Nonparametric estimation of volatility models with serially dependent innovations

by

Christian M. Dahl
Department of Economics
Purdue University

and

Michael Levine
Department of Statistics
Purdue University

The "general" model of interest

$$y_t = g(\boldsymbol{x}_t) + \sigma(\boldsymbol{x}_t)\epsilon_t, \qquad t = 1, 2, ..., T. \tag{1}$$

- $y_t \in \mathbb{R},$ and $\boldsymbol{x}_t \in \mathbb{R}^k,$ and $\{(y_t, \boldsymbol{x}_t)\}$ is a strictly stationary process.

- $\mathsf{E}\left(y_t|\boldsymbol{x}_t\right) = g(\boldsymbol{x}_t)$ and $g : \mathbb{R}^k \to \mathbb{R}.$

- $\mathsf{var}\left(y_t|\boldsymbol{x}_t\right) = \sigma(\boldsymbol{x}_t)^2$ and $\sigma : \mathbb{R}^k \to \mathbb{R}_+.$

- ϵ_t is a strictly stationary process with $\mathsf{E}\left(\epsilon_t\right) = 0$ and $\mathsf{var}\left(\epsilon_t\right) = 1.$

1 Applications (short term interest rates)

The ARCH(-M) specification:

- $y_t = \Delta r_t$ and $\boldsymbol{x}_t = (1, \Delta r_{t-1}, \Delta r_{t-2}, ..., \Delta r_{t-k-1})'$.
- $g(\boldsymbol{x}_t) = \boldsymbol{x}_t'\boldsymbol{\beta} \qquad \left(+\sigma(\boldsymbol{x}_t)^2\right)$.
- $\sigma(\boldsymbol{x}_t)^2 = \alpha_0 + \alpha_1 \Delta r_{t-1}^2 + ... + \alpha_p \Delta r_{t-p}^2$.
- $\epsilon_t \sim i.i.d.$

The one-factor diffusion model (cont. time):

- Chan, Karolyi, Longstaff and Saunders (1992) generalized short rate specification

$$dr = (\alpha + \beta r)\, dt + \psi r^{\gamma} dW_t. \qquad (2)$$

- Euler-Maruyama discrete time approximation of (2)

$$\begin{aligned} \Delta r_t &= \alpha + \beta r_{t-1} + \psi r_{t-1}^{\gamma} \epsilon_t, \\ \mathsf{E}\left(\epsilon_t | F_{t-1}\right) &= 0, \qquad \mathsf{E}\left(\epsilon_t^2 | F_{t-1}\right) = 1. \end{aligned}$$

- Brenner, Harjes and Kroner (1996) and Andersen and Lund (1997) suggest

$$\psi_t^2 = \omega + \theta \Delta r_{t-1}^2 + \phi \psi_{t-1}^2.$$

The Markov regime switching model

- Hamilton and Susmel (1994), Cai (1994) and Gray (1996)

$$\begin{aligned} \Delta r_t | F_{t-1} &\sim \begin{cases} \mathsf{N}\left(g_{1t}, \sigma_{1t}^2\right) & \text{w.p. } p_{1t} \\ \mathsf{N}\left(g_{2t}, \sigma_{2t}^2\right) & \text{w.p. } 1 - p_{1t} \end{cases}, \\ p_{1t} &= \mathsf{Pr}\left(S_t = 1 | F_{t-1}\right) \end{aligned}$$

Focus of the paper

- Consider the following process for the time series of interest denoted $y_t \in \mathbb{R}$, $t = 1, 2, ..., T$ where

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, & (3) \\ \epsilon_t &= \phi_0 \epsilon_{t-1} + v_t, & (4) \end{aligned}$$

  - $v_t \sim i.i.d.\,(0,1)$, $\mathsf{E}\left(|v_t|^{l+\gamma}\right) < \infty$ for $l = 1, ..., 4$ and $\gamma > 0$
  - $\phi_0 \in \Theta = (-1; 1)$,
  - $\sigma_t^2 \equiv \sigma(x_t)^2 \in \mathcal{F} = C^2[0,1]$ and $P\left(\sigma_t^2 > 0\right) = 1$
  - ϵ_t is a $\alpha-$mixing sequence with mixing coefficient $-\left(1 + 2/\delta\right)$ for $\delta > 0$.

Econometric/economic implications

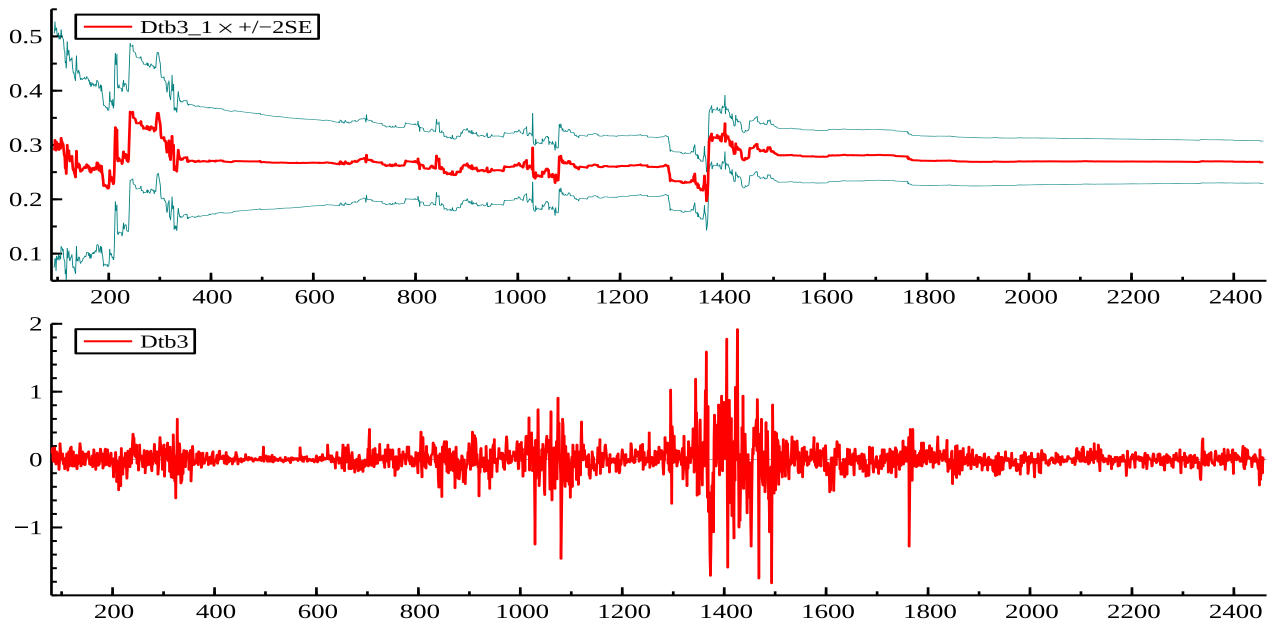
- A time varying (functional) coefficient representation of (3)-(4)

$$y_t = \frac{\sigma_t}{\sigma_{t-1}} \phi y_{t-1} + \sigma_t v_t. \tag{5}$$

- Change in relative volatility $\left(\frac{\sigma_t}{\sigma_{t-1}}\right) \rightarrow$ change in expected return $(\mathsf{E}\left(y_t | y_{t-1}\right))$.

- Direction of this effect depends on sgn(ϕ) (LEVELS) or sgn(ϕy_{t-1})

Figure 1: Modelling the change (w/w) in the 3 month US t-bill ($Dtb3$). Stability of ρ in regression $Dtb3_t = \gamma + \rho Dtb3_{t-1} + \sigma_t v_t$.



- Ignoring $\phi \neq 0$ can result in "poor" second-step estimators of σ_t^2.
  - Estimation properties of σ_t^2 relies on consistency of $\widehat{g}\,(x_t)$.
  - Halunga and Orme (2004): Discuss the asymptotic sensitivity of (G)ARCH estimation in the case of misspecification of $g(x_t)$.
  - Hillebrand (2005): Neglecting parameter changes in (G)ARCH model implies that "the sum of the estimated autoregressive parameters is heavily biased towards unity in finite samples."

- $\left|\sqrt{\frac{\sigma_t^2}{\sigma_{t-1}^2}}\phi\right| > 1$ for some $t \rightarrow$ occasional explosive behavior of y_t.
  - Granger (2005) and Yoon (2003): Models with a similar property (STochastic Unit Roots, Granger and Swanson (1997) and Leybourne et al. (1996a,b)) can generate return series that resemble the behavior of returns (r_t) surprisingly well.
    * r_t have little or no serial correlation
    * The autocorrelations of r_t^2 and $|r_t|^d$ decline slowly with slowest decline for $d = 1$ (Taylor effect).
    * Autocorrelations of sgn(r_t) are all small, insignificant.
    * If one fits a GARCH(1,1) model to the series, then "$\alpha + \beta \approx 1$"

Two Cases of Interest

- Case 1: x_t has a fixed design on the unit interval.

- Case 2: x_t is a strictly stationary process with an $\alpha-$mixing base.

  - Case 2 encompasses the situation where $x_t = y_{t-1}$, hence allows for ARCH/LEVEL effects.

- Main simplifying assumptions

  - Simple AR(1) innovation term structure.

  - The conditional mean function is ignored, i.e., $g(x_t) \equiv 0$.

Remarks

- The model of interest can be viewed as a nonparametric regression with correlated innovation terms.

- Some existing results (e.g., Fan and Masry (1993), Tong (1990), Fan, Yao and Tong(1996)) concerning the estimation of the conditional mean, i.e., $g(x_t) = \mathsf{E}(y_t|x_t)$.

  - Assume that the process satisfies certain mixing/memory conditions. Then

    * The local linear/polynomial estimator $\hat{g}(x_t)$ is asymptotically consistent.

* It is asymptotically normal.

* The rate of convergence in mean squared is $O\left(\frac{1}{Th_T}\right)$, for any bandwidth $h_T = O\left(\frac{1}{T^{2p+3}}\right)$.

- Similar results on the estimation of the conditional variance, $\sigma_t^2 = \text{var}(y_t|x_t)$, is unknown to us!

  – Most general results available assume the innovation process to be a martingale difference sequence (series of recent working papers by O. Linton and coauthors) .

Goals

- Estimate variance-covariance structure of the model

- Provide asymptotic characterization of resulting estimators: all parametric as well as nonparametric estimators of the model.

Characterization of $\widehat{\phi}$ (MINPIN) and its asymptotics

- Let $\widehat{\sigma}_t^2$ denote a (preliminary nonparametric) estimator of σ_t^2.

- Then we can define

$$\widehat{\epsilon}_t = \widehat{\sigma}_t^{-1} y_t$$

and an estimator of ϕ_0 could (perhaps) be obtained by applying OLS to

$$\widehat{\epsilon}_t = \phi \widehat{\epsilon}_{t-1} + \widetilde{v}_t$$

- More formally we define the (2-step) estimator of ϕ_0 as

$$\begin{aligned}\widehat{\phi} &= \underset{\phi \in \Theta}{\arg\min} (2T)^{-1} \sum_{t=1}^{T} d\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}; \phi\right) \\ &= \left((1/T) \sum_{t=1}^{T} \widehat{\sigma}_{t-1}^{-2} y_{t-1}^2 \right)^{-1} \left((1/T) \sum_{t=1}^{T} \widehat{\sigma}_t^{-1} \widehat{\sigma}_{t-1}^{-1} y_t y_{t-1} \right).\end{aligned}$$

- Note: If $\widehat{\sigma}_t$ was a parametric estimator, the asymptotics could be established using the "nice" theory on two-step M-estimators, see, e.g., Section 12.4 in Wooldridge (2002).

- However, as $\widehat{\sigma}_t$ is an infinite dimensional estimator a mean value expansion of, say, $d\left(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}; \phi\right)$ about $\widehat{\sigma}_t$ is not feasible.

- Andrews (1994) provides solutions/tools: MINPIN estimators (estimators that MINimize as criterion function that may depend on a Preliminary Infinite dimensional Nuisance parameter estimator.)

Theorem 1 Let data be generated according to model (3)-(4) under assumptions i)-iv) with x_t defined as in Case 1 or Case 2. Let $\widehat{\sigma}_t^2$ be a nonparametric estimator of σ_t^2 and suppose a) $\sup_{\phi \in \Theta} \left\| \widehat{\sigma}_t^2 - \sigma_t^2 \right\| \xrightarrow{p} 0$ for some $\sigma_t^2 \in \mathcal{F}$ and b) $P(\widehat{\sigma}_t^2 \in \mathcal{F}) \xrightarrow{p} 1$. Then $\hat{\phi} \xrightarrow{p} \phi_0$.

Theorem 2 Let the Assumptions of Theorem 1 hold. Then under Case 1 and 2, $\sqrt{T}\left(\widehat{\phi} - \phi_0\right) \xrightarrow{d} \mathrm{N}\left(0, 1 - \phi_0^2\right)$.

Proof of Theorem 2 (sketch)

- Define $d\left(\sigma_t, \sigma_{t-1}; \phi\right) = m\left(\sigma_t, \sigma_{t-1}; \phi\right)^2$, where

$$\begin{aligned} m\left(\sigma_t, \sigma_{t-1}; \phi\right) &= \epsilon_{t-1}\left(\epsilon_t - \phi\epsilon_{t-1}\right) \\ &= \sigma_t^{-1}\sigma_{t-1}^{-1}y_t y_{t-1} - \sigma_{t-1}^{-2}\phi y_{t-1}^2 \end{aligned}$$

- Define $\overline{m}_T(\sigma; \phi) = (1/T)\sum_{t=1}^{T} m\left(\sigma_t, \sigma_{t-1}; \phi\right)$.

- Consider mean value expansion of $\sqrt{T}\overline{m}_T(\widehat{\sigma}; \widehat{\phi})$ about ϕ_0 given as

$$\sqrt{T}\overline{m}_T(\widehat{\sigma}; \widehat{\phi}) = \sqrt{T}\overline{m}_T(\widehat{\sigma}; \phi_0) + \frac{\partial}{\partial\phi}\overline{m}_T(\widehat{\sigma}; \phi^*)\sqrt{T}\left(\widehat{\phi} - \phi_0\right), \quad (6)$$

where ϕ^* lies between $\widehat{\phi}$ and ϕ_0.

- If a) $\lim_{T\to\infty}(\partial/\partial\phi)\overline{m}_T(\widehat{\sigma};\phi^*) \xrightarrow{p} M$, where M is nonsingular then it holds that

$$\sqrt{T}\left(\widehat{\phi}-\phi_0\right) = -M^{-1}\left(o_p(1)+\sqrt{T}\overline{m}_T(\widehat{\sigma};\phi_0)\right), \qquad (7)$$

since $\sqrt{T}\overline{m}_T(\widehat{\sigma};\widehat{\phi}) = o_p(1)$ as $\widehat{\phi}$ solves the first order condition $\overline{m}_T(\widehat{\sigma};\widehat{\phi}) = 0$ and $\widehat{\phi}$ belongs to the interior of Θ wp$\to$1.

- Asymptotic normality of $\sqrt{T}\left(\widehat{\phi}-\phi_0\right)$ follows if
  - b) $\sqrt{T}\overline{m}_T(\sigma;\phi_0)$ is asymptotically normally distributed
  - c) $\sqrt{T}(\overline{m}_T(\widehat{\sigma};\phi_0) - \overline{m}_T(\sigma;\phi_0)) \xrightarrow{p} 0$.

- Let $\overline{m}_T^*(\sigma;\phi) = (1/T)\sum_{t=1}^T \mathsf{E}\left(m\left(\sigma_t, \sigma_{t-1};\phi\right)\right)$ and

$$\upsilon_T(\sigma) = \sqrt{T}((1/T)\sum_{t=1}^{T} m\left(\sigma_t, \sigma_{t-1};\phi\right) - \overline{m}_T^*(\sigma;\phi))$$

such that

$$\sqrt{T}\left(\overline{m}_T(\widehat{\sigma};\phi_0) - \overline{m}_T(\sigma;\phi_0)\right) = \upsilon_T(\widehat{\sigma}) - \upsilon_T(\sigma) - \sqrt{T}\overline{m}_T^*(\widehat{\sigma};\phi).$$

- Condition c) is true if $\upsilon_T(\widehat{\sigma})$ is stochastic equicontinuous at σ, i.e.,

$$\lim_{T\to\infty} P\left(\sup_{\rho_{\mathcal{F}}(\widehat{\sigma},\sigma)<\delta'} |\upsilon_T(\widehat{\sigma}) - \upsilon_T(\sigma)| > \eta\right) \to 0$$

and $\sqrt{T}\overline{m}_T^*(\sigma;\phi) \xrightarrow{p} 0$.

Properties of the MINPIN estimator:

- $\widehat{\phi}$ converges at $\sqrt{T}$-rate

- $\widehat{\phi}$ is asymptotically efficient

- $\widehat{\phi}$ is adaptive

Difference based estimator of σ_t^2 (Case 1)

- Consider (for unknown $g(\cdot)$ and $e_t \sim i.i.d.$(0,1))

$$y_t = g(x_t) + \sigma e_t, \tag{8}$$

- Hall et al. (1990) suggest estimating σ^2 by

$$\hat{\sigma}_T^2 = (T - (m_1 + m_2))^{-1} \sum_{t=m_1+1}^{T-m_2} \Delta_{t,r}^2$$

$$\Delta_{t,r}^2 = \left(\sum_{j=0}^{r} d_j y_{j+t} \right)^2 \text{ where } \sum_{j=0}^{r} d_j = 0, \text{ and } \sum_{j=0}^{r} d_j^2 = 1.$$

- For $r = 1$: $\Delta_{t,1}^2 = \frac{(y_t - y_{t-1})^2}{2}$

- $\hat{\sigma}_T^2$ enjoys the parametric rate of convergence.

- Consider next (for unknown $g(\cdot)$, $\sigma(\cdot)^2$ and $e_t \sim$i.i.d.(0,1))

$$y_t = g(x_t) + \sigma(x_t)e_t,$$

- Levine (2003) suggests applying a local kernel smoother to all $\Delta_{t,1}^2 = \frac{(y_t - y_{t-1})^2}{2}$ in order to produce the estimator

$$\widehat{\sigma}_t^2 = \frac{\sum_{t=1}^{T-r} \Delta_{t,1}^2 K\left(\frac{x - x_t}{h}\right)}{\sum_{t=1}^{T-r} K\left(\frac{x - x_t}{h}\right)}. \tag{9}$$

- $\widehat{\sigma}_t^2$ has the optimal rate of convergence $T^{-\frac{2p}{2p+1}}$, if $\sigma_t^2 \in C^p[0,1]$ and $g(x) \in C^{p-1}[0,1]$ for $p > 1$.

- Crucial question: Can we obtain a consistent estimator of σ_t^2 in (2)-(3) using the estimator (9)? No! In this case $\widehat{\sigma}_t^2 \stackrel{p}{\longrightarrow} \sigma_t^2/(1+\phi)$.

- Idea: If ϵ_t follows an AR(1) process then $\text{var}(\epsilon_t) - \text{cov}(\epsilon_t, \epsilon_{t-2}) = 1$

- Define $\eta_t = (1/2)^{1/2}\,(y_t - y_{t-2})$ and consider the local linear estimator $\widehat{\sigma}_t^2 = \widehat{\sigma}\,(x_t)^2$ given as $\hat{a}_t$ that solves the problem

$$(\hat{a}_t, \hat{b}_t) = \arg\min_{a_t, b_t} \sum_{s=2}^{T} \left(\eta_s^2 - a_t - (x_s - x_t) b_t\right)^2 K_h\,(x_s - x_t) \qquad (10)$$

Theorem 3 (Consistency of $\widehat{\sigma}_t^2$) Let data be generated according to the model (3)-(4) under Case 1. Suppose that $K(u)$ is a second order non-negative kernel function satisfying: $K(u) \geq 0$ for any $u \in [-1, 1]$, $\mu_1 = \int K(u)\, du = 0$, $\sigma_K^2 \equiv \mu_2 = \int u^2 K(u)\, du \neq 0$ and $R_K = \int K(u)^2\, du$. Then the estimator given by (10) is consistent in mean square with convergence rate $O(T^{-4/5})$. Furthermore, the optimal (in the sense of Parzen (1962) and Rosenblatt (1956)) bandwidth is $h = O(T^{-1/5})$.

- Why will this idea work? Consider the Taylor series expansions (about x)

$$\begin{aligned}\eta_t^2 &= \frac{1}{2}\left(\sigma_t^2\epsilon_t^2 + \sigma_{t-2}^2\epsilon_{t-2}^2 - 2\sqrt{\sigma_t^2\sigma_{t-2}^2}\epsilon_t\epsilon_{t-1}\right)\\ \sigma_t^2 &= \sigma^2 - \mathsf{D}\sigma^2(x-x_t) + \mathsf{D}^2\sigma^2(x-x_t)^2/2 + o(h^2)\\ \sqrt{\sigma_t^2\sigma_{t-1}^2} &= \sigma^2 + \text{higher order expansion terms}\end{aligned}$$

using $\sqrt{1+x} = 1 + x/2 + o(x)$.

- Then

$$\begin{aligned}\mathsf{E}\left(\eta_t^2\right) &= \frac{1}{2}\left(\sigma_t^2 + \sigma_{t-2}^2\right)\mathsf{var}(\epsilon_t) - \sqrt{\sigma_t^2\sigma_{t-2}^2}\,\mathsf{cov}(\epsilon_t, \epsilon_{t-2})\\ &= \left(\mathsf{var}(\epsilon_t) - \mathsf{cov}(\epsilon_t, \epsilon_{t-2})\right)\sigma^2 + \text{higher order expansion terms}\\ &= \sigma^2 + \text{ higher order expansion terms}\end{aligned}$$

as $\text{var}(\epsilon_t) - \text{cov}(\epsilon_t, \epsilon_{t-2}) = (\gamma_0 - \gamma_2) = 1.$

Theorem 4 (Asymptotic Normality of $\widehat{\sigma}_t^2$)   Let the Assumptions of Theorem 1 hold. Then,

$$\widehat{\sigma}_t^2 \xrightarrow{d} \mathsf{N}\left(\sigma_t^2 + Bias(\widehat{\sigma}_t^2),\ R_K C(\phi_0)\sigma^4\,(Th)^{-1} + o((Th)^{-1})\right).$$

as $T \to \infty$, $h \to 0$ and $Th \to \infty$, where

$$Bias(\widehat{\sigma}_t^2) \;=\; \left[\mathsf{D}^2\sigma^2/4 - \gamma_2(\mathsf{D}\sigma^2)^2/\sigma^2\right] h^2\sigma_K^2/2 + o(h^2).$$

Fan and Yao's estimator of σ_t^2 (Case 2)

- Fan and Yao prove consistency and asymptotic normality of the estimators of $\mathsf{E}(y_t|\boldsymbol{x}_t)$ and $\mathsf{var}(y_t|\boldsymbol{x}_t)$ denoted $\widehat{a}_t$ and $\widehat{\alpha}_t$ respectively and given as

$$
\begin{aligned}
(\hat{a}_t, \widehat{\boldsymbol{b}_t}) &= \arg\min_{a_t, \boldsymbol{b}_t} \sum_{s=3}^{T} (y_s - a_t - (\boldsymbol{x}_s - \boldsymbol{x}_t)\boldsymbol{b}_t)^2 K_H(\boldsymbol{x}_s - \boldsymbol{x}_t), \\
(\widehat{\alpha}_t, \widehat{\beta}_t) &= \arg\min_{\alpha_t, \beta_t} \sum_{s=3}^{T} \left(\widehat{r}_s^2 - \alpha_t - (x_s - x_t)\beta_t\right)^2 K_h(x_s - x_t),
\end{aligned}
$$

where $\widehat{r}_t = y_t - \hat{a}_t$, and $\boldsymbol{x}_t = (x_t, x_{t-1}, y_{t-1})$.

Simulation results

Table 1: Alternative data generating processes. $\varphi(\cdot)$ is the standard normal c.d.f.

	Specifications
Model 1	$y_t = \sqrt{0.1 + 0.5x_t^2}\epsilon_t$
Model 2	$y_t = \sqrt{0.4\exp(-2x_t^2) + 0.2}\epsilon_t$
Model 3	$y_t = \sqrt{\varphi(x_t + 1.2) + 1.5\varphi(x_t - 1.2)}\epsilon_t$

Figure 2: MSE from the Difference based variance function estimator (solid line) and the Fan-Yao estimator (dotted line) under alternative variance function specifications and alternative values of ϕ_0. $T = 1000$ and the number of Monte Carlo replications equals 1000.

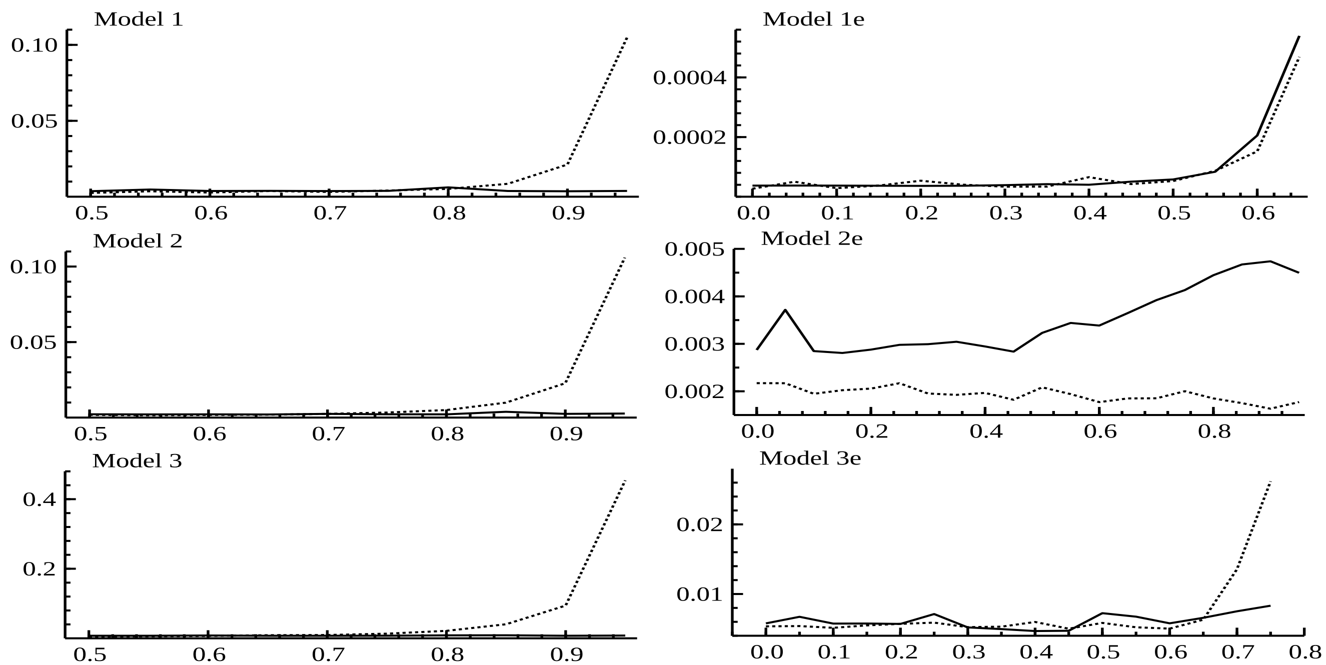
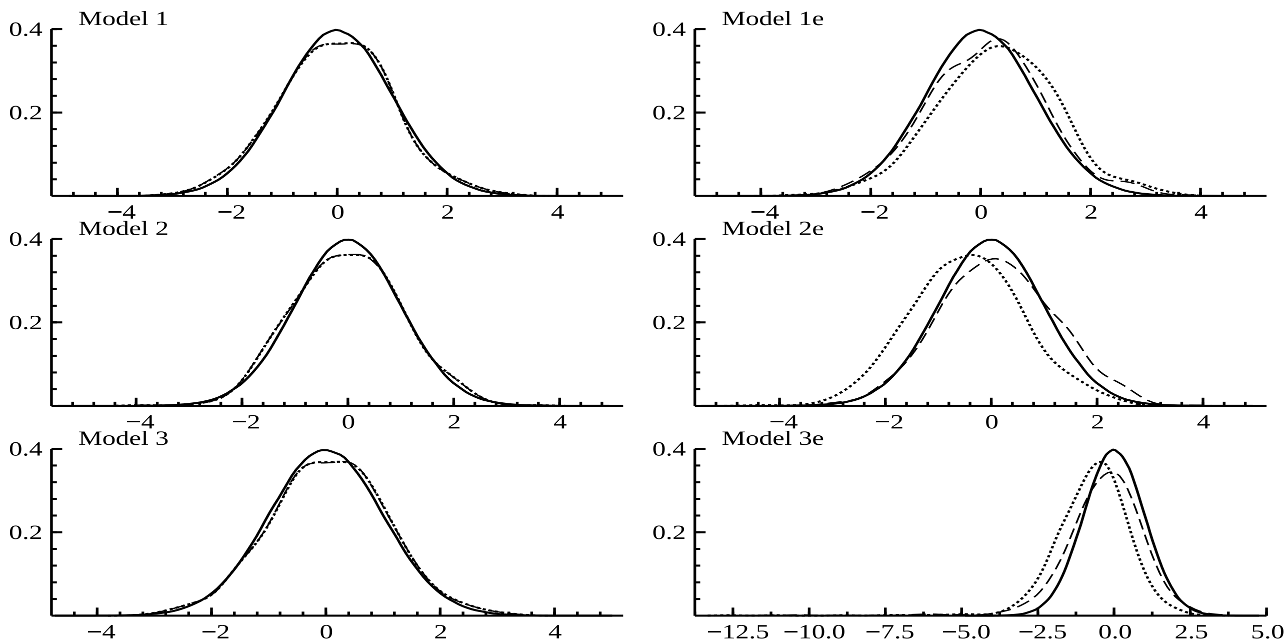


Figure 3: Finite sample (simulated) densities and the asymptotic density of $\sqrt{T}\left(\widehat{\phi}-\phi_0\right)/\sqrt{1-\phi_0^2}$ under alternative variance function specifications for T $= 1000$ and $\phi_0 = 0.5$. Solid line: N$(0,1)$. Dashed Line: Fan-Yao. Dotted line: Difference based. The number of Monte Carlo replications equals 1000.



Empirical illustration (3m t-bill)

- GARCH-M

$$
\begin{aligned}
\Delta r_t &= \gamma + \rho r_{t-1} + \delta\sigma_t + \sigma_t v_t, \qquad v_t \sim i.i.d(0,1), \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \Delta r_{t-1}^2 + \beta\sigma_{t-1}^2.
\end{aligned}
$$

- GARCH with AR(1) innovations

$$
\begin{aligned}
\Delta r_t &= \gamma + \rho r_{t-1} + \phi\left(\frac{\sigma_t}{\sigma_{t-1}}\right)\Delta r_{t-1} + \kappa_1\left(\frac{\sigma_t}{\sigma_{t-1}}\right) r_{t-2} + \kappa_2\left(\frac{\sigma_t}{\sigma_{t-1}}\right) + \sigma_t v_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \Delta r_{t-1}^2 + \beta\sigma_{t-1}^2, \\
\kappa_1 &= -\phi\rho, \\
\kappa_2 &= -\phi\gamma.
\end{aligned}
$$

- Nonparametric model (NonP a.)

$$r_t = g(r_{t-1}, r_{t-2}; \phi) + \sigma_t v_t,$$

- Nonparametric model (NonP b.)

$$\begin{aligned} \Delta r_t &= \gamma + \rho r_{t-1} + \phi \frac{\widehat{\sigma}_t^{np}}{\widehat{\sigma}_{t-1}^{np}} \Delta r_{t-1} + \kappa_1 \frac{\widehat{\sigma}_t^{np}}{\widehat{\sigma}_{t-1}^{np}} r_{t-2} + \kappa_2 \frac{\widehat{\sigma}_t^{np}}{\widehat{\sigma}_{t-1}^{np}} + \sigma_t v_t. \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \Delta r_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

Table 2: Estimation results. Sample period: 01/08/1954-12/31/1999.

	GARCH-M	GARCH-AR(1)	NonP (a)	NonP (b)
γ	0.016 (0.005)	-0.044 (0.045)	·	-0.398 (0.329)
ρ	-0.003 (0.001)	0.011 (0.009)	·	0.111 (0.084)
δ	0.007 (0.045)	·	·	·
α_0	0.000 (0.000)	0.001 (0.000)	·	0.001 (0.000)
α_1	0.169	0.156	·	0.154
β	0.830 (0.036)	0.844 (0.018)	·	0.845 (0.018)
ϕ	·	0.177 (0.021)	0.252 (0.020)	0.212 (0.037)
κ_1	·	-0.014 (0.009)	·	-0.115 (0.084)
κ_2	·	0.060 (0.045)	·	0.415 (0.329)

Directions for ongoing/future research

- Include $g(x_t)$ in (3)-(4): Estimation parametrically/non-parametrically?

- Generalization of the innovations process to AR(p).

- Capturing GARCH effects (nonparametrically)

- Evaluate the predictive ability of the model based on simulated/actual data

- Evaluate the significance of ϕ (failure of the martingale difference hypothesis)

Include $g(x_t)$

- Case 1: Under weak regularity conditions the effect of $g(x_t)$ will be removed by the "difference" operation.

- Case 2: As long as k is not too large $g(x_t)$ is consistently estimated in the first step of the Fan and Yao (1998) estimation procedure. If additional assumptions on $g(\cdot)$ is imposed, such as additivity, k can be large without affecting the asymptotics.

Generalization of the innovations process: AR(p) - structure

- Consider now the following model

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t && (11) \\
\epsilon_t &= \sum_{i=1}^{p} \phi_i \epsilon_{t-i} + v_t && (12)
\end{aligned}
$$

Case 1:

- Define $\widetilde{\eta}_t = (1/2)^{1/2}(y_t - y_{t-1})$ and $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)'$. By using the same expansion-techniques as above we have

$$\begin{aligned} \mathsf{E}\left(\widetilde{\eta}_t^2\right) &= \frac{1}{2}\left(\sigma_t^2 + \sigma_{t-1}^2\right)\mathsf{var}(\epsilon_t) - \sqrt{\sigma_t^2\sigma_{t-1}^2}\,\mathsf{cov}(\epsilon_t, \epsilon_{t-2}) \\ &= \left(\gamma_0(\boldsymbol{\phi}) - \gamma_1(\boldsymbol{\phi})\right)\sigma^2 + \text{higher order expansion terms} \end{aligned}$$

and consequently one can show that

$$\widehat{\sigma}_t^2 \xrightarrow{p} \left(\gamma_0(\boldsymbol{\phi}) - \gamma_1(\boldsymbol{\phi})\right)\sigma^2$$

where $\widehat{\sigma}_t^2$ is obtained using local linear regression to $\widetilde{\eta}_t^2$.

- Consider the following estimator of $\boldsymbol{\phi}$ given as

$$\widehat{\boldsymbol{\phi}} = \boldsymbol{A}\left(\widehat{\sigma}_{t-1}, ..., \widehat{\sigma}_{t-p}\right)^{-1}\boldsymbol{B}\left(\widehat{\sigma}_t, ..., \widehat{\sigma}_{t-p}\right)$$

where

$$\boldsymbol{A}\left(\widehat{\sigma}_{t-1},...,\widehat{\sigma}_{t-p}\right)=\begin{bmatrix} \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-1}^2}{\widehat{\sigma}_{t-1}^2} & \cdots & \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-1}y_{t-p}}{\widehat{\sigma}_{t-1}\widehat{\sigma}_{t-p}} \\ \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-2}y_{t-1}}{\widehat{\sigma}_{t-2}\widehat{\sigma}_{t-1}} & \cdots & \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-2}y_{t-p}}{\widehat{\sigma}_{t-2}\widehat{\sigma}_{t-p}} \\ \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-p}y_{t-1}}{\widehat{\sigma}_{t-p}\widehat{\sigma}_{t-1}} & \cdots & \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-p}^2}{\widehat{\sigma}_{t-p}^2} \end{bmatrix}$$

and

$$\boldsymbol{B}\left(\widehat{\sigma}_{t},...,\widehat{\sigma}_{t-p}\right)=\begin{bmatrix} \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-1}y_t}{\widehat{\sigma}_{t-1}\widehat{\sigma}_t} \\ \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-2}y_t}{\widehat{\sigma}_{t-2}\widehat{\sigma}_t} \\ \vdots \\ \frac{1}{T}\sum_{t=p+1}^{T}\frac{y_{t-p}y_t}{\widehat{\sigma}_{t-p}\widehat{\sigma}_t} \end{bmatrix}$$

Proposition E1 Let data be generated according to the model given by (3)-(4). Then

$$\boldsymbol{A}\left(\widehat{\sigma}_{t-1}, ..., \widehat{\sigma}_{t-p}\right) \xrightarrow{p} \left(\gamma_0(\boldsymbol{\phi}) - \gamma_1(\boldsymbol{\phi})\right)^{-1} \boldsymbol{A}\left(\sigma_{t-1}, ..., \sigma_{t-p}\right)$$

Proposition E2 Let data be generated according to the model given by (3)-(4). Then

$$\boldsymbol{B}\left(\widehat{\sigma}_{t}, ..., \widehat{\sigma}_{t-p}\right) \xrightarrow{p} \left(\gamma_0(\boldsymbol{\phi}) - \gamma_1(\boldsymbol{\phi})\right)^{-1} \boldsymbol{B}\left(\sigma_{t}, ..., \sigma_{t-p}\right)$$

Theorem E1 Let data be generated according to the model given by (3)-(4). Then

$$\widehat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}.$$

- A consistent estimator of σ^2 can then be obtained as

$$\left(\gamma_0(\widehat{\phi}) - \gamma_1(\widehat{\phi})\right)^{-1} \widehat{\sigma}_t^2$$

where $\widehat{\sigma}_t^2$ denotes the preliminary (and inconsistent) nonparametric estimator

Case 2:

- Rewrite the model as

$$y_t = g(y_{t-1}, ..., y_{t-p-1}) + \sigma_t v_t$$

where

$$g(y_{t-1}, ..., y_{t-p-1}) = \phi_1 \frac{\sigma_t}{\sigma_{t-1}} y_{t-1} + \phi_2 \frac{\sigma_t}{\sigma_{t-2}} y_{t-2} + ... + \phi_p \frac{\sigma_t}{\sigma_{t-p}} y_{t-p}$$

- Although convergence will be slow an estimator of $g(y_{t-1}, ..., y_{t-p-1})$ can be obtained by the Fan and Yao (1998) approach.

- Based on simulations the following "iterated" Fan and Yao estimator seems promising:

1. Obtain a preliminary consistent estimates of $g(y_{t-1}, ..., y_{t-p-1})$ and σ_t^2 using Fan-Yao.

2. Obtain $\widehat{\boldsymbol{\phi}} = \boldsymbol{A}\left(\widehat{\sigma}_{t-1}, ..., \widehat{\sigma}_{t-p}\right)^{-1} \boldsymbol{B}\left(\widehat{\sigma}_{t}, ..., \widehat{\sigma}_{t-p}\right)$

3. Form $\widehat{r}_t = y_t - \widehat{\phi}_1 \frac{\widehat{\sigma}_t}{\widehat{\sigma}_{t-1}} y_{t-1} - \widehat{\phi}_2 \frac{\widehat{\sigma}_t}{\widehat{\sigma}_{t-2}} y_{t-2} - ... - \widehat{\phi}_p \frac{\widehat{\sigma}_t}{\widehat{\sigma}_{t-p}} y_{t-p}$ and

update the estimator of σ_t^2 (given as $\widehat{\alpha}_t$) from

$$(\widehat{\alpha}_t, \widehat{\beta}_t) = \arg\min_{\alpha_t, \beta_t} \sum_{s=3}^{T} \left(\widehat{r}_s^2 - \alpha_t - (y_{s-1} - y_{t-1})\beta_t \right)^2 K_h \left(x_s - x_t \right)$$

4. Repeat steps 2. and 3. until convergence.

- Alternatively to the "iterated" Fan-Yao estimator one could apply the estimation algorithms proposed by Fan, Yao and Cai (2003) for so-called Varying Coefficent Linear Models.

Capturing GARCH effects

- Consider the following semi-parametric GARCH specification

$$\sigma_t^2 = m(y_{t-1}) + \beta\sigma_{t-1}^2$$

or equivalently (provided that $|\beta| < 1$)

$$\sigma_t^2 = \sum_{j=1}^{\infty} \beta^{j-1} m(y_{t-j})$$

- Next, consider the approximation (after truncation and replacing σ_t with $\widehat{r}_t = y_t - \hat{a}_t$)

$$\widehat{r}_t^2 = \sum_{j=1}^{J} \beta^{j-1} m(y_{t-j}) + \omega_t \tag{13}$$

or

$$\widehat{r}_t^2 = \sum_{j=1}^{J} m_j(y_{t-j}) + \omega_t \tag{14}$$

- Note that (14) is an ordinary additive model where the component functions $m_j(\cdot)$ are linked by β, i.e., $m_j(x) = \beta^{j-1} m_1(x)$.

- Carroll, Haerdle and Mammen (2002) suggest estimating $m(\cdot)$ and β by the following procedure

1. Estimate $m_j(\cdot)$ for $j = 1, ..., J$, based on (14) and note that $m_1(x) = m(x)$

2. Obtain the estimator of β as

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t=J+1}^{T} \sum_{j=2}^{J} w_j(y_{t-j}) \left[\widehat{m}_j(y_{t-j}) - \beta^{j-1} \widehat{m}_1(y_{t-1}) \right]^2$$