Time Series Prediction via Density Ratio Modeling

Benjamin Kedem
Department of Mathematics & ISR University of Maryland, College Park, MD

Joint with: Richard E. Gagnon and Haiming Guo.

A Density Ratio

Our starting point is a well known fact regarding k-parameter exponential families:

$\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$

$$g(x, \boldsymbol{\theta}) = d(\boldsymbol{\theta}) S(x) \exp\left\{\sum_{i=1}^{k} c_i(\boldsymbol{\theta}) T_i(x)\right\}$$

Define:
$\alpha = \log[d(\boldsymbol{\theta}_1)/d(\boldsymbol{\theta}_2)]$
$\boldsymbol{\beta} = (c_1(\boldsymbol{\theta}_1) - c_1(\boldsymbol{\theta}_2), ..., c_k(\boldsymbol{\theta}_1) - c_k(\boldsymbol{\theta}_2))'$
$\boldsymbol{h} = (T_1(x), ..., T_k(x))'$
Then

$$\frac{g_1(x)}{g_2(x)} \equiv \frac{g(x, \boldsymbol{\theta}_1)}{g(x, \boldsymbol{\theta}_2)} = \exp\{\alpha + \boldsymbol{\beta}' \boldsymbol{h}(x)\} \tag{1}$$

or

$$g_1(x) = \exp\{\alpha + \boldsymbol{\beta}' \boldsymbol{h}(x)\} g_2(x) \tag{2}$$

Normal distribution:
$\boldsymbol{\theta} = (\mu, \sigma^2)$,

$$\begin{aligned}
\alpha &= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\mu_2^2}{2\sigma_2^2} - \frac{\mu_1^2}{2\sigma_1^2} \\
\boldsymbol{\beta} &= \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}, \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)' \\
\boldsymbol{h}(x) &= (x, x^2)'
\end{aligned}$$

Gamma distribution:
with shape r and scale λ, $\boldsymbol{\theta} = (r, \lambda)$,

$$\begin{aligned}
\alpha &= \log\frac{\lambda_1^{r_1}\Gamma(r_2)}{\lambda_2^{r_2}\Gamma(r_1)} \\
\boldsymbol{\beta} &= (\lambda_2 - \lambda_1, r_1 - r_2)' \\
\boldsymbol{h}(x) &= (x, \log x)'
\end{aligned}$$

Rayleigh distribution:
$\alpha = \log\frac{\theta_2^2}{\theta_1^2}, \quad \beta = \frac{1}{2\theta_2^2} - \frac{1}{2\theta_1^2}, \quad h(x) = x^2$

A Density Ratio Model for Time Series

Assume we have m regressions:

$$\begin{aligned} x_{1t} &= f_1(\boldsymbol{z}_{1,t-1}) + \epsilon_{1t}, \quad t = 1, ..., n_1 \\ &. \\ &. \\ &. \\ x_{qt} &= f_q(\boldsymbol{z}_{q,t-1}) + \epsilon_{qt}, \quad t = 1, ..., n_q \\ x_{mt} &= f_m(\boldsymbol{z}_{m,t-1}) + \epsilon_{mt}, \quad t = 1, ..., n_m \end{aligned} \tag{3}$$

The x_{jt} may be nonstationary, and the f_i may be nonlinear.

In applications the f_i are estimated and the ϵ_{kt} are the residuals. They may be dependent.

Suppose that for each t,

$$\epsilon_{jt} \sim g_j(x), \quad j = 1, ..., q, m$$

Define the *reference* density:

$$g(x) = g_m(x)$$

Emboldened by (2), we assume:

$$g_j(x) = \exp\{\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(x)\} g(x) \qquad j = 1, ..., q \tag{4}$$

Define *the combined data*

$$\boldsymbol{\tau} \equiv \{(\epsilon_{1t}, ..., \epsilon_{1n_1}), ..., (\epsilon_{q1}, ..., \epsilon_{qn_q}), (\epsilon_{m1}, ..., \epsilon_{mn_m})\}$$

Problem:
From *the combined data* estimate all the $\alpha_j, \boldsymbol{\beta}_j$, and $g(x)$,$G(x)$, for the purpose of predicting the future reference value $x_{m,t+1}$.

Remark

$g(x)$ is estimated from all the combined residual set of size $n = n_1 + n_2 + \cdots + n_m$ and not just from "its own" residuals $(\epsilon_{m1}, ..., \epsilon_{mn_m})$ of size $n_m < n$.

Remark

The exponential "tilt" relationships (4) relative to a reference or baseline density g enable a semiparametric inference about the $\alpha_j, \boldsymbol{\beta}_j$, and about g, G based on the combined data set.

Remark

Using the estimator $\widehat{G}$ of G we can estimate future probabilities of events formulated in terms of the "reference" $x_{m,t+1}$ conditional on $\boldsymbol{z}_{m,t}$.

Example: Bivariate AR.

The linear system

$$\begin{aligned} x_t &= a_1 x_{t-1} + a_2 y_{t-1} + \epsilon_t \\ y_t &= b_1 x_{t-1} + b_2 y_{t-1} + \eta_t \end{aligned}$$

$t = 1, ..., N$, with independent Gaussian noise components $\epsilon_t \sim \mathrm{N}(0, \sigma_1^2)$ and $\eta_t \sim \mathrm{N}(0, \sigma_2^2)$, satisfies the density ratio model. We have

$$\begin{aligned} \frac{g_\epsilon(x)}{g_\eta(x)} &= \exp\left\{\log\frac{\sigma_2}{\sigma_1} + \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)x^2\right\} \\ &\equiv e^{\alpha+\beta x^2} \end{aligned}$$

or $m = 2, q = 1$, and (4) reduces to

$$g_\epsilon(x) = e^{\alpha+\beta x^2} g_\eta(x) \tag{5}$$

Semiparametric Estimation

A maximum likelihood estimator of $G(x)$ can be obtained by maximizing the empirical likelihood over the class of step cdf's with jumps at the values $\tau_1, ..., \tau_n$: Qin and Lawless(1994), Qin and Zhang(1997), GLV(1999), FKQS(2001).

Let $w_j(\tau) = \exp\left\{\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(\tau)\right\}$, $j = 1, ..., q$, and $p_i = dG(\tau_i)$, $i = 1, .., n$.

Then the *empirical likelihood* becomes (Owen(2001)),

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} w_1(\epsilon_{1j}) \cdots \prod_{j=1}^{n_q} w_q(\epsilon_{qj}) \tag{6}$$

Define:

$$\widehat{w}_j(\tau) = \exp\left\{\widehat{\alpha}_j + \widehat{\boldsymbol{\beta}}_j' \boldsymbol{h}(\tau)\right\}$$

We obtain FKQS(2001),

$$\widehat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 \exp(\widehat{w}_1(\tau_i)) + \cdots + \rho_q \exp(\widehat{w}_q(\tau_i))}$$

Therefore, with $I(B)$ the indicator of B,

$$\widehat{G}(\tau) = \sum_{i=1}^{n} \widehat{p}_i I(\tau_i \leq \tau) =$$

$$\frac{1}{n_m} \cdot \sum_{i=1}^{n} \frac{I(\tau_i \leq \tau)}{1 + \rho_1 \exp(\widehat{w}_1((\tau_i)) + \cdots + \rho_q \exp(\widehat{w}_q((\tau_i))}$$

Asymptotic properties of $\widehat{G}$, and its optimality over the empirical distribution function obtained only from the reference sample

$$\epsilon_{m1}, ..., \epsilon_{mn_m}$$

ignoring all the other samples, are discussed in a sequence of papers by Zhang (2000abc).

Prediction

Since

$$x_{m,t+1} = f_m(\boldsymbol{z}_{m,t}) + \epsilon_{m,t+1}$$

and $\epsilon_{m,t+1} \sim G$, we have the useful approximation of the predictive probability at $t+1$ conditional on $\boldsymbol{z}_{m,t}$,

$$\begin{aligned} P(x_{m,t+1} \leq x \mid \boldsymbol{z}_{m,t}) &= G(x - f_m(\boldsymbol{z}_{m,t})) \\ &\approx \widehat{G}(x - f_m(\boldsymbol{z}_{m,t})) \end{aligned} \tag{7}$$

From (7) we can get the predicted values

$$\widehat{E}(x_{m,t+1} \mid \boldsymbol{z}_{m,t})$$

Bivariate AR Example

The bivariate AR(1) system (5) was simulated with parameters

$$(a_1, a_2, b_1, b_2, \sigma_1, \sigma_2) = (0.6, -0.5, 0.4, 0.5, 1, 0.5)$$

and $n_1 = n_2 = 500$.

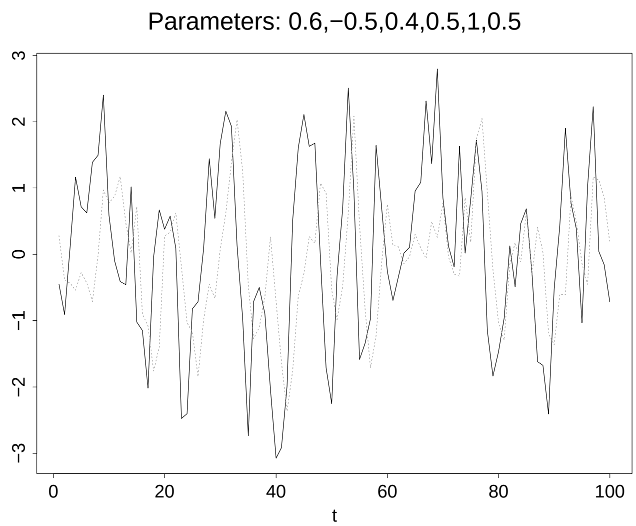
The true “tilt” parameters are:

$$(\alpha_1, \beta_1) = (-0.693, 1.5)$$

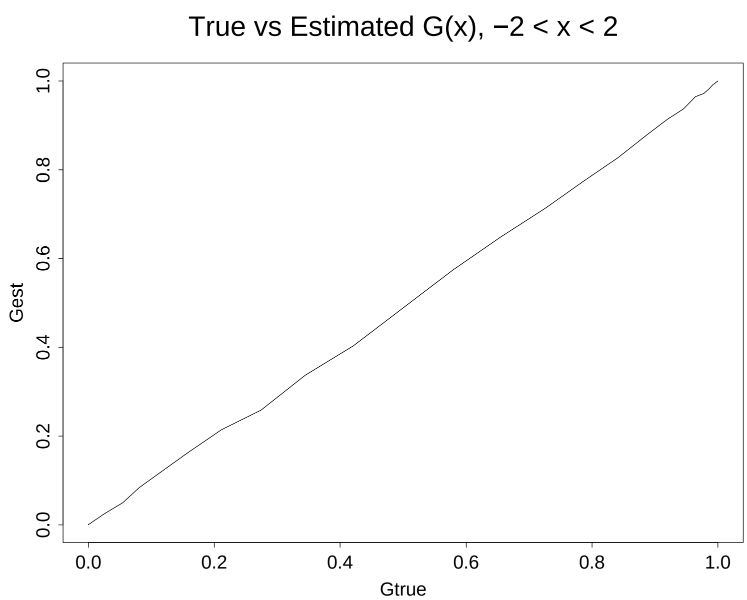
and the reference g is the pdf of η_t,

$$\eta_t \sim \mathrm{N}(0, 0.5^2)$$

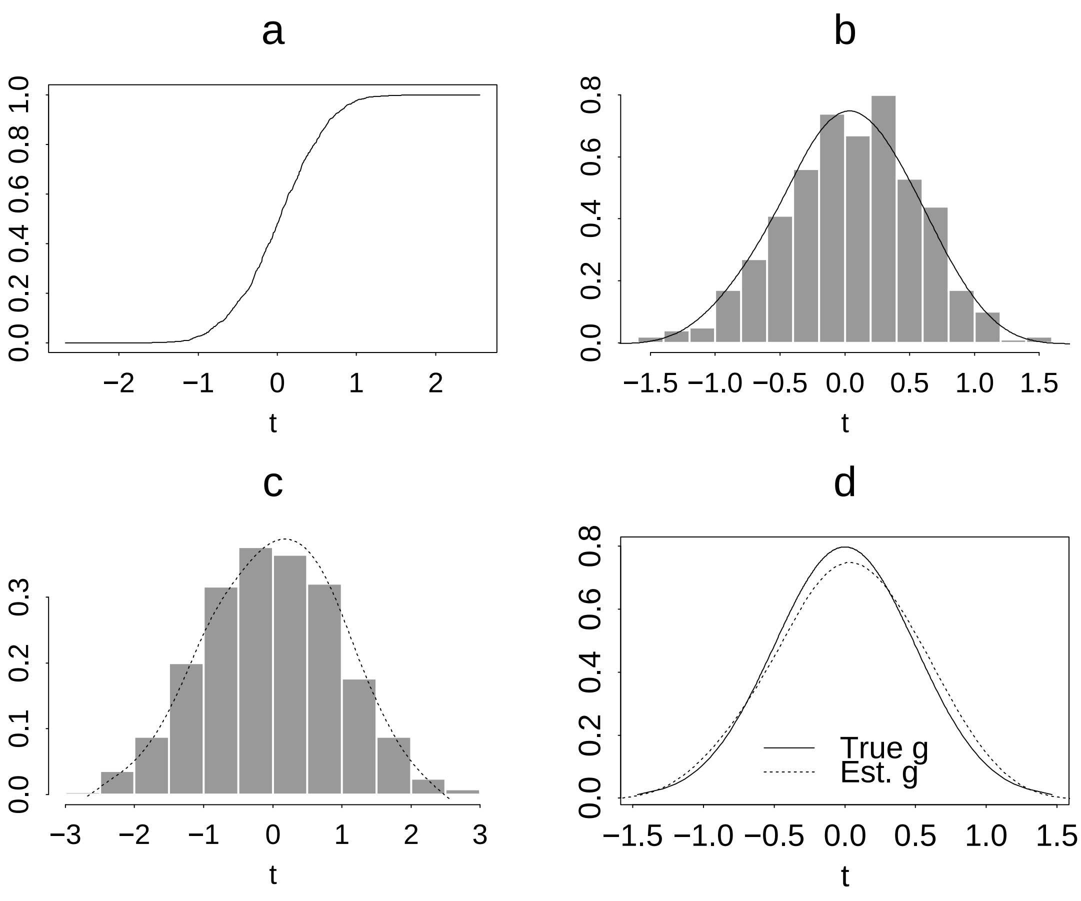
Plots of x_t (solid) and y_t (dashed), $t = 1, ..., 100$.



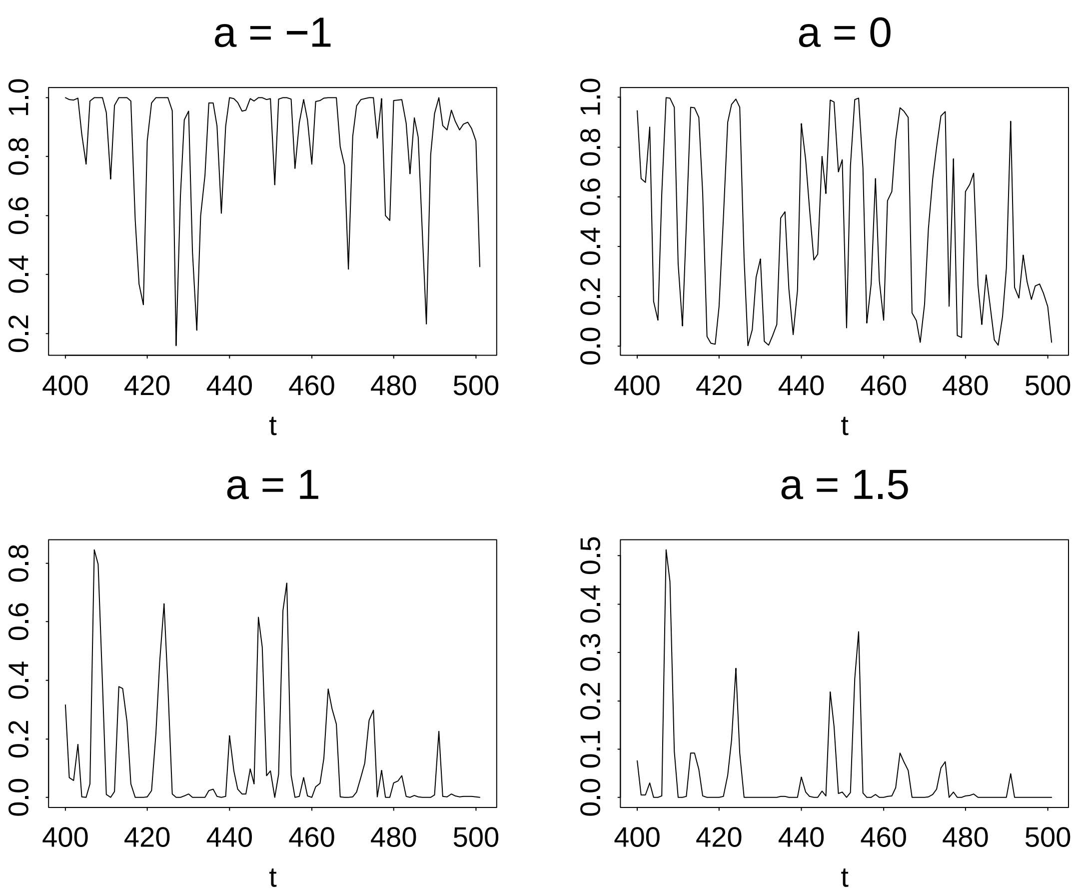
Pairs of $(G(x), \widehat{G}(x))$, $-2 < x < 2$.



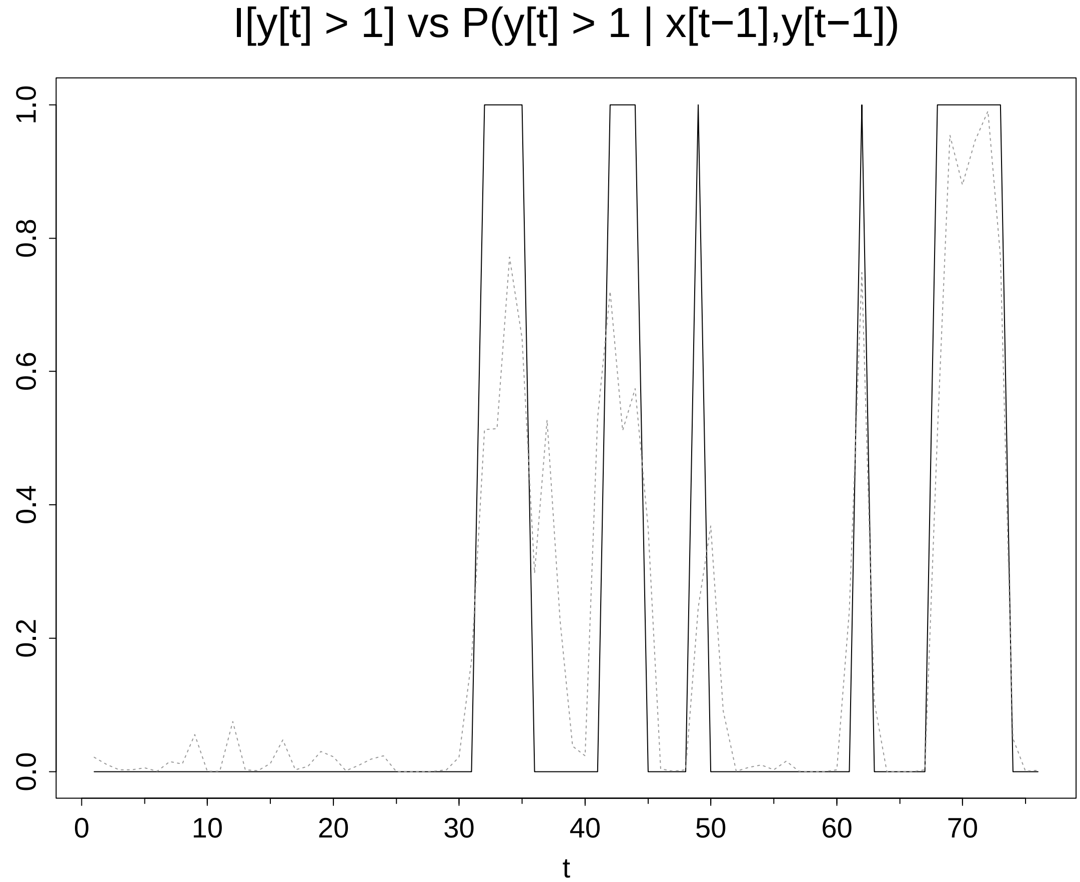
Bivariate AR: a. Estimated reference cdf $\widehat{G} = \widehat{G}_\eta$. b. Estimated spline-smoothed reference pdf $\widehat{g} = \widehat{g}_\eta$ and a histogram from η_t. c. Estimated spline-smoothed tilted pdf $\widehat{g}_\epsilon$ and a histogram from ϵ_t. d. Estimated spline-smoothed reference pdf (dashed line) and the true $g = g_\eta$ (solid line).



Prediction: $\widehat{P}(y_t > a \mid x_{t-1}, y_{t-1})$, $t = 400, ..., 501$.

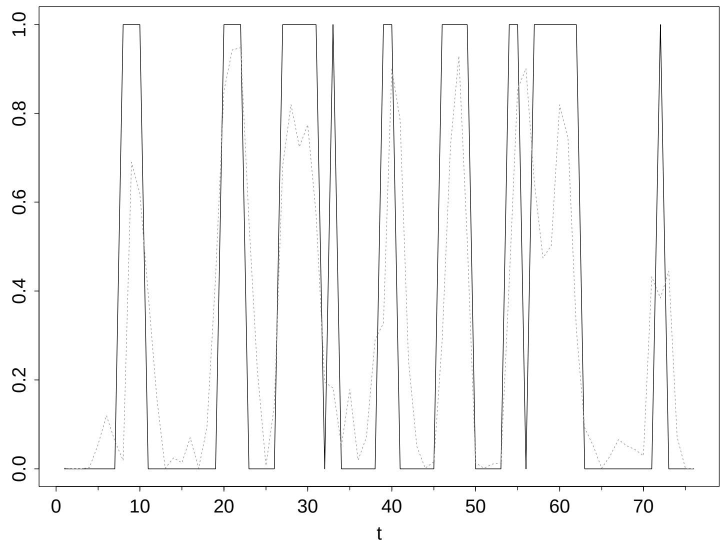


Prediction: $\widehat{P}(y_t > 1 \mid x_{t-1}, y_{t-1})$ versus $I[y_t > 1]$, $t = 400, ..., 501$.



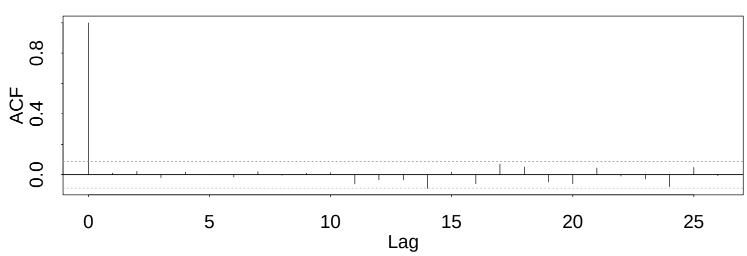
Prediction: $\widehat{P}(y_t > 0.5 \mid x_{t-1}, y_{t-1})$ versus $I[y_t > 0.5]$, $t = 400, ..., 501$.

I[y[t] > 0.5] vs P(y[t] > 0.5 | x[t−1],y[t−1])

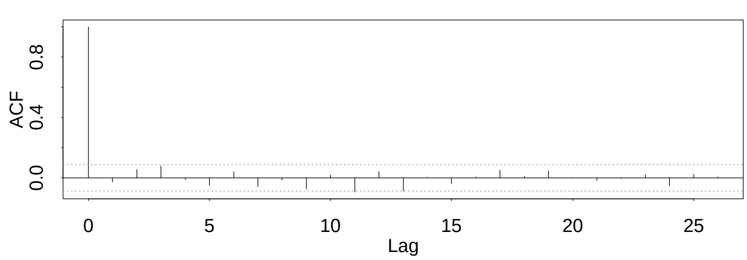


Estimated autocorrelation of the residual series resulting from the estimated bivariate AR model (5).

Series : r1

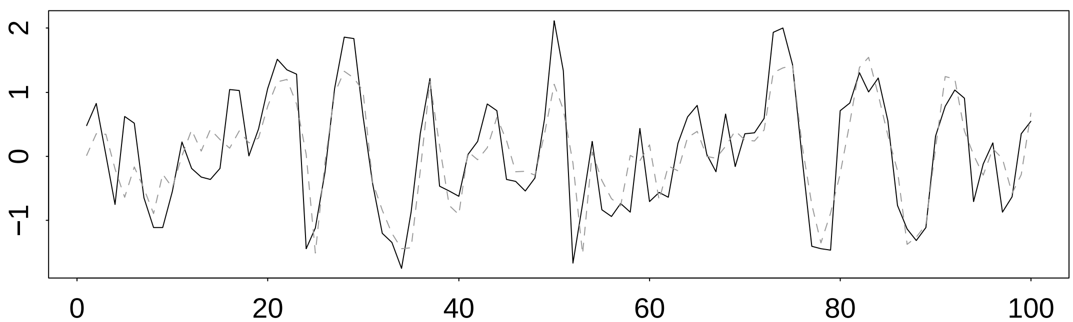


Series : r2

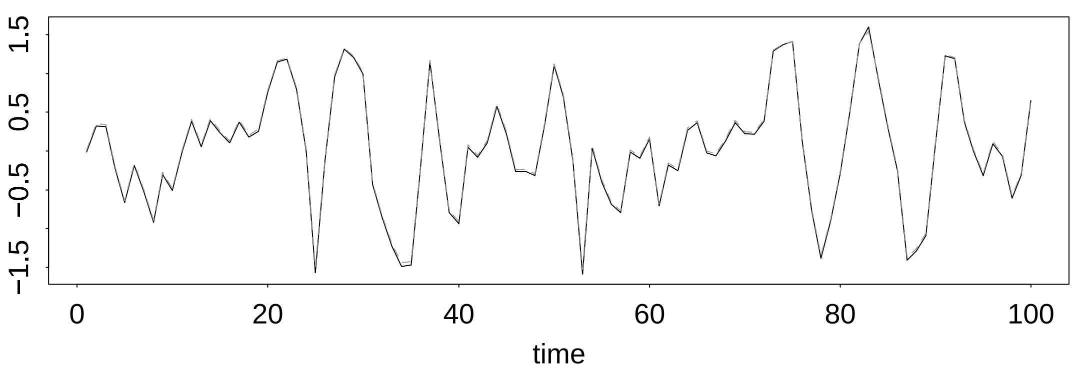


Top: Prediction (dashed line) of y_t from the bivariate AR model (5) using the conditional expectation computed from (7). Bottom: The conditional expectation $E[y_t \mid x_{t-1}, y_{t-1}]$ and its estimate (dashed line) obtained from (7).

Semiparametric Prediction



The Semiparametric Predictor vs E(y[t]|x[t−1],y[t−1])



LA Mortality Prediction

We shall now apply the semiparametric method to sampled filtered mortality data from Los Angeles County, 01.01.1970 to 12.31.1979, discussed in Shumway et al(1988). The original daily data, consisting of the response series (total mortality) and its covariate series (two weather and six pollution series), were filtered and then sampled weekly to produce series of length $N = 508$ each.

$y =$ Total mortality
$T =$ Temperature (centered at 74.26 F$^\circ$)
$CO =$ Carbon monoxide

Kedem and Fokianos(2002): **Model 1**

$$y_t = \exp\left\{\widehat{\beta}_0 + \widehat{\beta}_1 y_{t-1} + \widehat{\beta}_2 y_{t-2} + \widehat{\beta}_3 T_{t-1} + \widehat{\beta}_4 \log(CO_t)\right\} + \eta_t$$

$$(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4) = (4.505, 0.00189, 0.00184, -0.00133, 0.04683)$$

The exponential model outperforms several competitors, giving close to white noise residual η_t.

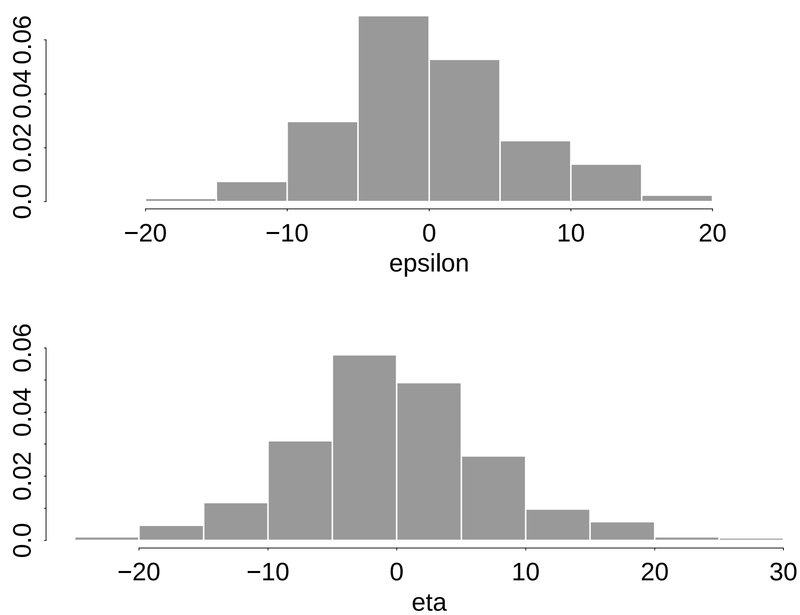
Additional regression: **Model 2**

$$T_t = \widehat{\phi}_1 T_{t-1} + \widehat{\phi}_2 T_{t-2} + \widehat{\phi}_3 T_{t-3} + \widehat{\phi}_4 T_{t-4} + \epsilon_t$$

$$(\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\phi}_3, \widehat{\phi}_4) = (0.272, 0.283, 0.096, 0.183)$$

The sample autocorrelation of the residual ϵ_t has somewhat significant values at lags 8,9,22, so that ϵ_t is probably not white noise.

LA Mortality: Histograms of ϵ_t and η_t residual series.



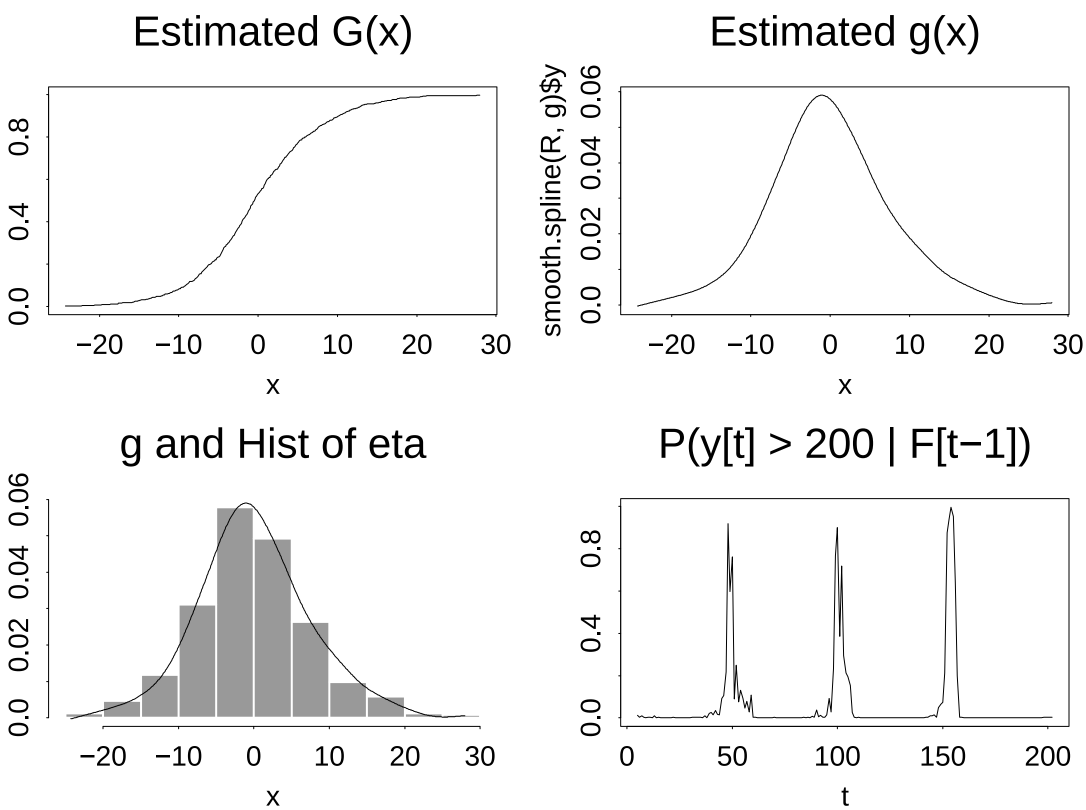
The figure suggests the tilt model (5),

$$g_\epsilon(x) = e^{\alpha+\beta x^2} g_\eta(x)$$

where η_t is the reference with density $g = g_\eta$ and distribution function G. In the present case

$$(\widehat{\alpha}, \widehat{\beta}) = (0.145(0.04), -0.003(0.00086))$$

Estimated g, G, comparison of g and the corresponding histogram from η_t, and prediction for the filtered/sampled mortality data.



Bottom left: Note the fit.
Bottom right: The estimated prediction probabilities obtained from $\widehat{G}$ of exceeding 200 "deaths",

$$P(Y_t > 200 \mid Y_{t-1}, Y_{t-2}, T_{t-1}, \log(CO_t)) \approx 1 -$$
$$\widehat{G}\left(200 - \exp\left\{\widehat{\beta}_0 + \widehat{\beta}_1 Y_{t-1} + \widehat{\beta}_2 Y_{t-2} + \widehat{\beta}_3 T_{t-1} + \widehat{\beta}_4 \log(CO_t)\right\}\right)$$

t=3,...,202, pointing to a strong annual cycle.

Semiparametric Forecasting of U.S. Mortality

a. For each age 1-85 we have a short ts: 1970-2001 (size 32).

b. Fusion of short time series 1970-2001 and predicting 2002.

c. Each age is predicted from 5 time series corresponding to 5 ages.

d. Autoregression for each age:

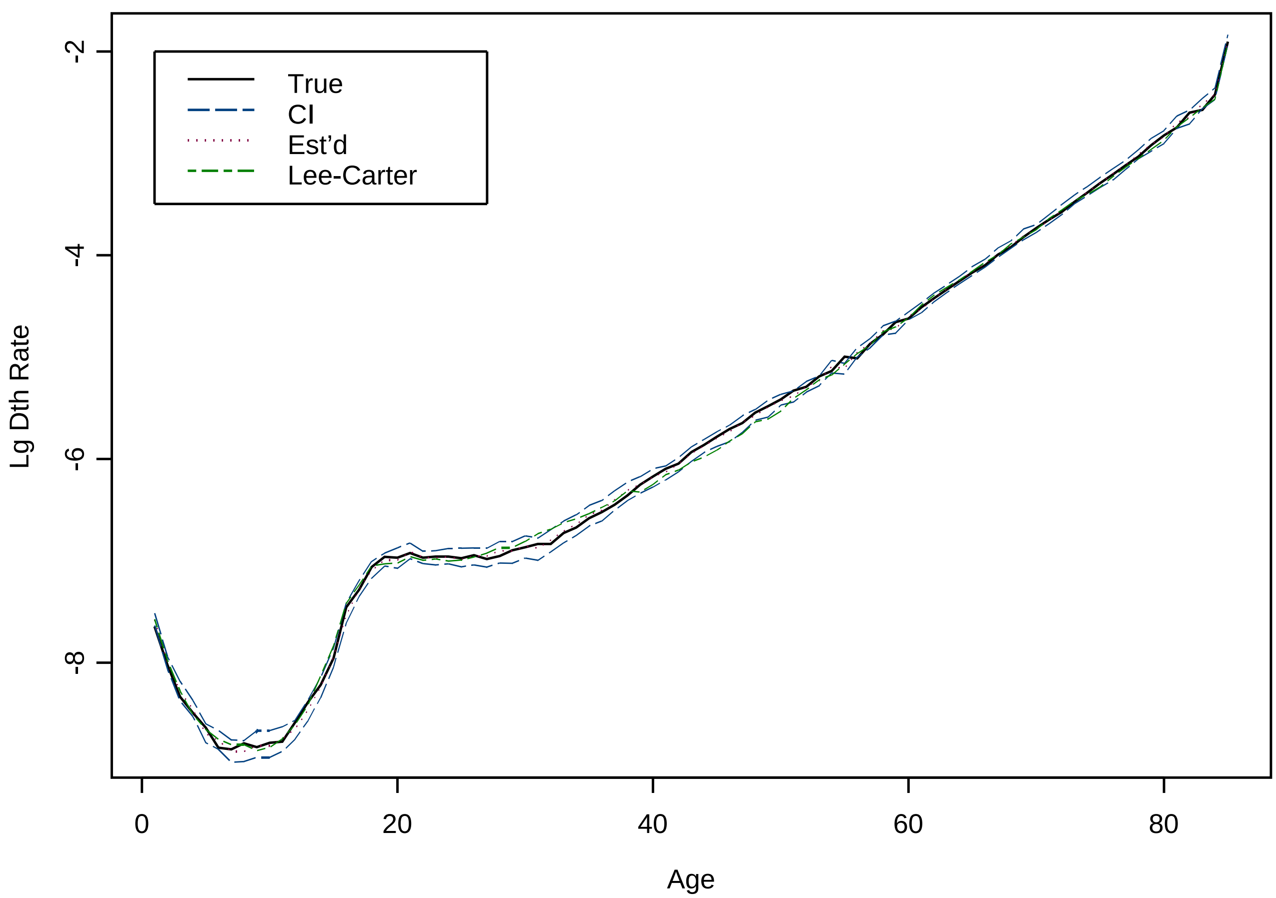
$$x_t = bx_{t-1} + c + \epsilon_t, \quad t = 1970, ..., 2001 \tag{8}$$

e. All Age MSE: Lee & Carter (1992) SVD vs SP.

Prediction	Case	LC	SP
1 Step	Total Pop	0.297	0.104
1 Step	Female	0.619	0.187
1 Step	White Female	0.645	0.249
2 Step	Total Pop	0.389	0.180

Prediction for year 2002.

True (solid), SP Model (dotted), L-C model (dash-dot), 95% CI bounds (dash).



VaR: SP vs GARCH(1,1)

GARCH-Normal case:

$$\widetilde{r}_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

TGARCH-Gamma case:

$$\widetilde{r}_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathrm{Gamma}(4,2).$$

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & : \quad \epsilon_t > 0 \\ \alpha_0 + \alpha_2 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & : \quad \epsilon_t \leq 0 \end{cases}$$

GARCH uses $\widetilde{r}_t$ only, assuming it is normal.
SP Uses $r_t, \widetilde{r}_t$, assuming only a quadratic tilt':

$$g(x) = \exp(\alpha + \beta x^2)\widetilde{g}(x)$$

Define: $r_t = 0.8 * \widetilde{r}_t + 0.2 * y_t$, with y_t a normal GARCH(1,1).
The tilt assumption is not satisfied!

Problem: Estimate VaR of the reference $\widetilde{r}_t$.

Normal GARCH case:

Given p		SP	Historical	GARCH
0.01	MSE	0.0709	0.0709	0.0044
0.05	MSE	0.0121	0.0158	0.0020
0.01	BIAS	1.3295	0.9745	0.0032
0.05	BIAS	0.0604	0.0680	0.0000
0.01	ER	0.0097	0.0131	0.0254
0.05	ER	0.0513	0.0541	0.0650

TGARCH Case:

Given p		SP	Historical	GARCH
0.01	MSE	0.0066	0.0129	0.0280
0.05	MSE	0.0046	0.0047	0.0072
0.01	BIAS	0.1243	0.2356	1.2696
0.05	BIAS	0.0005	0.0028	0.2836
0.01	ER	0.0111	0.0113	0.0036
0.05	ER	0.0577	0.0515	0.0256

We close by noting the curious fact that the tilt model (4) is reminiscent of the celebrated relationship between the spectral densities of the input, $f(\omega)$, and output, $f_H(\omega)$, in a linear system with squared gain $|H^2(\omega)|$,

$$f_H(\omega) = |H(\omega)|^2 f(\omega)$$

If the input series is operated on by a sequence of m linear filters, then with output spectral densities $f_1(\omega), ..., f_m(\omega)$

$$\begin{array}{rcl} f_1(\omega) & = & |H_1(\omega)|^2 f(\omega) \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ f_m(\omega) & = & |H_m(\omega)|^2 f(\omega) \end{array}$$

That is, m distortions of the same f as in (4).

<u>Derivation of SE's</u>

The asymptotic covariance matrix of the parameter estimates is

$$\boldsymbol{\Sigma} = \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1} \tag{9}$$

The matrices $\boldsymbol{S}, \boldsymbol{V}$ are derived by repeated differentiation of the profile log-likelihood.

First define

$$\nabla \equiv \left(\frac{\partial}{\partial\alpha_1}, ..., \frac{\partial}{\partial\alpha_q}, \frac{\partial}{\partial\boldsymbol{\beta}_1} ..., \frac{\partial}{\partial\boldsymbol{\beta}_q}\right)' \tag{10}$$

Then $E[\nabla l(\alpha_1, ..., \alpha_q, \boldsymbol{\beta}_1, .., \boldsymbol{\beta}_q)] = \boldsymbol{0}$. To obtain the score second moments it is convenient to define $\rho_m \equiv 1,\ w_m(t) \equiv 1$,

$$E_j[\boldsymbol{h}(t)] \equiv \int \boldsymbol{h}(t) w_j(t) dG(t) \tag{11}$$

and,

$$A_0(j, j') \equiv \int \frac{w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{12}$$

$$\boldsymbol{A}_1(j, j') \equiv \int \frac{\boldsymbol{h}(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{13}$$

$$\boldsymbol{A}_2(j, j') \equiv \int \frac{\boldsymbol{h}(t)\boldsymbol{h}'(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{14}$$

for $j, j' = 1, ..., q$.

Then, the entries in

$$\boldsymbol{V} \equiv Var\left[\frac{1}{\sqrt{n}}\nabla l(\alpha_1,...,\alpha_q,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_q)\right] \tag{15}$$

are,

$$\frac{1}{n}Var\left(\frac{\partial l}{\partial \alpha_j}\right) = \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j) - \sum_{r=1}^m \rho_r A_0^2(j,r)\} \tag{16}$$

$$\begin{aligned}\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j},\frac{\partial l}{\partial \alpha_{j'}}\right) &= \frac{\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j')\\ &- \sum_{r=1}^m \rho_r A_0(j,r)A_0(j',r)\}\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j},\frac{\partial l}{\partial \boldsymbol{\beta}_j}\right) &= \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j)E_j[\boldsymbol{h}'(t)]\\ &- \sum_{r=1}^m \rho_r A_0(j,r)\boldsymbol{A}_1'(j,r)\}\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j},\frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) &= \frac{\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j')E_{j'}[\boldsymbol{h}'(t)]\\ &- \sum_{r=1}^m \rho_r A_0(j,r)\boldsymbol{A}_1'(j',r)\}\end{aligned} \tag{19}$$

$$\begin{aligned}\frac{1}{n}Cov\left(\frac{\partial l}{\partial \boldsymbol{\beta}_j},\frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) &= \frac{\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{-\boldsymbol{A}_2(j,j') + E_j[\boldsymbol{h}(t)]\boldsymbol{A}_1'(j,j')\\ &+ \boldsymbol{A}_1(j,j')E_{j'}[\boldsymbol{h}'(t)]\\ &- \sum_{r=1}^m \rho_r \boldsymbol{A}_1(j,r)\boldsymbol{A}_1'(j',r)\}\end{aligned}$$

$$+ \quad \frac{1}{n}\sum_{i=1}^{n_j}\sum_{k=1}^{n_{j'}} Cov[\boldsymbol{h}(\epsilon_{ji}), \boldsymbol{h}(\epsilon_{j'k})] \tag{20}$$

The last term is 0 for $j \neq j'$ and $(n_j/n)Var[\boldsymbol{h}(\epsilon_{j1})]$ for $j = j'$.

Next, as $n \to \infty$,

$$-\frac{1}{n}\nabla\nabla' l(\alpha_1, ..., \alpha_q, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q) \to \mathbf{S} \tag{21}$$

where $\mathbf{S}$ is a $q(1+p) \times q(1+p)$ matrix with entries corresponding to $j, j' = 1, ..., q$,

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j^2} \quad \to \quad \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{22}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \alpha_{j'}} \quad \to \quad \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{23}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \partial \boldsymbol{\beta}_j'} \quad \to \quad \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{24}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \partial \boldsymbol{\beta}_{j'}'} \quad \to \quad \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{25}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_j'} \quad \to \quad \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{26}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_{j'}'} \quad \to \quad \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \tag{27}$$

Aitchison, J. and Dunsmore, I.R. (1975). *Statistical Prediction Analysis*. Cambridge: Cambridge University Press..

Chen, M.H. and Ibrahim, .J.G. (2000). Bayesian predictive inference for time series count data. *Biometrics*, 56, 678-85.

De Oliveira, V., Kedem, B., and Short, D.A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, 92, 1422-1433.

Efron, B. and Tibshirani, R. (1996). Using specially designed exponential families for density estimation. *The Annals of Statistics*, **24**, 2431-2461.

Fokianos, K. (2004). Merging information for semiparametric density estimation. *JRSS*, B, **66**, 941-958.

Fokianos, K. and Kaimi, I. (2004). On the effect of misspecifying the density of ratio model. Submitted.

Fokianos, K., B. Kedem, J. Qin, J. Haferman, and D. A. Short (1998). On Combining Instruments. *Journal of Applied Meteorology*, **37**, 220-226.

Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, **43**, 56-65.

Gagnon, R. (2005). Certain Computational Aspects of Power Efficiency and State Space Models. PhD Dissertation, Mathematics Department, University of Maryland, College Park

Geisser, S. (1993). *Predictive inference : an introduction*. New York : Chapman & Hall.

Gilbert, P.B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *Annals of Statistics*, **28**, 151-194.

Gilbert, P.B., Lele, S.R., Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, 86, 27-43.

Kay, R., and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, **74**, 495-501.

Kedem, B., Wolff, D.B., and Fokianos, K. (2004). Statistical Comparison of Algorithms. *IEEE Trans. Instrum. Meas.*, Vol. 53, 770-776.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, New York: Wiley.

Kedem, B., Gagnon, R., and Guo, H. (2005). Time series prediction via density ratio modeling. Submitted.

Kedem, B., Lu, G., Wei, R, and Williams, P.D. (2005). Forecasting US mortality rates. Unpublished.

Lee, R.D. and Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of American Statistical Association.*, Vol. 87, 659-671.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, New York.

Patil, G.P., Rao, C.R., and Zelen, M. (1988). Weighted distributions. In *Encyclopedia of Statistical Sciences*, 9, 565-71, S. Kotz and N.L. Johnson Eds., New York: Wiley.

Qin, J. (1993). Empirical likelihood in biased sampling problems. *Ann. Statist.*, 21, 1182-1186.

Qin, J., and J.F. Lawless (1994), "Empirical likelihood and general estimating equations," *The Annals of Statistics*, **22**, 300-325.

Qin, J., and Zhang, B. (1997). A goodness of fit test for logistic regression models based on case-control data. in *Biometrika*, **84**, 609-618.

Shumway, R.H, Azari, A.S., and Pawitan, Y (1988). Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. *Environmental Research*, 45, 224-241.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Annals of Statistics*, Vol. 10, 616-20.

Vardi, Y. (1985). Empirical distribution in selection bias models. *Annals of Statistics*, Vol. 13, 178-203.

Zhang, B. (2000a). M-estimation under a two sample semiparametric model. *Scand. Jour. of Statistics*, 27, 263-280.

Zhang, B. (2000b). Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6, 491-511.

Zhang, B. (2000c). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica*, 10, 839-865.