Fast Computational Approaches for Predictive Inference for Time Correlated Data Streams

Nalini Ravishanker
Dept. of Statistics, University of Connecticut, Storrs
e-mail: nalini.ravishanker@uconn.edu
URL: http:// www.stat.uconn.edu/ nalini

SAMSI Computational Advertising Workshop Aug 2012

Outline

I. Dynamic Models for Discrete-valued Time Series

II. Fast Likehood Evaluation using the Multivariate Preconditioned Conjugate Gradient (MPCG) Algorithm

III. Duration Models via Martingale Estimating Equations

Relevance of these predictive modeling approaches to problems in Computational Advertising —

I. Dynamic Models for Discrete-valued Time Series

In several applications, there is a need for accurate modeling of univariate and multivariate discrete-valued time series

- on several subjects,
- as functions of relevant covariates (subject-specific and time-varying),
- incorporating dependence over time, and
- incorporating dependence between the components of the response vector in MV case.

Three model frameworks:

- DGLM: Dynamic Generalized Linear Models
- HDGLM: Hierarchical Dynamic Generalized Linear Models
- HDNLM: Hierarchical Dynamic Non-Linear Models

Applications to Time Series of Counts

- Transportation engineering (crash counts on CT highways)
- Ecology (snails populations in Puerto Rico)
- Marketing (monthly prescriptions by US physicians)
- Collaborators on projects: Shan Hu, John Ivan, Michael Willig (UConn); Raj Venkatesan (U. Virginia)

DGLM

Observation Equation:

$$p(y_t|\tilde{\theta}_t) \propto \exp\left\{\frac{y_t\tilde{\theta}_t - b(\tilde{\theta}_t)}{\phi_t}\right\} \tag{1}$$

$$g(\lambda_t) = \eta_t = \mathbf{F}_t'\boldsymbol{\beta}_t$$

System or State Equation:

$$\boldsymbol{\beta}_t = \mathbf{G}_t\boldsymbol{\beta}_{t-1} + \mathbf{w}_t \tag{2}$$

- y_t: univariate count time series, $t = 1, \ldots, T$; $\lambda_t = E(y_t|\tilde{\theta}_t)$;
- $\mathbf{F}_t$: known vector of explanatory variables at time t ;
- $\boldsymbol{\beta}_t$: p-dim state parameter vector;
- $\mathbf{G}_t$: $p \times p$ state transition matrix; $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W})$

Examples of Latent Gaussian Models: Poisson or Negative Binomial DGLM, with log link function and Gaussian latent state vector

DGLM - Example of Highway Safety

- Observation Equation:

$$y_t|\lambda_t \sim NegBin(\lambda_t, \delta)$$
$$\log(\lambda_t) = \log(\alpha_t) + \beta_{1t}\log(V_t)$$

where $\beta_{0t} = \log(\alpha_t)$ and β_{1t} are two state parameters and δ is the dispersion parameter.

- System/State Equation:

$$\beta_{0t} = \beta_{0,t-1} + w_{0t}$$
$$\beta_{1t} = \beta_{1,t-1} + w_{1t}$$

where $w_{0t} \sim N(0, W_0)$ and $w_{1t} \sim N(0, W_1)$.

DGLM - MCMC

Priors: $\boldsymbol{\beta}_1 \sim N(\mathbf{m}, \mathbf{R})$; $\mathbf{W} \sim IW(v, \mathbf{S})$; $\delta \sim Gamma(a, b)$.

Joint Posterior: $\pi(\boldsymbol{\beta}, \delta, \mathbf{W}) \propto$

$$
\begin{aligned}
& \prod_{t=1}^{T} \frac{\Gamma(y_t + \delta)}{\Gamma(\delta)} \exp\left\{ y_t \log\left(\frac{\lambda_t}{\delta + \lambda_t}\right) + \delta \log\left(\frac{\delta}{\delta + \lambda_t}\right)\right\} \\
\times \quad & \prod_{t=2}^{T} |\mathbf{W}|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})'\mathbf{W}^{-1}(\boldsymbol{\beta}_t - \mathbf{G}\boldsymbol{\beta}_{t-1})\right\} \\
\times \quad & \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta}_1 - \mathbf{m})'\mathbf{R}^{-1}(\boldsymbol{\beta}_1 - \mathbf{m})\right\} \\
\times \quad & \delta^{a-1} \exp\{-\delta / b\} \\
\times \quad & |\mathbf{W}|^{-v/2} \exp\left\{ -\frac{1}{2} tr(\mathbf{W}^{-1}\mathbf{S})\right\}
\end{aligned}
$$

DGLM - MCMC

- Gamerman, 1998 Bmka
- Full conditional distribution of $\mathbf{W}$ has closed form. Sample $\mathbf{W}^{(new)}$ directly from an Inverse Wishart draw.
- Full conditional distribution of $\boldsymbol{\beta}_t$: Metropolis-Hastings Algorithm with modification
  Full conditional distribution of dispersion param: Metropolis-Hastings Algorithm
- MCMC computations can be time consuming!

DGLM - INLA

- Fast Approximate Bayesian Inference via Integrated Nested Laplace approximations (INLA): Rue *et al.*, 2009 JRSSB - in latent Gaussian models.
- Canned templates for use with a large class of statistical models available in R-INLA, see http://www.r-inla.org/

DGLM - INLA

Assumptions:

- Response variable belongs to an exponential family of distributions; includes Normal, Poisson, Negative Binomial, Binomial, Multinomial
- Mean response is linked to structured additive predictors
- Hyperparameters enable us to handle the variance terms of the latent Gaussian system and additional parameters from the distribution of the response variable (e.g., the dispersion parameter in the NegBin distribution).

DGLM - INLA

- Given response $\mathbf{y}$, latent Gaussian variables $\mathbf{x}$, and hyperparameters $\boldsymbol{\theta}$.
- Likelihood: $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$
- Joint Posterior $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$
- Here, $\pi(\mathbf{x}|\boldsymbol{\theta})$ is Gaussian by model assumption

DGLM - INLA

- Posterior marginals of interest

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}$$

- The INLA approach constructs nested approximations:

$$\tilde{\pi}(x_i|\mathbf{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}$$

DGLM - INLA

The INLA approach consists of three main steps:

- Approximate $\pi(\theta_j|\mathbf{y})$
- Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ using either a Gaussian approximation, simplified Laplace approximation, or Laplace approximation.
- Use numeric integration

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_j \tilde{\pi}(x_i|\theta_j, \mathbf{y})\tilde{\pi}(\theta_j|\mathbf{y})\Delta_j$$

DGLM - INLA

Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ by $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}$, evaluated at $\mathbf{x} = \mathbf{x}(\boldsymbol{\theta}^*)$

- Locate mode of $\pi(\boldsymbol{\theta}|\mathbf{y})$, call it $\boldsymbol{\theta}^*$
- Compute the Hessian $\mathbf{H}$ at θ^*; let $\boldsymbol{\Sigma} = \mathbf{H}^{-1} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$; reparametrize to correct for scale and rotation by $\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$
- explore the density concentration for $\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ using the transformation
- Approximate the posterior marginals $\tilde{\pi}(\theta_j|\mathbf{y})$ from $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ by numerical integration through interpolants selected in above steps.

DGLM - INLA

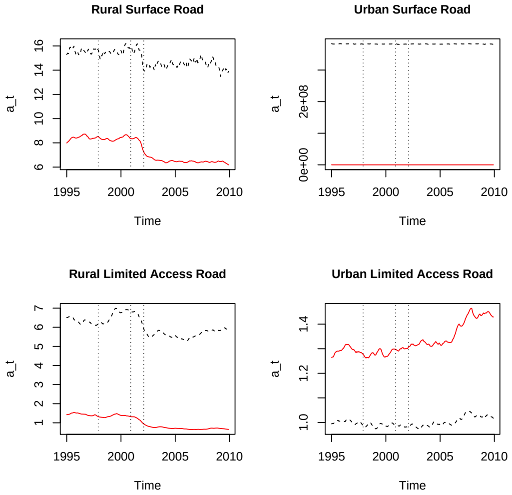
Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

- Gaussian approximation: $\tilde{\pi}_G(x_i|\boldsymbol{\theta}, \mathbf{y})$, need to determine mean $\mu_i(\boldsymbol{\theta})$ and variance $\sigma_i^2(\boldsymbol{\theta})$
- Laplace approximation: $\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_{GG}(x_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})}$ evaluated at $x_{-i} = x_{-i}(x_i, \boldsymbol{\theta}^*)$
- Simplified Laplace approximation by a series expansion of $\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})$ around $x_i = \mu_i(\boldsymbol{\theta})$

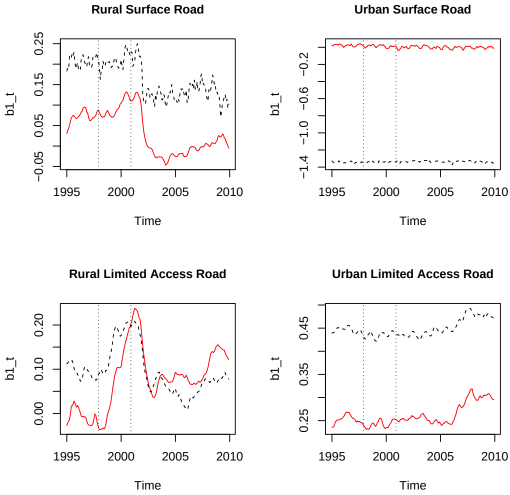
DGLM - INLA - Example of Highway Safety

- We used R-INLA package, with log-gamma priors for hyperparameters $\theta_1 = \log(W_0^{-1})$, $\theta_2 = \log(W_1^{-1})$, and $\theta_3 = \log(\delta)$
- We ran Bayesian analysis via INLA for each combination of demographic group, area type and access type, and different counts responses defined by crash severity level (K, KA, and KAB).
- Converged results from INLA were similar to converged MCMC results - but much faster!
- Question: Time to run a million DGLMs - optimized and in parallel?

Dynamic Intercept



Dynamic Slope for log VMT



HDGLM

Observation equation

$$
\begin{aligned}
p(Y_{it} = y_{it} | \lambda_{it}) &= e^{-\lambda_{it}} \frac{\lambda_{it}^{y_{it}}}{y_{it}!} \\
\log(\lambda_{it}) &= \mathbf{F}'_{1_{it}} \boldsymbol{\beta}_{1_{it}}
\end{aligned} \tag{3}
$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where

- $\mathbf{F}_{1_{it}}$: p-dim vector of explanatory variables for subject i at time t, and
- $\boldsymbol{\beta}_{1_{it}}$: p-dim subject-specific and time-varying param vector.

HDGLM

- Let $\mathbf{Y}_t = (Y_{1t}, \cdots, Y_{nt})'$
- $\boldsymbol{\lambda}_t = (\lambda_{1t}, \cdots, \lambda_{nt})'$
- $\mathbf{F}_{1_t}$: an $N \times Np$ obs. predictor matrix (we assume there are p predictors whose values we observe across all N customers at each time t)
- $\boldsymbol{\beta}_{1_t} = (\boldsymbol{\beta}'_{1_{1t}}, \cdots, \boldsymbol{\beta}'_{1_{Nt}})'$: Np-dim vector of coefficients for all N subjects on p predictors, where, for $i = 1, \cdots, N$, $\boldsymbol{\beta}_{1_{it}} = (\beta_{1_{0it}}, \beta_{1_{1it}}, \cdots, \beta_{1_{pit}})'$ is a p-dim vector of coefficients at time t for the ith subject.

HDGLM

In vector notation, at time t:

$$\log(\boldsymbol{\lambda}_t) = F_{1_t}\boldsymbol{\beta}_{1_t} \qquad (4)$$

HDGLM

Structural equation

- Write the p-dim subject-specific parameter vector $\boldsymbol{\beta}_{1_{it}}$ as a function of a pooled p-dim state parameter $\boldsymbol{\beta}_{2_t}$ and an additive normal error:

$$\boldsymbol{\beta}_{1_{it}} = \boldsymbol{\beta}_{2_t} + \mathbf{v}_{it} \tag{5}$$

- $\mathbf{v}_{it}$ is a p-dim error vector assumed to be $N(\mathbf{0}, \mathbf{V}_{it})$, and $\mathbf{V}_{it}$ is a $p \times p$ matrix.

HDGLM

In vector notation, at time t

$$\boldsymbol{\beta}_{1_t} = \mathbf{F}_{2_t}\boldsymbol{\beta}_{2_t} + \mathbf{v}_t \tag{6}$$

- where $\mathbf{F}_{2_t}$ is an $np \times p$ matrix which maps elements of the np-dim vector $\boldsymbol{\beta}_{1_t}$ to elements of the p-dim vector $\boldsymbol{\beta}_{2_t}$, and
- $\mathbf{v}_t$ is an Np- dim vector.

HDGLM

From (5), we see that the matrix $\mathbf{F}_{2_t}$ can be thought of as an n-dim array of $p \times p$ matrices each of which is $\mathbf{I}_p$, so that we could write for $i = 1, \cdots, N$,

$$\boldsymbol{\beta}_{1_{it}} = \mathbf{I}_p \boldsymbol{\beta}_{2_t}$$

It is possible to include covariates on the right side of (5).

HDGLM

System equation:

$$\boldsymbol{\beta}_{2_t} = \mathbf{G}_t \boldsymbol{\beta}_{2_{t-1}} + \mathbf{w}_t \tag{7}$$

- where $\mathbf{G}_t$ is a $p \times p$ state transition matrix (often assumed to be the identity matrix)
- the p-dim state error vector $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t)$, where $\mathbf{W}_t$ is a $p \times p$ matrix.
- Also, we assume that the initial state vector $\boldsymbol{\beta}_0 \sim N(\mathbf{a}, \mathbf{R})$, and we assume that $\mathbf{v}_t$ and $\mathbf{w}_t$ are independent.
- We may assume for simplicity that $\mathbf{v}_{it} \sim N(\mathbf{0}, \mathbf{V}_i)$ and that $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W})$, i.e., they are not time-varying.

Priors:

- MVN priors for initial state vector β_{2_0}, and
- Inverse Wishart priors for variance terms $\mathbf{V}_i$ and $\mathbf{W}$

- Joint Posterior Density is proportional to product of Likelihood and Prior
- Sampling from the conditional distribution of β_{2_t}: FFBS algorithm proposed by Carter and Kohn 1994.
- INLA: Work in progress consists of writing source code for implementing INLA for application in modified models.

HDNLM

- Let $\mathbf{y}_t = (y_{1it}, y_{2it}, \ldots, y_{mit})'$: vector time series of counts, $t = 1, \ldots, T$ modeled by a finite mixture of two-way covariance structured MV Poissons, which allows for overdispersion in marginal distributions and negative associations; Karlis & Meligkotsidou 2005; 2007
- $p(\mathbf{Y}|\mathbf{\Phi}) = \sum_{h=1}^{H} \pi_h MP_m(\mathbf{Y}|\boldsymbol{\lambda}_h)$

HDNLM

- We have a fast MV Poisson pmf evaluation scheme
- We have fast mixture MVP computation using Diebolt & Robert 1994
- We have implemented MCMC - slow!
- Currently, we are investigating faster approaches - open problem...

II. Fast Gaussian Likehood Evaluation via MPCG Algorithm

MPCG: Multivariate Preconditioned Conjugate Gradient.

- Vector Linear Long Memory VARFIMA Models
- Periodic Autoregressions with long memory
- MV Stochastic Volatility (SV) Models with long memory
- Collaborators on projects: Jeffrey Pai (U. Manitoa), Scott Holan (U. Missouri), Robert Lund (Clemson U)

Form of the Gaussian Likelihood

- $\mathbf{X}_t = (X_{t,1}, \cdots, X_{t,k})'$ for $t = 1, \cdots, n$ is a k-variate time series.
- Let $\mathbf{X} = (\mathbf{X}_1', \cdots, \mathbf{X}_n')'$
- Assume model errors ε_t are i.i.d. $N_k(\mathbf{0}, \mathbf{\Sigma})$;

Form of the Gaussian Likelihood

$$f(\mathbf{X};\boldsymbol{\Psi}) = (2\pi)^{-kn/2}|\boldsymbol{\Omega}|^{-1/2} \times \exp\left[\frac{-(\mathbf{X}-\mu)'\boldsymbol{\Omega}^{-1}(\mathbf{X}-\mu)}{2}\right]$$

- $\boldsymbol{\Omega}$: Cov($\mathbf{X}$) with dim $kn \times kn$
- it is a function of $\boldsymbol{\Phi}, \boldsymbol{\Theta}, \mathbf{d}, \boldsymbol{\Sigma}$

- $k = 1$, spectral density function $f(\omega) = |1 - \exp(-i\omega)|^{-2d} f^*(\omega)$, $\omega \in [-\pi, \pi]$. where $d \in (-0.5, 0.5)$.
- $\mathbf{\Omega}$ is an $n \times n$ Toeplitz matrix, and for $j, k = 1, \cdots, n$

$$\Omega_{j,k} = \int_{-\pi}^{\pi} \exp(i(j-k)\omega f(\omega) d\omega$$

  $\mathbf{\Omega}$ is ill-conditioned, with condition number $= O(n^{2|d|})$.
- Inverting $\mathbf{\Omega}$ is time consuming for large n;
- This makes exact Likelihood based inference very slow, especially for long time series

- Given data $\mathbf{X}_n$ and Fourier frequencies $\omega_j = 2\pi j/n$;

$$\text{Forward DFT: } D_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \exp(-i\omega_j t)$$

$i = \sqrt{-1}$

- Backward DFT: replace $\exp(-i\omega_j t)$ by $\exp(i\omega_j t)$

- Fast Inference using Multivariate Preconditioned Conjugate Gradient (MPCG) Algorithm: Pai & Ravishanker 2009, SPL
- Basic idea: Use of Fast Fourier Transform (FFT) yields fast solution of Toeplitz block systems.
- Conjugate Gradient: iteratively gives numerical solution for some systems (with symmetric p.d. matrix, say) of linear equations, $\mathbf{Ax} = \mathbf{b}$.
- Preconditioning: A preconditioner is a matrix $\mathbf{C}$ such that $\mathbf{C}^{-1}\mathbf{A}$ has a smaller condition number than $\mathbf{A}$; so solving $\mathbf{C}^{-1}\mathbf{Ax} = \mathbf{C}^{-1}\mathbf{b}$ is faster than solving $\mathbf{Ax} = \mathbf{b}$.

- A Toeplitz-block matrix; $\mathbf{T}_{ij}$ is Toeplitz

$$\mathbf{TB} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1k} \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2k} \\ & \vdots & & \\ \mathbf{T}_{k1} & \mathbf{T}_{k2} & \cdots & \mathbf{T}_{kk} \end{pmatrix}$$

Using MPCG algorithm with the Chan 1988 preconditioner, total number of operations for one evaluation of the likelihood has order

$$nk^2 \left[1 + \frac{k}{3} + \log(n)\right] + nk\,[5 + 3k + 4k\log(2) + (2 + 4k)\log(n)] \times k \log^{3/2}(n)$$

— much lower than $O(k^3 n^2)$ from Levinson algorithm

Basic Idea: Steps to Multiply **C** and a vector **y**

- Forward FFT of **y**
- Forward FFT of Col 1 of **C**
- Pointwise multiplication of both FFTs
- Backward FFT of product yields **Cy**

- Circulant block matrix, Chan and Olkin 1994
- Let $\gamma_{a,ii}(h)$ for lags $h = 0, \cdots, n-1$ be the autocovariances for the ith series, $i = 1, \cdots, k$.
- Let $\gamma_{a,ij}(h)$ for lags $h = -(n-1), \cdots, -1, 0, 1, \cdots, n-1$ be the cross-covariances between the ith and jth series, $i \neq j = 1, \cdots, k$.
- The first column of $circ_F(.)$ is defined by

$$e_k = \frac{1}{n}\{k\gamma_{a,ij}(-(n-k)) + (n-k)\gamma_{a,ij}(k)\}$$

  from which the circulant block matrices can be constructed.
- Note: the matrix C_F is itself not a circulant matrix, but is a matrix with k^2 blocks, each $n \times n$ block being a circulant matrix whose first column is denoted by elements $e_0, e_{n-1}, \cdots, e_1$

Steps for VARFIMA model

- Adjust VARFIMA$(0, \mathbf{d}, 0)$ likelihood to that of the VARFIMA$(p, \mathbf{d}, q)$ likelihood Ravishanker & Ray 1997
- Numerically integrate out "latent" vectors through MC simulation
- Determinant term in likelihood is approximated using Böttcher and Silbermann 1999
- We are investigating further on reducing times.
- This approach could also be applied to a Gaussian function for a "latent" process in a hierarchical model with a higher level non-Gaussian setup.

III. Duration Models via Martingale Estimating Equations

Generalized Duration model:

$$x_i = h(\mathcal{F}_{i-1}^{x,\theta}, \psi_i, z_i)\varepsilon_i$$

- $x_i = t_i - t_{i-1}$
- t_i: time of the ith transaction (or event)

Collaborators on projects: A. Thanaveswaran, You Liang (U. Manitoba); Volodymyr Serhiyenko (UConn)

- ψ_i: conditional expectation of the adjusted duration x_i, and is a process based on the data history
- $\{z_i\}$ is a random process independent of the history $\mathcal{F}_{i-1}^{x,\theta}$
- ε_i: i.i.d. non-neg random variables with density function $f(\cdot)$ and unit mean,
- $\mathcal{F}_{t_i-1}^{x}$: information available at the $(i-1)$th event.
- We also assume that ε_i is independent of $\mathcal{F}_{i-1}^{x}$.

Basic Autoregressive Conditional Duration (ACD) Models: Engle and Russell 1998

$$x_i = \psi_i \varepsilon_i,$$
$$\psi_i = \mathrm{E}[x_i | \mathcal{F}^x_{t_i - 1}],$$

Log-ACD Models: Bauwens and Giot 2000

Log-ACD_1 model

$$x_i = \psi_i \varepsilon_i,$$

$$\ln \psi_i = \omega + \sum_{j=1}^{p} \alpha_j \ln x_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j},$$

where $\sum_{j=1}^{\max(p,\ q)} (\alpha_j + \beta_j) < 1$

Log-ACD$_2$ model

$$x_i = \psi_i \varepsilon_i,$$
$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \frac{x_{i-j}}{\psi_{i-j}} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j},$$

where $\sum_{j=1}^{q} \beta_j < 1$

Quadratic Stochastic Conditional Duration (SCD) model:

$$\begin{aligned} x_i &= \exp(cz_i + dz_i^2)\varepsilon_i \\ z_i &= \sum_{j=0}^{\infty} w_j a_{i-j}, \end{aligned} \tag{8}$$

- where $\sum_{j=0}^{\infty} w_j^2 < \infty$
- $a_i \sim NID(0, \sigma_a^2)$.

Inference via Estimating Equations

- Godambe 1985 Bmka; Thavaneswaran & Abraham 1988 JTSA
- Conditional moments of Durations $\{x_i\}$ are
  $\mu_t(\boldsymbol{\theta}) = \mathrm{E}[x_t|\mathcal{F}^x_{(t-1)}]$; $\sigma^2_t(\boldsymbol{\theta}) = \mathrm{Var}[x_t|\mathcal{F}^x_{(t-1)}]$;

  $\gamma_t(\boldsymbol{\theta}) = \mathrm{E}[(x_t - \mu_t(\boldsymbol{\theta}))^3|\mathcal{F}^x_{(t-1)}]$;
  $\kappa_t(\boldsymbol{\theta}) = \mathrm{E}[(x_t - \mu_t(\boldsymbol{\theta}))^4|\mathcal{F}^x_{(t-1)}]$
- To estimate the parameter $\boldsymbol{\theta}$ based on the observations $x_1, \ldots, x_n$

Inference via Estimating Equations

Two classes of martingale differences

- $\{m_i(\boldsymbol{\theta}) = x_i - \mu_i(\boldsymbol{\theta}), i = 1, \ldots, n\}$
- $\{M_i(\boldsymbol{\theta}) = q(m_i(\boldsymbol{\theta})) - \mathrm{E}[q(m_i(\boldsymbol{\theta}))|\mathcal{F}^x_{i-1}], i = 1, \ldots, n\}$

- Quadratic variation of $m_i(\boldsymbol{\theta})$ is $\langle m \rangle_i = \mathrm{E}[m_i^2(\boldsymbol{\theta})|\mathcal{F}_{t-1}^x] = \sigma_i^2(\boldsymbol{\theta})$
- Quadratic variation of $M_i(\boldsymbol{\theta})$ is $\langle M \rangle_i = \mathrm{E}[q^2(m_i(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] - \left(\mathrm{E}[q(m_i(\boldsymbol{\theta}))|\mathcal{F}_{i-1}^x]\right)^2$
- Quadratic covariation of $m_i(\boldsymbol{\theta})$ and $M_i(\boldsymbol{\theta})$ is $\langle m, M \rangle_i = \mathrm{E}[m_i(\boldsymbol{\theta})q(m_i(\boldsymbol{\theta}))|\mathcal{F}_{i-1}^x]$
- q is any differentiable function wrt $\boldsymbol{\theta}$ s.t. $\langle M \rangle_i$ and $\langle m, M \rangle_i$ exist.

Linear and Quadratic Estimating functions:

$$\mathbf{g}_m^*(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{m_i}{\langle m \rangle_i}$$

and

$$\mathbf{g}_M^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(\mathsf{E}\left[\frac{\partial q(m_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial \mathsf{E}\left[q(m_i(\boldsymbol{\theta})) \middle| \mathcal{F}_{i-1}^x \right]}{\partial \boldsymbol{\theta}} \right) \frac{M_i}{\langle M \rangle_i}.$$

Quadratic:

$$\mathbf{g}_M^*(\boldsymbol{\theta}) = -\sum_{t=1}^{n} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{M_t}{\langle M \rangle_t}$$

Information associated with $\mathbf{g}_m^*(\boldsymbol{\theta})$ and $\mathbf{g}_M^*(\boldsymbol{\theta})$ are obtained.

We derive an optimal combined estimating function:
$\mathbf{g}_C^*(\boldsymbol{\theta}) = \sum_{i=1}^n \left(\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}^*(\boldsymbol{\theta}) M_i(\boldsymbol{\theta})\right)$, where

$$\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) = \left(1 - \frac{\langle m, M\rangle_i^2}{\langle m\rangle_i \langle M\rangle_i}\right)^{-1} \left(-\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\langle m\rangle_i} - \left(\mathsf{E}\left[\left.\frac{\partial q(m_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right| \mathcal{F}_{i-1}^x\right] - \frac{\partial \mathsf{E}\left[\left. q(m_i(\boldsymbol{\theta}))\right| \mathcal{F}_{i-1}^x\right]}{\partial \boldsymbol{\theta}}\right) \frac{\langle m, M\rangle_i}{\langle m\rangle_i \langle M\rangle_i}\right)$$

$$\mathbf{b}_{i-1}^* \boldsymbol{\theta} = \left(1 - \frac{\langle m, M\rangle_i^2}{\langle m\rangle_i \langle M\rangle_i}\right)^{-1} \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\langle m, M\rangle_i}{\langle m\rangle_i \langle M\rangle_i} + \left(\mathsf{E}\left[\left.\frac{\partial q(m_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right| \mathcal{F}_{i-1}^x\right] - \frac{\partial \mathsf{E}\left[\left. q(m_i(\boldsymbol{\theta}))\right| \mathcal{F}_{i-1}^x\right]}{\partial \boldsymbol{\theta}}\right) \frac{1}{\langle M\rangle_i}\right);$$

- We derive the associated information
- we derive a recursive estimate for $\boldsymbol{\theta}$
- we derive the formulas for the duration models discussed in the literature.
- applications to financial transactions data

Thank you