A NOTE ON OPTIMAL SMOOTHING FOR TIME VARYING COEFFICIENT PROBLEMS

BY THOMAS F. COOLEY, BARK ROSENBERG, AND KENT D. WALL

An algorithm is presented which provides a complete solution to the optimal estimation problem for time-varying parameters when no proper prior distribution is specified. The key ideas involve a combination of the information-form Kalman filter with the two-filter interpretation of the optimal smoother. The algorithm produces efficient estimates of the parameter trajectories over the entire sample and is equally applicable when a proper prior distribution has been specified.

1. INTRODUCTION

In recent years, a significant body of literature has appeared that is addressed to the problem of estimating time varying regression coefficients (Pagan [1974], Rosenberg [1973a, 1973b], Sarris [1973], Cooley and Prescott [1973, 1976]). The most fashionable approach to these problems has been to apply Kalman filtering theory to the estimation of the coefficient trajectories. The application of filtering theory, however, requires either a priori knowledge or the estimation of an initial coefficient vector.

One approach to the starting problem is to use an empirical Bayesian technique like that suggested by Kaminsky et al. [1974] or Garbade [1975]. This approach involves selecting a subsample of the observations and using them to compute a prior distribution. This, in turn, is used to initialize the filter for the remaining observations.

The components of a theoretically complete solution to the initialization and smoothing problems have existed in the control theory literature for some time. Fraser [1967], Rosenberg [1967, 1968], and Fraser and Potter [1968] developed solutions to the smoothing problem which do not require proper prior distributions for the parameter vector. The purpose of this note is to integrate these components and provide an algorithm for optimal smoothing which does not require dichotomization of the sample. The approach is a simple one based on a combination of the information form of the filter and the two-filter interpretations of the smoother.

1 If the transition and covariance parameters of the underlying coefficient process are known, the empirical Bayes approach has computational advantages. However, this is unlikely to be the case in most applications.

2 Smoothed estimates are those that use all of the information in the sample to estimate the coefficient at each point in time.
The time varying coefficient problem is characterized by a regression equation:

\[ y_i = x_i^T \beta_i + \epsilon_i, \]
\[ \beta_{i+1} = \Phi \beta_i + u_i. \]

The variables \( y \) and \( x \) represent the observables of the system, \( \Phi \) is a \( k \times k \) matrix which governs the transitions of the \( k \) component parameter vector \( \beta \). The disturbances \( \epsilon \) and \( u \) are independently and identically distributed random variables with zero means and covariance matrices \( \Sigma^e \) and \( \Sigma^u \), respectively. The problem is to obtain estimates of the \( \beta \) based on the observations \( \{y_1, \ldots, y_T\} \). Let \( \beta_{i|T} \) be an estimate of \( \beta \) based on observations \( \{y_1, \ldots, y_i\} \) and let \( P_{i|T} \) be the variance-covariance matrix of the estimated coefficients.

\[ P_{i|T} = E[(\beta_i - \beta_{i|T})(\beta_i - \beta_{i|T})^T]. \]

**Information Filters**

We now let \( P_{i+1|T} \) and \( P_{i+1} \) be the variance-covariance matrices before and after making an observation (the "predicted" and "corrected" matrices) and define the corresponding "information" matrices \( H_{i+1} = P_{i+1|T}^{-1} \) and \( H_{i} = P_{i}^{-1} \), where "\( E \)" denotes the pseudoinverse. Finally, define the forward information variables:

\[ f_{i+1} = H_{i+1} \beta_{i+1}, \]
\[ f_{i} = H_{i} \beta_{i}, \]

which play the role of the estimates in the information form of the filter. Estimation proceeds by assuming a diffuse prior distribution for the \( \beta \)'s expressed by initializing the problem with

\[ h_{1|0} = 0 \quad \text{and} \quad f_{1|0} = 0. \]

Letting \( \tilde{Q} = \Phi^{-1} \Sigma^u \), the prediction and correction formulas are:\(^1\)

**Prediction**

\[ \beta_{i+1} = (\Phi^{-1} \Sigma^u) H_{i+1} \]

\(^1\)Equation (6) has appeared in the literature in slightly different form

\[ h_{i+1} = h_{i} + \tilde{Q} (\beta_{i+1} - \beta_{i+1|T}) \]

This has led some people to the conclusion that the information form of the filter cannot accommodate singular \( \tilde{Q} \) matrices. A similar observation applies to equation (11).
\[ K_{t-1} = [I - H_{t-1}(I + \hat{Q}H_{t-1})]^{-1} \hat{Q} ](\Phi^{-1}) = (\Phi' + H_{t-1}^{'}Q)^{-1} \]

\[ H_{t-1} = K_{t-1}H_{t-1}^{'} + \Phi' \]

\[ f_{t-1} = K_{t-1}f_{t-1} \]

**Correction**

\[ f_{t-1} = f_{t+1} + \frac{x_i y_i}{\sigma^2} \]

\[ H_{t-1} = H_{t+1} + \frac{x_i y_i}{\sigma^2} \]

In addition to the forward filtered estimates, we require the backward or reverse time filtered estimates which evolve in \( r = T - t \). Denote by \( G_{t+1} \) and \( \hat{G}_{t+1} \), the predicted and corrected reverse time information matrices, and let the corresponding filtered variables be \( r_{t+1} \) and \( \hat{r}_{t+1} \). The reverse time filter is initialized with:

\[ G_{T+1} = 0 \quad \text{and} \quad \hat{r}_{T+1} = 0. \]

and the prediction and correction formulas are:

**Prediction**

\[ J_{t+1} = \Phi' [I - G_{t+1}^{'}(1 + QG_{t+1})']Q = \Phi'(I + G_{t+1}^{'}Q)^{-1} \]

\[ G_{t+1} = J_{t+1}G_{t+1}^{'}[J_{t+1}]^{-1} \]

**Correction**

\[ r_{t+1} = r_{t+1} + \frac{x_i y_i}{\sigma^2} \]

\[ G_{t+1} = G_{t+1}^{'} + \frac{x_i y_i}{\sigma^2} \]

**The Optimal Smoother**

It is clear that in econometric applications interest should focus on the most efficient estimates of the parameters which use all of the information available. These are the smoothed estimates, \( h_{t, t} \), and they may be computed as a weighted combination of \( f_{t, t} \) and \( r_{t, t} \) (Liebelt [1967]):

\[ P_{t, t} = [H_{t, t} + G_{t, t}]^{-1} \]

\[ h_{t, t} = P_{t, t} [f_{t, t} + r_{t, t}] \]

The algorithm outlined above provides the best estimates of the parameter trajectories and their associated variances, without resorting
to ad hoc procedures. It also is equally applicable when a proper prior distribution is specified, since then one merely sets $H_1 = P_1$ and $f_{2,3} = H_{1,0} b_{1,0}$.

NBER and University of California, Santa Barbara: University of California, Berkeley; NBER and Systems Control, Inc.

Submitted July 1976
Revised April 1977

REFERENCES


