Journal of Forecasting, Vol. 6, 21–40 (1987)

Structural Change and the Combination of Forecasts

FRANCIS X. DIEBOLD
Board of Governors of the Federal Reserve System, Washington

PETER PAULY*
University of Pennsylvania

ABSTRACT

Forecasters are generally concerned about the properties of model-based predictions in the presence of structural change. In this paper, it is argued that forecast errors can under those conditions be greatly reduced through systematic combination of forecasts. We propose various extensions of the standard regression-based theory of forecast combination. Rolling weighted least squares and time-varying parameter techniques are shown to be useful generalizations of the basic framework. Numerical examples, based on various types of structural change in the constituent forecasts, indicate that the potential reduction in forecast error variance through these methods is very significant. The adaptive nature of these updating procedures greatly enhances the effect of risk-spreading embodied in standard combination techniques.

KEY WORDS Structural change Forecast combination Varying-parameter models

'... in reality there are constant changes, structural shifts in the economy, changes in attitudes, political moves that alter established trends, new technological developments, and the like, which cause existing patterns to change and existing relationships to shift. Forecasting must, therefore, accept that structural changes in the data are and will be taking place. Otherwise, it will not be a relevant and practical field. The major question, then, becomes how the various methods perform under a continuously changing environment. There is little interest in knowing which methods perform the best in fitting a model to a set of data. The most important and relevant aspect of forecasting is to know the methods which can minimize the post-sample forecasting errors.'

MAKRIDAKIS *et al.* (1984)

It has been said that the ultimate test of any econometric model is its performance as a forecasting device. Forecasting and simulation are the two most important uses of econometric models, and any new technique which significantly enhances their performance in one or both of these areas is a welcome addition to the economist's tool kit.

* We would like to thank the participants of the May 1985 meeting of the Institute for Applied Systems Analysis (Working Group on Statistical Analysis of Economic Structural Change), and the September 1985 meeting of Project LINK, for valuable comments. All errors are our own. Parts of this work draw on a companion paper, Diebold and Pauly (1986).

0277-6693/87/010021–10$10.00

Received April 1986
Revised December 1986

The recent boom in the area of time series analysis has greatly advanced our ability to forecast economic time series. Economic time series analysis is particularly powerful because the 'typical spectral shape' (Nerlove, 1964; Granger, 1966) of economic variables is well described by certain parsimonious classes of models, such as the Box and Jenkins (1970) multiplicative seasonal ARIMA model or the popular 'unobserved components' (trend, cycle, seasonal, irregular) model as developed by Persons (1919, 1925) and refined by Nerlove (1967), Engle (1978), Nerlove, Grether and Carvalho (1979), Harvey (1984), among others.

Short-term forecasts from these models have generally been found to be superior to conventional econometric (structural) forecasts in terms of mean squared prediction error and other reasonable criteria. However, this fact does not mean that structural forecasts should be discarded in favour of nonstructural time series approaches. First, it is well known that structural econometric models enable behavioural simulation and the study of policy issues which are impossible to analyze with a nonstructural model. Second, it has been argued (McNees, 1982) that the 'economic fundamentals' and nonlinearities contained in most econometric models have the most impact in the medium to long run, making those models most useful for forecasting over longer horizons. Finally, recent results in the theory of combining forecasts suggest that, even for very short term forecasting, various candidate models such as nonstructural/time series, structural/econometric, and expert consensus may *all* prove valuable. Specifically, in their pioneering work, Bates and Granger (1969) showed that if a number of unbiased forecasts of the same future variable are available, then it is rarely (if ever) optimal to seek out the 'best' of the competing forecasts and use it alone. Rather, the forecasts can always be combined in such a way that the composite forecast has variance less than or equal to any of the competing forecasts.

We view the explicit modeling of 'drift' in the combining weights as an attempt to compensate for the poor performance of the primary forecasts in situations of structural change of unknown form. In many situations, such an approach yields powerful increases in forecasting performance because the available primary forecasts do not adequately account for structural change. Furthermore, even if it is desired to explicitly model structural change in the primary forecasts, it is often difficult (or impossible) to locate and compensate for the changing structure, particularly in an ongoing forecasting organization when timely forecasts must be produced.

In section 1, we consider time-varying coefficient combining methods within the context of the so-called 'variance–covariance' approach. In section 2, we develop a number of regression-based combining methods which are more general than, and include as special cases, the time-varying variance–covariance methods. The dramatic improvements in forecasting performance which our methods deliver are illustrated by a numerical example in section 3. Conclusions and directions for future research are given in section 4.

1. TIME-VARYING WEIGHTS WITH THE VARIANCE–COVARIANCE METHOD

Consider two competing forecasts, f_t^1 and f_t^2, of y_t made at time $t-1$, and suppose we restrict ourselves to combined forecasts of the form

$$C_t = \phi f_t^1 + (1-\phi) f_t^2.$$

The 1-step ahead combined prediction error satisfies the same equality:

$$e_t^c = \phi e_t^1 + (1-\phi) e_t^2,$$

because

$$e_t^c = y_t - C_t.$$

Thus,

$$\operatorname{var} e_t^c = \phi^2\sigma_1^2 + (1-\phi)^2\sigma_2^2 + 2\phi(1-\phi)\sigma_{12}.$$

The ϕ weight which minimizes this expression is given by

$$\phi^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

We will refer to the calculation of the optimal weights in this fashion as the 'variance–covariance method'. Note the intuitive results that

$$\lim_{\sigma_1^2 \to \infty} \phi^* = 0 \qquad \lim_{\sigma_2^2 \to \infty} \phi^* = 1 \qquad \lim_{\sigma_1^2 \to 0} \phi^* = 1 \qquad \lim_{\sigma_2^2 \to 0} \phi^* = 0$$

Thus, the more reliable f_1 is, the more weight placed on it, and vice versa. Note also that the *covariance* σ_{12} plays an important role, and that the restriction that combining weights sum to unity guarantees that the combined forecast will be unbiased if the primary forecasts are unbiased.

Maintaining the restriction of unbiasedness of the primary forecasts, but allowing for an arbitrary finite number (m) of them, it has been shown by Reid (1969) and Granger and Newbold (1974) that the optimal combining weight vector $\boldsymbol{\phi}$ is given by

$$\underset{(m\times 1)}{\boldsymbol{\phi}} = (\Sigma^{-1}i)/(i'\Sigma^{-1}i),$$

where Σ is the variance–covariance matrix of 1-step ahead forecast errors and i is a conformable column vector of ones.

In practice we estimate $\boldsymbol{\phi}$ by replacing Σ with an estimate $\hat{\Sigma}$, where

$$\hat{\Sigma}_{ij} = \sum_{t=1}^{T} e_{it}e_{jt}.$$

Thus, the elements of Σ are viewed as fixed, but unknown, quantities to be estimated from the T sample observations. Even in a real-time forecasting environment in which $\hat{\boldsymbol{\phi}}_T$ is updated recursively and therefore changes as $T \to \infty$, the change is not structural but rather represents the convergence in probability of $\hat{\boldsymbol{\phi}}_T$ to $\boldsymbol{\phi}$.

A number of authors have recognized that the true but unknown matrix Σ, and hence the vector $\boldsymbol{\phi}$, may not be fixed over time. In such situations, the use of $\hat{\boldsymbol{\phi}}$ may be severely suboptimal. This suboptimality of fixed-weight combinations occurs for many reasons. For example, differential learning speeds of different forecasting groups and/or forecasting techniques may lead to a particular forecast becoming progressively better over time, relative to others. In such a situation, a truly optimal combining procedure should weight that particular forecast progressively more heavily over time. Second, the design of various forecasting models may make them relatively better forecasting tools in some situations than in others. Thus, a structural model with a highly developed wage-price sector may substantially outperform a simpler model during times of high inflation. In such times, the more sophisticated model should receive higher weight, and so on. Third, the macroeconomic environment changes, or 'drifts' over time, and certain forecasting techniques may be relatively more vulnerable to such change. Moreover, if many different types of structural change are simultaneously occurring, we would expect them to have differential effects on forecasts produced by different methods. Finally, the nonlinearity in the underlying economic structure leads directly to nonconstant forecast error variances, and hence to the desirability of nonconstant combining weights, as forcefully argued by Greene, Howrey and Hymans (1985). Most of the procedures that have been proposed to deal with the drift problems for the variance–covariance combining method are adaptive 'real-time' algorithms for calculating the combining weights. These methods make use of a moving data subset (e.g. the V most recent observations) to calculate the weights. Thus,

$$\hat{\boldsymbol{\phi}}_T = (\hat{\Sigma}_T^{-1} i)/(i'\hat{\Sigma}_T^{-1} i) \tag{1}$$

where

$$\hat{\Sigma}_{ij,T} = V^{-1} \sum_{t=T-V+1}^{T} e_{it} e_{jt}.$$

This has the desirable properties of giving the most weight to those forecasts which have performed best in the recent past and allowing for the possibility of a nonstationary relationship over time between the primary forecasts. On the other hand, the choice of V is arbitrary, and its value will have substantial effects on the estimated combining weights. Furthermore, as noted by Bessler and Brandt (1981) most of these methods not only lead to convex combining weights (as opposed to weights that simply sum to unity), but also force each weight to lie in the interval $(0, 1/m - 1)$, where m is the number of primary forecasts. This limitation is particularly severe if one primary forecast is substantially better than the others.

Granger and Newbold (1977) suggest the following possibilities, in addition to (1), which we list here for comparison with later results. Assuming that m forecasts are to be combined, we have

$$\hat{\phi}_{iT} = \left(\sum_{t=T-V+1}^{T} e_{it}^2\right)^{-1} \bigg/ \left(\sum_{j=1}^{m} \left(\sum_{t=T-V+1}^{T} e_{jt}^2\right)^{-1}\right) \qquad i = 1, \ldots, m \tag{2}$$

$$\hat{\phi}_{iT} = a\hat{\phi}_{i,T-1} + (1-a)\left(\sum_{t=T-V+1}^{T} e_{it}^2\right) \bigg/ \left(\sum_{j=1}^{m} \left(\sum_{t=T-V+1}^{T} e_{jt}^2\right)^{-1}\right) \qquad i = 1, \ldots, m \ (0 < a < 1) \tag{3}$$

$$\hat{\boldsymbol{\phi}}_T = (\hat{\Sigma}_T^{-1} i)/(i'\hat{\Sigma}_T^{-1} i), \tag{4}$$

where

$$\hat{\Sigma}_{ij,T} = \left(\sum_{t=1}^{T} \lambda^t e_{it} e_{jt}\right) \bigg/ \left(\sum_{t=1}^{T} \lambda^t\right), \qquad \lambda \geqslant 1$$

$$\hat{\phi}_{iT} = \left(\sum_{t=1}^{T} \lambda^t e_{it}^2\right)^{-1} \bigg/ \left(\sum_{j=1}^{m} \left(\sum_{t=1}^{T} \lambda^t e_{jt}^2\right)^{-1}\right), \qquad \lambda \geqslant 1. \tag{5}$$

Clearly, (1) represents a moving sample approach using all variance and covariance information, (2) uses the same moving sample but ignores covariance information, (3) is an 'adaptive' scheme which ignores covariance information, (4) uses the full sample but weights recent observations more heavily, and (5) is like (4) but ignores covariance information.

In a recent development, Engle, Granger and Kraft (1985) used the model of autoregressive conditional heteroskedasticity (ARCH), due to Engle (1982), to actually model the evolution of prediction error variances and covariances over time. This approach makes use of the full sample to produce a sequence of time-varying weights in a rigorous and systematic fashion, rather than simply (and artificially) basing the weight calculations on a recent subset of observations. While this approach represents a notable contribution, it has problems of its own. First, it produces an extremely noisy weight sequence, as opposed to the smoothly changing weights argued for by Granger and Newbold (1974). Second, although their ARCH-combined forecast does improve upon the individual forecasts, it does not compare favourably with a fixed-weight combination. They note that this may be due to misspecification of the diagonal bivariate ARCH model which they use, and that further research in this area is needed.

It should be noted that the Engle–Granger–Kraft approach requires the modeling of an entire conditional covariance matrix over time, which is a formidable task. Combination by a regression approach with time-varying parameters, on the other hand, may be more tractable since the evolution of only one parameter must be modeled. The regression estimator, while of course depending on all available variances and covariances, models their evolution *implicitly* rather than explicitly. We now consider such an approach in detail.

2. TIME-VARYING WEIGHTS AND THE REGRESSION METHOD

A weighted least squares approach

Granger and Ramanathan (1984) show that the above forecast combination theory has a regression interpretation, by estimating ϕ from

$$y_t = \beta_1 f_t^1 + \beta_2 f_t^2 + e_t \qquad \text{s.t. } \beta_1 + \beta_2 = 1.$$

(For simplicity we restrict ourselves to the case of two competing forecasts. The extension to the general case is immediate.) The restriction may be conveniently imposed by writing

$$(y_t - f_t^2) = \phi(f_t^1 - f_t^2) + e_t^c.$$

Estimation of ϕ by OLS yields a result numerically identical to ϕ^* above. To see this, recall that ϕ^* is estimated by the variance–covariance method as

$$\begin{aligned}\phi^* &= \frac{\Sigma(e_t^2)^2 - \Sigma e_t^1 e_t^2}{\Sigma(e_t^1)^2 + \Sigma(e_t^2)^2 - 2\Sigma e_t^1 e_t^2}\\ &= \frac{\Sigma(y_t - f_t^2)^2 - \Sigma(y_t - f_t^1)(y_t - f_t^2)}{\Sigma(y_t - f_t^1)^2 + \Sigma(y_t - f_t^2)^2 - 2\Sigma(y_t - f_t^1)(y_t - f_t^2)}\\ &= \frac{\Sigma(f_t^1 - f_t^2)(y_t - f_t^2)}{\Sigma(f_t^1 - f_t^2)^2},\end{aligned}$$

which is just $\hat{\phi}_{\text{OLS}}$.

Failure to impose the $\Sigma\beta_i = 1$ constraint in the regression method leads to a combined forecast that is biased unless:

(1) $Ef_t^i = y_t \,\forall\, i$

(2) $\Sigma\hat{\beta}_i = 1$

which is highly unlikely. However, it is well known that under quadratic loss there is nothing *necessarily* undesirable about a biased forecast, and, as mentioned above, the SSE will be lower than if the constraint had been imposed. In addition, any bias that may be present in the component forecasts may be eliminated by including an intercept in the combining regression; the resulting combined forecast will be unbiased and have smaller SSE than the forecast obtained by any other combining method. (It should be noted, however, that the variable being forecast must be stationary. Otherwise, an appropriate stationarity rendering transformation should be performed prior to analysis.)

The general success of time-varying weights constructed by the variance–covariance method should extend to weights produced by the regression method. As noted above, the relaxation of the restriction that the weights sum to unity and the ability to handle biased forecasts are strong advantages of the regression approach, so that it is particularly desirable to explore the possibilities for time-varying weights in that framework.

The most obvious approach, which is a direct analog of the earlier discussed adaptive variance–covariance method, is simply to base the combining regression on the most recent V observations. Furthermore, we can iterate over V to produce the V^* that minimizes mean squared combined prediction error.

Our goal of using the entire sample, while still weighting recent observations more heavily, can be met by using the familiar technique of weighted least squares. Instead of choosing β to minimize

$$e'e = \sum_{t=1}^{T} \left(y_t - \sum_{i=0}^{m} \beta_i f_t^i \right)^2$$

we instead choose it to minimize the matrix weighted average $e'We$, or

$$\sum_{t=1}^{T} \sum_{t'=1}^{T} w_{t,t'} e_t e_{t'},$$

where the $(T \times T)$ matrix of the quadratic form is given by $W = (w_{t,t'})$. For most applications it will be adequate to assume that the weighting matrix is diagonal, i.e.

$$W = \operatorname{diag}(w_{11}, w_{22}, \ldots, w_{T,T})$$

which means that we minimize the weighted sum of squares:

$$\sum_{t=1}^{T} w_{tt} \left(y_t - \sum_{i=0}^{m} \beta_i f_t^i \right)^2.$$

The least squares estimator is of course

$$\hat{\beta}_{\mathrm{WLS}} = (X'WX)^{-1} X'WY.$$

Note that a 'moving sample' estimator, analagous to the moving variance–covariance estimator discussed earlier, emerges as a special case when

$$W = \operatorname{diag}(w_a, w_b) \quad \text{and} \quad \underset{((T-V)X1)}{w_a} = (0, \ldots, 0)' \qquad \underset{(VX1)}{w_b} = (1, \ldots, 1)'.$$

A simple method for ensuring that the influence of past observations declines with their distance from the present is to specify

$$W = \operatorname{diag}(w_1, w_2, \ldots, w_T),$$

where

$$w_t \geqslant w_{t-1} \; \forall t = 2, \ldots, T.$$

There are, of course, insufficient degrees of freedom to maintain such generality, so that explicit parameterizations such as linearly or geometrically declining elements of the weighting matrix may prove extremely useful. Extracting a factor k $(k > 0)$, the linear specification is given by

$$W = k \begin{bmatrix} 1 & & & \\ & 2 & & \\ & & \ddots & 0 \\ & 0 & & T \end{bmatrix}. \tag{6}$$

Similarly, a geometric specification is

$$W = k\begin{bmatrix} \lambda & & & \\ & \ddots & 0 & \\ & & \lambda^{T-1} & \\ 0 & & & \lambda^{T} \end{bmatrix} \quad \text{for } \lambda \geqslant 1. \tag{7}$$

A more general nonlinearly declining weight specification, closely related to the Box and Cox (1964) transformation, is given by

$$W = k\begin{bmatrix} 1^{\lambda} & & \\ & 2^{\lambda} & 0 \\ 0 & & \ddots\, T^{\lambda} \end{bmatrix}$$

or $W = \text{diag}[w_{tt}] = [kt^{\lambda}]$, where $k, \lambda > 0$. Note that

$$\frac{\mathrm{d}w_{tt}}{\mathrm{d}t} = k\lambda t^{(\lambda-1)} > 0,$$

which guarantees that the recent past is weighted more heavily than the distant past. Furthermore,

$$\frac{\mathrm{d}^2 w_{tt}}{\mathrm{d}t^2} = k\lambda(1-\lambda)t^{(\lambda-2)} \quad \begin{matrix} >0 \text{ if } \lambda > 1 \\ <0 \text{ if } \lambda < 1. \end{matrix}$$

Thus, the sign of $(\lambda - 1)$ determines whether the weights decline at an increasing rate or at a decreasing rate (as we go farther back into the past). A full Box–Cox transformation may also be undertaken by letting $W = kW^*(\lambda)$, where

$$w_{tt}^*(\lambda) = \begin{cases} (t^{\lambda}-1)/\lambda & \text{if } 0 < \lambda \leqslant 1 \\ \ln t & \text{if } \lambda = 0. \end{cases}$$

To recap, we have considered five basic weighting schemes:

(1) Equal weight (standard regression-based combining): $w_{tt} = 1$ for all t.
(2) Linear: $w_{tt} = t$ for all t.
(3) Geometric: $w_{tt} = \lambda^{T-t}$, $0 < \lambda \leqslant 1$, or $w_{tt} = \lambda^{t}$, $\lambda \geqslant 1$.
(4) t^{λ} ('t-lambda'): $w_{tt} = t^{\lambda}$, $\lambda \geqslant 0$.
(5) Box–Cox: $w_{tt} = \begin{cases} (t^{\lambda}-1)/\lambda & \text{if } 0 < \lambda \leqslant 1 \\ \ln t & \text{if } \lambda = 0. \end{cases}$

Since schemes (1) and (2) emerge as special cases of one or more of schemes (3)–(5), we need not consider them further. Geometric weights for various λ values, normalized by λ^{T} so that $w_{TT} = 1$, are shown below in Figure 1 for a sample of size fifty. The case of $\lambda = 1$ corresponds to equal weights, and for all other λ the weights increase at an increasing rate as we get closer to the present (time T). The fundamental characteristic of the geometric scheme is its ability to produce weights that die out very quickly, which should be useful in an extremely unstable environment. Note also that the geometric scheme has the appealing property that it sweeps through the entire (w_{tt}, t) rectangle as λ ranges through $[1, \infty)$. However, the fact that the geometric weights always increase at an increasing rate as we get closer to the present is not necessarily desirable.

The weights obtained as a Box–Cox transformation of time are rather limited for our purposes, as shown in Figure 2. With the Box–Cox weight structure, we can obtain weights in the region bounded by the linear and log-linear schemes, but others are excluded. This is a severe limitation, and the Box–Cox weights will not be further considered.

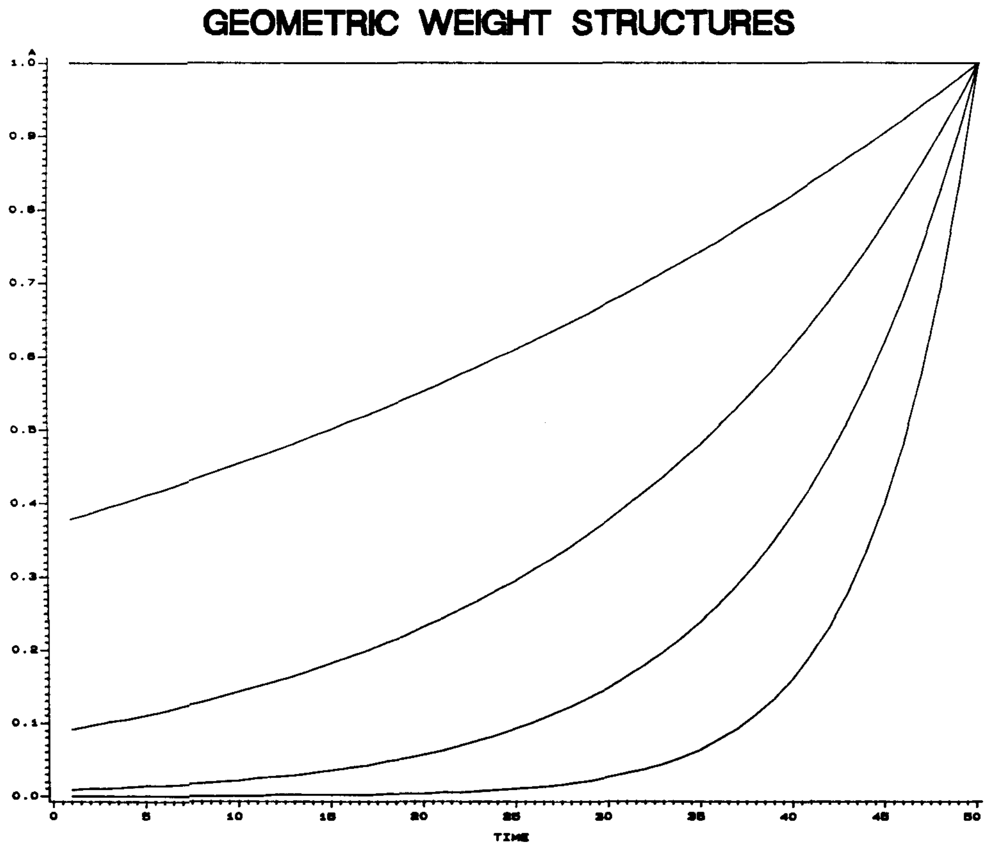


Figure 1

The t^λ specification, on the other hand, appears quite attractive. First, note that constant weights emerge for $\lambda = 0$ and linear weights emerge for $\lambda = 1$. Furthermore, unlike the geometric specification, this specification can produce weights that decrease either at an increasing or a decreasing rate, which increases its potential usefulness and applicability. In Figure 3, we see that t^λ weights as λ ranges from 0 to 7. (Again, without loss of generality, the weights are normalized by T^λ, so that $w_{TT} = 1$. Like geometric weights, t^λ weights are capable of dying out very quickly, for large λ.)

Finally, we need not pick W arbitrarily; rather, it too can be estimated. If $W = kW^*(\lambda)$, then we simply choose $\hat{\lambda}$ and $\hat{\beta}$ to solve

$$\min_{\lambda,\beta} e'We \quad \text{or} \quad \min_{\lambda,\beta} \sum_{t=1}^{T} w_{tt}^*(\lambda)\left(y_t - \sum_{i=0}^{m} \beta_i f_t^i\right)^2.$$

(Note that there is no need to choose k, since it cancels from the expression for the WLS estimator.)

This analysis highlights the extreme restrictions imposed by the 'moving sample' approach, since it restricts the weights on the V most recent observations to be constant and equal to unity, while all other weights are restricted to be constant and equal to zero. The WLS approach, on the other hand, uses *all* the data and requires only that the weights be decreasing. Furthermore, these facts should lead the WLS method to produce a non-noisy sequence of combining weights.

We have already seen that the moving sample approach (1) to the variance–covariance method

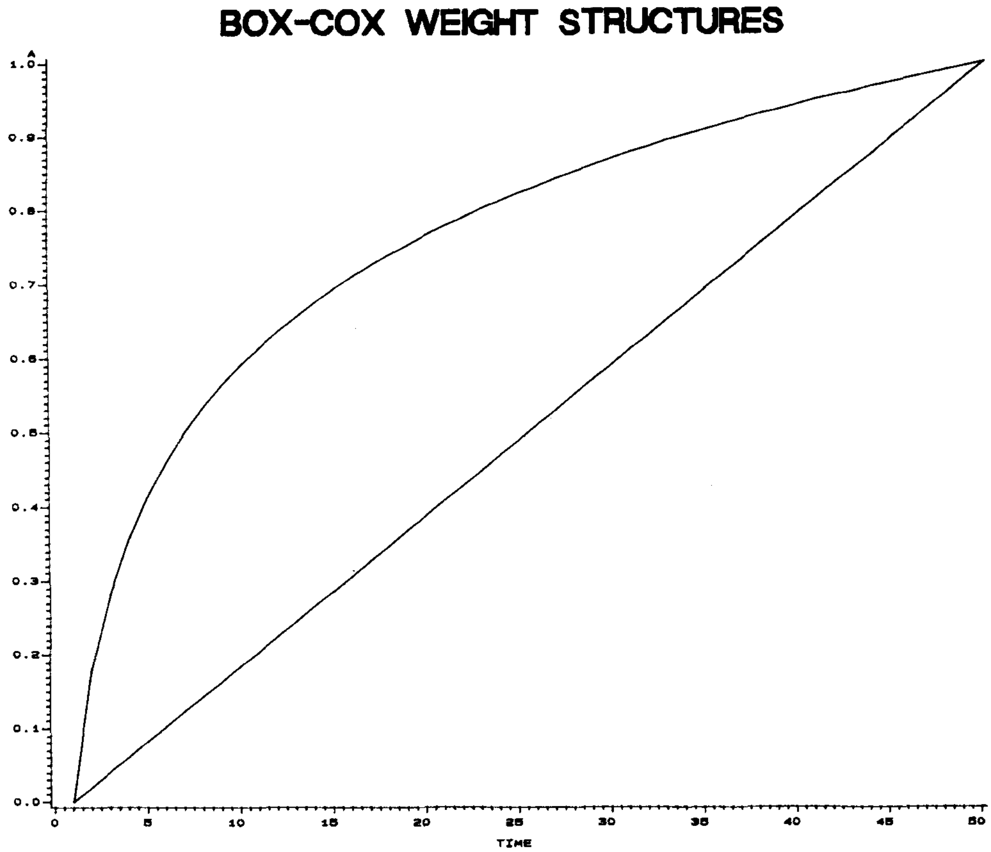


Figure 2

emerges as a special case of the WLS regression method for a particular W matrix. The WLS regression approach also sheds light on the 'weighted' variance–covariance approach (4). In particular, it is equivalent to geometric WLS, with no intercept and subject to the restriction that $\Sigma\beta_i = 1$. To see this, note that we can rewrite (4) as

$$\hat{\phi}_T = \frac{\Sigma\lambda^t(f_t^1 - f_t^2)(y_t - f_t^2)}{\Sigma\lambda^t(f_t^1 - f_t^2)^2}.$$

But this is just the expression for the OLS estimator for data that have been transformed by $(\lambda^t)^{1/2}$, i.e. it is our geometric WLS estimator.

The WLS regression approach also highlights the large amount of information that is lost in (2), (3) and (5) by ignoring covariance information, and the *convenience* of the WLS approach in terms of not having to explicitly compute all the elements of $\hat{\Sigma}$.

A regression-based approach with deterministic time-varying parameters

While the WLS regression-based approach offers substantial benefits relative to the moving sample variance–covariance approach, we may also want to consider a regression-based systematically time-varying parameter model; like our earlier WLS models, such a model makes use of the full sample. The simplest and straightforward member of this class has deterministically

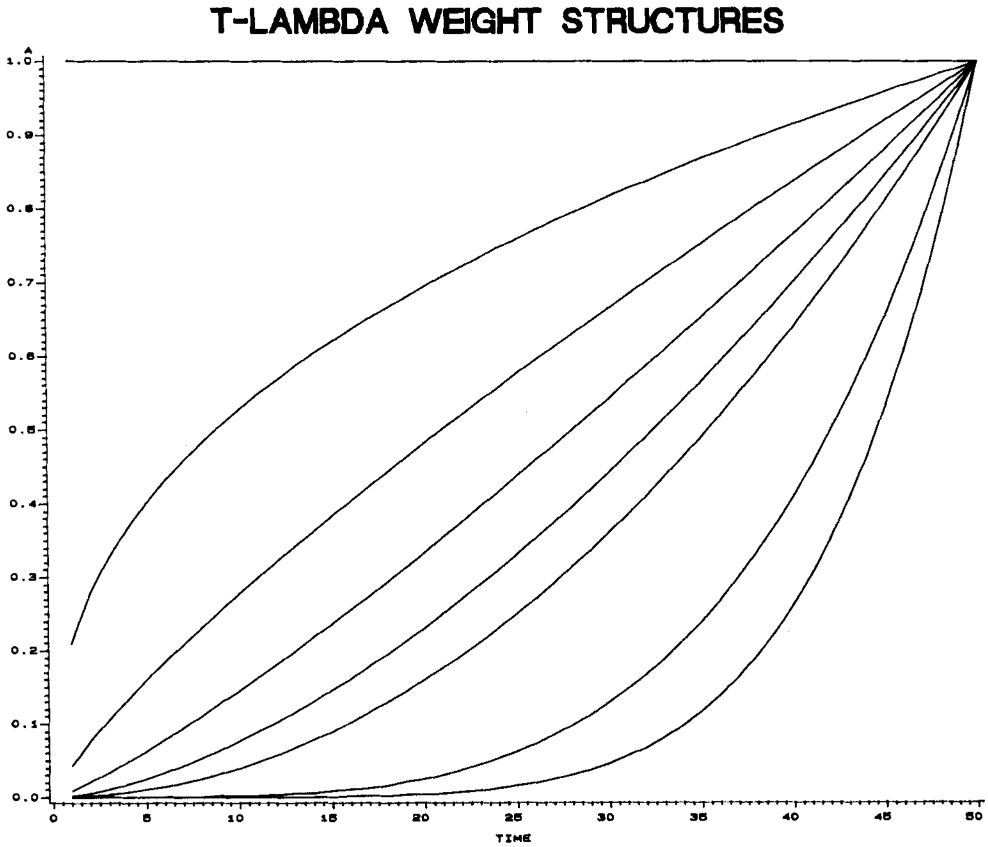


Figure 3

time-varying parameters. This gives the combining equation

$$\underset{(T\times 1)}{y} = \underset{(T\times(m+1))}{f} \underset{((m+1)\times 1)}{\beta} + \underset{(T\times 1)}{\varepsilon},$$

where

$$\beta^i = P^i(t) \qquad P^i(t) = p_0^i + p_1^i t + \cdots + p_r^i t^r \qquad i = 0, \ldots, m.$$

$f_t = (1, f_t^1, f_t^2, \ldots, f_t^m)$, $\beta = (\beta^0, \ldots, \beta^m)'$, and f is the matrix with t-th row f_t. Thus, the smoothly varying combining weights will be deterministic nonlinear (polynomial) functions of time.

If the evolution of the elements of β (due to the evolution of underlying forecast error variances and covariances) is well described by low-order deterministic time trends, then exploitation of that fact may yield substantial increases in forecasting performance. The advantage of this approach relative to our earlier WLS regression-based approach is that it enables us to explicitly *model* any parameter evolution in the combining equation, and to *project* that evolution when combining the forecasts. For example, consider the simple restricted combining equation:

$$(y_t - f_t^2) = \phi(f_t^1 - f_t^2).$$

At this point, to avoid confusion and emphasize the fact that $f_t^{\cdot}$ is a forecast of y_t made at time $t-1$, we will increase the notational burden and write ${}_{t-1}f_t^1$ and ${}_{t-1}f_t^2$ rather than f_t^1 and f_t^2. Now, at time T, the following data will be available:

$$\{y_t\}_{t=1}^{T} \qquad \{{}_{t-1}f_t^1\}_{t=1}^{T+1} \qquad \{{}_{t-1}f_t^2\}_{t=1}^{T+1}.$$

For constant parameter combination, $\hat{\phi}$ is obtained from the T-observation combining regression, and then the forecast of y_{T+1} is obtained as

$$\hat{y}_{T+1} = \hat{\phi}_T f^1_{T+1} + (1-\hat{\phi})_T f^2_{T+1}.$$

With the (linearly) deterministically time-varying parameter model, on the other hand, the combining regression is

$$(y_t - {}_{t-1}f_t^2) = (\phi_0 + \phi_1 t)({}_{t-1}f_t^1 - {}_{t-1}f_t^2) = \phi_0({}_{t-1}f_t^1 - {}_{t-1}f_t^2) + \phi_1 t({}_{t-1}f_t^1 - {}_{t-1}f_t^2).$$

The estimated parameters $\hat{\phi}_0$ and $\hat{\phi}_1$ are then used to produce the forecast as

$${}_T\hat{y}_{T+1} = (\hat{\phi}_0 + \hat{\phi}_1(T+1))_T f^1_{T+1} + (1 - \hat{\phi}_0 - \hat{\phi}_1(T+1))_T f^2_{T+1}.$$

The extension to general polynomial trends and unrestricted regression-based combination is immediate.

For example, the unrestricted regression-based analog of the above example is

$$\begin{aligned} y_t &= (p_0^0 + p_1^0 t) + (p_0^1 + p_1^1 t)_{t-1} f_t^1 + (p_0^2 + p_1^2 t)_{t-1} f_t^2 \\ &= p_0^0 + p_1^0 t + p^1_{0\,t-1} f_t^1 + p_1^1(t_{t-1} f_t^1) + p^2_{0\,t-1} f_t^2 + p_1^2(t_{t-1} f_t^2). \end{aligned}$$

After estimation of the parameters p_0^i and p_1^i $(i = 0, 1, 2)$, the predictor is obtained as

$${}_t\hat{y}_{t+1} = (\hat{p}_0^0 + \hat{p}_1^0(T+1)) + (\hat{p}_0^1 + \hat{p}_1^1(T+1))_T f^1_{T+1} + (\hat{p}_0^2 + \hat{p}_1^2(T+1))_T f^2_{T+1}.$$

Finally, while the use of time-varying parameters lessens the need to weight recent observations more heavily, it does not eliminate it. Thus, a WLS approach, together with time-varying parameters, may prove very useful.

A regression-based approach with stochastic time-varying parameters

It may be more realistic to make the regression-based combining weights stochastic, rather than deterministic, functions of time. If we view the disturbance term in the standard combining regression

$$y_t = f_t\beta + \varepsilon_t$$

as arising from a random intercept, i.e.

$$y = f\beta = \sum_{i=0}^{m} \beta^i f_t^i \text{ (where } f_t^0 = 1 \text{ for all } t) \quad \text{and} \quad \beta_t^0 = \bar{\beta}^0 + \mu_t^0.$$

Realizing, however, that there is no reason why only the *intercept* of the combining regression should be random, we are led to the logical extension

$$\beta_t^i = \bar{\beta}^i + \mu_t^i \qquad E\mu_t^i = 0 \qquad \operatorname{var}(\mu_t^i) = \gamma^i \quad \text{for all } i = 0, \ldots, m \text{ and all } t.$$

This gives the (heteroskedastic) combining equation:

$$y_t = \sum_{i=0}^{m} f_t^i(\bar{\beta}^i + \mu_t^i) = \sum_{i=0}^{m} \bar{\beta}^i f_t^i + v_t^i$$

where $v_t^i = \sum_{i=0}^{m} f_t^i \mu_t^i$. This model was studied by Hildreth and Houck (1968) and further refined by Crockett (1985). The model as stated represents purely random coefficient variation, so it is inadequate for our purposes. However, making use of the results of Singh *et al.* (1976), we can produce a stochastic systematically varying parameter model for the combining equation.

We retain $y_t = f_t\beta_t$ and write $\beta_t^i = g^i(t) + \mu_t^i$, where $g^i(t)$ is a function of time. Thus,

$$y_t = \sum_{i=0}^{m} (g^i(t) + \mu_t^i) f_t^i.$$

Rewrite this as

$$y_t = \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t$$

where $\omega_t = \sum_{i=0}^{m} f_t^i \mu_t^i$. Again, we assume that $E\mu_t^i = 0$ and $\operatorname{var} \mu_t^i = \gamma^i$. Thus, we have $E\omega_t = 0$, all t, and

$$\Omega = \operatorname{cov}(\omega) = \begin{bmatrix} \sum_{i=0}^{m} (f_1^i)^2 \gamma^i & & 0 \\ & \ddots & \\ 0 & & \sum_{i=0}^{m} (f_T^i)^2 \gamma^i \end{bmatrix}.$$

By applying a result of Breusch and Pagan (1979) as modified by Koenker (1981) we can easily develop a test for the stochastic systematic parameter variation discussed above.

Consider the heteroskedastic alternative:

$$\Omega_{tt} = h(z_t' \alpha),$$

where

$$\underset{((m+1)x1)}{z_t} = (1, z_t^{*\prime})' \quad \text{and} \quad \underset{((m+1)x1)}{\alpha} = (\sigma^2, \alpha^{*\prime})'.$$

The m-dimensional null hypothesis is that $\alpha^* = 0$, or $\Omega_{tt} = \sigma^2$ for all t. Under the assumption of normal disturbances, the Lagrange Multiplier test statistic does not depend on h and is given by

$$\mathrm{LM} = \frac{q' z (z'z)^{-1} z' q}{2\hat{\sigma}^4},$$

where $\quad q = \dot{\hat{\omega}} - \hat{\sigma}^2 i \qquad \underset{(Tx(m+1))}{Z} = (z_1, \ldots, z_T)' \qquad \hat{\sigma}^2 = \dfrac{\hat{\omega}' \hat{\omega}}{T}$

$i = (Tx1)$ column vector of ones

$\hat{\omega} =$ OLS residual vector, obtained under the null

and the 'dot' operator squares all elements of a vector or matrix.

Conveniently, the numerator of LM is equal to the explained sum of squares in a regression of $\dot{\hat{\omega}}$ on Z. Under the null, $\mathrm{LM} \overset{asy}{\sim} \chi_m^2$.

Koenker (1981) shows that the size and power of this test are extremely sensitive to the normality assumption, and he develops a robust LM test by replacing $2\hat{\sigma}^4$ with $1/T \sum (\hat{\omega}_t^2 - \hat{\sigma}^2)^2$. The reason for the robustness of the Koenker test to nonnormality is that while both $2\hat{\sigma}^4$ and $1/T \sum (\hat{\omega}_t^2 - \hat{\sigma}^2)^2$ may be viewed as estimates of $\operatorname{var}(\omega_t^2)$, the former estimate is valid only under normality since only in that situation is $\operatorname{var}(\omega_t^2) = 2\sigma^4$. On the other hand, $1/T \sum (\hat{\omega}_t^2 - \hat{\sigma}^2)^2$ is consistent for $\operatorname{var}(\omega_t^2)$ under much more general conditions, which Koenker specifies. Furthermore, the modified LM test may be calculated as TR^2 in a regression of $\dot{\hat{\omega}}$ on Z.

Thus, to implement the test, we proceed as follows. First, note that because in our case the functions $g^i(t)$ are time polynomials, we can write under the null:

$$y_t = \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t = (g_0^0 + g_1^0 t + \cdots + g_p^0 t^p) + (g_0^1 + g_1^1 t + \cdots + g_p^1 t^p) f_t^1 + \cdots$$
$$+ (g_0^m + g_1^m t + \cdots + g_p^m t^p) f_t^m + \omega_t.$$

Thus, y_t is regressed on an intercept, $t, t^2, \ldots, t^p, f_t^1, tf_t^1, \ldots, t^p f_t^1, f_t^2, tf_t^2, \ldots, t^p f_t^2, \ldots, f_t^m, tf_t^m, \ldots, t^p f_t^m$. The residuals from this regression are retained, and squared. $\dot{\hat{\omega}}_t$ is then regressed on (f_t^i),

$i = 1, \ldots, m$. The LM statistic, given by the uncentered coefficient of determination from this regression multiplied by sample size, is then distributed as χ^2_m under the null.

If the null is rejected, then we may estimate the stochastic systematically varying parameter model quite simply by recalling that

$$y_t = \sum_{i=0}^{m} g^i(t) f_t^i + \omega_t \quad \text{or} \quad y = Xg + \omega,$$

where $x_t = (1, t, \ldots, t^p, f_t^1, \ldots, t^p f_t^1, \ldots, f_t^m, \ldots, t^p f_t^m)'$ and $g = (g_0^0, \ldots, g_p^0, \ldots, g_0^m, \ldots, g_p^m)'$. As usual, $\hat{\omega} = My$, where $M = I - X(X'X)^{-1}X'$, the fundamental idempotent matrix of least squares. Note that

$$\dot{\hat{\omega}} = \dot{M}\dot{X}\gamma^{1/2} + \eta \equiv G\gamma^{1/2} + \eta$$

(where a 'dot' indicates squaring all elements, and $\gamma = (\gamma^1, \ldots, \gamma^m)'$ is the vector of parameter disturbance variances). Thus, $\hat{\gamma}$ is immediately obtained as

$$\hat{\gamma} = (G'G)^{-1}G'\dot{\hat{\omega}}.$$

This enables us to obtain $\hat{\Omega}$ and then finally

$$\hat{B}_{\text{GLS}} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y.$$

It should be noted that the above model of parameter variation is but a first step. More sophisticated analysis will allow for correlated parameter innovations, as in Swamy and Mehta (1975) and serially correlated deviations from parameter trend, as in Swamy and Tinsley (1980).

3. EXAMPLES

In order to illustrate the results, four artificial datasets, each with different stochastic properties, were generated. Each of these datasets, which we will refer to as case 1–case 4, respectively, consists of 80 observations on three variables, y_t, ${}_{t-1}f_t^{1i}$ and ${}_{t-1}f_t^{2i}$. The y_t variables are the same in each case (to aid in variance reduction) and are realizations of the stationary AR(1) process:

$$(1 - 0.9L)(y_t - 20) = \varepsilon_t \qquad e_t \overset{iid}{\sim} N(0, 1).$$

To obtain these, we set $y_0 = 0$ and generated 480 observations on the process; the last 80 observations were then used as our sample, to guarantee that the initial condition had absolutely no effect. In case i, ${}_{t-1}f_t^{1i}$ and ${}_{t-1}f_t^{2i}$ are two different forecasts of y_t, made at time $t-1$. For each case, f^{1i} and f^{2i} are generated as

$${}_{t-1}f_t^{\cdot i} = 1 + {}_{t-1}e_t^{\cdot i} + y_t.$$

Thus, each forecast (in each case) has a unit bias, and is equal to the true realized value y_t plus a one step ahead prediction error. It is the variance–covariance structures of e^{1i} and e^{2i} that are changed across the cases.

In case 1, the 1-step ahead prediction errors are uncorrelated and have constant, but different, variances throughout the sample ($\sigma_1 = 1$ and $\sigma_2 = 2$). Clearly, although we would expect f^{11} to receive more weight in the combination than f^{21}, we would not expect our time-varying coefficient methods to outperform the traditional constant coefficient methods here.

In case 2, we again have $\sigma_2 = 2$ throughout the sample, but now $\sigma_{1_t} = 1$, $t = 1, \ldots, 50$, at which point it begins to grow linearly until it achieves a value of 5 by $t = 80$. The forecast errors are again independent throughout the entire sample. In such a situation, we would expect our time-varying methods to lead to substantial forecasting improvements.

Case 3 is identical to case 2 for $t = 1, \ldots, 50$, but σ_{1t} then grows linearly from $t = 50, \ldots, 65$, at which point it reaches its maximum of 5. It then retreats linearly back to 1 by $t = 80$. Again we would expect our time-varying methods, particularly those with quickly decreasing weights and/or nonlinearly deterministically varying combining weights, to perform well.

In case 4 we explore the possibility of a changing covariance between the forecast errors. We hold $\sigma_1 = 3$ and $\sigma_2 = 4$ throughout the entire sample, while the covariance is held at 0 for $t = 1, \ldots, 50$ but grows linearly to 11.7 by $t = 80$.

To generate the forecasts, we proceed as follows. Recall that

$$_{t-1}f_t^{\cdot i} = y_t + {}_{t-1}v_t^{\cdot i},$$

where

$$_{t-1}v_t^{\cdot i} \equiv 1 + {}_{t-1}e_t^{\cdot i}.$$

Thus the prediction error v is composed of a systematic bias and the nonsystematic error e. In vector form

$$\begin{bmatrix} _{t-1}v_t^{1i} \\ _{t-1}v_t^{2i} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} _{t-1}e_t^{1i} \\ _{t-1}e_t^{2i} \end{bmatrix}$$

or $\mathbf{v}_t^i = \mathbf{1} + \mathbf{e}_t^i$, $i = 1, \ldots, 4$; $t = 1, \ldots, 80$. Thus,

$$\operatorname{cov} \mathbf{v}_t^i = \operatorname{cov} \mathbf{e}_t^i = \begin{bmatrix} \sigma_{1t}^2 & \sigma_{12t} \\ \sigma_{12t} & \sigma_{2t}^2 \end{bmatrix} \equiv \Sigma_t.$$

Our goal is to construct a sequence of vectors $\mathbf{v}_t^i$, $t = 1, \ldots, 80$, with covariance matrix Σ_t. We begin by drawing 80 vector observations $N_t = (N_{1t}, N_{2t})'$ from the bivariate normal distribution $N(0, I_2)$, and then we obtain the sequence of $\mathbf{e}_t^i$ by transforming the N_t vectors by the matrix S_t:

$$S_t = \begin{bmatrix} S_{11t} & S_{12t} \\ 0 & S_{22t} \end{bmatrix}.$$

Thus, $\mathbf{e}_t^i = S_t N_t$. Immediately,

$$\operatorname{cov} \mathbf{e}_t^i = \begin{bmatrix} S_{11t}^2 + S_{12t}^2 & S_{12t}S_{22t} \\ S_{12t}S_{22t} & S_{22t} \end{bmatrix} \equiv \Sigma^*.$$

Equating elements of Σ and Σ^*, we obtain

$$S_{22t} = \sigma_{2t} \qquad S_{12t} = \sigma_{12t}/\sigma_{2t} \qquad S_{11t} = (\sigma_{11t}^2 - \sigma_{12t}/\sigma_{2t}^2)^{1/2}.$$

This gives the desired transformation matrix S_t, so that the sequence of $\mathbf{e}_t^i$ generated as $\mathbf{e}_t^i = S_t N_t$ will have covariance matrix Σ_t at time t. Finally, we construct the $\mathbf{v}_t^i$ as $\mathbf{v}_t^i = \mathbf{1} + \mathbf{e}_t^i$, and then

$$\begin{bmatrix} _{t-1}f_t^{1i} \\ _{t-1}f_t^{2i} \end{bmatrix} = \begin{bmatrix} y_t \\ y_t \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} _{t-1}e_t^{1i} \\ _{t-1}e_t^{2i} \end{bmatrix}$$

or $\mathbf{f}_t^i = \mathbf{y}_t + \mathbf{v}_t^i$.

For each case, the following eight methods were used to produce time-varying weights:

(M1) WLS, geometric weights (λ^t)
(M2) WLS, geometric weights, linear deterministic time-varying parameters
(M3) WLS, geometric weights, quadratic deterministic time-varying parameters
(M4) WLS, t^λ weights
(M5) WLS, t^λ weights, linear deterministic time-varying parameters

(M6) WLS, t^{λ} weights, quadratic deterministic time-varying parameters
(M7) OLS (simple unrestricted regression-based combination)
(M8) Restricted OLS (variance–covariance combination).

Note that the unrestricted OLS case (i.e. equal weights) occurs in method 1 for $\lambda = 1$ and in methods 2 for $\lambda = 0$. Note also that the simple methods M3 and M4 also produce time-varying weights due to the expanding sample.

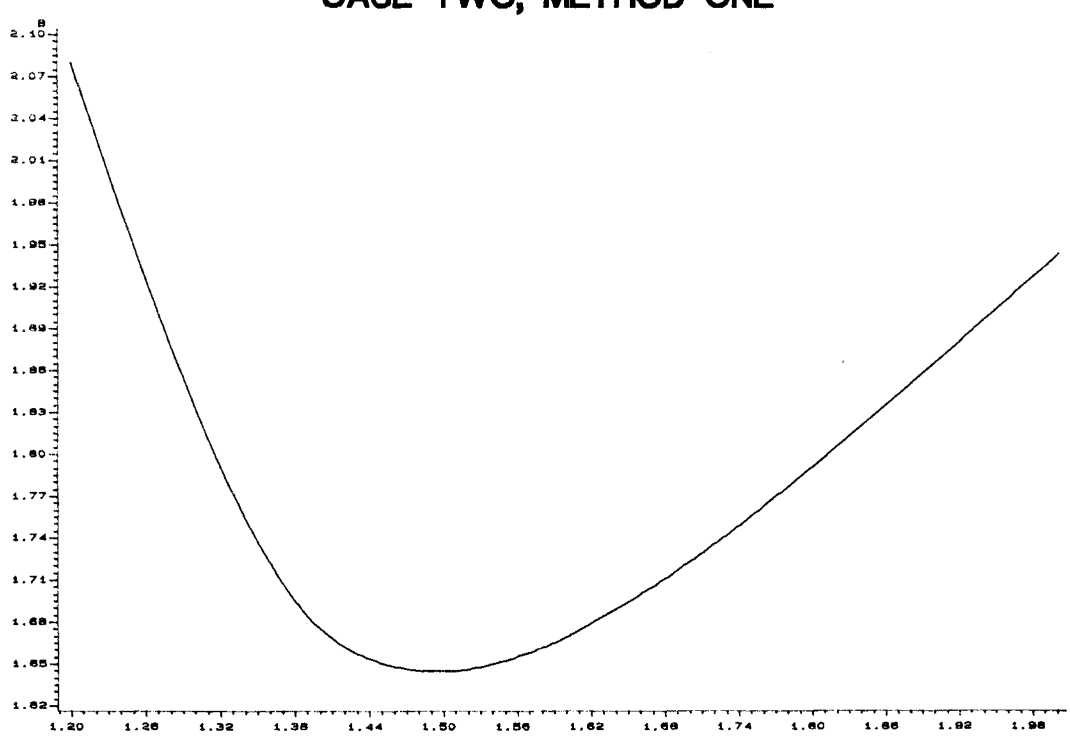
We begin the exercise in period 50, in which $\{y_t\}_{t=1}^{50}$, $\{_{t-1}f_t^{1i}\}_{t=1}^{51}$ and $\{_{t-1}f_t^{2i}\}_{t=1}^{51}$ are available, $i = 1, \ldots, 4$. These 50 observations on y are regressed on the first 50 observations of the two forecasts, and the combined forecast $\hat{y}_{51}$ is obtained as

$$\hat{y}_{51}^{i} = \hat{\beta}_0 + \hat{\beta}_{1\,50}f_{51}^{1i} + \hat{\beta}_{2\,50}f_{51}^{2i}, \qquad i = 1, \ldots, 4.$$

This process is then repeated recursively until the entire sample is exhausted. The end result, then, is four sets (corresponding to $i = 1, \ldots, 4$) of four forecasts (corresponding to the different methods). The mean squared 1-step ahead prediction errors of these forecasts (for optimal λ, calculated by a grid search) are given in Table 1.

Some general characteristics of the results are at once apparent. First, the standard (i.e. unrestricted OLS) regression-based combined forecast absolutely dominates the primary forecasts f^{1i} and f^{2i} (as well as the restricted variance–covariance combination), cutting the MSPE by

CASE TWO, METHOD ONE



MSPE AS A FUNCTION OF LAMBDA

Figure 4

Table 1. MSPE results, with associated optimal λ

	MSPE (λ*)			
	Case 1	Case 2	Case 3	Case 4
M1: WLS, λ^t	0.650 (1.20)	1.645 (1.50)	1.501 (1.55)	1.715 (1.55)
M2: λ^t, lin	0.704 (0.110)	1.525 (1.25)	1.354 (1.30)	1.352 (1.35)
M3: λ^t, qd	0.799 (1.05)	1.552 (1.15)	1.404 (1.15)	1.289 (1.25)
M4: WLS, t^λ	0.638 (9.00)	1.595 (23.00)	1.461 (24.50)	1.654 (26.00)
M5: t^λ, lin	0.700 (3.50)	1.453 (12.00)	1.290 (14.00)	1.305 (17.00)
M6: t^λ, qd	0.828 (1.00)	1.497 (6.00)	1.347 (7.00)	1.244 (10.00)
M7: OLS	0.658	3.336	3.374	3.009
M8: var–cov	2.130	8.010	7.240	10.690
f^1 alone	2.018	15.980	12.625	9.193
f^2 alone	7.569	7.569	7.569	24.475

approximately 60%. In addition, our time-varying combination procedures led to substantial further reductions in MSPE in cases 2–4.

Recall that in case 1, in addition to the usual bias of 1.0 for both f^{11} and f^{21}, the forecasts are uncorrelated and $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, for all t. The MSPE of f^{11} is 2.018, which is very close to the variance plus squared bias, while the MSPE of f^{21} is 7.569, which is somewhat above the expected MSPE of 5. The equally weighted combined forecast has MSPE of 0.658, which is less than 9% of

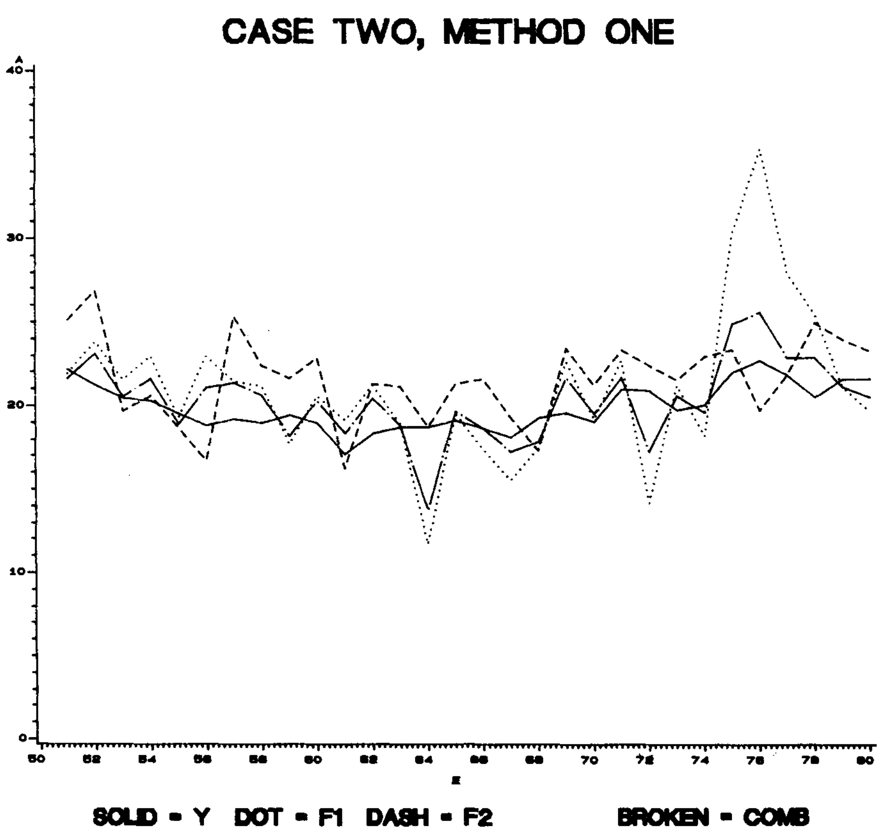


Figure 5

the MSPE of f^{11} and less than 33% of the MSPE of f^{21}. Furthermore, none of the WLS methods are helpful, which is what we expected for case 1. Moving to case 2, we see that the MSPE of the first forecast is 15.980 while that of the second forecast is 7.569, and the MSPE of the unrestricted regression-based combined forecast is a greatly reduced 3.336. More importantly, the time-varying combining methods enable *further* reductions in MSPE of approximately 50%! Consider first WLS with geometric weights. The optimal λ was found to be 1.50 and led to a combined MSPE of 1.645. This large value of λ implies quickly declining combining weights, which are needed to capture the quickly changing prediction error variance of f^{12}.

When we allow for linear deterministic time-varying combining weights in addition to the geometric WLS scheme (M2), the MSPE drops form 1.645 to 1.525, and λ^* drops from 1.50 to 1.25. The drop in MSPE is expected since the use of linearly time-varying combining weights enables us to model and forecast the structural change which is occurring. The drop in λ^* is also to be expected, because once we *model* the evolution of the combining weights there is less need to heavily discount the past. Allowing for quadratic time-varying weights with geometric WLS (M3) yields a slightly higher MSPE, which is explained by the estimation inefficiency incurred by including the unnecessary quadratic term. (Recall that σ_{1t}^2 is simply growing linearly over time.) The results for the three t^λ WLS methods (M4–M6) are much the same (and slightly better). The large λ^* values again reflect quickly declining weights.

The results for cases 3 and 4 are very similar, both in terms of the substantial decrease in MSPE

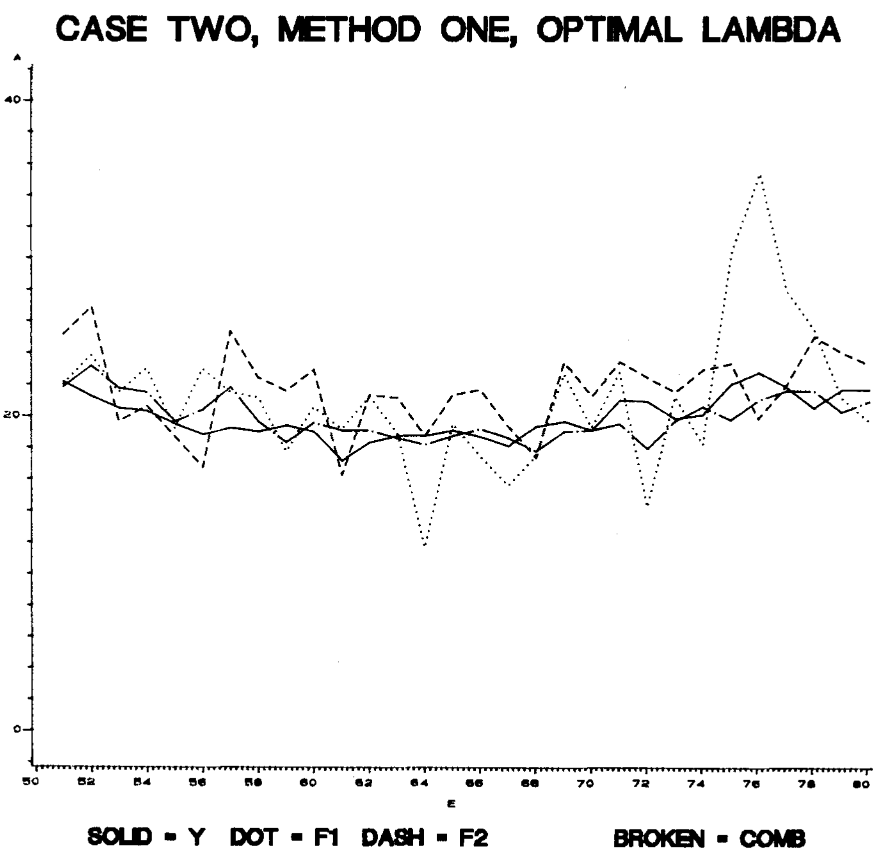


Figure 6

which is obtained by the use of our time-varying coefficient methods and the slight superiority of the t^{λ} approach (M4–M6) relative to the λ^{t} approach (M1–M3).

Finally, to develop greater intuition, the results for case 2, method 1, are explored in detail in Figures 4 to 6. First, in Figure 4 we show MSPE as a function of λ, which clearly displays a minimum at $\lambda = 1.50$. Next, in Figure 5, we show the actual series, the two primary forecasts, and the unrestricted OLS regression-based combined forecast. Note the substantial improvement in forecasting accuracy due to combining. The coefficient vector $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ changes from an initial value of (4.5, .66, .15) to a final value of (5.6, .31, .28). The reduction in the weight placed on the first constituent forecast correctly reflects the increase in its variance over the forecast period.

Figure 6 is analogous to Figure 5, except that we now use M1 with $\lambda = \lambda^* = 1.50$. The predictive improvement is clear.

4. CONCLUSIONS AND DIRECTIONS OF FUTURE RESEARCH

We have developed and illustrated the potential usefulness of regression-based WLS methods of forecast combination. It was shown that all of the earlier methods of forecast combination, based on the variance–covariance approach, emerge as special cases of the WLS regression approach, and that the suppression of explicit modeling of variances and covariances, facilitated by our approach, is particularly useful. In effect, our time-varying parameter models replace the explicit modeling of the evolution of variances and covariances, as in Engle *et al.* (1985), with the simpler modeling of time-varying regression parameters, for which a well-developed theory is available. In the example which we presented, our combined forecasts had MSPE of as little as 10% of that of the worst primary forecast, and 40% of the unrestricted OLS regression-based combined forecast.

The research is currently proceeding in a number of directions. First, we are studying the usefulness of our methods for combining forecasts of real macroeconomic time series. While we have illustrated their potential usefulness, their actual usefulness has yet to be determined. The end result will hopefully be the recommendation of a particular time-varying parameter model and weighting matrix which performs well for a wide range of variables. We are also considering other systematically time-varying parameter models, such as the random walk parameter model, which can be conveniently estimated using the Kalman filter. The Kalman filter approach also facilitates real-time parameter 'updating' and can readily handle both stationary (e.g. ARMA) and nonstationary (e.g. integrated ARMA) parameter drift. (The theory of statistical inference when Kalman filtering with nonstationary state vectors is in the early stages of development, however, but recent work by Ansley and Kohn (1983a,b) appears promising.)

Second, while combining weights obtained by our methods enable quick adaptation to structural change, they may be unduly influenced by outliers, so that robust estimation methods, such as least absolute deviations or *m*-estimation, may prove useful for the combining equation.

Third, nonlinear combining equations may lead to large decreases in MSPE, particularly if a number of linear forecasts are being combined, but the true (and unknown) process is nonlinear. Some obvious possibilities are viewing the standard combining equation as a first-order Taylor expansion, and therefore proceeding to include higher order terms, or using a Box–Cox transformation on the combining equation. Both of these methods have the advantage of including the standard linear specification as a special case. More generally, if we view forecast combination as a 'production' process with the primary forecasts as 'inputs', the use of flexible functional forms may prove worthwhile.

Finally, there is no guarantee that the disturbances of the combining equation are white, so that modeling them as a stationary ARMA process may prove very useful. In fact, Diebold (1985) has shown that such serial correlation may occur quite frequently, even if all primary forecast errors are 'white'. The k-step ahead forecast of the disturbance process then enters additively into the k-step ahead combined forecast, which provides further MSPE reductions directly related to the degree of serial correlation in the disturbance of the combining equation.

REFERENCES

Ansley, C. F. and Kohn, R., 'State-space models with diffuse initial conditions I: Filtering and likelihood', Technical Report #13, University of Chicago, 1983a.

Ansley, C. F. and Kohn, R., 'State-space models with diffuse initial conditions II: Smoothing and continuous-time processes', Technical Report #15, University of Chicago, 1983b.

Bates, J. M. and Granger, C. W. J., 'The combination of forecasts', *Operations Research Quarterly*, **20** (1969), 451–468.

Bessler, D. A. and Brandt, J. A., 'Forecasting livestock prices with individual and composite methods', *Applied Economics*, **13** (1981), 513–522.

Box, G. E. P. and Cox, D. R., 'An analysis of transformations', *Journal of The Royal Statistical Society, Series B*, **26** (1964), 211–243.

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day, 1970.

Breusch, T. S. and Pagan, A. R., 'A simple test for heteroskedasticity and random coefficient variation', *Econometrica*, **47** (1979), 1287–1294.

Crockett, P. W., 'Asymptotic distribution of the Hildreth–Houck estimator', *Journal of The American Statistical Association*, **80** (1985), 202–204.

Dickenson, J. P., 'Some statistical results on the combination of forecasts', *Operations Research Quarterly*, **24** (1975), 253–260.

Diebold, F. X., 'Serial correlation and the combination of forecasts', Unpublished Manuscript, University of Pennsylvania, 1985.

Diebold, F. X. and Pauly, P., 'The combination of forecasts', *Prevision et Analyse Economique*, **7** (1986), in press.

Engle, R. F., 'Estimating structural models of seasonality', in A. Zellner, ed., *Seasonal Analysis of Economic Time Series*, Washington, D.C.: Bureau of Census, 1978.

Engle, R. F., 'Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation', *Econometrica*, **50** (1982), 987–1008.

Engle, R. F., Granger, C. W. J. and Kraft, D. F., 'Combining competing forecasts of inflation using a bivariate ARCH model', *Journal of Economic Dynamics and Control*, **9** (1985), 67–85.

Granger, C. W. J., 'The typical spectral shape of an economic variable', *Econometrica*, **34** (1966), 150–161.

Granger, C. W. J., *Forecasting in Business and Economics*, New York: Academic Press, 1980.

Granger, C. W. J. and Newbold, P., 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of The Royal Statistical Society, Series A*, **137** (Part 2) (1974), 131–164.

Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, New York: Academic Press, 1977.

Granger, C. W. J. and Ramanathan, R., 'Improved methods of combining forecasts', *Journal of Forecasting*, **3** (1984), 197–204.

Greene, M. N., Howrey, E. P. and Hymans, S. H., 'The use of outside information in econometric forecasting', in E. Kuh and D. A. Belsley (eds.), *Model Reliability*, Cambridge, Mass.: MIT Press, 1985.

Harvey, A. C., 'A unified view of statistical forecasting procedures', *Journal of Forecasting*, **3** (1984), 245–275.

Hildreth, C. and Houck, J. P., 'Some estimators for a linear model with random coefficients', *Journal of American Statistical Association*, **63** (1968), 584–595.

Koenker, R., 'A note on studentizing a test for heteroskedasticity', *Journal of Econometrics*, **17** (1981), 107–112.

Longbottom, J. A. and Holly, S., 'The role of time series analysis in the evaluation of econometric models', *Journal of Forecasting*, **4** (1985), 75–87.

Makridakis, S., 'Forecasting accuracy and the assumption of constancy', *Omega*, **9** (1981), 307–311.

Makridakis, S. and Winkler, R. L., 'Averages of forecasts: some exact results', *Management Science*, **29** (1983), 983–996.

Makridakis, S. *et al.*, *The Forecasting Accuracy of Major Time Series Methods*, New York: Wiley, 1984.

McNees, S. K., 'The role of macroeconometric models in forecasting and policy analysis in the U.S.', *Journal of Forecasting*, **1** (1982), 37–48.
Nerlove, M., 'Spectral analysis of seasonal adjustment procedures', *Econometrica*, **32** (1964), 241–286.
Nerlove, M., 'Distributed lags and unobserved components in economic time series', in *Ten Economic Essays in the Tradition of Irving Fisher* (W. Fellner *et al.*, eds.), New York: Wiley, 1967.
Nerlove, M., Grether, D. M. and Carvalho, J. L., *Analysis of Economic Time Series: A Synthesis*, New York: Academic Press, 1979.
Persons, W. M., 'Indices of business conditions', *Review of Economic Statistics*, **1** (1919), 5–107.
Persons, W. M., 'Statistics and economic theory', *Review of Economic Statistics*, **7** (1925), 179–197.
Reid, D. J., 'A comparative study of time-series prediction techniques on economic data', Ph.D. Thesis, Department of Mathematics, University of Nottingham, 1969.
Singh, B., Nagar, A. L., Choudhry, N. K. and Raj, B., 'On the estimation of structural change: a generalization of the random coefficients regression model', *International Economic Review*, **17** (1976), 340–361.
Swamy, P. A. V. B. and Mehta, J. S., 'Bayesian and Non-Bayesian Analysis of Switching Regressions and Random Coefficient Models', *Journal of the American Statistical Association*, **70** (1975), 593–602.
Swamy, P. A. V. B. and Tinsley, P. 'Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients', *Journal of Econometrics*, **11** (1980), 259–274.

Authors' biographies:
Francis X. Diebold is with the Special Studies Section of the Board of Governors of the Federal Reserve System. He holds a B.A. and a Ph.D. in economics from the University of Pennsylvania. His current research interests focus on problems of applied time series analysis.

Peter Pauly is Associate Professor of Economics at the University of Pennsylvania. He is also one of the principal investigators of Project Link. His research interests are macroeconomics and applied econometrics, and he has published widely in both areas.