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

Volume Title: Errors In Provisional Estimates Of Gross National Product

Volume Author/Editor: Cole, Rosanne

Volume Publisher:

Volume URL: http://www.nber.org/books/cole69-1

Publication Date: 1969

Chapter Title: Appendix and Index

Chapter Author: Rosanne Cole

Chapter URL: http://www.nber.org/chapters/c3995

Chapter pages in book: (p. 97 - 108)

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 . More specifically, suppose that the trend values of ψ and χ in period i, denoted ψ^T_i and χ^T_i , respectively, are estimated as

(1)
$$\psi_i^T = \psi_0 + \frac{i}{N} (\psi_N - \psi_0)$$
 and $\chi_i^T = \chi_0 + \frac{i}{N} (\chi_N - \chi_0), i = 0, \dots, N.$

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

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

$$\hat{\psi}_i = \psi_i^T + b v_i.$$

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

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

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 $k = 0, N, \dots$ and $i = 0, \dots, N, \dots$, 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)
$$E_{i} = \left[\xi_{0} + \frac{i}{N}(\xi_{N} - \xi_{0})\right] + b\left[\eta_{i} - \eta_{0} - \frac{i}{N}(\eta_{N} - \eta_{0})\right] + [(b - \beta)\nu_{i} - w_{i}].$$

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

and the error would equal

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

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

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

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

²If 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,

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

³ See footnote 1 on p. 97.

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

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

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 $\chi(X^0)$, 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^0 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

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

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

the revision in ΔY_{i+1}° would equal

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

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 $Cov(\epsilon, \epsilon(Y)_N - b\epsilon(X)_N)$ and $Cov(\epsilon_{i+1}, \epsilon_i)$ are zero. Thus M would exceed M_Δ if the prediction errors were the major component, and

more specifically if
$$\sigma^2(\epsilon(Y)_N - b\epsilon(X)_N) > \frac{2N^2}{i(i+2)}b^2\sigma^2(\epsilon)$$
.

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)
$$\sigma^2[\epsilon(Y)] = (1 - \rho_Y^2)\sigma^2(Y)$$
 and $\sigma^2[\epsilon(X)] = (1 - \rho_X^2)\sigma^2(X)$,

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.

Anticipation series, 28	and forecasts of seasonal patterns, 84
Auto and appliance maintenance and	Business inventories estimates
repairs, 10-11	errors in, 25t
_	compared with naive model pro-
Balance of payments, net exports data	jection errors, 43t
and, 13	and gains over time, 58t
Benchmark estimates	and successive revisions, 50t, 52t
errors in, 14, 18	and GNP estimates, 12–13
revisions of, 49	and seasonal patterns of GNP, 82,
Bias in estimates of changes in GNP, 94-95	83 <i>t</i> , 84
Biennial Survey of Education, 11	Census Bureau, see U.S. Bureau of the
Budget of the U.S. government, 13	Census
Building permits, new construction esti-	Cloos, George W., 80n
mates and, 11	Cole, Rosanne, 26n, 96n
Bureau of the Census, see U.S. Bureau	Commodities, benchmark estimates for,
of the Census	8–10
Bureau of Labor Statistics, see U.S.	Commodity flow estimates, 8-9
Bureau of Labor Statistics	of producers' durables, 12
Business cycles, 23	Construction, see New construction;
• •	Public utility construction; Non-
and GNP estimates, 6, 69t, 70, 71t,	farm residential construction
72t, 73, 74t-79t, 80-81	Consumer durables estimates
bias in, 94-95	errors in, 24t
errors in, 6, 22–23, 23t, 26, 91–92	compared with errors in naive
and postwar trends, 84-90	model projections, 41t
and seasonal factors, 81	and gains over time, 57t
Business and Defense Administration,	and successive revisions, 53t
11–12	•
Business fixed investment, construction	Consumer goods expenditures, see Per-
components of, 11	sonal consumption expenditures
Business forecasts of GNP	Consumer nondurables estimates
accuracy of, 92–93	errors in, 24t
compared with provisional estimates,	compared with errors in naive
•	model projections, 42t
29, 30t, 31t, 32t, 32-33, 34t, 35,	and gains over time, 57t
92–93	and successive revisions, 53t
error statistics for, 22t	Consumer services estimates
errors in, 36t	errors in, 24t
overestimation and underestimation	compared with errors in naive
and, 26	model projections, 42t

penditures Contract awards, construction estimates and, 11 Contractions, see Business cycles Council of Economic Advisers, 16, 29n Cyclical changes, see Business cycles Cyclical errors, see Business cycles and gross national product estimates Deane, Phyllis, 4n Definitional changes in GNP, 35n DeJanosi, Peter E., 4n Direction-of-change errors, 23, 24t, 25t, 25-27 Dodge series, 11 Domestic investment, see Gross private domestic investment Domestic service expenditure estimates. Durables, see Consumer durables Dutch GNP forecasts, 48 Economic Indicators, 16 Economic Report of the President, 16, 29n, 32n Employment data, 11 Error model, 18, 97 Error statistics for business forecasts of GNP, 22t for estimates of GNP, 20t compared with forecasts, 29, 30t, 31t, 32t, 32-33, 34t, 35 and successive revisions of GNP estimates and forecasts, 48-56 See also Errors in GNP estimates Errors in GNP estimates and business cycles, 94-95 calculation methods, 100-103

compared with errors in forecasts and

and errors in expenditures and income

use of statistical discrepancy to meas-

estimates, 62, 63t, 63-64 and gains over time, 56-61

and use of preliminary data, 96

trend and cyclical, 90-92

ure, 64-68

naive model projections, 36t, 41t

and successive revisions, 53t

See also Personal consumption ex-

See also Error statistics: Gross national product estimates, errors Expenditures estimates errors in, 62, 63t, 63-64 and GNP estimates, 95-96 See also Government expenditures: Personal consumption expenditures Exports, net, see Net exports of goods and services Extrapolation errors, 15n and errors in early GNP data, 56 and errors in GNP estimates, 17-22 and GNP revisions, 28, 62, 64-68 Farm inventory estimates, 13 Farm and public utility construction data, 11 Federal Aviation Administration, 10 Federal Reserve Board Index of Industrial Production, 80 Survey of Consumer Finances, 11 Final expenditures estimates, 63t Final purchases, season pattern of GNP and, 83t, 84 Forecaster, task of, 29 Forecasts of GNP, see Business forecasts of GNP Friedman, Milton, 14n, 97n F. W. Dodge Corporation, 11 GNP, see Gross national product; Business forecasts of GNP: Gross national product estimates; Gross national product extrapolations Goldberg, Simon, 4n Goods and services, see Government expenditures on goods and services; Net exports of goods and services Government expenditures on goods and services error statistics for estimates of, 20t error statistics for forecasts of, 22t

errors in estimates of, 25t

ter covered, 23t

and gains over time, 58t

and cyclical characteristics of quar-

Government expenditures on goods and cyclical characteristics of, 22-23, services, errors in estimates of 23t and errors in expenditures and (cont.) and successive revisions, 48, 50t, income estimates, 62, 63t, 63-64 52t, 54t, 89t major sources of, 14 errors in forecasts and estimates commeasurement of, 64-68, 100-103 pared, 31t, 36t overestimation, underestimation, errors in naive model projections and and direction-of-change errors. estimates compared, 36t, 44t 23, 24t, 25t, 25-27 and GNP estimates, 13-14 resemblance to extrapolation errors, overestimates of, 32 17 - 22Gross national product (GNP) revised estimates as measure of, 4-5 average rates of change of, 86-87, 88 decline in, 68-81 types of, potential of revisions and, exclusive of statistical discrepancy, 14-17 87t establishment of, 7 peak to trough changes in components evaluation of, 91-92 of, 72t, 74t and federal government purchases of postwar trends in, 84-90 goods and services, 13 seasonal pattern of, 82, 83t, 84 first, 16, 87t See also Business forecasts of GNP; and gains over time, 56-61 Gross national product estimates and gains through successive re-Gross national product estimates visions, 48-56 accuracy of, 4-5, 92-93 and incorporation of new benchbased on expenditures and income marks, 16-17 data, 64-68 and major patterns of change, 6 bias in, 94-95 major revisions in, 17 and business cycles, 69t, 70, 71t, 72t, minor revisions in, 15 73, 74*t*-79*t*, 80-81 and net exports of goods and services, and change in business inventories, 13 12-13 and new construction, 11-12 of changes and levels compared, peak to trough decline in, 6 99-100 and personal consumption expendicompared with extrapolations, 35-47 tures, 9-11 compared with GNP forecasts, 29, and producer's durable equipment 30t, 31t, 32-33, 34t, 35, 92-93 expenditures, 12 and comparison of trend and cyclical and revisions of seasonal factors, errors, 90-92 81-82, 83t, 84 construction of, 97n and state and local purchases of data and methods used in, 8-14 goods and services, 13-14 data sources, 7-8 and successive revisions, 93-94 and definitional changes, 17 turning-point dates in, 78t-79t error statistics for, 20t and use of expenditures and income errors in, 38t-39t, 41t-45t data compared, 95-96 characteristics of, 18 and use of preliminary data, 96-100 uses of, 3 compared with errors in naive model projections, 41t Gross national product extrapolations created by limitations of concepts, compared with provisional estidefinitions, and coverage, 3 mates, 35-47

Gross national product forecasts, see Magnitudes of business cycles, 69-70 Business forecasts of GNP Maintenance and repairs, see Auto and appliance maintenance and re-Gross private domestic investment error statistics for average business pairs forecasts of, 22t Mean square errors error statistics for estimates of, 20t in forecasts of GNP, 29 errors in estimates of, 24t in GNP estimates, business cycles compared with forecast errors and and, 23t naive model projection errors, in naive projections of GNP, 34t 31t, 36t, 42t See also Error statistics: Errors in and cyclical characteristics of quar-GNP estimates ter covered, 23t Mincer, Jacob, 26n Moore, Geoffrey H., 40t, 81 and gains over time, 57t and successive revisions, 50t, 52t Morgenstern, Oskar, 4n first estimates compared with revised Naive model projections errors, 34t, 36t average annual rates of increase Nassimbene, Raymond, 4 in, 89t National Housing Inventory, 12 long-term changes in, 88-89 National income estimates, 14n successive revisions in estimates of, Net exports of goods and services esti-53t mates, 13, 32n-33n, 52t error statistics for, 20t Home rental value estimates, 10 errors in, 25t Household utilities expenditure esticompared with forecast errors, 32t mates, 10 compared with forecast and naive Housing statistics, 11-12 model projection errors, 37t compared with naive model pro-Income estimates jection errors, 45t compared with consumption estiand cyclical characteristics of quarmates, 10n ter covered, 23t errors in, 62, 63t, 63-64 and gains over time, 58t and gross national product estimates, and successive revisions, 50t, 54t 95-96 Net exports of goods and services busias predictions of final expenditures, ness forecasts, 32t 63t error statistics for, 22t and product estimates, 7-8 New construction estimates Input-output data, 17 errors in, 24t Interest paid by consumers, 17n compared with naive model pro-Internal Revenue Service tax return jection errors, 43t data, 12-13 and gains over time, 57t Interstate Commerce Commission, 10 and successive revisions, 50t, 52t Investment, see Gross private domestic and GNP estimates, 11-12 investment successive revisions of, 54t Noncorporate sector business invento-Jaszi, George, 4, 4n ries data, 12 Nonfarm residential construction, 11 Kuznets, Simon, 8 Overestimation, 86-87 Laundry expenditures estimates, 10-11 of GNP components, 88-90

of GNP decline, 70, 73

LIFO inventories, 13

108 Index
Overestimation (cont.)

ed with in-
d with in-
d with in-
7n
ditures esti-
stimates, 11
les
-11
••
litures esti-
intuics esti-
, 93–94
in GNP
naive pro-
timates, 59,
timates, 55,
error meas-
6, 80
16, 32 <i>n</i>
construction
construction
ports of, 10
ports of, 10
ports of, 10 estimates,
ports of, 10 estimates,
ports of, 10 estimates,