FUNCTIONAL DATA ANALYSIS
FOR
VOLATILITY PROCESS

Rituparna Sen

Monday, July 31

10:45am-12:30pm Classroom 228
St-C5 Financial Models

Joint work with Hans-Georg Müller and Ulrich Stadtmüller
1. INTRODUCTION

What are functional data?

Univariate Data ⇒ Multivariate Data (MDA) ⇒ Infinite-dimensional Data (FDA)

Per subject or experimental unit, one samples one or several functions $X(t), t \in T$ (interval)

Order, neighborhood and smoothness are important concepts in FDA in contrast to Multivariate Data Analysis

Modeling data as independent realizations of a stochastic process with smooth trajectories
2. DECOMPOSING NOISY FUNCTIONAL DATA

into a smooth random process $S$ and additive noise:

$$X_{ij} = S_i(t_{ij}) + R_{ij}, \quad i = 1, \ldots n, j = 1, \ldots m.$$  

$i$ denotes the subject or experimental unit and $j$ denotes the time.

The $R_{ij}$ are additive noise:

$R_{ij}, R_{i'k}$ are independent for all $i \neq i'$,

$E(R_{ij}) = 0, \text{var}(R_{ij}) = \sigma_R^2 < 1.$

Note that the noise $R_{ij}$ within the same subject or item $i$ may be correlated.
3. MODELING NOISE COMPONENTS

\[
\log(R_{ij}^2) = V(t_{ij}) + W_{ij}
\]

where \( V \) is the \textbf{functional variance process} which is smooth, i.e., it has a smooth mean function \( \mu_V \), and a smooth covariance structure

\[
G_V(s, t) = \text{cov}(V(s), V(t)), \quad s, t \in \mathcal{T}.
\]

The \( W_{ij} \) are \textbf{'white noise'}:

\[
\begin{align*}
E(W_{ij}) &= 0, \quad \text{var}(W_{ij}) = \sigma_W^2, \\
W_{ij} \perp W_{ik} & \quad \text{for} \quad j \neq k, \quad W \perp V, W \perp S.
\end{align*}
\]
4. MODELING FINANCIAL RETURNS

Black-Scholes model for stock price

\[ d \log X(t, \omega) = \mu dt + \sigma dW(t, \omega), \quad t \geq 0 \]

Here \( W \) is a standard Wiener process, \( \sigma > 0 \) is called the volatility, \( \mu \) is the drift.

Generalizations (Barndorff-Nielsen & Shephard 2002)

\[ d \log X(t, \omega) = \mu(t, \omega) dt + \sigma(t, \omega) dW(t, \omega) \]

where \( \mu(t, \omega), \sigma(t, \omega), W(t, \omega) \) are independent stochastic processes, \( \mu(\cdot), \sigma(\cdot) \) with smooth paths.

Lemma 1 Assume that \( (\mu(t))_{t \in [0, T]} \) and \( (\sigma(t))_{t \in [0, T]} \) denote stochastic processes with paths being Lipschitz continuous of order 1. Then we have that

i) \[ \frac{1}{\sqrt{\Delta}} \int_{t}^{t+\Delta} \mu(v) dv = \mu(t) \sqrt{\Delta} + O_{a.s.}(\Delta^{3/2}) \text{ uniformly in } t \]

ii) \[ \frac{1}{\sqrt{\Delta}} \int_{t}^{t+\Delta} \sigma(v) dW(v) = \sigma(t) \frac{W(t+\Delta) - W(t)}{\sqrt{\Delta}} + O_{p}(\Delta^{1/2}) \text{ uniformly in } t \]
5. VOLATILITY PROCESS

For
\[ X_\Delta(t) = \frac{(\log X(t + \Delta) - \log(X(t)))}{\sqrt{\Delta}} \]
and
\[ W_\Delta(t) = \frac{(W(t + \Delta) - (W(t)))}{\sqrt{\Delta}}, \]

obtain in the limit as \( \Delta \to 0 \)
\[ X_\Delta(t) = \sigma(t)W_\Delta(t) \]

Let \( \eta = \text{E} \log(N(0, 1)^2) \).

With \( R(t) = X_\Delta(t) \), obtain
\[
\log(R^2(t)) = \log \sigma^2(t) + \log W^2_\Delta(t) \\
= (\log \sigma^2(t) + \eta) + (\log W^2_\Delta(t) - \eta) \\
= \tilde{V}(t) + \tilde{W}(t)
\]

Let \( V(t) = \tilde{V}(t) - \eta \) providing justification for defining \( V \) as the volatility process.
The decomposition

\[ Z_{ij} = \log(R_{ij}^2) = V(t_{ij}) + W_{ij} \]

implies

\[ \mathbb{E}(Z_{ij}) = \mathbb{E}(V(t_{ij})) = \mu_V(t_{ij}) \]

\[ \text{cov}(Z_{ij}, Z_{ik}) = \text{cov}(V_i(t_{ij}), V_i(t_{ik})) = G_V(t_{ij}, t_{ik}), \quad j \neq k \]

for the functional variance process \( V \).
Auto-covariance operator associated with the symmetric kernel $G_V$:

$$G_V(f)(s) = \int_T G_V(s, t)f(t)dt,$$

has smooth eigenfunctions $\psi_k$ with nonnegative eigenvalues $\rho_k \Rightarrow$

Representations

$$G_V(s, t) = \sum_k \rho_k \psi_k(s)\psi_k(t), \quad s, t \in \mathcal{T}$$

$$V(t) = \mu_V(t) + \sum_k \zeta_k \psi_k(t),$$

with functional principal component scores $\zeta_k, k \geq 1$ with

$E\zeta_k = 0,$

$\text{var}(\zeta_k) = \rho_k,$

$\zeta_k = \int_T (V(t) - \mu_V(t))\psi_k(t)dt,$

$\zeta_k$ uncorrelated,

$\sum \rho_k < \infty$
7. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS (FPCA)

Estimation of mean function $\mu_X$ by cross-sectional averaging, of eigenfunctions and eigenvalues from carefully smoothed covariance surface

$$\Gamma(s, t) = \text{cov}(V(s), V(t))$$

(Rice & Silverman 1991 JRSS-B, Yao et al. 2005 JASA)

How many components $M$ to include?

Cross-validation ok, pseudo-AIC, BIC work better.

Asymptotically, $M = M(n) \to \infty$
8. ESTIMATION OF MODEL COMPONENTS

First step: Smoothing of observed trajectories in the model

\[ X_{ij} = S_i(t_{ij}) + R_{ij} \]

Obtain smoothed trajectories \( \hat{S}_i(t_{ij}) \) and estimated errors and transformed residuals

\[ \hat{R}_{ij} = X_{ij} - \hat{S}_i(t_{ij}) \]

\[ \hat{Z}_{ij} = \log(\hat{R}_{ij}^2) = \log(X_{ij} - \hat{S}_i(t_{ij}))^2 \]

This is finite sample correction and needs to be done because \( \Delta \) is finite.
**Second step:** Apply functional principal component analysis (Principal Analysis of Random Trajectories - PART algorithm) to the sample of transformed residuals $\hat{Z}_{ij}$:

- Estimate mean function $\mu_V$ (smoothing of cross-sectional averages)
- Estimate smooth covariance surface by smoothing of empirical covariances (omitting the diagonal)
- Obtain eigenvalues/eigenfunctions; choosing number of components $M$ by cross-validation (or AIC, BIC)
- From diagonal of covariance surface, obtain $\text{var}(W_{ij}) = \sigma_W^2$
- Obtain individual FPC scores $\zeta_{ij}$ by integration.
9. ASYMPTOTIC RESULTS

**Basic Lemma:** Under regularity conditions, with smoothing bandwidths $b_S$, $m$ no. of measurements per trajectory,

$$
\mathbb{E}\left( \sup_{t \in T} | \hat{S}_i(t) - S_i(t) | \right) = O\left( b_S^2 + \frac{1}{\sqrt{mb_S}} \right).
$$

**Results for mean and covariance structure:**

$$
\sup_{t \in T} | \hat{\mu}_V(t) - \mu_V(t) | = O_p(\alpha_n)
$$

$$
\sup_{s,t \in T} | \hat{G}_V(s,t) - G_V(s,t) | = O_p(\beta_n)
$$

$$
| \hat{\sigma}_W^2 - \sigma_W^2 | = O_p(\gamma_n)
$$

for sequences $\alpha_n, \beta_n, \gamma_n \to 0$. 
Results for eigenvalues/eigenfunctions of functional variance process:

\[
\sup_{t \in T} | \hat{\psi}_k(t) - \psi_k(t) | \overset{p}{\to} 0
\]

\[
\hat{\rho}_k \overset{p}{\to} \rho_k.
\]

Results for individual trajectories of functional variance process:
As \( M = M(n) \to \infty \),

\[
\sup_{1 \leq k \leq M} | \hat{\zeta}_{ik} - \zeta_{ik} | \overset{p}{\to} 0
\]

\[
\sup_{t \in T} | \hat{V}_i(t) - V_i(t) | \overset{p}{\to} 0.
\]
10. REAL DATA APPLICATIONS

INTRA-DAY TRADING PATTERNS:

Five-minute closing prices of Affymetrix stock and SOXX (PHLX Semiconductor Sector Index) observed for 40 days. Data are log-returns, log($X(t)/X(t - 1)$)
Patterns of volatility?
Apply two-step procedure to obtain trajectories of volatility process.
Figure 1: Raw curves for 20 days of Affymetrix data
Figure 2: SOXX log returns with smooth trends
Figure 3: Estimated mean(left) and covariance surface(right) for the variance process of Affymetrix data
Figure 4: Estimated mean(left) and covariance surface(right) for the variance process of SOXX data
Figure 5: Estimated eigenfunctions of the variance process of Affymetrix data
Figure 6: Estimated eigenfunctions of the variance process of SOXX data
Figure 7: Estimated transformed residuals for the days with volatility time courses most aligned with the 3 eigenfunctions for SOXX data
11. Regression of second half day on first half day for SOXX data
Figure 8: Left top: Estimated and actual transformed residuals for a typical second half day: The blue line is prediction using only the mean, green line is prediction using the first half day and red line is the estimate using the second half day. Right Top: Beta function . Bottom: Scatter plot of estimated principal component scores of second half day vs first half day for first(left) and second(right) principal
12. Jump detection
Figure 9: Density estimate of $z_{TP}$ (red curve) superimposed with standard normal density (blue curve) for Japanese Yen and US Dollar exchange rate: The left panel uses Huang-Tauchen method. The right panel uses FDA. The figures on top are for 5 min data and those on bottom are for 1 min data.
Figure 10: Density estimate of $z_{TP}$ (red curve) superimposed with standard normal density (blue curve) for Euro and US Dollar exchange rate: The left panel uses Huang-Tauchen method. The right panel uses FDA. The figures on top are for 5 min data and those on bottom are for 1 min data.