = \psi_i - \psi_i^T \quad \text{and} \quad v_i = \chi_i - \chi_i^T,$$

and if the unknown u_i were estimated from the known v_i as in

(3) $$u_i = bv_i,$$

then $\hat{\psi}_i$, the estimated value of ψ in period i, would equal

(4) $$\hat{\psi}_i = \psi_i^T + bv_i.$$

If the true relation between u_i and v_i were not an exact one as in (3), but equal instead to

(3′) $$u_i = \beta v_i + w_i,$$

then the error in estimating ψ_i would equal

(5) $$\hat{\psi}_i - \psi_i = (b - \beta)v_i - w_i.$$

Thus according to (5), the error in $\hat{\psi}_i$ would consist of the error that would arise if the relation between the movements in the two variables were incorrectly estimated (i.e., if $b \neq \beta$) and because the relation is inexact ($w \neq 0$).

Suppose we now assume that the true values, ψ_0, ψ_N, and χ_i are unknown, but that they are estimated as

(6) $Y_k = \psi_k + \xi_k$ and $X_i = \chi_i + \eta_i$ for $k = 0, N, \cdots$ and $i = 0, \cdots, N, \ldots$, where ξ_k denotes the error in the benchmark estimates (Y_k), and η_i, the measurement error in the related series (X_i). Using these data, the estimate of ψ_i would then become

(7) $$Y_i = \left[Y_0 + \frac{i}{N}(Y_N - Y_0)\right] + b\left[X_i - X_0 - \frac{i}{N}(X_N - X_0)\right]$$

and the error, defined as $E_i = Y_i - \psi_i$, would equal

(8) $$E_i = \left[\xi_0 + \frac{i}{N}(\xi_N - \xi_0)\right] + b\left[\eta_i - \eta_0 - \frac{i}{N}(\eta_N - \eta_0)\right] + [(b - \beta)v_i - w_i].$$

The error in estimating ψ_i would thus consist of errors in the benchmark estimates (ξ), measurement errors in the related variable (η), and errors due to an inexact or incorrectly estimated relationship between the movements in the two variables.

Let ξ_k and η_i be stationary to the second order with means μ_ξ and μ_η and variances $\sigma^2(\xi)$ and $\sigma^2(\eta)$. The expected value (denoted E) of E_i would then equal

(9) $E(E_i) = \mu_\xi$, provided that $E(v_i)$ and $E(w_i)$ are zero for all i.

As (9) shows, Y_i would be a biased estimate of ψ_i unless μ_ξ equals zero. But even if $\mu_\xi \neq 0$, it is readily shown that the implicit estimate of the change in ψ from period i to period $i + 1$ would be unbiased. Since Y_{i+1} would equal

$$Y_{i+1} = \left[Y_0 + \frac{i+1}{N}(Y_N - Y_0)\right] + b\left[X_{i+1} - X_0 - \frac{i+1}{N}(X_N - X_0)\right],$$

the implicit estimate of change would equal

$$(10) \qquad \Delta Y_{i+1} = \frac{1}{N}(Y_N - Y_0) + b\left[X_{i+1} - X_i - \frac{1}{N}(X_N - X_0)\right],$$

and the error would equal

$$(11) \quad E_{\Delta i+1} = \frac{1}{N}(\xi_N - \xi_0) + b\left[\eta_{i+1} - \eta_i - \frac{1}{N}(\eta_N - \eta_0)\right] + [(b - \beta)(v_{i+1} - v_i) - (w_{i+1} - w_i)].$$

On the previous assumption of stationarity, the expected value of $E_{\Delta i+1}$ is

(12) $E(E_{\Delta i+1}) = 0$, provided $E(v)$ and $E(w)$ are zero for all i.

The change estimates would therefore be unbiased even if the level estimates were not.

It might be tempting to conclude that this result supports the widespread notion that estimates of changes (say, in GNP) are more reliable than estimates of levels. However, this is not necessarily true. The accuracy of change estimates relative to that of levels depends on the relative importance of the error components and bias is only one element of over-all error. If, for example, benchmark errors were by far the most important source of error, then the mean square error levels would exceed the mean square error of changes. On the other hand, if the relation between the movements in ψ and χ were not strong and if it were incorrectly estimated, then benchmark errors (and errors in measuring χ) could be relatively minor components of the error in estimating levels and changes in ψ. In this case, the mean square error of changes could exceed that of levels.[2]

Errors in Provisional Estimates

Thus far we have considered the errors in estimating ψ when movements in a related series X are used to interpolate between two known benchmarks, Y_0 and Y_N. These, however, are the errors that could be expected in "final" or benchmark revised estimates rather than in the estimates that are published on a current basis. Provisional estimates of the current value of ψ in each of the periods between *0* and N must be made without benefit of Y_N. Such estimates are essentially extrapolations rather than interpolations, and their errors will differ accordingly.

To illustrate, suppose that a provisional estimate, made in period i, of the value of ψ_i were constructed by first extrapolating the last known benchmark period estimate and then using the deviation in the related series from its extrapolated trend to estimate the corresponding deviation in ψ_i.[3] In addition, suppose that only preliminary estimates of χ, denoted X°, are available at the time the provisional estimates of ψ are prepared. The provisional estimate, Y_i°, would then be

$$(13) \quad Y_i^\circ = \left[Y_0 + \frac{i}{N}(\psi_N^* - Y_0) \right] + b\left[X_i^\circ - X_0 - \frac{i}{N}(\chi_N^* - X_0) \right],$$

where ψ_N^* and χ_N^* denote predictions of ψ_N and χ_N, respectively. If the prediction errors and the errors in the preliminary estimates of χ are defined as

$$\epsilon(\psi)_N = \psi_N^* - \psi_N; \quad \epsilon(\chi)_N = \chi_N^* - \chi_N; \quad \text{and } \eta^\circ = X_i^\circ - \chi_i,$$

the error in the provisional estimate, defined as $E_i^\circ = Y_i^\circ - \psi_i$, would equal

$$E_i^\circ = \left[\xi_0 + \frac{i}{N}(\epsilon(\psi)_N - \xi_0) \right] + b\left[\eta_i^\circ - \eta_0 - \frac{i}{N}(\epsilon(\chi)_N - \eta_0) \right] + (b - \beta)v_i - w_i.$$

[2] If ξ and η were relatively unimportant, the mean square errors would be

$$M(E_i) \approx (b - \beta)^2\sigma^2(v_i) + \sigma^2(w_i)$$
$$M(E_{\Delta i}) \approx (b - \beta)^2\sigma^2(v_{i+1} - v_i) + \sigma^2(w_{i+1} - w_i).$$

Unless there is strong positive serial correlation in v or in w,

$$M(E_{\Delta i}) > M(E_i).$$

[3] See footnote 1 on p. 97.

Using (8), E_i° can also be expressed as

$$(13) \quad E_i^\circ = E_i + \frac{i}{N}[\epsilon(\psi)_N - \xi_N) - b(\epsilon(\chi)_N - \eta_N)] + b(\eta_i^\circ - \eta_i).$$

With the aid of (13), the differences in accuracy between the provisional and the revised estimates of ψ become fully visible. The error in the provisional estimate (E_i°) would exceed the error in the revised estimate (E_i as long as: (1) the errors in the predictions, made in period i, of the values of ψ_N and χ_N exceed the error in the benchmark estimate, Y_N, and the measurement error in X_N (i.e., if $\epsilon(\psi)_N$ and $\epsilon(\chi)_N$ exceed ξ_N and η_N, respectively) and; (2) the error in the estimate based on preliminary data exceeds the error in the revised estimate of χ_i (i.e., $\eta_i^\circ > \eta_i$).

It is often contended that merely the fact that national accounts estimates are revised offers no guarantee that the revised estimates are more accurate than the initial figures. We have seen, however, that within the framework of the present model, the major revisions could be viewed as replacing predictions, made before period N occurs, with estimates, based on data from period N, of ψ_N and χ_N. In order for the revised estimates to be only as accurate as the initial figures, the predictions, ψ_N^* and χ_N^*, would have to be as accurate as the later estimates, Y_N and X_N. Since the predictions would rely to some extent on the last known benchmark estimate, Y_0, and on preliminary estimates of χ (X°), any errors in these data would be transmitted to the predictions and become a component of their errors. Thus in the absence of evidence that the accuracy of the benchmark estimates and data for the related series has deteriorated over time (such that Y_0 and X° are more accurate than Y_N and X), it is reasonable to suppose that predictions would be less accurate than estimates of ψ_N and χ_N and hence to reject the contention that the revised estimates are no more accurate (let alone *less* accurate) than the provisional estimates

Errors Measured by the Revisions

The revisions are defined as the difference between provisional and revised estimates. Both Y_i° and Y_i can be expressed as

$$Y_i^\circ = \psi_i + E_i^\circ \quad \text{and } Y_i = \psi_i + E_i,$$

the sum of the true value and the respective errors of estimate. The revision would then equal

$$Y_i^\circ - Y_i = E_i^\circ - E_i,$$

which, using (13), can be expressed

(14) $$Y_i^\circ - Y_i = \frac{i}{N}[\epsilon(Y)_N - b\epsilon(X)_N] + b\epsilon_i,$$

where $\epsilon(Y)_N = \epsilon(\psi)_N - \xi_N$; $\epsilon(X)_N = \epsilon(\chi)_N - \eta_N$; and $\epsilon_i = \eta_i^\circ - \eta_i$. Since $\epsilon(Y)_N$ and $\epsilon(X)_N$ would also equal $\psi_N^* - Y_N$ and $\chi_N^* - \chi_N$, respectively,[4] the revisions would be a measure of the errors in predicting Y_N and X_N and of the reduction in measurement errors in the preliminary data on χ. The prediction errors $\epsilon(Y)$ and $\epsilon(X)$ would be a common component of the revision in the provisional estimates of ψ in each period between 0 and N, and thus a source of positive serial correlation in the revisions. Moreover, the importance of the prediction error would increase with i and hence the revisions would grow larger as i increases.

It is readily shown that the revisions in estimates of period-to-period changes would not depend on i and that they would be smaller than the revisions in level estimates, provided prediction errors were a major component of the revisions. Since

$$Y_{i+1}^\circ - Y_i = \frac{i+1}{N}[\epsilon(Y)_N - b\epsilon(X)_N] + b\epsilon_{i+1},$$

the revision in ΔY_{i+1}° would equal

(15) $$\Delta Y_{i+1}^\circ - \Delta Y_{i+1} = \frac{1}{N}[\epsilon(Y)_N - b\epsilon(X)_N] + b(\epsilon_{i+1} - \epsilon_i).$$

Let both the predictions and the preliminary data be unbiased. The mean square revision of the level would then be

(16) $$M = \left[\frac{i+1}{N}\right]^2 \sigma^2(\epsilon(Y)_N - b\epsilon(X)_N) + b^2\sigma^2(\epsilon),$$

and of the change,

(17) $$M_\Delta = \frac{1}{N^2}\sigma^2(\epsilon(Y)_N - b\epsilon(X)_N) + 2b^2\sigma^2(\epsilon),$$

provided $\mathrm{Cov}(\epsilon, \epsilon(Y)_N - b\epsilon(X)_N)$ and $\mathrm{Cov}(\epsilon_{i+1}, \epsilon_i)$ are zero. Thus M would exceed M_Δ if the prediction errors were the major component, and more specifically if $\sigma^2(\epsilon(Y)_N - b\epsilon(X)_N) > \dfrac{2N^2}{i(i+2)} b^2\sigma^2(\epsilon)$.

[4] This is seen by writing

$$\epsilon(Y)_N = \epsilon(\psi)_N - \xi_N = \epsilon(\psi_N^* - \psi_N) - (Y_N - \psi_N) = \psi_N^* - Y_N$$

and

$$\epsilon(X)_N = \epsilon(\chi)_N - \eta_N = (\chi_N^* - \chi_N) - (X_N - \chi_N) = \chi_N^* - X_N.$$

Finally, if the predictions were extrapolations of the last known benchmark period estimates, the variance of their errors would be

$$(18) \qquad \sigma^2[\epsilon(Y)] = (1 - \rho_Y^2)\sigma^2(Y) \quad \text{and} \quad \sigma^2[\epsilon(X)] = (1 - \rho_X^2)\sigma^2(X),$$

where ρ_Y and ρ_X are the coefficients of serial correlation. In this case then, the magnitude of the errors would depend on the variability and the strength of the serial correlation in the series to be predicted.