TECHNICAL WORKING PAPER SERIES

ASYMPTOTICALLY OPTIMAL SMOOTHING WITH ARCH MODELS

Daniel B. Nelson

Technical Working Paper No. 161

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1994

This is a revision of parts of an earlier working paper "Asymptotic Filtering and Smoothing Theory for Multivariate ARCH Models." I would like to thank Dean Foster, Boaz Schwartz, and seminar participants at Brigham Young, Chicago, M.I.T., the Multivariate Financial Time Series Conference, the NBER Asset Pricing Group, Northwestern, Princeton, Utah, Yale and Wisconsin (Madison) for helpful discussions. This material is based on work supported by the National Science Foundation under grants #SES-9110131 and #SES-9310683. The Center for Research in Security Prices provided additional research support. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #161
August 1994

ASYMPTOTICALLY OPTIMAL SMOOTHING WITH ARCH MODELS

ABSTRACT

Suppose an observed time series is generated by a stochastic volatility model-i.e., there is an unobservable state variable controlling the volatility of the innovations in the series. As shown by Nelson (1992), and Nelson and Foster (1994), a misspecified ARCH model will often be able to consistently (as a continuous time limit is approached) estimate the unobserved volatility process, using information in the lagged residuals. This paper shows how to more efficiently estimate such a volatility process using information in both lagged and led residuals. In particular, this paper expands the optimal filtering results of Nelson and Foster (1994) and Nelson (1994) to smoothing.

Daniel B. Nelson
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637
and NBER

This paper makes two extensions to the ARCH asymptotic filtering theory of Nelson and Foster (1994) and Nelson (1994) (henceforth NF and N respectively). First, we allow a random initial condition for the filtering error. Once this extension is made, we are able to consider using both leads and lags of observed state variables to estimate unobserved state variables—i.e., smoothing. In econometric practice, the conditional variances generated by an ARCH model are usually treated as "true" apart from parameter estimation error (see, for example, the survey papers of Bollerslev, Chou, and Kroner (1992) or of Engle, Bollerslev, and Nelson (1994)). NF and N treat them simply as estimates of unobservable state variables. Clearly, if the ARCH variances *are* true, then there is no error in the estimate (conditional on the system parameters) and no room for smoothing—i.e., all information is contained in the *lagged* residuals and none in the *led* residuals. To the extent that ARCH conditional variances are noisy estimates, however, there is a role for smoothing, and two-sided ARCH models can be employed to improve estimates of historical volatilities.

The basic setup in N takes the data generating process to be a random step function with jumps at times h, 2h, 3h, ...[2]

$$(1)\quad \begin{bmatrix} X_{t+h} \\ Y_{t+h} \end{bmatrix} = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} + h^{\delta}\cdot \begin{bmatrix} \mu(X_t, Y_t, t) \\ \kappa(X_t, Y_t, t) \end{bmatrix} + h^{1/2}\cdot \begin{bmatrix} \xi_{X,t+h} \\ \xi_{Y,t+h} \end{bmatrix},$$

δ is set equal to either 3/4 or 1. X_t is an $n \times 1$ *observable* process. Y_t is an $m \times 1$ *unobservable* process. The $\{X_t, Y_t\}$ process is assumed markovian. $\xi_{X,t}$ and $\xi_{X,t}$ are respectively, $n \times 1$ and $m \times 1$

2) Notice that the scale factor $h^{1/2}$ on the ξ terms in (1) is missing in the univariate case presented in NF equations (5.1) and (5.1'). These scale terms are correct here, but were inadvertently omitted in NF. In addition, the "h^{-1}" terms in NF (5.8)-(5.9) should be deleted.

martingale difference sequences, with joint conditional density $f_h(\xi_{X,t+h},\xi_{Y,t+h}|X_t,Y_t,t)$, which we assume is continuously differentiable in (X_t,Y_t). Existence and differentiability of the conditional densities $f_h(\xi_{Y,t+h}|\xi_{X,t+h},X_t,Y_t,t)$, and $f_h(\xi_{X,t+h}|X_t,Y_t,t)$ is also assumed. Our interest is in using information in the sample path of the observed $\{X_t\}$ process to estimate the $\{Y_t\}$ process. The analysis is asymptotic, in that we approach continuous time, i.e., we take limits as $h \downarrow 0$. When $\delta = 1$ and some mild regularity conditions are satisfied, $\{X_t,Y_t\}$ converge weakly to a diffusion as $h \downarrow 0$, leading us to term $\{X_t,Y_t\}$ a *near-diffusion*.

The ARCH models used to estimate $\{Y_t\}$ take the form

(2) $\hat{Y}_{t+h} = \hat{Y}_t + h^{\delta}\cdot\hat{\kappa}(X_t,\hat{Y}_t,t) + h^{1/2}G(\hat{\xi}_{X,t+h},X_t,\hat{Y}_t,t,h)$, where

$$\hat{\xi}_{X,t+h} \equiv \frac{X_{t+h} - X_t - h^{\delta}\hat{\mu}(X_t,\hat{Y}_t,t)}{h^{1/2}}.$$

$\hat{\mu}$, $\hat{\kappa}$, and G are functions selected by the econometrician. $\hat{\mu}$ and $\hat{\kappa}$ are drift terms corresponding to μ and κ in (1). G is a noise term, a counterpart to the ξ_Y term in (1). Accordingly, we make the normalizing assumption that for all X_t, Y_t, t, and h, $E_t[G(\xi_{X,t+h},X_t,Y_t,t,h)] = 0_{m\times 1}$. In this note we will present results on asymptotically optimal filtering and smoothing. For reasons explained in NF and N, μ, $\hat{\mu}$, κ, and $\hat{\kappa}$ do not appear in the asymptotic distribution of the normalized measurement error $Q_t \equiv h^{-1/4}(\hat{Y}_t-Y_t)$ when $\delta = 1$, prompting N and NF to focus on the case $\delta = 3/4$, in which an asymptotic bias is introduced in $(\hat{Y}_t-Y_t)$ unless $\mu = \hat{\mu}$, and $\kappa = \hat{\kappa}$. Since in this paper we focus on asymptotically optimal filters and smoothers, which eliminate this bias by setting $\mu = \hat{\mu}$, $\kappa = \hat{\kappa}$, we will take $\delta=1$, allowing us to effectively ignore these terms.

As in the earlier papers on ARCH filtering, $\{Q_t\}$ oscillates very rapidly as $h \downarrow 0$, becoming heteroskedastic white noise in the limit. To obtain a *diffusion* limit for Q_t we must focus on increasingly short intervals of time—e.g., $[T, T+h^{1/2}M_h]$, where $M_h \to \infty$ slowly as $h \downarrow 0$. We then 'stretch' this time scale into an interval $[0, M_h]$ on a new 'fast' time scale. Over the shrinking time interval $[T, T+h^{1/2}M_h]$, X_t and Y_t move more and more slowly, becoming asymptotically constant at their time T values as $h \downarrow 0$. $\{Q_t\}$, on the other hand, converges to a diffusion limit. To accommodate the change in time scales, we define

(3) $Q_{T,\tau} \equiv Q_{T+\tau h^{1/2}}$, $X_{T,\tau} \equiv X_{T+\tau h^{1/2}}$, $Y_{T,\tau} \equiv Y_{T+\tau h^{1/2}}$.

T indexes the start point of the shrinking interval, while τ indexes time on the transformed 'fast' time scale. To transform from the fast to the standard time scale, we note that $\tau = h^{-1/2}(t-T)$. For further discussion of the passage to continuous time and the changing of the time scales, see NF and N.

The two time scales can make the notation cumbersome, so we make some simplifications: first, we define t as a function of T, τ, and h by

(4) $t \equiv t(T,\tau,h) \equiv T + h^{1/2}\tau$,

and suppress the arguments of t. We frequently drop time indices in the arguments of conditional expectations, so, for example, "$E_t[f]$" means "$E[f_{t+h} | X_t, Y_t, Q_t]$." We will also often write f_{t+h} instead of $f(\xi_{X,t+h}, X_t, Y_t, t, h)$ (so the ξ_X term is evaluated h time units after X and Y arguments). We often drop T subscripts and write, for example, ω_τ instead of $\omega_{T,\tau}$ (so if there is a single subscript τ, we are always indicating the 'fast' time scale, holding T fixed). Define the vector/matrix norm $\|A\| \equiv [\text{Trace}(AA')]^{1/2}$. For Λ positive semidefinite, let $\Lambda^{1/2}$ be the unique positive semidefinite matrix satisfying $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$, and let $\Lambda^{\underline{1/2}}$ be any real-valued matrix for which $\Lambda^{\underline{1/2}}\Lambda^{\underline{1/2}\prime} = \Lambda$. $\Lambda^{\underline{1/2}}$ need not be positive definite or symmetric and is not generally unique.

We make several assumptions. For discussion and motivation, see NF or N.

ASSUMPTION 1: *Define t as a function of T and τ by $t \equiv T + h^{1/2}\tau$. The following functions are well defined, continuous in t, and twice differentiable in X_t and Y_t:*

$$(5)\ B(X,Y,T,\tau) \equiv -\lim_{h\downarrow 0} E[\partial G(\xi_{X,t+h}, X_t, Y_t, t, h)/\partial Y \mid (X_t, Y_t) = (X,Y)],$$

$$(6)\ C(X,Y,T,\tau) \equiv \lim_{h\downarrow 0} E[(G(\xi_{X,t+h}, X_t, Y_t, t, h) - \xi_{Y,t+h})$$

$$(G(\xi_{X,t+h}, X_t, Y_t, t, h) - \xi_{Y,t+h})' \mid (X_t, Y_t) = (X,Y)].$$

Further,

$$(7)\ h^{-1/2}E[Q_{t+h} - Q_t \mid (X_t, Y_t, Q_t) = (X,Y,Q)] \rightarrow -B(X,Y,T,\tau)\cdot Q, \text{ and}$$

$$(8)\ h^{-1/2}Cov[Q_{t+h} - Q_t \mid (X_t, Y_t, Q_t) = (X,Y,Q)] \rightarrow C(X,Y,T,\tau)$$

as $h \downarrow 0$ uniformly on every bounded (X,Y,Q,T,τ) set.

We will simplify notation further by writing $B_{T,\tau}$ and $C_{T,\tau}$ for $B(X,Y,T,\tau)$ and $C(X,Y,T,\tau)$.

ASSUMPTION 2: *For some $\delta > 0$*

$$(9)\ E[\| h^{-1/2}(Y_{t+h} - Y_t)\|^{2+\delta} \mid (X_t, Y_t) = (X,Y)],$$

$$(10)\ E[\| h^{-1/2}(X_{t+h} - X_t)\|^{2+\delta} \mid (X_t, Y_t) = (X,Y)], \text{ and}$$

(11) $E[\| h^{-1/2}G(\hat{\xi}_{X,t+h}, X_t, \hat{Y}_t, t, h)\|^{2+\delta} \mid (X_t, Y_t, Q_t) = (X,Y,Q)]$ are bounded as $h \downarrow 0$, uniformly on every bounded (X,Y,Q,T,τ) set.

ASSUMPTION 3: *Given X_T and Y_T, $Q_{T,0} \Rightarrow N(0_{m\times 1}, V_{T,0})$ as $h \downarrow 0$, where $V_{T,0}$ may depend on X_T, Y_T, and T.*

Assumption 3 differs from the corresponding assumptions in N and NF, which conditioned at time T on the values of X_T, Y_T, and Q_T. Here, we condition on X_T and Y_T, but allow a random initial condition for Q_T, in which Q_T is asymptotically normal with a covariance matrix that may depend on X_T, Y_T, and T. Why do this? First, ARCH models of the form (2) require some "startup" estimate of Y_0 to initialize the filter. If $\hat{Y}_0$ is fixed arbitrarily for all h and Y_0 is also fixed for all h, Q_0 explodes to ∞ as $h \downarrow 0$ unless $\hat{Y}_0 = Y_0$. This, of course, makes conditioning of $Q_{T,0}$ in passing to the fast time scale problematic. Foster and Nelson (1994), however, show that it is often possible to obtain a $\hat{Y}_{T,0}$ such that $Q_{T,0} = h^{-1/4}(\hat{Y}_{T,0} - Y_{T,0})$ is asymptotically mean zero and normal using a rolling regression. $V_{T,0}$ in this case can be interpreted as the asymptotic error covariance matrix for $Q_{T,0}$ from this rolling regression. Alternatively, the ARCH filter may have been running for some time prior to date T, long enough to settle into the steady state error covariance matrices delivered in N and NF.

N (Theorem 2.1) shows that under Assumptions 1-3, conditional on (X_T, Y_T), $\{X_{T,\tau}, Y_{T,\tau}, Q_{T,\tau}\}$ converge weakly to the solution of the stochastic integral equation

$$(12) \quad \begin{aligned} X_{T,\tau} &= X_T, \quad Y_{T,\tau} = Y_T \\ Q_{T,\tau} &= Q_{T,0} - \int_0^\tau B_{T,\tau} Q_{T,s} ds + \int_0^\tau [C_{T,\tau}]^{1/2} dW_s, \end{aligned}$$

where $Q_{T,0} \sim N(0_{m\times 1}, V_{T,0})$. (12) is a vector version of an Ornstein-Uhlenbeck process, a continuous time Gaussian VAR(1) with τ-dependent coefficients.[3] (12) implies that for fixed τ,

$$(13) \quad [Q_{T,\tau} \mid (X_T, Y_T, Q_T)] \rightarrow N[0_{m\times 1}, V_{T,\tau}],$$

where $V_{T,\tau}$ solves the matrix differential equation

$$(14) \quad \frac{dV_{T,\tau}}{d\tau} = -B_{T,\tau} V_{T,\tau} - V_{T,\tau} B_{T,\tau} + C_{T,\tau}$$

with initial condition $V_{T,0}$ given by Assumption 3. (See Karatzas and Shreve (1988, Section 5.6)). This weak convergence is uniform on $0 \leq \tau \leq M$ for all finite M. This implies (see Helland (1982, Lemma 5.2) that it is also uniform on $0 \leq \tau \leq M_h$, where $M_h \rightarrow \infty$, provided that M_h increases sufficiently slowly with h. This makes it meaningful to consider the limiting

3) N looked at the case in which $Q_{T,0}$ was given, and in which $B_{T,\tau}$ and $C_{T,\tau}$ were independent of τ. The extension to our case is straightforward.

case $\tau \to \infty$.

Note that on the transformed (fast) time scale, $X_{T,\tau}$ and $Y_{T,\tau}$ are constant at their time T values. This implies that when evaluating continuous functions of (X_t,Y_t,t) for $T \le t \le T+h^{1/2}M_h$, the functions evaluated at (X_t,Y_t,t) are asymptotically equal to those evaluated at (X_T,Y_T,T).

NF and N defined optimality by minimizing (in a matrix sense) the steady state error covariance matrix[4] $\lim_{\tau\to\infty} V_{T,\tau}$, in particular minimizing $\lim_{\tau\to\infty} u'V_{T,\tau}u$ for arbitrary $m\times 1$ vector u or minimizing $\lim_{\tau\to\infty}$ Trace$[V_{T,\tau}]$. We now consider the more general task of minimizing $V_{T,\tau}$ for every $\tau > 0$:

ASSUMPTION 4: *For every h, the conditional densities $f(\xi_X,\xi_Y|X,Y,t,h)$ and $f(\xi_X|X,Y,t,h)$ are well defined and continuous in X, Y, t, and h and $f(\xi_X|X,Y,t,h)$ is continuously differentiable in Y almost everywhere, with one sided partial derivatives with respect to Y everywhere. Define the $m\times 1$ vectors*

(15) $P(\xi_X,X,Y,t) \equiv E[\xi_{Y,t+h}|(\xi_{X,t+h},X_t,Y_t)=(\xi_X,X,Y)]$, *and*

(16) $S(\xi_X,X,Y,t) \equiv \partial \ln[f(\xi_{X,t+h}|X,Y)]/\partial Y$.

For some $\delta > 0$

(17) $E[\|h^{-1/2}P(\xi_{X,t+h},X_t,Y_t,t)\|^{2+\delta}|X_t=X,Y_t=Y]$, *and*

(18) $E[\|h^{-1/2}S(\xi_{X,t+h},X_t,Y_t,t)\|^{2+\delta}|X_t=X,Y_t=Y]$

are bounded as $h\downarrow 0$, uniformly on every bounded (X,Y,T,τ) set.

ASSUMPTION 5: *There exists a unique solution to the matrix Riccati equation*

4) The existence of $\lim_{\tau\to\infty} V_{T,\tau}$ required N and NF to make additional assumptions that we do not need to make, in particular, the assumption that the real parts of the eigenvalues of $B_{T,\tau}$ are positive (see NF Theorem 3.1 and N Theorem 2.1). NF and N's definition of optimality also required that an asymptotic bias term disappear. Since we are looking at the standard drift case (i.e., $\delta=1$, not $\delta=3/4$) this asymptotic bias term does not appear.

(19) $d\omega_\tau = -E_T[PS']\omega_\tau - \omega_\tau E_T[SP'] - \omega_\tau E_T[SS']\omega_\tau + E_T[(\xi_Y - P)(\xi_Y - P)']$

with initial condition $\omega_{T,0} = V_{T,0}$ *such that* ω_τ *is positive semidefinite for all* $\tau \geq 0$.

Anderson and Moore (1971, Section 15.2) show that a sufficient condition for Assumption 5 is that $E_T[SS']$ is positive definite and that $\omega_{T,0} = V_{T,0}$ is positive semidefinite.

THEOREM 1: *Let Assumptions 1-5 be satisfied. Then for any* $R > 0$, *a set of sufficient conditions for* $Trace[V^*_{T,R}]$ *to be minimized is that for all* $0 \leq \tau \leq R$,

(20) $G(\xi_{X,t+h}, X_t, Y_t, t, \tau, h) = P(\xi_{X,t+h}, X_t, Y_t, t) + \omega_\tau S(\xi_{X,t+h}, X_t, Y_t, t)$.

If (20) defines $G(\cdot)$, *then* $V_{T,\tau} = \omega_\tau$.

Let $\tilde{G}(\xi_X, X, Y, T, \tau, h)$ *satisfy Assumptions 1-2, and let* $\tilde{V}_{T,\tau}$ *be the asymptotic covariance matrix delivered at* τ *by (13) using this* $G(\cdot)$ *function. Then* $\tilde{V}_{T,\tau} - \omega_\tau$ *is positive semidefinite and non null unless* $E_s[(\tilde{G}-G)(\tilde{G}-G)'] = 0_{m\times m}$ *for almost all s,* $T \leq s \leq T+h^{1/2}\tau$.

PROOF: *Our strategy is to guess an optimal* $G(\cdot)$ *and verify its optimality. To arrive at the guess (20), treat the dynamic minimization of* $Trace[V^*_{T,R}]$ *as an optimal control problem (see, e.g., Kamien and Schwartz (1981, Part II, Section 5)) naively proceeding as if, given X, Y, T, and h,* $G(\xi_X, X, Y, T, \tau, h)$ *were a separate control variable for each* ξ_X *(so there are an uncountably infinite number of control variables.) (20) is one of the first-order conditions.*

To verify the guess, let $G(\cdot)$ *be given by (20). As in the proof of (N, Theorem 2.2), one can show that* $B_{T,\tau} = E_s[PS' + \omega_\tau SS']$ *and* $C_{T,\tau} = E_s[(P + \omega_\tau S - \xi_Y)(P + \omega_\tau S - \xi_Y)']$. *Now let* $\tilde{G}_{t+h} \equiv G_{t+h} + H_{t+h}$, *where* $H_{t+h} = H(\xi_{X,t+h}, X_t, Y_t, t, \tau)$, *and* $\tilde{G}$ *satisfies Assumptions 1 and 2. Then, as in the proof of (N, Theorem 2.2),* $\tilde{B}_{T,\tau} = E_s[\tilde{G}S'] = E_s[PS'] + \omega_\tau E_s[SS'] + E_s[HS']$, *and* $\tilde{C}_{T,\tau} = E_s[(P + \omega_\tau S + H - \xi_Y)(P + \omega_\tau S + H - \xi_Y)']$. *Let* $\tilde{V}_{T,\tau}$ *be the asymptotic covariance matrix delivered by (13) using* $\tilde{G}(\cdot)$. *Subtracting the expressions for* dV_τ *and* $d\omega_\tau$ *and simplifying,*

(21) $d(\tilde{V}_\tau - \omega_\tau) = -\bar{B}_\tau(\tilde{V}_\tau - \omega_\tau) - (\tilde{V}_\tau - \omega_\tau)\bar{B}_\tau' + E_\tau[HH']$.

Comparing Karatzas and Shreve (1988, Chapter 5) equations (6.1) and (6.13)), we see that $(\tilde{V}_\tau - \omega_\tau)$ *is the time* τ *covariance matrix of the vector gaussian stochastic differential equation*

(22) $d\lambda_\tau = -\bar{B}_{T,\tau}\lambda_\tau dt + E_\tau[HH']^{1/2}_{\_}dW_\tau$,

where λ_0 *is fixed, and* W_τ *is an* $m\times 1$ *standard Brownian motion. Since* $(\tilde{V}_\tau - \omega_\tau)$ *is a covariance matrix, it is, perforce, positive semidefinite. That it is non-null unless* $E_\tau[HH']$ *is a matrix of zeros for (almost) all* τ, $0 \leq \tau \leq R$ *follows from Karatzas and Shreve (1988, Chapter 5, equations 6.3 and 6.11).* $\tilde{V}_\tau$ *therefore exceeds* ω_τ *for all* τ *by a positive semidefinite non-null matrix, completing the proof.*

The Breakdown of Moment Matching

The 'large τ' optimality criterion of NF and N led to the interesting property that the asymptotically optimal filter matched the first two conditional moments in the ARCH model considered as a data generating process to the first two conditional moments of the $\{X_t, Y_t\}$ process. (See NF, pp. 18-19 and N Theorem 2.5). This property does not necessarily hold for the filter of Theorem 1. Suppose, for example, that the filter is initialized at $\tau=0$ with a "bad" initial guess (i.e., $V_{T,0}$ is large). If the filter of (N, Theorem 2.2) is utilized, then $\omega_{T,\tau}$ eventually (i.e., as $\tau\to\infty$) settles down to the ω_T delivered by (N, Theorem 2.2). For finite τ, however, the filter of our Theorem 1 delivers a smaller error covariance matrix than the filter of N Theorem 2.2. Since $V_{T,0}$ is high, however, the filter of Theorem 1 puts more weight on the score term for small τ than does the filter of (N, Theorem 2.2), and therefore does not match moments. This provides another motivation for the random initial condition of Assumption 3: if the filter has not yet settled down into its large τ steady state, then the filter of N, Theorem 2.2 is inefficient.

Solving the Matrix Riccati Equation

Anderson and Moore (1971, Section 15.2) show that the matrix Riccati equation has a unique, positive semidefinite solution ω_τ for all $\tau \geq 0$, provided that $\omega_{T,0}$ is positive semidefinite, and that $E_T[SS']$ is positive definite. Anderson and Moore and Gelb et al (1974, Section 4.6) describe a procedure for solving the (nonlinear) Riccati equation by converting it into a system of linear equations. Following Gelb et al, define

$$\text{(23)} \quad \begin{aligned} \lambda_\tau &\equiv \omega_\tau Z_\tau, \\ dZ_\tau/d\tau &= E_T[SP']Z_\tau + E_T[SS']\lambda_\tau \end{aligned}$$

where $\lambda_0 = V_{T,0}$ and $Z_0 = I_{m\times m}$. We then have

$$\text{(24)} \quad \frac{d}{d\tau}\begin{bmatrix} Z_\tau \\ \lambda_\tau \end{bmatrix} = \begin{bmatrix} -E_T[SP'] & E_T[SS'] \\ E_T[(\xi_Y-P)(\xi_Y-P)'] & E_T[PS'] \end{bmatrix}\begin{bmatrix} Z_\tau \\ \lambda_\tau \end{bmatrix}, \text{ so}$$

$$\text{(25)} \quad \begin{bmatrix} Z_\tau \\ \lambda_\tau \end{bmatrix} = \exp\left[\begin{bmatrix} -\tau E_T[SP'] & \tau E_T[SS'] \\ \tau E_T[(\xi_Y-P)(\xi_Y-P)'] & \tau E_T[PS'] \end{bmatrix}\right]\begin{bmatrix} I_{m\times m} \\ V_{T,0} \end{bmatrix}, \text{ and } \omega_\tau = \lambda_\tau Z_\tau^{-1}.$$

Smoothing

Development of fixed-point ARCH smoothers is quite similar to their development in the Kalman filter (see, e.g., Anderson and Moore (1979)): for example, for time $t \geq T$ the state variables in Y_T are retained in the state vector along with Y_t. The ARCH filter now updates both $\hat{Y}_t$ and $\hat{Y}_t^T$, the latter the ARCH estimate at time t of Y_T, so $G(\cdot)$ and $Q_{T,\tau}$ are now $(2m)\times 1$ vectors, and $V_{T,\tau}$ is $2m\times 2m$.

By Theorem 1, the asymptotically optimal G(·) for $[\hat{Y}_T{}',\hat{Y}_t^T{}']'$ is

$$(26)\quad \begin{bmatrix} G_t \\ G_{t+h}^T \end{bmatrix} = \begin{bmatrix} P_{t+h} \\ 0_{m\times 1} \end{bmatrix} + \begin{bmatrix} \omega_{11,\tau} & \omega_{12,\tau} \\ \omega_{12,\tau}' & \omega_{22,\tau} \end{bmatrix} \begin{bmatrix} S_{t+h} \\ 0_{m\times 1} \end{bmatrix}$$

where S_{t+h} and P_{t+h} are the score and prediction components of Assumption 4. The matrix Riccati equation becomes

$$\frac{d\omega_{11,\tau}}{d\tau} = -E_T[PS']\,\omega_{11,\tau} - \omega_{11,\tau}E_T[SP'] - \omega_{11,\tau}E_T[SS']\,\omega_{11,\tau} + E_T[(\xi_Y-P)(\xi_Y-P)']$$

$$(27)\quad \frac{d\omega_{12,\tau}}{d\tau} = -(E_T[PS'] + \omega_{11,\tau}E_T[SS'])\,\omega_{12,\tau}$$

$$\frac{d\omega_{22,\tau}}{d\tau} = -\omega_{12,\tau}'E_T[SS']\,\omega_{12,\tau}$$

where $\omega_{11,0} = \omega_{12,0} = \omega_{22,0} = V_T$. The top line of (27) is, not surprisingly, the Riccati equation for the original system. The behavior of $\omega_{12,\tau}$ and $\omega_{22,\tau}$ are also of interest, since $G_{t+h}^T = \omega_{12,\tau}'S_{t+h}$ supplies the update for $\hat{Y}_{t+h}^T$, and $\omega_{22,\tau}$ is the asymptotic covariance matrix of $h^{-1/4}[\hat{Y}_t^T - Y_T]$.

To interpret the smoother in terms of the optimal filter of Theorem 1, recall first that since Y_T doesn't change after time t (though of course Y_t does) its corresponding prediction component is a vector of zeros. Similarly, Y_T doesn't enter the conditional density of X_t after time T, so the corresponding score component is also a vector of zeros. What makes smoothing possible is that the measurement errors in Y_T and Y_t are positively correlated (in fact they are identical at time T) so information regarding Y_t in the score term S_{t+h} can be used to update $\hat{Y}_t^T$ via the second term in (26). As Y_t moves steadily farther from Y_T, the value of the marginal updates diminishes, so $\omega_{12,\tau} \to 0_{m\times m}$ as $\tau \to \infty$.

Explicit results are possible when $E_T[PS'] = 0_{m\times m}$. Fortunately, this is a leading special case: If the conditional distribution of $[X_{t+h}', Y_{t+h}']$ is elliptically symmetric for all t (N Theorem 2.3), or if $\{X_t, Y_t\}$ is generated by discretely observed diffusion (N, Theorem 3.1), $E_t[PS'] = 0_{m\times m}$. Suppose that the optimal filter for Y_t has been running for a 'long' time before date T—i.e., it was initialized at a time $T-h^{1/2}M_h$ (where, as usual, M_h increases slowly to ∞ as $h\downarrow 0$). The filter has settled into its steady state, so $\omega_{T,0}$ is given by

(28) $\lim_{\tau\to\infty} \omega_\tau = E_T[SS']^{-1/2}[E_T[SS']^{1/2}E_T[(\xi_Y-P)(\xi_Y-P)']E_T[SS']^{1/2}]^{1/2}E_T[SS']^{-1/2}$

(See N, Theorem 2.3) Now initialize the smoother at time T. $\omega_{11,0} = \omega_{12,0} = \omega_{22,0} = \omega_{T,0}$. Further $\omega_{11,\tau} = \omega_{T,0}$ for all $\tau \geq 0$. This drastically simplifies (27):

THEOREM 2: *Let $E_T[PS'] = 0_{m\times m}$, let $E_T[SS']$ and $E_T[(\xi_Y-P)(\xi_Y-P)']$ be positive definite, and let the filter/smoother $[G_t', G_T^{t\prime}]'$ be given by (26) with $\omega_{11,0} = \omega_{12,0} = \omega_{22,0} = \omega_{T,0}$. Then $\omega_{11,\tau} = \omega_{T,0}$ for all $\tau\geq 0$, and*

(29) $\omega_{12,\tau} = \exp[-\omega_{T,0}E_T[SS']\tau]\omega_{T,0}$

(30) $\omega_{22,\tau} = \omega_{T,0} - \omega_{T,0}(\int_0^\tau \exp[-E_T[SS']\omega_{T,0}s]E_T[SS']\exp[-\omega_{T,0}E_T[SS']s]ds)\omega_{T,0}$, and

(31) $\lim_{\tau\to\infty} \omega_{22,\tau} = \omega_{T,0}/2,\ \lim_{\tau\to\infty} \omega_{12,\tau} = 0_{m\times m}$.

PROOF: *Under the assumptions of the theorem, (27) becomes*

$$
(27')\quad
\begin{aligned}
\omega_{11,\tau} &= \omega_{T,0}\\
\frac{d\omega_{12,\tau}}{d\tau} &= -(\omega_{T,0}E_T[SS'])\,\omega_{12,\tau}\\
\frac{d\omega_{22,\tau}}{d\tau} &= -\omega_{12,\tau}'\,E_T[SS']\,\omega_{12,\tau}
\end{aligned}
$$

The middle line of (27') leads immediately to (29). Since $\omega_{T,0}$ and $E_T[SS']$ are positive definite (recall that positive definiteness of $E_T[SS']$ and $E_T[(\xi_Y-P)(\xi_Y-P)']$ implies the positive definiteness of $\omega_{T,0}$), their product has positive eigenvalues (Taussky (1968, p. 177)). This in turn implies the second half of (31) (see Bellman (1970, Chapter 13, Theorem 1)). The last line in

(27') becomes

(32) $d\omega_{22,\tau}/d\tau = -\omega_{12,\tau}'E_T[SS']\omega_{12,\tau} = -\omega_{T,0}\exp[-E_T[SS']\omega_{T,0}\tau]E_T[SS']\exp[-\omega_{T,0}E_T[SS']\tau]\omega_{T,0}$,

which is equivalent to (30), since $\omega_{22,\tau} = \omega_{22,0} + \int_0^\tau d\omega_{22,s}$. *Since the eigenvalues of* $\omega_{T,0}E_T[SS']$ *are positive we may use Lancaster and Tismenetsky (1985, Chapter 12.3, Theorem 3) to recognize* $\int_0^\infty \exp[-E_T[SS']\omega s]E_T[SS']\exp[-\omega E_T[SS']s]ds$ *as the unique solution X to*

(33) $E_T[SS']\omega_{T,0}X + X\omega_{T,0}E_T[SS'] = E_T[SS']$.

It is clear by inspection that (33) has the solution $X = \omega^{-1}/2$. *Substituting this into (30) and setting* $\tau = \infty$ *yields the first half of (31).*

The first part of (31) has an interesting interpretation: under the conditions of Theorem 3, as $\tau\to\infty$, the covariance matrix of $h^{-1/4}[\hat{Y}^T_{T+h^{1/2}\tau}-Y_T]$ falls to 1/2 the covariance matrix of $h^{-1/4}[\hat{Y}_T-Y_T]$, indicating that the information content of data prior to time T is the same as the information content of data after time T. Note that when $\{X_t,Y_t\}$ are conditionally multivariate normal, the $G(\cdot)$ function of the optimal filter depends only on their conditional covariance matrix. By (N, Theorem 3.1) this is true also when the data are generated by discretely observed diffusion. As Haussmann and Pardoux (1986) show, the instantaneous covariance matrix of a diffusion and its time reversed version are the same, so that the accuracy of 'filtering forward' and 'filtering backward' should be the same.[5] Theorem 3 extends this intuition to the case of conditionally elliptically symmetric $\{X_t,Y_t\}$. This is definitely not true in for all near-diffusions: for example if GARCH generates the data σ_t^2 can be extracted without error from lagged X_t's, but not from led X_t's—see Foster and Nelson (1994).

---

5) The drift terms, however, are not the same when time is reversed. If we considered the $\delta=3/4$ case—we have not in this paper—the asymptotic bias terms in the filter would differ when time is reversed. The optimal filter, however, would eliminate this bias.

An Example

We next consider smoothing for a stochastic volatility model from the options pricing literature (Wiggins (1987), Hull and White (1987), and Scott (1987)), which has also received considerable attention in the econometrics literature (e.g., Melino and Turnbull (1990), Jacquier, Polson, and Rossi (1992), Ruiz (1994), Harvey and Shephard (1993), Kim and Shephard (1993)). In this model, S_t is a stock price and σ_t is its instantaneous returns volatility. We observe $\{S_t\}$ at discrete intervals of length h. The model sets

(34) $dS_t = \mu S_t dt + S_t \sigma_t dW_{1,t}$, and

(35) $d[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]dt + \psi \cdot dW_{2,t}$,

where $W_{1,t}$ and $W_{2,t}$ are standard Brownian motions independent of (S_0, σ_0^2) with

(36) $$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \left[\, dW_{1,t} \;\, dW_{2,t} \right] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} dt.$$

μ, ψ, β, and α are constants with $\psi > 0$. NF derived the asymptotically optimal filter for this model: Defining $x_t \equiv \ln(S_t)$ and $y_t \equiv \ln(\sigma_t^2)$, we first rewrite the model as

(34') $dx_t = (\mu - \exp(y_t)/2)h^{\delta-1}dt + \exp(y_t/2)dW_{1,t}$,

(35') $dy_t = -\beta[y_t - \alpha]h^{\delta-1}dt + \psi \cdot dW_{2,t}$.

We have added the $h^{\delta-1}$ terms to allow the possibility of 'fast' drift.[6] The steady state optimal filter was developed in NF Theorem 4.4. Since this is a diffusion model satisfying the conditions on N Theorem 3.1, the asymptotic results are the same as for the model

(34") $x_{t+h} = x_t + (\mu - \exp(y_t)/2) \cdot h^{\delta} + h^{1/2}\xi_{x,t+h}$

---

6) This illustrates the arbitrariness in the fast drift case $\delta = 3/4$: if we had introduced fast drift in (34) and then transformed, the drift in (34') would have been $[\mu h^{\delta-1} - \exp(y_t/2)]$ instead of $[\mu - \exp(y_t/2)]h^{\delta-1}$.

(35") $y_{t+h} = y_t - \beta[y_t - \alpha]\cdot h^{\delta} + h^{1/2}\xi_{y,t+h}$, where

(37) $$\begin{bmatrix}\xi_{x,t+h} \\ \xi_{y,t+h}\end{bmatrix} \Big| \, x_t, y_t \sim N\left[\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\exp(y_t) & \psi\rho\exp(y_t/2) \\ \psi\rho\exp(y_t/2) & \psi^2\end{bmatrix}\right]$$

Routine calculations now yield $P_{t+h} = \psi\rho\xi_{x,t+h}\cdot\exp(-y_t/2)$, $S_{t+h} = [\xi^2_{x,t+h}\cdot\exp(-y_t)-1)]/2$, $E_t[S^2] = 1/2$, $E_t[(\xi_y\text{-}P)^2] = \psi^2(1\text{-}\rho^2)$, and $E_t[PS] = 0$. If we initialize the filter at time T with asymptotic variance V_T, the optimal $G(\cdot)$ function is

(38) $G(\xi_{x,t+h},x_t,y_t,T,\tau) \equiv \psi\rho\xi_{x,t+h}\cdot\exp(-y_t/2) + \omega_\tau[\xi^2_{x,t+h}\cdot\exp(-y_t)-1)]/2$,

where $\tau = h^{-1/2}(t\text{-}T)$. The solution ω_τ of the Riccati equation can be shown to be

(39) $$\omega_\tau = 2M\frac{2M\sinh(M\tau)+\cosh(M\tau)V_T}{2M\cosh(M\tau)+\sinh(M\tau)V_T}$$

where $M = \psi[(1\text{-}\rho^2)/2]^{1/2}$. As $\tau\to\infty$, $\omega_\tau \to \psi[2(1\text{-}\rho^2)]^{1/2}$, the steady state solution derived in NF Theorem 4.4. (NF reported the asymptotic variance of $h^{-1/4}[\hat{\sigma}^2_t\text{-}\sigma^2_t]$, but it is easy to arrive at our ω_∞ by the delta method.) If we initialize the filter at time T with the steady state covariance $[2\psi^2(1\text{-}\rho^2)]^{1/2}$, then by Theorem 2,

$$\omega_{11,\tau} = [2(1-\rho^2)]^{1/2}\psi,$$

(40) $$\omega_{12,\tau} = [2(1-\rho^2)]^{1/2}\psi\exp(-[(1-\rho^2)/2]^{1/2}\psi\tau), \text{ and}$$

$$\omega_{22,\tau} = [(1-\rho^2)/2]^{1/2}\psi[1 + \exp[-[2(1-\rho^2)]^{1/2}\psi\tau]] \to \omega_{11,0}/2 = \psi[(1\text{-}\rho^2)/2]^{1/2} \text{ as } \tau\to\infty,$$

so after time T,

(41) $$\hat{y}_{t+h} = \hat{y}_t - \beta[\hat{y}_t-\alpha]\cdot h^{\delta} + h^{1/2}[\psi\rho\hat{\xi}_{x,t+h}\cdot\exp(-\hat{y}_t/2) + (2(1-\rho^2))^{1/2}\psi[\hat{\xi}^2_{x,t+h}\cdot\exp(-\hat{y}_t)-1)]/2]$$

(42) $$\hat{y}^T_{t+h} = \hat{y}^T_t + h^{1/2}\omega_{12,\tau}[\hat{\xi}^2_{x,t+h}\cdot\exp(-\hat{y}_t)-1)]/2] = \hat{y}^T_t + h^{1/2}[2(1-\rho^2)]^{1/2}\psi\exp(-[(1-\rho^2)/2]^{1/2}\psi h^{-1/2}(t\text{-}T))[\hat{\xi}^2_{x,t+h}\cdot\exp(-\hat{y}_t)-1)]/2]$$

REFERENCES

ANDERSON, B. D. O. and J. B. MOORE (1971): *Linear Optimal Control*. Englewood Cliffs, NJ: Prentice Hall.

ANDERSON, B. D. O. and J. B. MOORE (1979): *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall.

BELLMAN, R. (1970): *Introduction to Matrix Analysis*, second edition. New York: McGraw Hill.

FOSTER, D. P., and NELSON, D. B. (1994): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," Working Paper, The Wharton School.

GELB, A., J. F. KASPER, Jr., R. A. NASH Jr., C. F. PRICE, A. A. SUTHERLAND Jr. (1974): *Applied Optimal Estimation*, Cambridge, MA: M.I.T. Press.

HARVEY, A. C., and N. SHEPHARD (1993): "The Econometrics of Stochastic Volatility," forthcoming, *Review of Economic Studies*.

HAUSSMANN, U. G. and E. PARDOUX (1986): "Time Reversal of Diffusions," *Annals of Probability*, 14, 1188-1205.

HELLAND, I. S. (1982): "Central Limit Theorems for Martingales with Discrete or Continuous Time." *Scandinavian Journal of Statistics*, 9, 79-94.

HULL, J. and A. WHITE (1987): "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance*, 42, 281-300.

JACQUIER, E., N. G. POLSON, and P. E. ROSSI (1992): "Bayesian Analysis of Stochastic Volatility Models," forthcoming, *Journal of Business and Economic Statistics*.

KAMIEN, M., I., and N. L. SCHWARTZ (1981): *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*. New York: North Holland.

KARATZAS, I. and S. E. SHREVE (1988): *Brownian Motion and Stochastic Calculus*. New York: Springer Verlag.

KIM, S. and N. G. SHEPHARD (1993): "Stochastic Volatility: New Models and Optimal Likelihood Inference," working paper, Princeton University.

LANCASTER P., and M. TISMENETSKY (1985): *The Theory of Matrices*, second edition. San Diego, CA: Academic Press.

MELINO, A. and S. M. TURNBULL (1990): Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239-265.

NELSON, D. B. (1992): "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 25, 61-90.

NELSON, D. B. (1994): "Asymptotic Filtering Theory for Multivariate ARCH Models," working paper, University of Chicago GSB.

NELSON, D. B., and D. P. FOSTER (1994): "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica* 62, 1-41.

RUIZ, E. (1994): "Quasi-maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289-306.

SCOTT, L. O. (1987): "Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419-438.

TAUSSKY, O. (1968): "Positive-Definite Matrices and Their Role in the Study of the Characteristic Roots of General Matrices," *Advances in Mathematics*, 2, 175-186.

WIGGINS, J. B. (1987): "Option Values under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics*, 19, 351-372.