A RANDOM WALK, MARKOV MODEL FOR THE
DISTRIBUTION OF TIME SERIES

Robert B. Litterman

Federal Reserve Bank of Minneapolis

ABSTRACT

This paper describes a technique for distributing quarterly time
series across monthly values. The method generalizes an approach
described by Fernandez (1981). The paper also presents results of
a test of the accuracy of these two approaches and two standard
procedures suggested by Chow and Lin (1971).

1. INTRODUCTION

This paper describes a new technique for distributing time series, that is, for estimating the unobserved monthly movements in data for which only quarterly averaged observations are available. The method relies on estimating the relationship between the quarterly series and some related monthly series. The problem confronted here is not new; the standard solution is a procedure described by Chow and Lin (1971). In a test reported here, however, a recent alternative suggested by Fernández (1981) performs significantly better than the Chow-Lin procedure. The new solution I propose is a slight modification of the Fernández procedure. My new method takes more general account of serial correlation of errors than do other methods. My test suggests that this new technique will be the most accurate in many distributions of time series.

2. THE PROBLEM AND PREVIOUSLY PROPOSED SOLUTIONS

Empirical researchers frequently face the related problems of distribution and interpolation, both of which concern the estimation of intraperiod values of a variable whose actual values are observed only once per period. Distribution is a problem which arises with flow variables or time averages of stock variables. It is the estimation of several values, the sum or average of which equals the observed value over the longer interval. Interpolation refers to the estimation of unobserved values of a stock variable whose actual values are observed less frequently. Interpolation is handled in a fashion parallel to the distribution

techniques described here; Chow and Lin (1971) provide the details of how to translate a technique for distribution into one for interpolation.

In an early survey and critique of standard techniques, Friedman (1962) pointed out that simple linear interpolation and most methods that commonly use related monthly series to interpolate quarterly series could be improved by estimating the degree of correlation between the related monthly and quarterly series. He showed that the optimal linear unbiased interpolation of a quarterly series using a related monthly series could be solved by a standard regression technique. [Optimal here and hereafter refers to minimum variance.] Chow and Lin (1971) generalized and extended Friedman's work. Their framework for analysis provides the basis for the method I propose in Section 3.

To set up that framework, assume that observations are available on a variable of interest, y, only on a quarterly basis. We want to estimate monthly values of that variable such that their average is equal to the quarterly value. [All variables with monthly time units are indicated by asterisks.] Let there be n quarterly observations: y_1, y_2, ..., y_n. For each t = 1, ..., n, we want to estimate the monthly values $y_{t,1}^*$, $y_{t,2}^*$, and $y_{t,3}^*$ such that

$$y_t = (y_{t,1}^* + y_{t,2}^* + y_{t,3}^*) / 3. \qquad (1)$$

In estimating the monthly values, assume further that the series satisfies a linear stochastic relationship with a set of p observed monthly variables. That is,

$$y^*_{t,i} = x^{1*}_{t,i}\beta_1 + x^{2*}_{t,i}\beta_2 + \ldots + x^{p*}_{t,i}\beta_p + u^*_{t,i} \qquad (2)$$

for month i of quarter t. The 3nx1 vector $U^* = (u^*_{1,1}\ u^*_{1,2}\ \ldots$ $u^*_{n,3})$ defined by this relationship is assumed to have mean zero and covariance matrix V*.

The nx3n distribution matrix B plays an important role in the estimation of the β_j's. B takes the form

$$B = (1/3) \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 \end{bmatrix} . \qquad (3)$$

Let $Y = (y_1\ y_2\ \ldots\ y_n)'$ be the nx1 vector of quarterly observations of y, and let $Y^* = (y^*_{1,1}\ y^*_{1,2}\ \ldots\ y^*_{n,3})'$ be the 3nx1 vector of unobserved monthly values. Then

$$Y = BY^*. \qquad (4)$$

We desire an optimal linear unbiased estimator of Y*. In the Chow-Lin procedure, the solution to this problem is the estimator

$$\hat{Y}^* = X^*\hat{\beta} + V^*B'(BV^*B')^{-1}\hat{U} \qquad (5)$$

where

$$X^* = \begin{bmatrix} x^{1*}_{1,1} & x^{1*}_{1,2} & \cdots & x^{1*}_{n,3} \\ x^{2*}_{1,1} & x^{2*}_{1,2} & \cdots & x^{2*}_{n,3} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x^{p*}_{1,1} & x^{p*}_{1,2} & \cdots & x^{p*}_{n,3} \end{bmatrix} \qquad (6)$$

is the 3nxp matrix of monthly explanatory variables;

$$\hat{\beta} = [X'(BV*B')^{-1}X]^{-1}X'(BV*B')^{-1}Y \qquad (7)$$

is the generalized least-squares estimate of the coefficients in a regression of Y on the quarterly averaged data, X, given by

$$X = BX*; \qquad (8)$$

and

$$\hat{U} = Y - X\hat{\beta} \qquad (9)$$

is the nx1 vector of residuals in the quarterly regression.

The intuition behind this solution is that the monthly estimates of y* are based on two components, the first a linear function of the monthly movements in the related x* variables and the second a distribution of the quarterly residuals so that the monthly values average to the quarterly observations.

In most cases, the relationship between short-run movements in y* and x*β is fairly stable, but the levels of y* and x*β may vary over time. The quarterly residuals will then exhibit serial correlation, and the Chow-Lin procedure with V* proportional to the identity will be inadequate. In particular, this procedure will lead to step discontinuities of the monthly estimates between quarters because it allocates each quarterly residual equally among the three monthly estimates.

Chow and Lin (1971) proposed a method of estimating a V* matrix associated with errors that are generated by a first-order Markov process. This technique is probably an improvement over

the estimates based on an assumption of no serial correlation, but it is adequate only when the error process is stationary. The test results presented below suggest that this procedure is often inadequate.

Fernandez (1981) recently proposed a generalization of the Chow-Lin procedure that corrects for this problem. The Fernandez solution derives from a model in which u* follows a random walk. That is,

$$u^*_{t,i} = u^*_{t,i-1} + \varepsilon^*_{t,i} \tag{10}$$

where $u^*_{t,0} = u^*_{t-1,3}$ and $\varepsilon^*_{t,i}$ is a white noise process with variance σ^2. As an initial condition, Fernandez assumed that $u^*_{0,3} = 0$.

In this case the formula for the Chow-Lin estimator is used, except V* is replaced by $(D'D)^{-1}$ where the 3nx3n matrix D is given by

$$D = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ -1 & 1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & -1 & 1 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & -1 & 1 \end{bmatrix}. \tag{11}$$

Because under these assumptions var $(U^*) = (D'D)^{-1}\sigma^2$, the Fernandez estimator is best linear unbiased.

3. A CORRECTION FOR SERIAL CORRELATION

I will show that the Fernandez suggestion of allowing random drift in the error process often appears to improve estimates relative to either of the Chow-Lin estimators. However,

the Fernandez procedure is quite specific in its assumption about the error process. The random walk assumption for the monthly error term defines a filter that will remove all serial correlation in the quarterly residuals when the model is correct. In several applications of the Fernandez procedure, I found that this particular filter did not remove all of the serial correlation. Fernandez suggested that in such cases one should prefilter the data before applying his procedure. As an alternative to an ad hoc search for an appropriate filter, I suggest the following generalization of the Fernandez approach.

Assume that the monthly values of y* are generated by

$$y^*_{t,i} = x^{1*}_{t,i}\beta_1 + x^{2*}_{t,i}\beta_2 + \dots + x^{p*}_{t,i}\beta_p + u^*_{t,i} \tag{12}$$

where

$$u^*_{t,i} = u^*_{t,i-1} + \varepsilon^*_{t,i} \tag{13}$$

and

$$\varepsilon^*_{t,i} = \alpha\varepsilon^*_{t,i-1} + e^*_{t,i} \tag{14}$$

where $e^*_{t,i}$ is a white noise process with variance σ^2.

As initial conditions, assume that $u^*_{0,3} = \varepsilon^*_{0,3} = 0$. [This assumption greatly simplifies the analysis. It could be relaxed by backcasting the initial residuals.] To derive the best linear unbiased estimator of Y* under these conditions, we need to derive the variance matrix for u*. Let the 3nx3n matrix H be given by

$$H = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 & 0 \\ -\alpha & 1 & 0 & . & . & . & 0 & 0 \\ 0 & -\alpha & 1 & . & . & . & 0 & 0 \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ 0 & 0 & 0 & . & . & . & -\alpha & 1 \end{bmatrix}. \tag{15}$$

Then

$$E^* = HDU^* \tag{16}$$

where

$$E^* = (e^*_{1,1} \ e^*_{1,2} \ \cdots \ e^*_{n,3}). \tag{17}$$

Thus, $U^* = D^{-1}H^{-1}E^*$, and

$$\text{var } (U^*) = (D'H'HD)^{-1}\sigma^2. \tag{18}$$

Replacing V^* in the Chow-Lin formula with the above expression gives the estimator

$$\hat{Y}^* = X^*\hat{\beta} + (D'H'HD)^{-1}B'[B(D'H'HD)^{-1}B']^{-1}\hat{U} \tag{19}$$

where

$$\hat{\beta} = \{X'[B(D'H'HD)^{-1}B']^{-1}X\}^{-1}X'[B(D'H'HD)^{-1}B']^{-1}Y. \tag{20}$$

Two problems arise when we use this estimator. First, we need to estimate the Markov parameter, α. Second, the matrix $D'H'HD$ may be too large (on the order of 400x400 for postwar data) to invert by conventional methods.

The first problem can be solved by the following steps. First, form the Fernández estimator and generate the quarterly residuals, $\hat{U}$, associated with it. Under the assumptions given

above, the $\hat{U}$ will be consistent estimates of the quarterly averages of the $u^*_{t,i}$'s.

To generate an estimate of α, notice that

$$QD = \Delta B \tag{21}$$

where D and B are as above, the nx3n matrix Q is given by

$$Q = \begin{bmatrix} 3 & 2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 3 & 2 & 1 \end{bmatrix} \tag{22}$$

and the nxn matrix Δ is given by

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & \cdots & & 0 & 0 \\ -1 & 1 & 0 & \cdots & & 0 & 0 \\ 0 & -1 & 1 & \cdots & & 0 & 0 \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & & -1 & 1 \end{bmatrix}. \tag{23}$$

Notice also that

$$\Delta U = \Delta BU^* = QDU^* = QH^{-1}E^*. \tag{24}$$

Thus, an estimate of α may be obtained by forming the first-order autocorrelation coefficient of the first difference of the quarterly residuals and solving for the value of α which when substituted in H leads to a covariance matrix $QH^{-1}H^{-1'}Q'$ in which the ratio of the off-diagonal element to the diagonal element equals this coefficient. For large n, this procedure amounts to solving the equation

$$(4 + 11\alpha + 16\alpha^2 + 19\alpha^3 + 16\alpha^4 + 10\alpha^5 + 4\alpha^6 + \alpha^7)$$

$$\div (19 + 32\alpha + 20\alpha^2 + 8\alpha^3 + 2\alpha^4) = q \qquad (25)$$

where q is the first-order serial correlation coefficient of the differenced quarterly residuals. Given q, this equation provides a unique solution for α. (The equation defines a one-to-one mapping between q and α with domain and range equal to the interval [-1, 1].)

The second problem, the inversion of D'H'HD, is solved by taking advantage of the structure of this matrix. It is easy to show that

$$\left[(D'H'HD)^{-1}\right]_{i,j} =$$

$$(1 - \alpha)^{-2} \sum_{\ell=1}^{\min(i,j)} \sum_{s=1}^{\ell} (1 - \alpha^{i-s+1})(1 - \alpha^{j-s+1}). \qquad (26)$$

4. A COMPARISON OF FOUR METHODS

A natural test of the merit of this procedure relative to others is to compare accuracy in interpolating quarterly averages of data for which the monthly values are observed. The results of such a test are reported here.

Four methods of distribution were tested. The first, labeled white noise, is the Chow-Lin estimator under the assumption that V* is the identity matrix. The second, labeled Markov, is the Chow-Lin estimator with the V* matrix estimated using Chow and Lin's procedure under the assumption that the monthly residuals are first-order Markov. [For details, see Chow and Lin 1971, pp. 374-375. To minimize expense, my implementation takes only

one pass through the suggested iteration of the estimation of the Markov parameter.] The third method, labeled random walk, is the Fernandez estimator. The final method, labeled random walk, Markov, is my procedure outlined in Section 3.

I used six sets of data to test the different methods. The time series chosen represent the kinds of data found in the National Income Accounts and the Flow of Funds Accounts for which monthly observations are not available. The related series were chosen on an ad hoc basis, and no attempt was made to improve the original specification for each series. Accuracy was measured in terms of the mean square error of the monthly distributed levels from the actual values and the mean square error of the changes in monthly distributed values from the actual changes. In all but the last test case, the ordering of the results was the same by both measures.

The test results, presented in the table, indicate that the random walk, Markov procedure is likely to be much more accurate than other methods in many distributions of time series. In four of the six data sets considered, this method had the smallest mean square error by both measures.

There are distributions, however, for which this procedure is likely to be less accurate than the other procedures. Fortunately, the test results suggest that these distributions can be detected. In the two test cases in which the random walk, Markov procedure performed worse than one of the other procedures (Cases 4 and 6), the estimated Markov parameters were -.7 and -.5. Such strong negative serial correlation in monthly data is

highly unlikely, and it probably indicates that the random walk model was misspecified. The random walk, Markov procedure should therefore probably be used only when the estimated Markov parameter is positive. When it is negative, the Chow-Lin, first-order Markov model is probably preferable.

Nevertheless, if my test results are representative, then many distributions of time series will benefit considerably from the use of the random walk, Markov procedure. In my three test cases which had positive Markov parameters (Cases 1, 2, and 3), the average reduction in the level mean square error over the best alternative was a promising 13 percent.

REFERENCES

CHOW, GREGORY C., and LIN, AN-LOH (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series," Review of Economics and Statistics, 53, 372-375.

FERNANDEZ, ROQUE B. (1981), "A Methodological Note on the Estimation of Time Series," Review of Economics and Statistics, 63, 471-476.

FRIEDMAN, MILTON (1962), "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association, 57, 729-757.

Testing the Accuracy of Four Methods of Interpolation

Nos.	Interpolated Quarterly Variables	Cases[a] Related Monthly Variables	Periods	Methods	Mean Square Errors		Markov Parameters
					Levels	Changes	
1.	Industrial Production Index	3-Month Treasury Bill Rate Manufacturing Shipments S&P Stock Price Index New Orders of Capital Goods	1948:1-1981:6	White Noise Markov Random Walk Random Walk, Markov	2.662000 0.878000 0.249000 0.172000	6.087000 1.988000 0.596000 0.416000	-- .85 -- .50
2.	Personal Income	Employees on Nonag Payrolls Avg. Hourly Earnings, Mfg. Industrial Production Index	1948:1-1979:12	White Noise Markov Random Walk Random Walk, Markov	19.617000 15.090000 11.970000 11.078000	46.924000 35.539000 28.067000 25.873000	-- .93 -- .33
3.	Unemployment	Industrial Production Index 3-Month Treasury Bill Rate	1948:1-1981:6	White Noise Markov Random Walk Random Walk, Markov	0.019650 0.013600 0.013640 0.013500	0.050010 0.034030 0.034100 0.033710	-- .98 -- .56
4.	Consumption of Nondurables	Disposable Personal Income Unemployment Rate	1959:1-1981:6	White Noise Markov Random Walk Random Walk, Markov	2.069700 1.758900 1.747900 1.935000	5.284100 4.462200 4.445600 5.195700	-- .89 -- -.70
5.	Personal Consumption Deflator	Consumer Price Index	1959:1-1981:6	White Noise Markov Random Walk Random Walk, Markov	0.017718 0.010353 0.009955 0.009855	0.035509 0.018281 0.017475 0.017207	-- .97 -- -.10
6.	Money Supply (M1B)	Monetary Base Federal Funds Rate	1959:1-1981:6	White Noise Markov Random Walk Random Walk, Markov	0.771600 0.518800 0.610500 0.607300	1.321700 0.866700 0.923600 1.002300	-- .90 -- -.50

[a]All equations include a constant and trend.