Semiparametric Multivariate Volatility Models

Christian M. Hafner

Econometric Institute
Erasmus University Rotterdam

June 2005

Introduction

Consider the model for a univariate time series $\{y_t\}, t \in \mathbb{Z}$,

$$y_t = \mu_t + \sigma_t \xi_t$$

where $\xi_t \sim i.i.d.(0,1)$, and μ_t and σ_t are measurable w.r.t. the information set at time $t-1$, $\mathcal{F}_{t-1}$.

In finance, a popular model for σ_t is GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

where $\varepsilon_t = y_t - \mu_t$, and $\omega > 0, \alpha, \beta \geq 0$ are parameters (Engle, 1982, Bollerslev, 1986)

Two types of semiparametric GARCH:

1. Keep the parametric form of σ_t but let the density of ξ_t be nonparametric: Engle and Gonzalez-Rivera (1991), Drost and Klaassen (1997) **(Type I)**

2. Specify a parametric density for ξ_t but model σ_t nonparametrically: e.g. Linton and Mammen (2005) **(Type II)**

This talk

discusses generalizations of semiparametric GARCH to the multivariate case for both, Type I and Type II models

First: Type I

C.M. Hafner and J.V.K. Rombouts (2005), Semiparametric multivariate volatility models, under revision for *Econometric Theory*.

A multivariate volatility model

Consider the model for the $N \times 1$ vector series $\{\varepsilon_t\}$,

$$\varepsilon_t = H_t(\theta)\xi_t$$

where $H_t(\theta)$ is an invertible matrix that is measurable w.r.t. $\mathcal{F}_{t-1}$, and $\theta \in \Theta \subset \mathbb{R}^K$

Let $\{\xi_t\}$ be i.i.d. with $E[\xi_t] = 0$, $E[\xi_t \xi_t'] = I_N$

Then, $\mathsf{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ and $\Sigma_t = \mathsf{V}[\varepsilon_t \mid \mathcal{F}_{t-1}] = H_t H_t'$

Estimation strategies

Maximum likelihood estimation maximizes the log likelihood function

$$L(\theta) = \sum_{t=1}^{n} l_t(\theta) = -\sum_{t=1}^{n} \ln \mathrm{abs}|H_t(\theta)| + \ln g(H_t^{-1}(\theta)\varepsilon_t)$$

where g is the density of the innovations v_t

1. ML: Assume a specific parametric distribution g
2. QML: Assume g is Gaussian
3. SP: Let g be nonparametric, with two possibilities:
   - g is only satisfying smoothness conditions
   - g is in a restricted class, e.g. spherical

Maximum Likelihood

One assumes a certain parametric distribution g. Under correct specification (and regularity conditions), parameter estimates $\tilde{\theta}$ are consistent and

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, V_{ml})$$

with $V_{ml} = \mathcal{I}^{-1}$, and

$$\mathcal{I} = \mathsf{E}\left[\frac{\partial l_t}{\partial \theta}\frac{\partial l_t}{\partial \theta'}\right]$$

Drawback: under misspecification, consistency is no longer retained in general.

Quasi Maximum Likelihood

The Gaussian likelihood is given by

$$l_t^{qml}(\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_t(\theta)| - \frac{1}{2}\varepsilon_t'\Sigma_t^{-1}(\theta)\varepsilon_t$$

- Advantage: QML provides consistent estimates even if the likelihood is misspecified (Bollerslev and Wooldridge, 1992)
- Disadvantage: If misspecified, then efficiency loss compared to ML can be substantial
- Proofs for particular multivariate GARCH models were provided by Jeantheau (1998, consistency) and Comte and Lieberman (2001, asymptotic normality)

QML asymptotics

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, V_{qml})$$

with $V_{qml} = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$, and

$$\mathcal{J} = -\mathsf{E}\left[\frac{\partial^2 l_t^{qml}}{\partial\theta\partial\theta'}\right], \quad \mathcal{I} = \mathsf{E}\left[\frac{\partial l_t^{qml}}{\partial\theta}\frac{\partial l_t^{qml}}{\partial\theta'}\right]$$

If g is indeed multinormal, then $\mathcal{J} = \mathcal{I}$, and $V_{qml} = \mathcal{I}^{-1}$ as for ML

Semiparametric ML

- Idea: regain efficiency while keeping consistency
- Fitting the model into the general time series framework of Drost, Klaassen and Werker, we can give conditions for Local Asymptotic Normality (LAN)
- We show that estimators can be constructed that attain the semiparametric lower bound
- This can be done by a particular one-step Newton-Raphson improvement of a consistent estimator $\tilde{\theta}$ (e.g. QML)

A useful factorization

The score vector takes the form

$$\dot{l}_t(\theta) = W_t \psi_t$$

where

$$W_t = W_t(\theta) = \frac{\partial \text{vec}(H_t)'}{\partial \theta}(I_N \otimes H_t'^{-1})$$

and

$$\psi_t = \psi_t(\xi_t(\theta)) = -\text{vec}\left(I_N + \frac{\partial \log g(\xi_t)}{\partial \xi_t}\xi_t'\right)$$

Let $M_{\psi\psi} = \mathsf{E}[\psi_t \psi_t']$ and assume $M_{\psi\psi} < \infty$

The tangent set

Define the set

$$\mathcal{T} = \{f : I\!R^N \rightarrow I\!R^K \mid \mathsf{E}[f(\xi_t)] = 0, \mathsf{E}[f(\xi_t)F_t'] = 0, \mathsf{E}[f(\xi_t)f(\xi_t)'] < \infty\}$$

where $F_t = (\xi_t, \text{vech}(\xi_t\xi_t' - I_N))$

This is the infinite dimensional Hilbert space containing all linear combinations of the score w.r.t. the 'nuisance parameter' g

Orthogonality to F_t is required because ξ_t is standardized (mean zero, variance identity)

The efficient score

Define

$$\dot{\ell}_t^*(\theta) = \dot{\ell}_t(\theta) - P_t(\theta)$$

where

$$P_t(\theta) = \mathcal{P}(\dot{l}_t(\theta) \mid \mathcal{T}) = E[W_t(\theta)] \left(\psi_t - M_{\psi F} M_{FF}^{-1} F_t\right).$$

and $M_{\psi F} = \mathsf{E}[\psi_t F_t'], M_{FF} = \mathsf{E}[F_t F_t']$.

The semiparametric estimator

Then the efficient semiparametric estimator is defined by

$$\hat{\theta} = \tilde{\theta} + \left(\sum_{t=1}^{n} \dot{\ell}_t^*(\tilde{\theta}) \dot{\ell}_t^{*\prime}(\tilde{\theta}) \right)^{-1} \sum_{t=1}^{n} \dot{\ell}_t^*(\tilde{\theta})$$

where $\dot{\ell}_t^*(\tilde{\theta})$ is replaced by a consistent estimator based on $W_t(\tilde{\theta})$ and $\hat{\psi}_t$, a nonparametric estimator of $\psi_t(\xi_t(\tilde{\theta}))$

It then can be shown that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, V_{sp})$$

with $V_{sp} = \mathsf{E}[\dot{\ell}_t^* \dot{\ell}_t^{*\prime}]^{-1}$

Furthermore,

$$V_{ml} \leq V_{sp} \leq V_{qml}$$

In particular,

$$V_{ml}^{-1} - V_{sp}^{-1} = \mathsf{E}[W_t(\theta)]Q\mathsf{E}[W_t(\theta)]'$$

where

$$Q = M_{\psi\psi} - M_{\psi F} M_{FF}^{-1} M_{F\psi}$$

Adaptive estimation

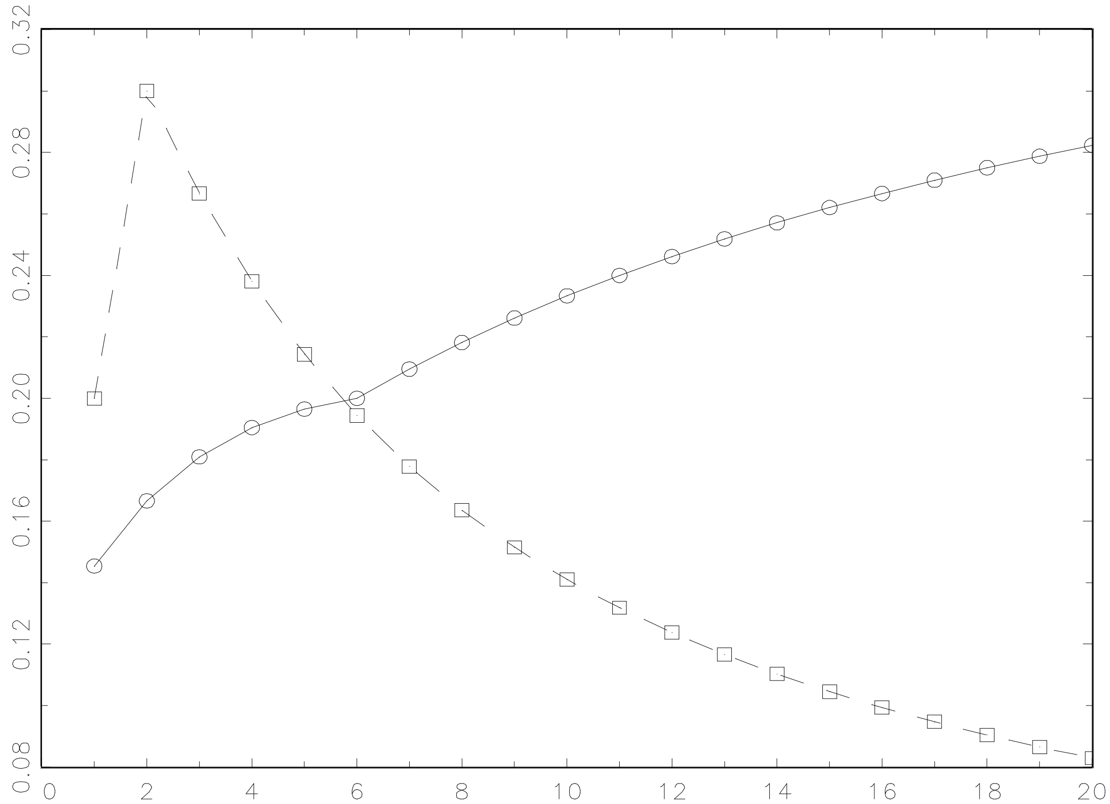
- If $V_{ml} = V_{sp}$, we say that the semiparametric estimator is adaptive.
- This happens if and only if $Q = 0$, and we show that this is equivalent to g being Gaussian. Hence, adaptive estimation is not possible.
- The spectral norm $\rho(Q)$ measures the distance from adaptivity and can be calculated for particular distributions.

Spectral norm of Q

(Measure of distance between V_{ml} and V_{sp})

Distribution	$N = 1$	$N = 2$	$N = 3$
t ($\nu = 5$)	0.7500	0.8889	0.9333
t ($\nu = 8$)	0.3117	0.3333	0.3590
t ($\nu = 12$)	0.1454	0.1667	0.1810
Laplace	0.2000	0.3000	0.2667
ES Logistic	11.4005	0.9629	0.7816

Spectral norm of Q as a function of N, for Laplace and t_{12}



Further issues

- In practice, restrictions of the class of densities may be desired if N is large
- For example, spherical distributions (dealt with in the paper)
- Possible future work: copula model for g

‘Type II’:

C.M. Hafner, D.J. van Dijk and Ph.H. Franses, Semiparametric Modelling of Correlation Dynamics, *Advances in Econometrics*, forthcoming 2005.

Parametric multivariate GARCH

- Vec model (Bollerslev, Engle and Wooldridge, 1988)
  - BEKK (Engle and Kroner, 1995)
  - Factor models (Diebold and Nason, 1989, Engle, Ng and Rothshild, 1991)
  - O-GARCH, GO-GARCH (Alexander, 2001, van der Weide, 2003)
- DCC type models
  - CCC (Bollerslev, 1990)
  - Standard DCC and extensions (Engle, 2002; Tse and Tsui, 2003)
  - Our semiparametric model

Problems of parametric models

- Difficult to balance parsimony with flexibility if the dimension is large
- Likelihood function often ill-conditioned
- This results in long estimation times and numerically unstable solutions
- Parameters difficult to interpret

Goals of this paper

- to provide an easy-to-estimate but sufficiently flexible model for large dimensions
- to answer important questions such as
  - Are conditional correlations constant over time?
  - Are conditional correlations higher in bear markets than in bull markets?
  - Are conditional correlations higher in states of high volatility than in states of low volatility?
- use one observable factor on which conditional correlations depend.

Correlation asymmetries

- increasing literature, e.g. Ang and Chen (2002), Cappiello et al. (2003), Longin and Solnik (2001)
- Often found empirically: ‘correlation breakdowns’
- However: may be spurious if correlations are compared over subperiods
- We need a model that nests the CCC model as special case and allows for testing against it.

Dynamic correlation models

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \Sigma_t), \quad \Sigma_t = D_t(\theta) R_t D_t(\theta),$$

where $D_t(\theta) = \mathsf{diag}(\sigma_{it})_{i=1}^N$ (e.g. univariate GARCH)

Written elementwise:

$$\Sigma_{ij,t} = \sigma_{it}\sigma_{jt}R_{ij,t}$$

and $R_{ij,t}$ is the conditional correlation of $(\varepsilon_{it}, \varepsilon_{jt})$.

The simplest model for R_t is to assume that it is constant, $R_t = R$ (CCC model, Bollerslev, 1990)

A semiparametric conditional correlation model

- Use an observable factor x_t such as the (lagged) market index, volatility or simply time
- First stage: obtain standardized residuals $\hat{v}_{it} = \varepsilon_{it}/\sigma_{it}(\hat{\theta})$ using a $\sqrt{T}$-consistent estimator of θ.
- Second stage:

$$\widehat{R}_{ij}(x) = \frac{\widehat{Q}_{ij}(x)}{\sqrt{\widehat{Q}_{ii}(x)\widehat{Q}_{jj}(x)}}$$

where

$$\widehat{Q}_{ij}(x) = \frac{\sum_{t=1}^{T} \hat{v}_{it}\hat{v}_{jt}K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)}$$

and $K_h(\cdot) = (1/h)K(\cdot/h)$, K is a kernel function and $h > 0$ a bandwidth

Comments

- This is a transformation of a Nadaraya-Watson kernel estimator
- Other nonparametric estimators such as local polynomials or splines could be used as well.
- If the bandwidth h is the same for all i, j, then $\widehat{R}$ is positive definite.

Some theory

- Under $h \to 0$ as $T \to \infty$ such that $Th^5 \to 0$, some standard assumptions on the kernel and weak temporal dependence of $\{\varepsilon_t\}$ (e.g. α-mixing) we can show that

$$\widehat{Q}_{ii}(x) \xrightarrow{p} 1, \quad i = 1, \ldots, N$$

- Hence, $\widehat{Q}$ and $\widehat{R}$ are asymptotically equivalent.

- In finite samples, $\widehat{R}(x)$ has the advantage of being a correlation matrix, whereas $\widehat{Q}(x)$ may not be one.

Some more theory

- Writing $\eta_t = \text{vech}(v_t v_t')$, $r(x) = \text{vech}(R(x))$ and $\hat{r}(x) = \text{vech}(\hat{R}(x))$, we can show

$$\sqrt{Tb}(\widehat{r}(x) - r(x)) \xrightarrow{\mathcal{L}} N(0, \Sigma(x))$$

  where

$$\Sigma(x) = \frac{\int K^2(u)du}{f(x)} \mathsf{V}(\eta_t \mid x_t = x)$$

- Under the conditional normality assumption, explicit expressions can be found for $\mathsf{V}(\eta_t \mid x_t = x)$

- However, without this assumption a consistent estimator can be derived:

An estimator of $\Sigma(x)$

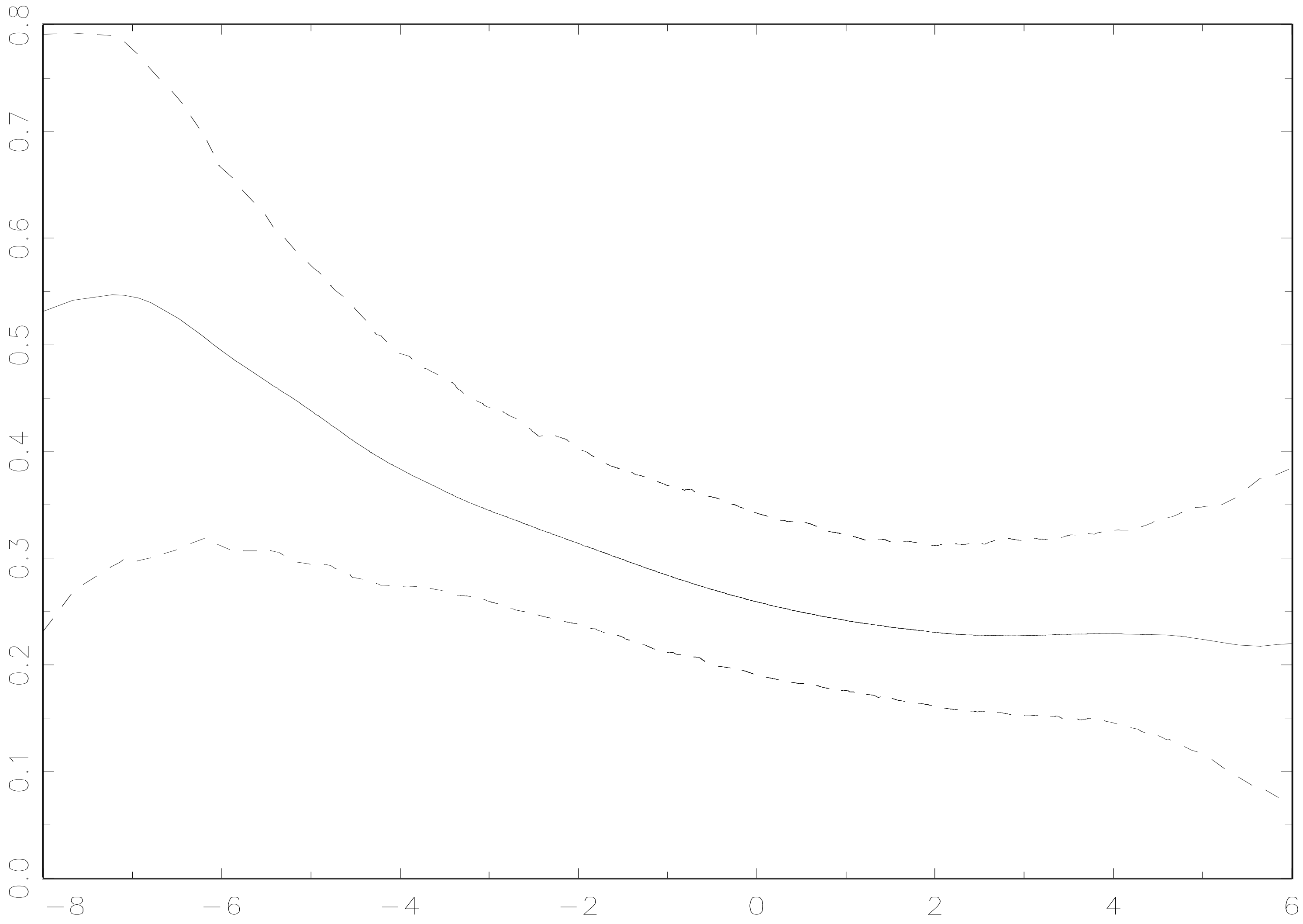
- A consistent estimator of the asymptotic covariance is given by

$$\widehat{\Sigma}(x) = \frac{\int K^2(u)du}{\hat{f}(x)} \frac{\sum_{t=1}^{T}(\eta_t - \widehat{r}(x))(\eta_t - \widehat{r}(x))' K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)}$$

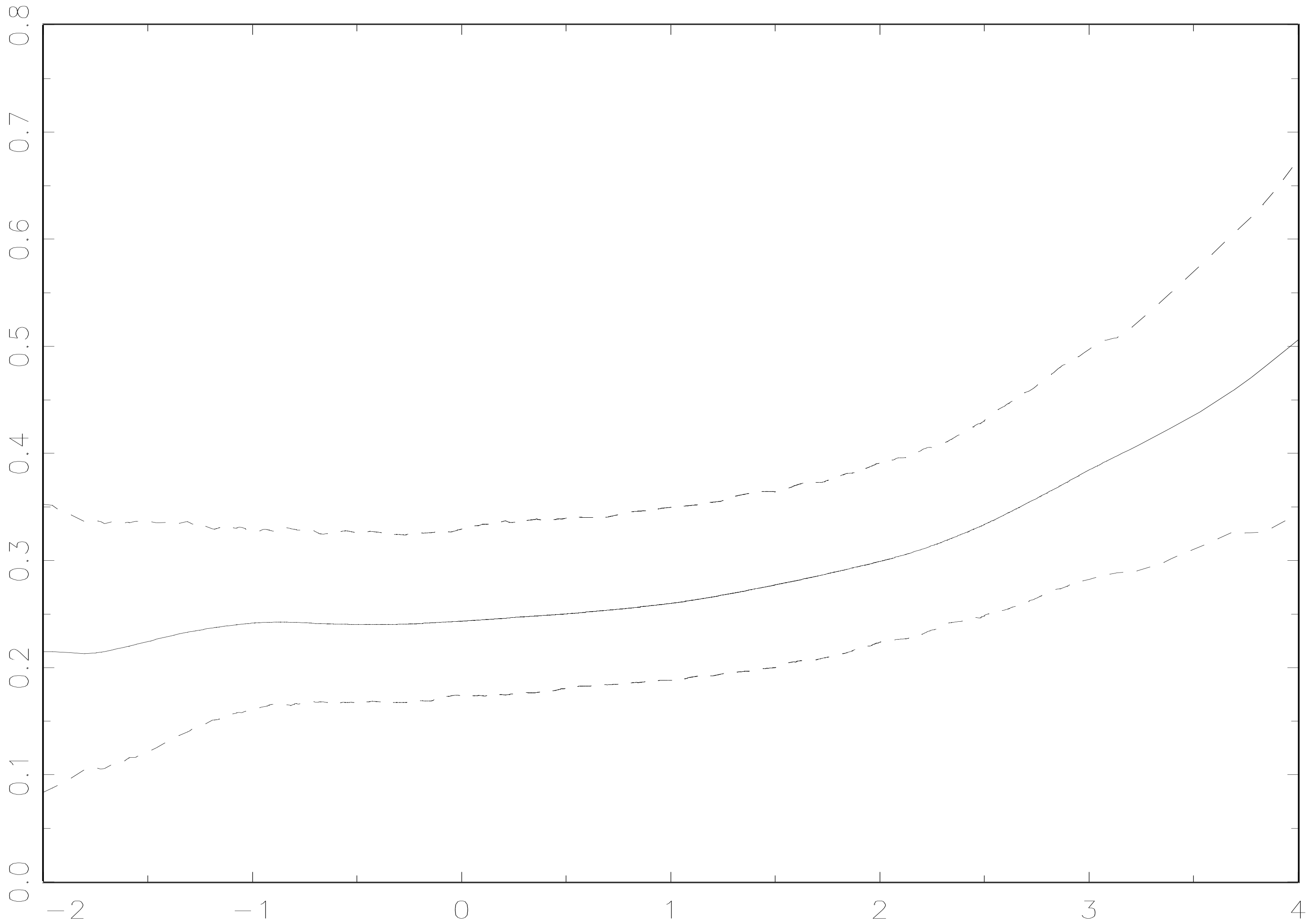
An empirical application

- In the paper, we provide an extensive comparison of various multivariate volatility models, applied to the 30 stocks of the DOW Jones IA, from Jan 1, 1989 until Dec 31, 2003. Sample size: $T = 3784$ trading days.
- As performance criteria we use likelihood, Value-at-Risk forecasts, and prediction of particular portfolio variances (e.g. the minimum-variance portfolio)
- Here, I will just show two typical examples of estimated correlation functions, one with market return, the other with market volatility as factor

Example: AT&T and Merck, Market return factor



Example: AT&T and Merck, Market volatility factor



Results

- Correlations typically increase in market volatility, and increase stronger in bear markets than in bull markets.
- But there are some pairs of stocks for which this does not hold. This information could be used by portfolio managers.

Conclusions and Outlook

- Type I and Type II semiparametric multivariate volatility models provide a good compromise between flexibility and efficiency
- We have a simple, sufficiently flexible Type II model that provides a good fit to the data
- Estimation time is far below that of even the simplest DCC model
- Future: develop more guidance with respect to bandwidth selection
- find ways to avoid the curse of dimensionality when more than one factor is considered