Profile Likelihood

version: 2020-01-21 · 11:02:16

Setup

- **Objective:** Maximize the loglikelihood function $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y}),\ \boldsymbol{\theta} \in \mathbb{R}^p$.
- **Dimension Reduction:** Suppose that we have $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\phi})$, where $\boldsymbol{\eta} \in \mathbb{R}^q$ and $q < p$, such that for any value of $\boldsymbol{\eta}$, the conditional MLE

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\phi}} \ell(\boldsymbol{\eta}, \boldsymbol{\phi} \mid \boldsymbol{Y})$$

  can be easily calculated.
- **Profile Likelihood:** Defined as the q-dimensional function

$$\ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y}) = \ell(\boldsymbol{\eta}, \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} \mid \boldsymbol{Y}).$$

- **Proposition:** Let $\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y})$ and $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\eta}}}$. Then

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\phi}}) = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta} \mid \boldsymbol{Y}).$$

  Proof:

$$\begin{aligned} \ell(\boldsymbol{\eta}, \boldsymbol{\phi} \mid \boldsymbol{Y}) \le \ell(\boldsymbol{\eta}, \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} \mid \boldsymbol{Y}) &= \ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y}) \\ &\le \ell_{\text{prof}}(\hat{\boldsymbol{\eta}} \mid \boldsymbol{Y}) = \ell(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\eta}}} \mid \boldsymbol{Y}) = \ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{Y}). \end{aligned}$$

Example: Regression-Like Models

- **Generalized Regression Model:** $M_R : \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{V})$.
  - **Loglikelihood:** Dropping only the $(2\pi)^{n/2}$ term we have

$$\begin{aligned}\log p(\boldsymbol{y} \mid \beta, \sigma, \boldsymbol{X}, \boldsymbol{V}) &= -\frac{1}{2}\left\{\frac{(\boldsymbol{y}-\boldsymbol{X}\beta)'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\beta)}{\sigma^2} + n\log(\sigma^2) + \log|\boldsymbol{V}|\right\} \\ &= g(\beta, \sigma \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{V}).\end{aligned}$$

  - **MLE:** For given $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{V}$ we have

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}, \qquad \hat{\sigma}^2 = \tfrac{1}{n}(\boldsymbol{y}-\boldsymbol{X}\hat{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}).$$

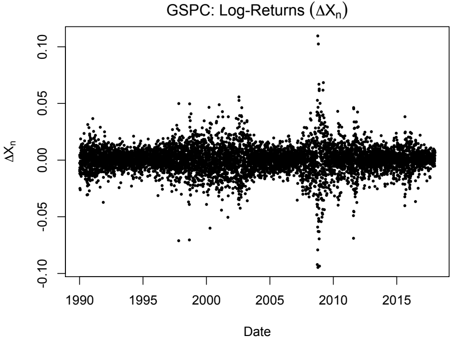
Example: Regression-Like Models

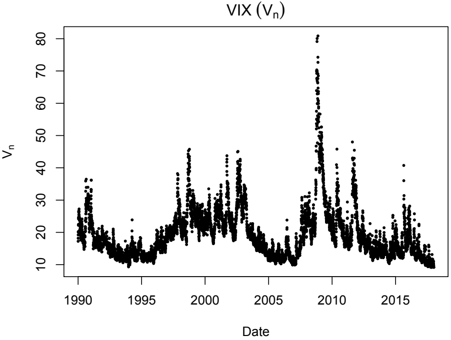
- **Generalized Regression Model:** $M_R : \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{V})$.
  - **Loglikelihood:**
$$\log p(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma, \boldsymbol{X}, \boldsymbol{V}) = -\frac{1}{2}\left\{\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^2} + n\log(\sigma^2) + \log|\boldsymbol{V}|\right\}$$
$$= g(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{V}).$$
- **Regression-Like Model:** $M : \boldsymbol{Y} \sim p(\boldsymbol{Y} \mid \boldsymbol{\theta})$, for which the loglikelihood function can be written as
$$\ell(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \ell(\boldsymbol{\eta}, \boldsymbol{\phi} \mid \boldsymbol{Y}) = g(\beta, \sigma \mid \boldsymbol{y}_{\boldsymbol{\eta}}, \boldsymbol{X}_{\boldsymbol{\eta}}, \boldsymbol{V}_{\boldsymbol{\eta}}),$$
for $\boldsymbol{\phi} = (\beta, \sigma)$ and given functions $\boldsymbol{y}_{\boldsymbol{\eta}}$, $\boldsymbol{X}_{\boldsymbol{\eta}}$, and $\boldsymbol{V}_{\boldsymbol{\eta}}$.
  - **Conditional MLE:** For fixed $\boldsymbol{\eta}$ we have
$$\hat{\beta}_{\boldsymbol{\eta}} = (\boldsymbol{X}_{\eta}'\boldsymbol{V}_{\boldsymbol{\eta}}^{-1}\boldsymbol{X}_{\eta})^{-1}\boldsymbol{X}_{\eta}'\boldsymbol{V}_{\boldsymbol{\eta}}^{-1}\boldsymbol{y}_{\boldsymbol{\eta}}, \qquad \hat{\sigma}^2_{\boldsymbol{\eta}} = \tfrac{1}{n}(\boldsymbol{y}_{\boldsymbol{\eta}} - \boldsymbol{X}_{\eta}\hat{\beta}_{\boldsymbol{\eta}})'\boldsymbol{V}_{\boldsymbol{\eta}}^{-1}(\boldsymbol{y}_{\boldsymbol{\eta}} - \boldsymbol{X}_{\eta}\hat{\beta}_{\boldsymbol{\eta}}).$$
  - **Profile Likelihood:** $\ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y}) = -\tfrac{1}{2}\left\{n + n\log\hat{\sigma}^2_{\boldsymbol{\eta}} + \log|\boldsymbol{V}_{\boldsymbol{\eta}}|\right\}.$

Example: Stochastic Volatility Modeling

- **Data:**
  - X_i: log GSPC value on day i
  - V_i: VIX value on day i (measure of implied volatility determined by CBOE)
- **Stochastic Volatility Model:**

$$\begin{aligned} \mathrm{d}X_t &= (\alpha - \tfrac{1}{2}\tau V_t)\,\mathrm{d}t + (\tau V_t)^{1/2}\,\mathrm{d}B_{1t}, \\ \mathrm{d}V_t &= -\gamma(V_t - \mu)\,\mathrm{d}t + \sigma V_t^{\lambda}\,\mathrm{d}B_{2t}. \end{aligned}$$





Example: Stochastic Volatility Modeling

- **Data:** $\boldsymbol{Y}_i = (X_i, V_i)$: (log-GSPC, VIX) pair on day i. $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$.
- **Loglikelihood:** Sum of normal log-PDFs obtained from Markov property + Euler approximation. Thus, $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \ell(\alpha, \tau \mid \boldsymbol{Y}) + \ell(\gamma, \mu, \sigma, \lambda \mid \boldsymbol{V})$, where

$$\ell(\alpha, \tau \mid \boldsymbol{Y}) = \sum_{i=1}^{N-1} \log \varphi(\Delta X_i \mid \underbrace{(\alpha - \tfrac{1}{2}\tau V_i)\Delta t}_{\text{mean}}, \underbrace{\tau V_i \Delta t}_{\text{variance}}) \tag{1}$$

$$\ell(\gamma, \mu, \sigma, \lambda \mid \boldsymbol{V}) = \sum_{i=1}^{N-1} \log \varphi(\Delta V_i \mid \underbrace{-\gamma(V_i - \mu)\Delta t}_{\text{mean}}, \underbrace{\sigma^2 V_i^{2\lambda} \Delta t}_{\text{variance}}), \tag{2}$$

and $\varphi(x \mid \mu, \sigma^2)$ is the PDF of $x \sim \mathcal{N}(\mu, \sigma^2)$.

Example: Stochastic Volatility Modeling

- **Data:** $\boldsymbol{Y}_i = (X_i, V_i)$: (log-GSPC, VIX) pair on day i. $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$.
- **Profile Likelihood:** (1) and (2) are equivalent to two independent regression-like models:

$$
\begin{aligned}
(2) \quad \ell(\gamma, \mu, \sigma, \lambda \mid \boldsymbol{V}) &= \sum_{i=1}^{N-1} \log \varphi(\Delta V_i \mid -\gamma(V_i - \mu)\Delta t,\ \sigma^2 V_i^{2\lambda} \Delta t) \\
&= \sum_{i=1}^{N-1} \log \varphi(\Delta V_i \mid \underbrace{\gamma(-V_i \Delta t) + (\gamma\mu)\Delta t}_{\boldsymbol{\beta} = (\gamma, \gamma\mu)},\ \sigma^2 V_i^{2\lambda} \Delta t)
\end{aligned}
$$

$$
\iff \qquad \boldsymbol{y}_\lambda \sim \mathcal{N}(\boldsymbol{X}_\lambda \beta, \sigma^2 \boldsymbol{V}_\lambda), \qquad \text{where}
$$

$$
\boldsymbol{y}_\lambda = \begin{bmatrix} \Delta V_1 \\ \vdots \\ \Delta V_{N-1} \end{bmatrix}, \quad \boldsymbol{X}_\lambda = \begin{bmatrix} -V_1 & 1 \\ \vdots & \vdots \\ -V_{N-1} & 1 \end{bmatrix} \Delta t, \quad \boldsymbol{V}_\lambda = \begin{bmatrix} V_1^{2\lambda} & & 0 \\ & \ddots & \\ 0 & & V_{N-1}^{2\lambda} \end{bmatrix} \Delta t, \quad \boldsymbol{\beta} = \begin{bmatrix} \gamma \\ \gamma\mu \end{bmatrix}.
$$

Example: Stochastic Volatility Modeling

- **Data:** $\boldsymbol{Y}_i = (X_i, V_i)$: (log-GSPC, VIX) pair on day i. $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$.
- **Profile Likelihood:** (1) and (2) are equivalent to two independent regression-like models:

$$(1) \quad \ell(\alpha, \tau \mid \boldsymbol{Y}) = \sum_{i=1}^{N-1} \log \varphi(\Delta X_i \mid (\alpha - \tfrac{1}{2}\tau V_i)\Delta t,\ \tau V_i \Delta t)$$

$$= \sum_{i=1}^{N-1} \log \varphi(\Delta X_i + \tfrac{1}{2}\tau V_i \Delta t \mid \alpha \Delta t,\ \tau V_i \Delta t)$$

(since $\varphi(x \mid \mu, \sigma^2) = \varphi(x + a \mid \mu - a, \sigma^2)$)

$$\iff \qquad \boldsymbol{y}_\lambda \sim \mathcal{N}(\boldsymbol{X}_\lambda \beta, \sigma^2 \boldsymbol{V}_\lambda), \qquad \text{where}$$

$$\boldsymbol{y}_\tau = \begin{bmatrix} \Delta X_1 + \frac{1}{2}\tau V_1 \Delta t \\ \vdots \\ \Delta X_{N-1} + \frac{1}{2}\tau V_{N-1} \Delta t \end{bmatrix}, \quad \boldsymbol{X}_\tau = \begin{bmatrix} \Delta t \\ \vdots \\ \Delta t \end{bmatrix}, \quad \boldsymbol{V}_\tau = \begin{bmatrix} \tau V_1 & & 0 \\ & \ddots & \\ 0 & & \tau V_{N-1} \end{bmatrix} \Delta t, \quad \begin{aligned} \boldsymbol{\beta} &= \alpha, \\ \sigma &= 1. \end{aligned}$$

Example: Stochastic Volatility Modeling

- **Data:** $\boldsymbol{Y}_i = (X_i, V_i)$: (log-GSPC, VIX) pair on day i. $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$.
- **Profile Likelihood:** (1) and (2) are equivalent to two independent regression-like models:

$$
\begin{aligned}
(1) \quad \ell(\alpha, \tau \mid \boldsymbol{Y}) &= \sum_{i=1}^{N-1} \log \varphi(\Delta X_i \mid (\alpha - \tfrac{1}{2}\tau V_i)\Delta t,\ \tau V_i \Delta t) \\
&= \sum_{i=1}^{N-1} \log \varphi(\Delta X_i + \tfrac{1}{2}\tau V_i \Delta t \mid \alpha \Delta t,\ \tau V_i \Delta t)
\end{aligned}
$$

$$
(\text{since } \varphi(x \mid \mu, \sigma^2) = \varphi(x + a \mid \mu - a, \sigma^2))
$$

$$
\iff \qquad \boldsymbol{y}_\lambda \sim \mathcal{N}(\boldsymbol{X}_\lambda \boldsymbol{\beta}, \sigma^2 \boldsymbol{V}_\lambda), \qquad \text{where}
$$

$$
\boldsymbol{y}_\tau = \begin{bmatrix} \Delta X_1 + \frac{1}{2}\tau V_1 \Delta t \\ \vdots \\ \Delta X_{N-1} + \frac{1}{2}\tau V_{N-1}\Delta t \end{bmatrix}, \quad \boldsymbol{X}_\tau = \begin{bmatrix} \Delta t \\ \vdots \\ \Delta t \end{bmatrix}, \quad \boldsymbol{V}_\tau = \begin{bmatrix} \tau V_1 & & 0 \\ & \ddots & \\ 0 & & \tau V_{N-1} \end{bmatrix} \Delta t, \quad \begin{aligned} \boldsymbol{\beta} &= \alpha, \\ \sigma &= 1. \end{aligned}
$$

$\implies \ell_{\text{prof}}(\lambda, \tau \mid \boldsymbol{Y})$ reduces 6-d optimization to 2-d.

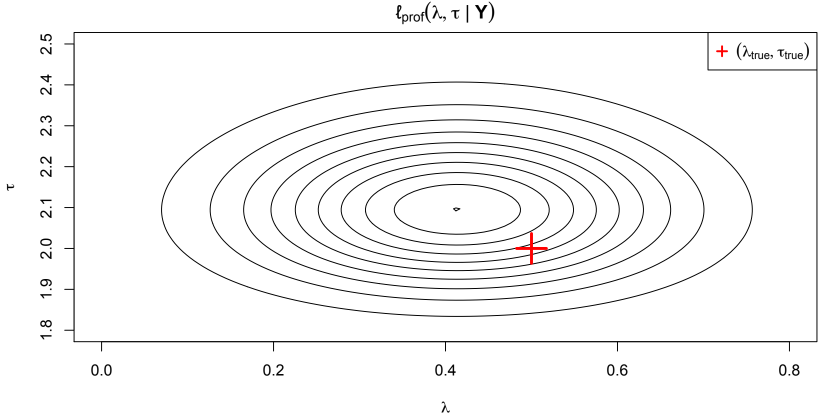
Example: Stochastic Volatility Modeling

- **Stochastic Volatility Model:**

$$\mathrm{d}X_t = (\alpha - \tfrac{1}{2}\tau V_t)\,\mathrm{d}t + (\tau V_t)^{1/2}\,\mathrm{d}B_{1t}$$
$$\mathrm{d}V_t = -\gamma(V_t - \mu)\,\mathrm{d}t + \sigma V_t^{\lambda}\,\mathrm{d}B_{2t}.$$

- **Simulation:**

Parameter	α	γ	μ	σ	λ	τ	X_0	V_0	N	Δt
Value	.1	.5	.2	.3	.5	2	6.5	.2	500	1/252



Profile Likelihood (Continued)

Confidence Intervals

- **Loglikelihood:** $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y})$
- **Profile Likelihood:** For $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\phi})$,

$$\ell_{\mathsf{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y}) = \ell(\boldsymbol{\eta}, \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} \mid \boldsymbol{Y}), \qquad \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\phi}} \ell(\boldsymbol{\eta}, \boldsymbol{\phi} \mid \boldsymbol{Y}).$$

- **MLE:** $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\eta}}})$, where $\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \ell_{\mathsf{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y})$.
- **Confidence Intervals:** Suppose $\boldsymbol{\theta} = (\eta, \boldsymbol{\phi})$. To construct the confidence interval $\hat{\eta} \pm 1.96 \cdot s_{\hat{\eta}}$, two options for the standard error $s_{\hat{\eta}}$:

  1. Use *full likelihood* Fisher Information: $\hat{\boldsymbol{\mathcal{I}}} = -\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{Y})$, $\quad s_{\hat{\eta}} = \sqrt{[\hat{\boldsymbol{\mathcal{I}}}^{-1}]_{11}}$
  2. Use *profile likelihood* Fisher Information: $\hat{\mathcal{I}}_{\mathsf{prof}} = -\frac{\mathrm{d}^2}{\mathrm{d}\eta^2} \ell_{\mathsf{prof}}(\hat{\eta} \mid \boldsymbol{Y})$, $\quad s_{\hat{\eta}} = \hat{\mathcal{I}}_{\mathsf{prof}}^{-1/2}$

  Question: Which to use? Does it matter?

Profile Likelihood: Confidence Intervals
Example

- **Model:** $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Weibull}(\gamma, \lambda)$, where

$$X \sim \text{Expo}(1) \implies Y = \lambda X^{1/\gamma} \sim \text{Weibull}(\gamma, \lambda).$$

- **Loglikelihood:**

$$\ell(\gamma, \lambda \mid \boldsymbol{Y}) = n\big[\log(\gamma) - \gamma \log(\lambda)\big] + \gamma \sum_{i=1}^{n} \log Y_i - \lambda^{-\gamma} \sum_{i=1}^{n} Y_i^{\gamma}.$$

- **Profile Likelihood:** $\hat{\lambda}_\gamma = \left(\frac{1}{n} \sum_{i=1}^{n} Y_i^{\gamma}\right)^{1/\gamma}.$

Profile Likelihood: Confidence Intervals

- **Profile Likelihood:** For $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y})$ and $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\phi})$, $\boldsymbol{\eta} \in \mathbb{R}^q$, $\boldsymbol{\theta} \in \mathbb{R}^p$, $q < p$,

$$\ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y}) = \ell(\boldsymbol{\eta}, \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} \mid \boldsymbol{Y}), \qquad \hat{\boldsymbol{\phi}}_{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\phi}} \ell(\boldsymbol{\eta}, \boldsymbol{\phi} \mid \boldsymbol{Y}).$$

$\implies$ MLE is $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\phi}}_{\hat{\boldsymbol{\eta}}})$, where $\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \ell_{\text{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y})$.

- **Confidence Intervals:** To construct variance estimate $\widehat{\text{var}}(\hat{\boldsymbol{\eta}}) \approx \text{var}(\hat{\boldsymbol{\eta}})$:

  1. *Full likelihood* Fisher Information: denoting $\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{ab}} = -\frac{\partial^2}{\partial \boldsymbol{a} \partial \boldsymbol{b}} \ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{Y})$, $\boldsymbol{a}, \boldsymbol{b} \in \{\boldsymbol{\eta}, \boldsymbol{\phi}\}$,

$$\hat{\boldsymbol{\mathcal{I}}} = \begin{bmatrix} \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta\eta}} & \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta\phi}} \\ \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\phi\eta}} & \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\phi\phi}} \end{bmatrix} \qquad \implies \qquad \begin{aligned} \widehat{\text{var}}_{\text{full}}(\hat{\boldsymbol{\eta}}) &= \text{top left } q \times q \text{ corner of } \hat{\boldsymbol{\mathcal{I}}}^{-1} \\ &= [\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta\eta}} - \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta\phi}} \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\phi\phi}}^{-1} \hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\phi\eta}}]^{-1}. \end{aligned}$$

  2. *Profile likelihood* Fisher Information: $\widehat{\text{var}}_{\text{prof}}(\hat{\boldsymbol{\eta}}) = \left[-\frac{\partial^2}{\partial \boldsymbol{\eta}^2} \ell_{\text{prof}}(\hat{\boldsymbol{\eta}} \mid \boldsymbol{Y})\right]^{-1}$.

- **Theorem:** $\widehat{\text{var}}_{\text{full}}(\hat{\boldsymbol{\eta}}) = \widehat{\text{var}}_{\text{prof}}(\hat{\boldsymbol{\eta}})$.

  $\implies$ If $\boldsymbol{\phi}$ are nuisance parameters, i.e., only $\boldsymbol{\eta}$ are parameters of interest, then profile likelihood is more efficient for calculating both MLE and confidence intervals for $\boldsymbol{\eta}$.

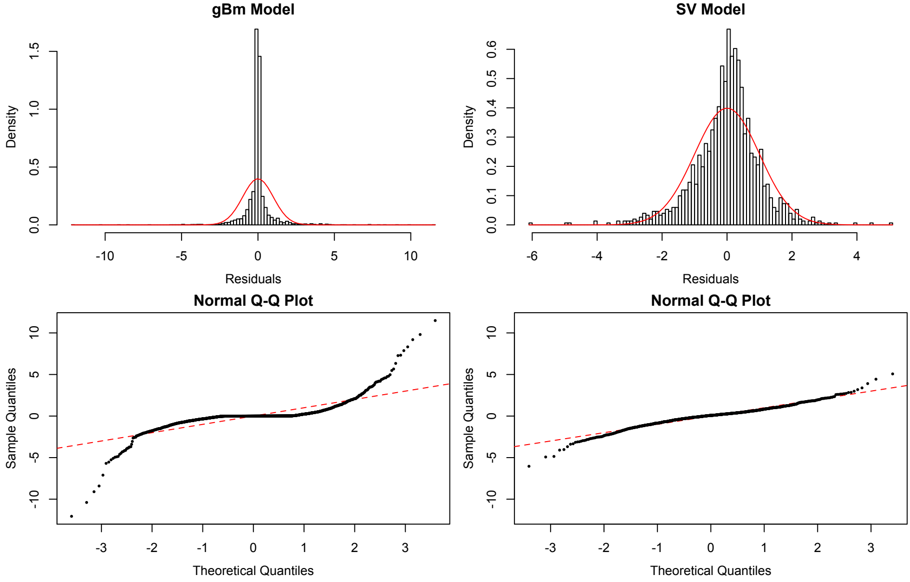
Stochastic Volatility Modeling (Continued)

Goodness-of-Fit

- **Data:** $\boldsymbol{Y}_i = (X_i, V_i)$: (log-GSPC, VIX) pair on day i. $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N)$.
- **Stochastic Volatility Model:**
$$\begin{aligned} \mathrm{d}X_t &= (\alpha - \tfrac{1}{2}\tau V_t)\,\mathrm{d}t + (\tau V_t)^{1/2}\,\mathrm{d}B_{1t} \\ \mathrm{d}V_t &= -\gamma(V_t - \mu)\,\mathrm{d}t + \sigma V_t^{\lambda}\,\mathrm{d}B_{2t}. \end{aligned}$$
- **Euler Approximation:** Given an initial value $\boldsymbol{Y}_0$, data simulated as
$$\Delta X_n = \underbrace{(\alpha - \tfrac{1}{2}\tau V_n)\Delta t}_{\mu_{1n}(\boldsymbol{\theta})} + \underbrace{(\tau V_n)^{1/2}}_{\sigma_{1n}(\boldsymbol{\theta})}\Delta B_{1n}, \qquad \Delta V_n = \underbrace{-\gamma(V_n - \mu)\Delta t}_{\mu_{2n}(\boldsymbol{\theta})} + \underbrace{\sigma V_n^{\lambda}}_{\sigma_{2n}(\boldsymbol{\theta})}\Delta B_{2n}.$$
- **Model Residuals:** Assuming Euler approximation, define
$$Z_{1n}(\boldsymbol{\theta}) = \frac{\Delta X_n - \mu_{1n}(\boldsymbol{\theta})}{\sigma_{1n}(\boldsymbol{\theta})\Delta t^{1/2}} = \frac{\Delta B_{1n}}{\Delta t^{1/2}}, \qquad Z_{2n}(\boldsymbol{\theta}) = \frac{\Delta V_n - \mu_{2n}(\boldsymbol{\theta})}{\sigma_{2n}(\boldsymbol{\theta})\Delta t^{1/2}} = \frac{\Delta B_{2n}}{\Delta t^{1/2}}.$$

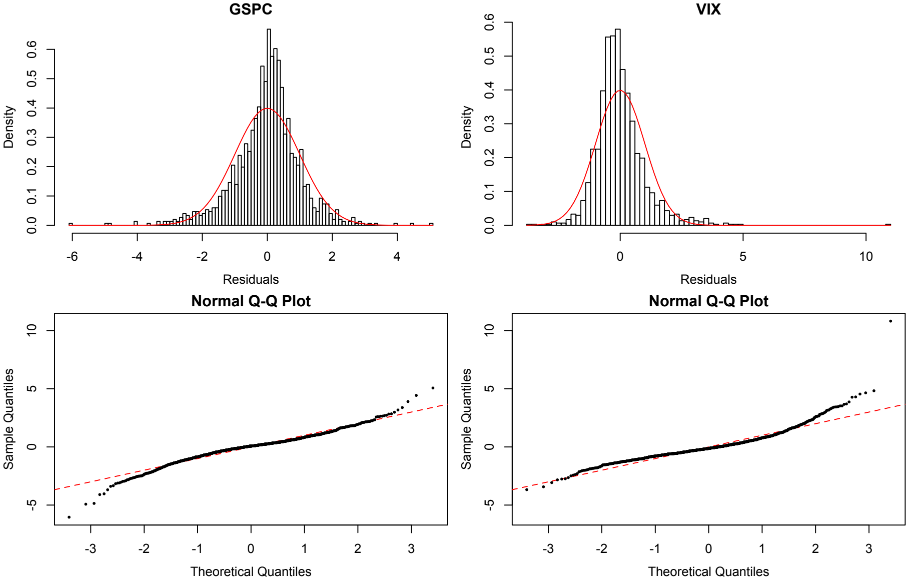
  By independent increments of Brownian motion, $Z_{in}(\boldsymbol{\theta}) \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Defining the residuals $\hat{Z}_{in} = Z_{in}(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is the MLE, approximately we have $\hat{Z}_{in} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

Stochastic Volatility: Goodness-of-Fit



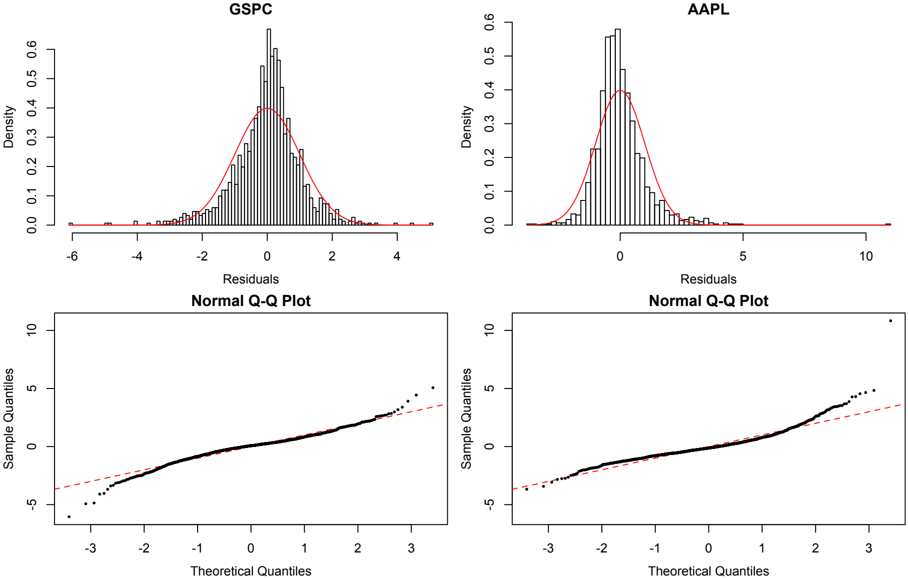
GSPC residuals for gBM and SV models.

Stochastic Volatility: Goodness-of-Fit



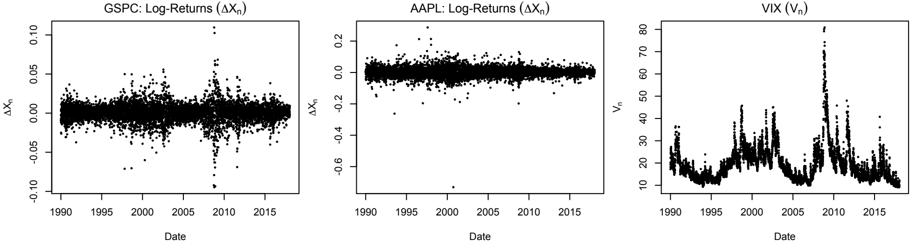
GSPC vs. VIX residuals. (Usually volatility is harder to model than price.)

Stochastic Volatility: Goodness-of-Fit



GSPC vs. AAPL residuals. (VIX is better proxy for volatility of aggregate than individual asset.)

Stochastic Volatility: Goodness-of-Fit



GSPC, AAPL, and VIX time series. (Huge drop in AAPL circa 2001 is due to 7-way stock split.)

Resources

- `quantmod`: R package to easily scrape web for financial data.
- `nloptr`: An R wrapper to `NLopt`, to nloptr an extremly powerful C++ library for nonlinear (possibly constrained) optimization.
- `LMN`: R package for profiling matrix-normal regression models of the form

  $$\boldsymbol{Y} \sim \mathrm{MatNorm}(\boldsymbol{X}_{\boldsymbol{\eta}}\beta, \boldsymbol{V}_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}),$$

  that is, for efficiently calculating $\ell_{\mathrm{prof}}(\boldsymbol{\eta} \mid \boldsymbol{Y})$ and $\hat{\phi}_{\boldsymbol{\eta}} = (\hat{\beta}_{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}})$.
- `TMB`: R wrapper to C++ library for automatic differentiation (autodiff), i.e., automatically get $\nabla \ell(\boldsymbol{\theta} \mid \boldsymbol{y})$ from $\ell(\boldsymbol{\theta} \mid \boldsymbol{y})$ without extra programming. `TMB` is extremely fast, well-documented, and the best way to do autodiff in R.
- `Autograd`: Excellent autodiff library for (pure) Python. An even faster (but less mature) library for this is `JAX`.