Federal Reserve Bank of Minneapolis
Research Department Working Paper

Historical

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

Robert B. Litterman

Working Paper 190
PACS File 2920

Revised October 1982

This paper describes a technique for the distribution of 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 related monthly series. The problem confronted here is not new; the standard solution is a procedure described by Chow-Lin (1971). In a set of test cases reported here, however, a recent alternative suggested by Fernández (1981) is found to be a significant improvement over Chow-Lin in many cases. The new solution to this problem proposed here is a slight modification of the Fernández procedure, which takes account of serial correlation of errors in a way which is more general than the previous methods. The tests suggest that this technique may be more accurate in many cases than any of the other solutions.

The distribution[1/] of time series is a problem frequently faced by empirical researchers for which several solutions have been proposed. In an early survey and critique of standard techniques Friedman [1962] pointed out that simple linear interpolation, and the bulk of commonly-used methods which relied on using related monthly series to interpolate quarterly series, could be improved upon by making use of an estimate of the degree

---

[1/] The related problem of interpolation of a time series refers to the estimation of unobserved values of a stock variable whose actual values are observed less frequently. Distribution refers to the problem with flow variables or time averages of stock variables of estimating a set of values, the sum or average of which equals the observed value over the longer interval. Interpolation is handled in a fashion parallel to the techniques described here; see Chow-Lin [1971] for details.

of correlation between the related monthly and quarterly series. He showed that the optimal[2/] linear unbiased interpolation of a quarterly series using a related monthly series can be solved by a standard regression technique. Chow and Lin [1971] generalized and extended Friedman's work. Their framework for analysis provides the basis for the method proposed in the next section.

It is assumed that observations are available on a variable of interest, y, only on a quarterly basis. It is desired to estimate monthly values of that variable such that their average is equal to the quarterly value. Let there be n quarterly observations, $y_1, y_2, \ldots, y_n$. For each $t = 1, \ldots, n$ we wish to estimate these monthly values $\underline{y_{t1}}$, $\underline{y_{t2}}$, and $\underline{y_{t3}}$,[3/] such that

(1) $$y_t = (\underline{y_{t1}}+\underline{y_{t2}}+\underline{y_{t3}}) \; / \; 3$$

In order to estimate the monthly values, it is further assumed that the series satisfies a linear stochastic relationship with a set of p observed monthly variables. That is,

(2) $$\underline{y_{ti}} = \underline{x^1_{ti}}\,\beta_1 + \underline{x^2_{ti}}\beta_2 + \cdots + \underline{x^p_{ti}}\beta_p + \underline{u_{ti}}$$

for month i of quarter t. The 3n x 1 vector $\underline{U} = [\underline{u_{11}}\ \underline{u_{12}} \cdots \underline{u_{n3}}]$ defined by this relationship is assumed to have mean 0 and covariance matrix $\underline{V}$.

[2/] Optimal throughout this paper refers to minimum variance.

[3/] Variables with monthly time units will be underscored throughout.

The n x 3n distribution matrix B plays an important role in the estimation of the $\hat{}_j$'s. B takes the form

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

Let $Y = [y_1\ y_2\ \cdots\ y_n]'$ be the n x 1 vector of quarterly observations of y, and $\underline{Y} = [\underline{y_{11}}\ \underline{y_{12}}\ \cdots\ \underline{y_{n3}}]'$ be the 3n x 1 vector of unobserved monthly values. Then

$$(4)\qquad Y = B\underline{Y}$$

We desire an optimal linear unbiased estimator of $\underline{Y}$. Chow-Lin show that the solution to this problem is the estimator

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

where

$$(6)\qquad \underline{X} = \begin{bmatrix} \underline{x^1_{11}} & \underline{x^1_{12}} & & \underline{x^1_{n3}} \\ \underline{x^2_{11}} & \underline{x^2_{12}} & & \underline{x^2_{n3}} \\ & & & \\ \underline{x^p_{11}} & \underline{x^p_{12}} & & \underline{x^p_{n3}} \end{bmatrix}$$

is the 3n x p matrix of monthly explanatory variables;

$$(7)\qquad \hat{\beta} = (X'(B\underline{V}B')^{-1}X)^{-1}X(B\underline{V}B')^{-1}Y$$

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

$$(8)\qquad X = B\,\underline{X};$$

and

$$\hat{U} = Y - X\hat{\beta} \tag{9}$$

is the n x 1 vector of residuals in the quarterly regression.

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

In most cases of interest, it is likely that the relationship between short run movements in $\underline{y}$ and in $\underline{x\beta}$ is fairly stable, but that over time the levels of $\underline{y}$ and $\underline{x\beta}$ may vary. In such a case, the quarterly residuals will exhibit serial correlation and the Chow-Lin procedure with $\underline{V}$ proportional to the identity will be inadequate. In particular, such a procedure, which allocates each quarterly residual equally among the three monthly estimates, will lead to step discontinuities of the monthly estimates between quarters.

Chow-Lin propose a method of estimating a $\underline{V}$ matrix associated with errors that are generated by a first order Markov process. This technique is likely to be an improvement over the estimates based on a no serial correlation assumption, but it will only be adequate when the error process is stationary. The test results presented below suggest this procedure is often inadequate.

Fernández [1981] recently proposed a generalization of Chow-Lin which corrects for this problem. The Fernández solution derives from a model in which $\underline{u}$ follows a random walk. That is,

$$(10) \qquad \underline{u_{ti}} = \underline{u_{t\ i-1}} + \underline{\varepsilon_{ti}}$$

where $\underline{u_{t0}} = \underline{u_{t-1\ 3}}$ and $\underline{\varepsilon_{ti}}$ is a white noise process with variance σ^2. As an initial condition, Fernández assumed that $\underline{u_{03}} = 0$.

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

$$(11) \qquad D = \begin{bmatrix} 1 & 0 & 0 & & 0 & 0 \\ -1 & 1 & 0 & & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ & & & & & \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix}$$

Since under these assumptions $\mathrm{Var}[\underline{U}] = (D'D)^{-1}\sigma^2$, the Fernández estimator is best linear unbiased.

A Correction for Serial Correlation

It will be shown below that the Fernández suggestion of allowing random drift in the error process often appears to improve estimates relative to either of the Chow-Lin estimators. However, the Fernández procedure is quite specific in its assumption about the error process. The random walk assumption for the monthly error term defines a filter which will remove all serial correlation in the quarterly residuals when the model is correct. In several applications of the Fernández procedure, I found that this particular filter did not remove all of the serial correlation. Fernández suggests that in such cases, one should prefilter the data before applying his procedure. As an alternative to such an ad hoc search for an appropriate filter, I suggest the following generalization of the Fernández approach.

Assume the monthly values of $\underline{y}$ are generated by

$$\underline{y_{ti}} = \underline{x_{ti}^1}\beta_1 + \underline{x_{ti}^2}\beta_2 + \cdots + \underline{x_{ti}^p}\beta_p + \underline{u_{ti}} \tag{12}$$

where

$$\underline{u_{ti}} = \underline{u_{t\ i-1}} + \underline{\varepsilon_{ti}} \tag{13}$$

and

$$\underline{\varepsilon_{ti}} = \alpha\underline{\varepsilon_{t\ i-1}} + \underline{e_{ti}}$$

where $\underline{e_{ti}}$ is a white noise process with variance σ^2.

As initial conditions, assume that $\underline{u_{03}} = \underline{\varepsilon_{03}} = 0$.[4/]

In order to derive the best linear unbiased estimator of $\underline{Y}$ under these conditions, we need to derive the variance matrix for $\underline{u}$. Let the 3n x 3n matrix H be given by

$$H = \begin{bmatrix} 1 & 0 & 0 & & 0 & 0 \\ -\alpha & 1 & 0 & & 0 & 0 \\ 0 & -\alpha & 1 & & 0 & 0 \\ & & & & & \\ & & & & & \\ 0 & 0 & 0 & & -\alpha & 1 \end{bmatrix} \tag{14}$$

Then

[4/] This assumption greatly simplifies the analysis. It could be relaxed by backcasting the initial residuals.

(15) $\underline{E} = HD\underline{U}$

where

(16) $\underline{E} = [\underline{e_{11}}\ \underline{e_{12}} \cdots \underline{e_{n3}}]$

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

(17) $\text{Var}\ [\underline{U}] = (D'H'HD)^{-1}\ \sigma^2$

Replacing $\underline{V}$ in the Chow-Lin formula with the above expression gives the estimator

(18) $\underline{\hat{Y}} = \underline{X}\ \hat{\beta} + (D'H'HD)^{-1}B'(B(D'H'HD)^{-1}B')^{-1}\ \hat{U}$

where

(19) $\hat{\beta} = (X'(B(D'H'HD)^{-1}B')^{-1}X)^{-1}X'(B(D'H'HD)^{-1}B')^{-1}Y$

Two problems arise in the implementation of the above estimator. First, one needs an estimate of the Markov parameter, α. Second, the matrix D'H'HD may be too large (on the order of 400 x 400 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 $\underline{u_{ti}}$'s.

In order to generate an estimate of α, notice that

(20) $Q\ D = \Delta B$

where D and B are as above,

the n x 3n matrix Q is given by

$$(21) \qquad Q = \begin{bmatrix} 3 & 2 & 1 & 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & & 0 & 0 & 0 \\ & & & & & & & & & \\ & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & & 3 & 2 & 1 \end{bmatrix},$$

and

the n x n matrix Δ is given by

$$(22) \qquad \Delta = \begin{bmatrix} 1 & 0 & 0 & & 0 & 0 \\ -1 & 1 & 0 & & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ & & & & & \\ & & & & & \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix}$$

Notice also that

$$(23) \qquad \Delta U = \Delta B\underline{U} = QD\underline{U} = QH^{-1}\underline{E}.$$

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 diagonal element equals this coefficient. For large n, this procedure amounts to solving the equation

$$(24) \qquad \frac{(4+11\alpha+16\alpha^2+19\alpha^3+16\alpha^4+10\alpha^5+4\alpha^6+\alpha^7)}{(19+32\alpha+20\alpha^2+8\alpha^3+2\alpha^4)} = q$$

where q is the first-order serial correlation coefficient of the differenced quarterly residuals. Given q, this equation provides a unique solution for α.[5/]

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

$$(25) \qquad [(D'H'HD)^{-1}]_{ij} = (1-\alpha)^{-2} \sum_{\ell=1}^{\min(i,j)} \sum_{s=1}^{\ell} (1-\alpha^{i-s+1})(1-a^{j-s+1})$$

A Comparison of Four Methods

A natural test of the merit of this procedure relative to others is to compare the accuracy of different methods for 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 used. The first method, labeled "White Noise," is the Chow-Lin estimator under the assumption that V is the identity matrix. The second method, labeled "Markov," is the Chow-Lin estimator with the V matrix estimated using their procedure under the assumption that the monthly residuals are first-order Markov.[6/] The third method, labeled "Random Walk," is the Fernández estimator. The final

---

[5/] Equation (24) defines a one-to-one mapping between q and α with domain and range equal to the interval [-1., 1.].

[6/] See Chow-Lin [1971] pages 374-375 for details. To minimize expense, my implementation takes only one pass through the suggested iteration of the estimation of the Markov parameter.

method, labeled "Random Walk, Markov," is the procedure outlined in the previous section.

Six sets of data were used to test the different methods. The time series chosen for the tests were selected to be representative of the kinds of data found in 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 mean square error of the monthly distributed levels from actual values, and mean square error of the changes in monthly distributed values from actual changes. In all but one case, the ordering of the results was the same by either measure. The results of the tests are presented in Table 1, below.

The test results indicate that the Random Walk, Markov procedure is likely to be much more accurate than other methods in many cases. In four of the six data sets considered, this method had the smallest mean square error by both measures.

There are cases, however, in which this procedure is not called for. Fortunately, the test results suggest that these cases can be detected. In the two cases for which the Random Walk, Markov procedure performs worse than one of the others, the estimated Markov parameters are -.7 and -.5. Strong negative serial correlation in monthly data seems highly unlikely, and is probably indicative that the random walk model is misspecified. Thus, in practice it may be desirable to use this procedure only when the estimated Markov parameter is positive. In cases where it is strongly negative the Chow-Lin first-order Markov model is probably preferable. Nevertheless, if the results considered here are representative, then considerable gain may be obtained in many cases through the Random Walk, Markov procedure. For the three cases in which the Markov parameter was positive, the average reduction in the level mean square error over the best alternative was a promising 13 percent.

Table 1

Accuracy Comparison of Four Methods of Interpolation

Case 1

Variable Interpolated: Industrial Production Index
Period: 1948:1 to 1981:6
Related Monthly Variables[7/]: 3-Month Treasury Bill Rate
Manufacturing Shipments
S&P Stock Price Index
New Orders of Capital Goods

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.85		.50
Level MSE	2.662	.878	.249	.172
Changes MSE	6.087	1.988	.596	.416

Case 2

Variable Interpolated: Personal Income
Period: 1948:1 to 1979:12
Related Monthly Variables: Employees on Nonagricultural Payrolls
Average Hourly Earnings, Manufacturing
Industrial Production Index

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.93		.33
Level MSE	19.617	15.090	11.970	11.078
Changes MSE	46.924	35.539	28.067	25.873

[7/] All equations include a constant and trend.

Case 3

Variable Interpolated: Unemployment
Period: 1948:1 to 1981:6
Related Monthly Variables: Industrial Production Index
3-Month Treasury Bill Rate

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.98		.56
Level MSE	.01965	.01360	.01364	.01350
Changes MSE	.05001	.03403	.03410	.03371

Case 4

Variable Interpolated: Consumption of Nondurable Goods
Period: 1959:1 to 1981:6
Related Monthly Variables: Disposable Personal Income
Unemployment Rate

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.89		-.70
Level MSE	2.0697	1.7589	1.7479	1.9350
Changes MSE	5.2841	4.4622	4.4456	5.1957

Case 5

Variable Interpolated: Personal Consumption Deflator
Period: 1959:1 to 1981:6
Related Monthly Variables: Consumer Price Index

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.97		-.10
Level MSE	.017718	.010353	.009955	.009855
Changes MSE	.035509	.018281	.017475	.017207

Case 6

Variable Interpolated: M1B
Period: 1959:1 to 1981:6
Related Monthly Variables: Monetary Base
Federal Funds Rate

	White Noise	Markov	Random Walk	Random Walk, Markov
Markov Parameter		.90		-.50
Level MSE	.7716	.5188	.6105	.6073
Changes MSE	1.3217	.8667	.9236	1.0023

Bibliography

Chow, Gregory and An-loh Lin, "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series," The Review of Economics and Statistics 53 (November 1971), 372-375.

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

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