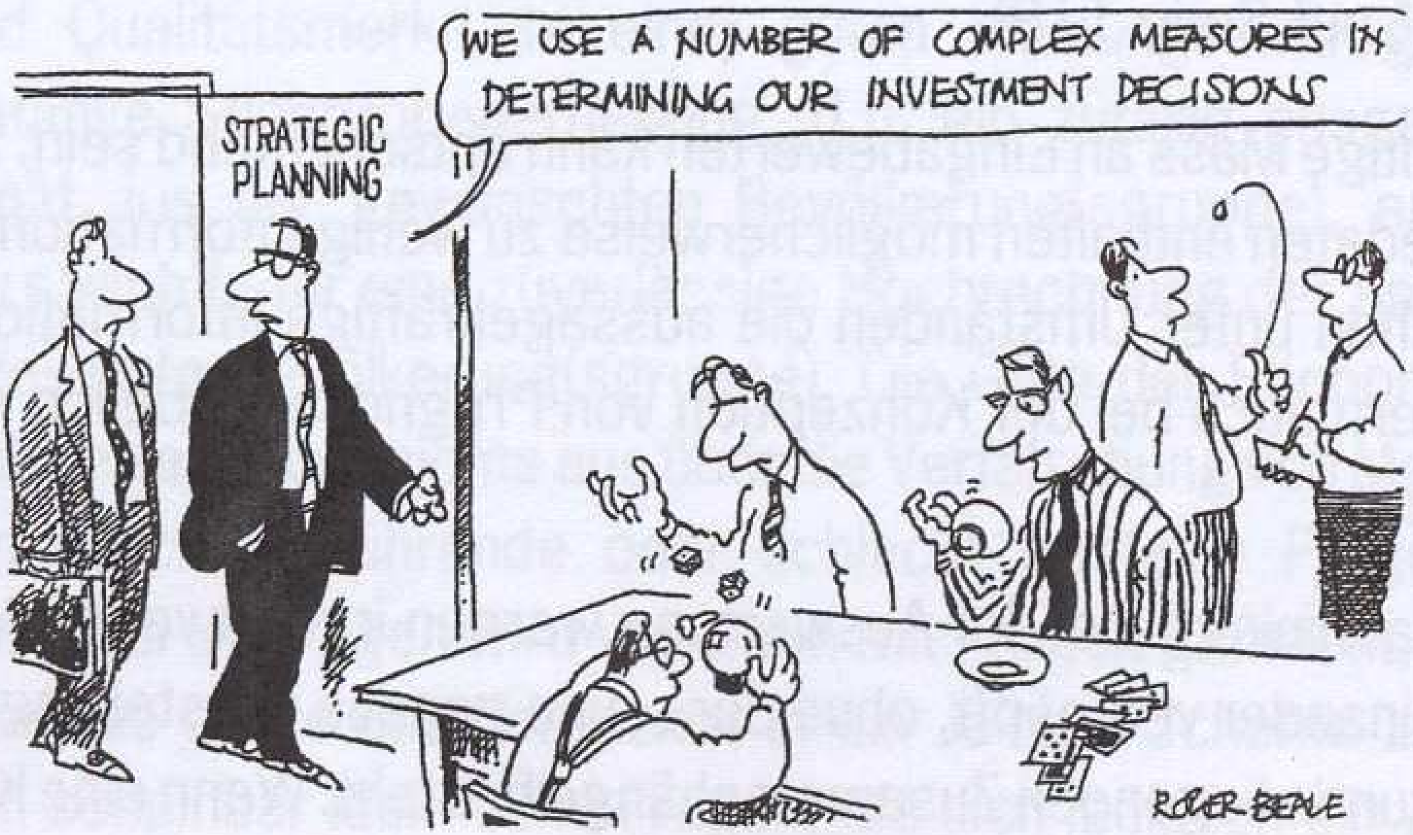
Robust Statistical Techniques for Financial Modelling

Elvezio Ronchetti

Department of Econometrics
University of Geneva
Switzerland

Elvezio.Ronchetti@metri.unige.ch

http://www.unige.ch/ses/metri/ronchetti/


WE USE A NUMBER OF COMPLEX MEASURES IN DETERMINING OUR INVESTMENT DECISIONS
STRATEGIC PLANNING
ROGER BEALE

♦ Outline

- Introduction
  - Stylized Facts

- Estimation and Inference for Single Factor Models
  - Robust Parametric Inference
  - Indirect Inference
  - Nonparametric Inference

- Robust Estimation and Inference for Dynamic Location-Scale Models
  - ARCH, GARCH, ...

- Other Topics and Outlook

♦ **Introduction**

Stylized Facts

- **Volatility clustering**
  Variance of daily price changes varies over time (heteroscedasticity)
  $\Rightarrow$ Model conditional distribution at time t given the information $\mathcal{F}_{t-1}$ up to time $t-1$

- **Low signal-to-noise ratio**

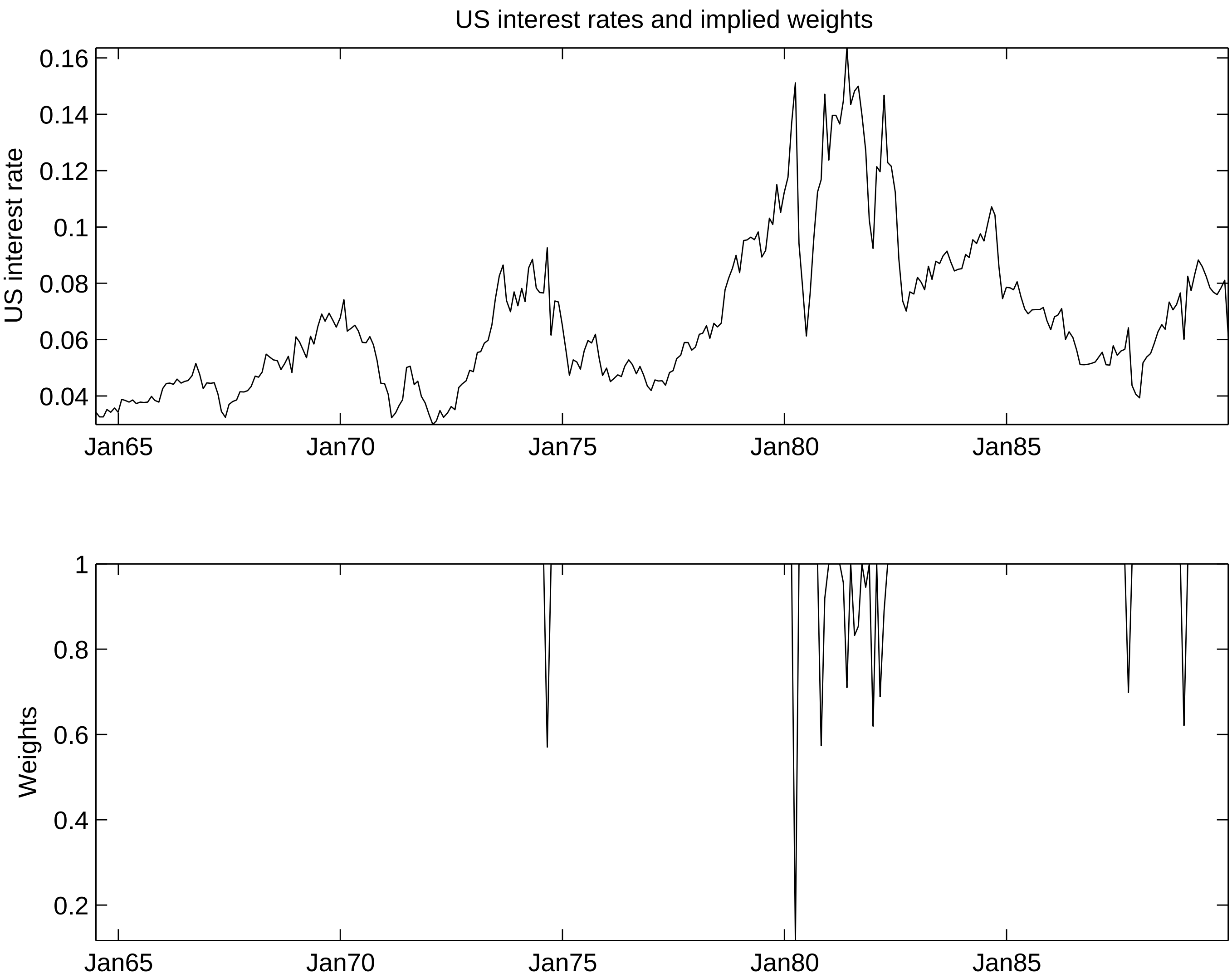
- **Robustness issue**
  Financial markets are complex; therefore **financial models** are at best only **approximate descriptions** of the underlying structure

♦ Estimation and Inference for Single Factor Models

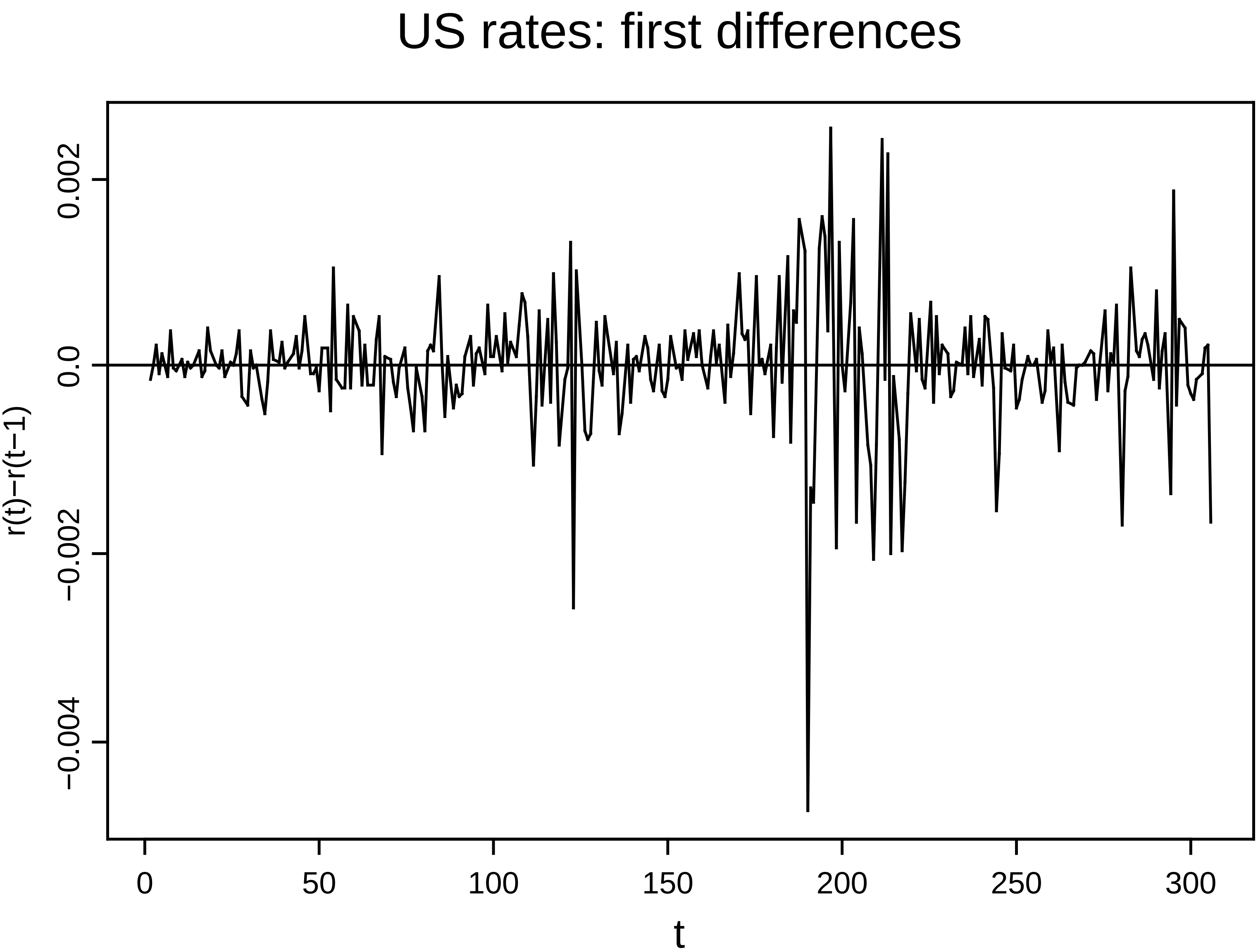
Data set of Chan, Karolyi, Longstaff, Sanders (1992):

One-month yields based on the average of bid and ask prices for Treasury bills normalized to reflect a standard month of 30.4 days. They are monthly observations covering the period from June 1964 to December 1989, for a total of 307 observations.

Ahn and Gao (1999): McCulloch and Kwon (1993) dataset over the period from December 1946 to February for comparability with the Ahn and Gao (1999) study.



US Short Term Rates


US rates: first differences
r(t)−r(t−1)
0.002
0.0
−0.002
−0.004
0
50
100
150
200
250
300
t

♦ Single Factor Short Rate Models

Basic Setting

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t \quad ,$$

where r_t is the short rate at time t and $(W_t)_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}$

- Chan, Karolyi, Longstaff, Sanders (1992) (CKLS) , linear drift term

$$\begin{aligned}\mu(r_t) &= \alpha + \beta r_t \\ \sigma(r_t) &= \sigma r_t^{\gamma}\end{aligned}$$

- Ahn and Gao (1999), quadratic drift term

$$\begin{aligned}\mu(r_t) &= \alpha + \alpha_1 r_t + \alpha_2 r_t^2 \\ \sigma(r_t) &= \sigma r_t^{3/2}\end{aligned}$$

- Ait Sahalia (1996)

$$\begin{aligned}\mu(r_t) &= \alpha + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1} \\ \sigma(r_t) &= \beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}\end{aligned}$$

Alternative Models of the Short Rate*

Model	α	β	σ	γ	**restrictions**
Merton		0		0	(0 **attainable**)
Vasicek				0	$\beta < 0$ (0 **attainable**)
Cox Ing. Ross				$\frac{1}{2}$	$\beta < 0$ **and** $2\alpha \geq \sigma^2$
Dothan	0	0		1	–
Geom. B. Mot.	0			1	$\beta < 0$, (0 **attainable**)
Brenn. Schw.				1	$\beta < 0$ **and** $\alpha > 0$
Variab. Rate	0	0		$\frac{3}{2}$	(0 **attainable**)
Const. El. Var.	0				$\beta < 0$, (0 **attainable**)

*Natural restrictions have to be imposed on the parameter values to ensure that the drift is mean-reverting at high interest rate values (infinity not attainable) and zero is unattainable; see Aït-Sahalia (1996).

♦ GMM Estimation of CKLS Models

Crude discretization

$$r_t - r_{t-1} = \alpha + \beta r_{t-1} + \epsilon_t$$

where $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 r_{t-1}^{2\gamma}$

Orthogonality conditions used in CKLS:

$$\begin{aligned} E(\epsilon_t) &= 0 \\ E(\epsilon_t r_{t-1}) &= 0 \\ E(\eta_t) &= 0 \\ E(\eta_t r_{t-1}) &= 0 \end{aligned}$$

where $\eta_t = \epsilon_t^2 - \sigma^2 r_{t-1}^{2\gamma}$

⧫ GMM Estimators

$\mathcal{X} := (X_n)_{n \in N}$ stationary ergodic sequence defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Parametric model
$\mathcal{P} := \{P_\theta, \theta \in \Theta\}$, $\theta \in \Theta \subset R^p$.
True parameter vector: θ_0.

Method of moments:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = E_\theta X_1 = g_1(\theta)$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E_\theta X_1^2 = g_2(\theta)$$

....

Equivalently:

$$\sum_{i=1}^{n}[X_i - g_1(\theta)] = 0$$
$$\sum_{i=1}^{n}[X_i^2 - g_2(\theta)] = 0$$

....

i.e.

$$E_\theta h(X_1;\theta) = 0$$

where $h(X_1;\theta) = (h_1(X_1;\theta), h_2(X_1;\theta), \dots)'$
Orthogonality conditions

GMM:

Estimate indirectly some function

$$a : \mathcal{P} \to \mathcal{A} := a(\mathcal{P}) \subset R^k$$

of *parameters of interest* by introducing a function

$$h : R^N \times \mathcal{A} \to R^H$$

enforcing a set of orthogonality conditions

$$E_{\theta_0} h(X_1; a(P_{\theta_0})) = 0 \quad , \tag{1}$$

on the structure of the underlying model.

$\mathcal{W} := (W_n)_{n \in N}$ sequence of weighting symmetric positive definite matrices
converging a.s to W_0, the inverse of the covariance matrix of $h(X_1, a(P_{\theta_0}))$,

Generalized method of moments estimator (GMME) associated with $\mathcal{W}$:
$(\tilde{a}(P_{\theta_n}))_{n \in N}$ solution to (Hansen, 1982)

$$\min_{a \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} h'(X_i; a) W_n \frac{1}{n} \sum_{i=1}^{n} h(X_i; a)$$
$$(= \min_{a \in \mathcal{A}} E_{\theta_n} h'(X_1; a) W_n E_{\theta_n} h(X_1; a))$$

where $P_{\theta_n} := \frac{1}{n} \sum \delta_{X_i}$ is the empirical distribution of $X_1, .., X_n$ and δ_x denotes the point mass distribution at $x \in R^N$).

Under appropriate regularity conditions the GMME exists, is strongly consistent and asymptotically normally distributed with asymptotic covariance matrix

$$\begin{aligned}\Sigma_{\theta_0}(W_0) \;=\; & [E_{\theta_0}\frac{\partial h'(X_1; a(P_{\theta_0}))}{\partial a} W_0 \times \\ & E_{\theta_0}\frac{\partial h(X_1; a(P_{\theta_0}))}{\partial a'} \quad]^{-1}.\end{aligned}$$

In our case:

$\mathcal{X} = (r_{t-1}, r_t)'_{t\geq 0}$
Orthogonality function h:

$$h(x, y; \alpha, \beta, \sigma, \gamma) =$$

$$\begin{array}{l} y - x - \alpha - \beta x \\ (y - x - \alpha - \beta x)x \\ (y - x - \alpha - \beta x)^2 - \sigma^2 x^{2\gamma} \\ ((y - x - \alpha - \beta x)^2 - \sigma^2 x^{2\gamma})x. \end{array}$$

Classical GMM Estimates of Alternative Models for the Short-Term Interest Rate

The parameters are estimated by the classical GMM induced by the original orthogonality function h t-statistics are in parentheses. The values of Hansen's statistics (ξ for short) are reported with p-values in parentheses.

Model	Classical GMM α	β	σ	γ	ξ
Unres.	0.0034	-0.0493	0.3656	1.4999	-
	(1.85)	(-1.55)	(1.56)	(5.95)	
Merton	0.0005	0	0.0062	0	6.76
	(1.44)		(14.54)		(0.034)
Vasicek	0.0005	-0.0013	0.0062	0	6.80
	(0.33)	(-0.04)	(14.33)		(0.009)
CIR	0.0011	-0.0102	0.0254	$\frac{1}{2}$	4.90
	(0.67)	(-0.36)	(15.28)		(0.027)
Dothan	0	0	0.0320	1	5.60
			(15.94)		(0.133)
GBM	0	0.0084	0.0993	1	3.16
		(1.50)	(16.07)		(0.206)
BS	0.0020	-0.0262	0.0994	1	2.21
	(1.24)	(-0.92)	(16.18)		(0.137)
VR	0	0	0.05	$\frac{3}{2}$	6.31
			(15.66)		(0.098)
CEV	0	0.0086	0.1554	1.1711	2.98
		(1.53)	(1.18)	(3.59)	(0.084)

Classical GMM Estimates

Classical Analysis

- Models that allow for values of $\gamma \geq 1$ are not rejected – using Hansen's statistic – while models where $\gamma \in [0, 1)$ are.

- The estimates γ in the corresponding models are strongly significant.

How stable is this analysis?

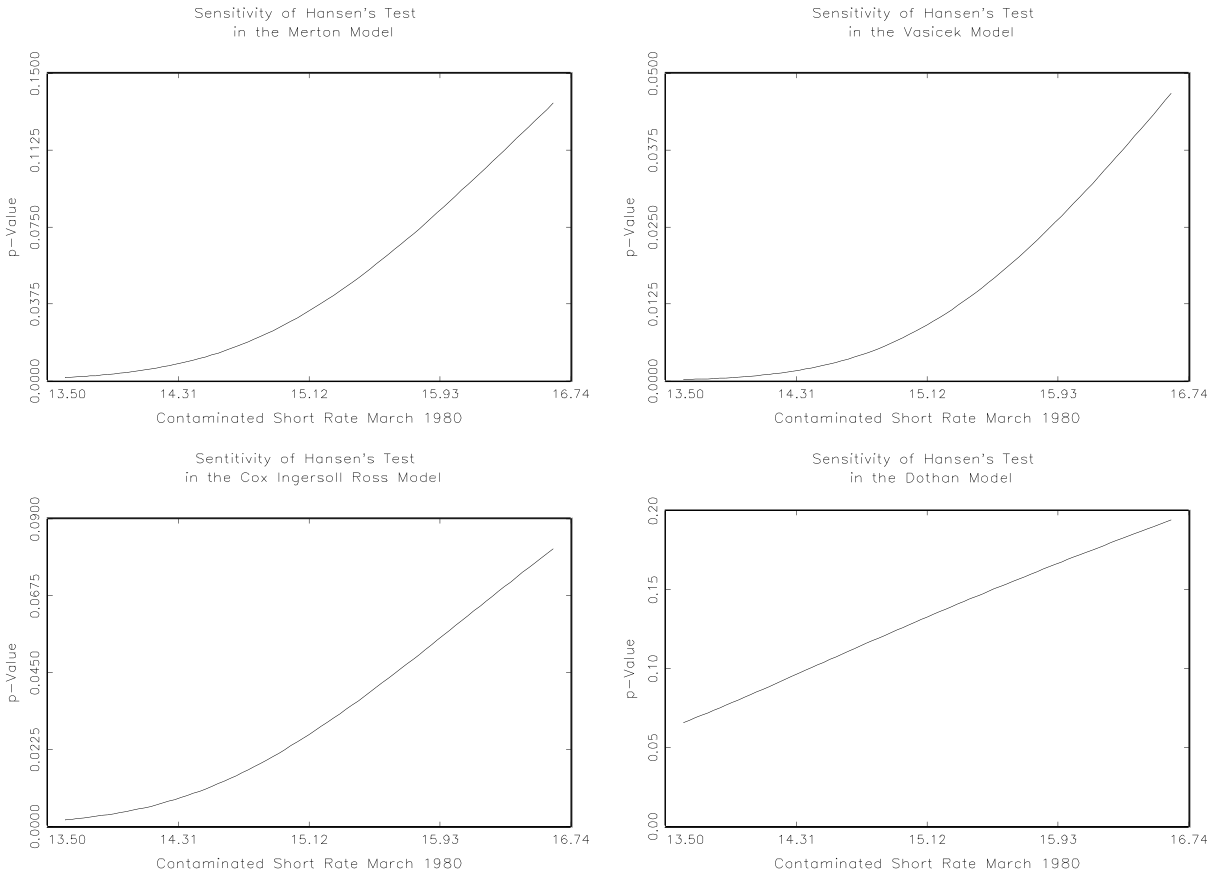
How reliable are these conclusions?

Sensitivity analysis of the $p-$value of Hansen's statistic

Change one observation: ϵ=1/306=0.3%. Vary the value corresponding to March 1980

(observed value: 0.1512) from 0.1350 to 0.1674 by steps of size 0.001.

Variability of the short-term rate changes around March 1980 is very high: a change from a 15% to a 9.5% interest rate level just before March 1980
$\Rightarrow$ the magnitude of the sensitivity analysis seems to be realistic with respect to the structure of the short-rate observations over this particular period.



Sensitivity Analysis
Merton, Vasicek, Cox et al., Dothan

Classical Hansen's test

- Very steep $p-$ values curves

- E.g. Merton and Cox et al. models: a change of the value of the observation of 40-90 basis points (100 basis points $= 0.01$) is sufficient for obtaining p-values *not rejecting* the model specification at a 5% significance value, as opposed to the results for the uncontaminated model.

This analysis shows that it is very difficult to distinguish – by means of the classical Hansen's test – the specification properties of the models, even when only a *single* observation is changed in the data.

♦ A Simulated Example

- Generate 200 paths of length 300 from a CIR process with the parameter chosen in order to produce a monthly autocorrelation of 0.98 and a variability comparable with the CKLS dataset.

- Take the largest observation and contaminate it with 300 and 400 basispoints (comparable to the movements around March 1980, e.g. 15% to 9.5% in March 1980).

Model	Non cont	Cont. 300	Cont. 400
cl. GMM: rej. CIR	8%	0%	0%
rob. GMM: rej. CIR	5%	5%	5%
cl. GMM: rej. BS	92%	51%	12%
rob. GMM: rej. BS	80%	64%	60%

- The classical Hansen's test loses rapidly power and has a vanishing size under contamination.

- The robust Hansen's test performs in a satisfactory way.

♦ Alternative Analysis: Robust Statistics

♦ Introduction

- deals with deviations from ideal models and their dangers for corresponding inference procedures

- primary goal is the development of procedures which are still reliable and reasonably efficient under deviations from the model used

♦ Robustness in Finance

Two main research fields on robustness:

- Modelling preferences for robustness and robust decision processes of agents that take into account some forms of model misspecification in their decisions

- Developing robust statistics for the econometric analysis of financial time series using models that are possibly misspecified

Wanted

Robust procedures that take into account model misspecifications both

- when determining optimal policies in financial models
- when estimating the parameter inputs for a financial model

Key tools

- Influence Function (local stability)
- Breakdown Point (global reliability)

The Influence Function (IF, Hampel (1968, 1974)) of a statistic (functional) T is defined by

$$IF(x;T,F) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon}$$

for all x where the limit exists. Δ_x is the distribution which puts mass 1 at x.

- The IF describes the normalized influence on the statistic of an infinitesimal observation at x.

- IF is the Gâteaux derivative of T at F or the integrand in the first term of the von Mises expansion.

- Examples of "interesting" statistics: an estimator, its expectation and variance, the power and the level of a test, a portfolio allocation, etc.

Wanted

Procedures with **bounded influence function**

- IF bounded implies a bounded bias of the statistic in a contaminated neighborhood of the model

- Many models in econometrics/finance imply optimal policies/statistics with unbounded IF

- Well-known examples: OLS–, TSLS–, NLLS–methods, many ML and GMM statistics; optimal portfolios and indirect utilities in mean variance optimization problems

Some general books

- Huber, P.J.(1981)
  Robust Statistics,
  Wiley (paperback 2004)

- Hampel,F.R., Ronchetti,E.M., Rousseeuw,P.J., Stahel, W.A. (1986)
  Robust Statistics: The Approach Based on Influence Functions,
  Wiley (paperback 2005)

- Rousseeuw, P.J., Leroy, A. M. (1987)
  Robust Regression & Outlier Detection,
  Wiley

Recent books

- Jureckova, J., Picek, J. (2005)
  Robust Statistical Methods with R, Chapman & Hall/CRC

- Maronna R. A., Martin, R.D., Yohai, V. J. (2006)
  Robust Statistics: Theory, and Methods, Wiley

Local stability properties of the GMME a:

$$IF(x; a, P_{\theta_0}) = -\Sigma_{\theta_0}(W_0) E_{\theta_0} \frac{\partial h'(X_1; a(P_{\theta_0}))}{\partial a} W_0 \mathbf{h}(\mathbf{x}; \mathbf{a}(\mathbf{P}_{\theta_0})) \quad .$$

The IF of a GMME *is proportional to* the orthogonality function of the model

and is

bounded if and only if the function inducing the orthogonality conditions of the model is bounded.

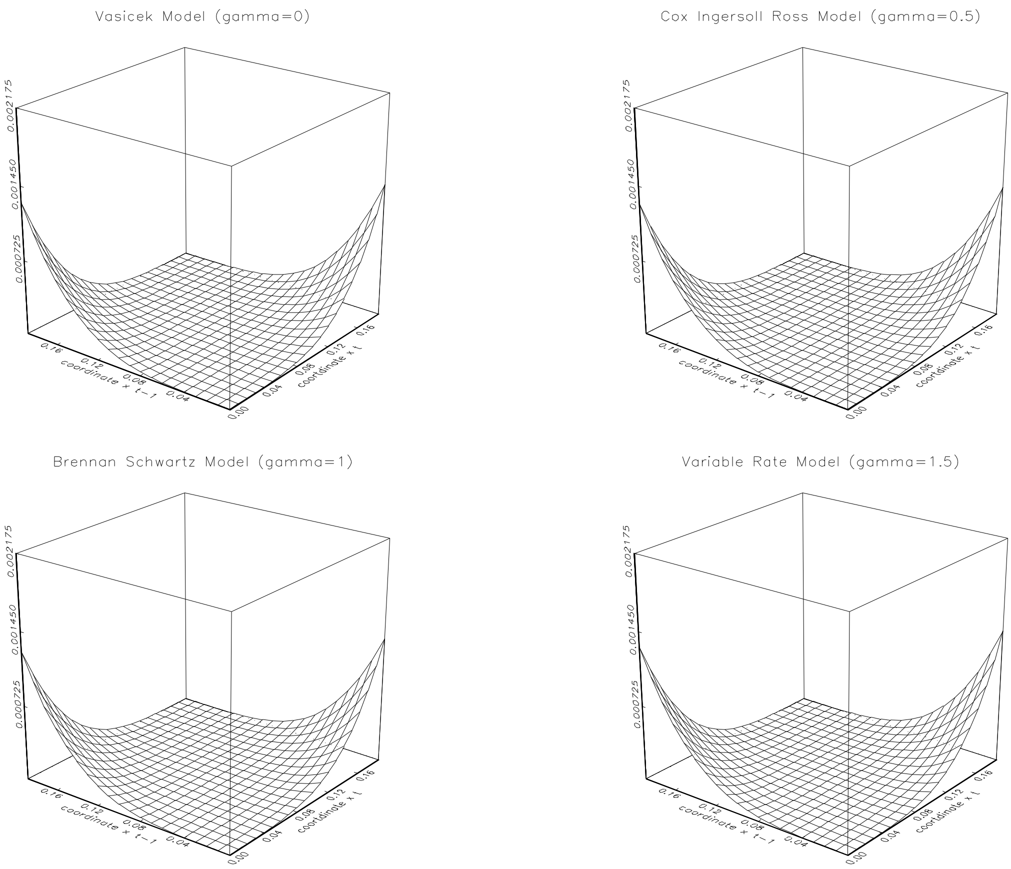
Examples of GMME with **unbounded** orthogonality conditions:

- linear and nonlinear LS

- instrumental variables estimators

In our model, the orthogonality function h is unbounded in x and y

The classical GMM estimator and tests corresponding to this orthogonality function are therefore not robust.



Function h for Vasiceck, Cox Ing. Ross, Brenn. Schw., Variab. Rate model

♦ Robust GMM Estimators

GMME with influence bounded by c

Huber function:

$$\mathcal{H}_c : R^H \to R^H ; y \mapsto y w_c(y),$$

where $w_c(y) := \min(1, \frac{c}{\|y\|})$
New orthogonality function:

$$h_c^{A,\tau} : R^N \times \mathcal{A} \to R^H$$

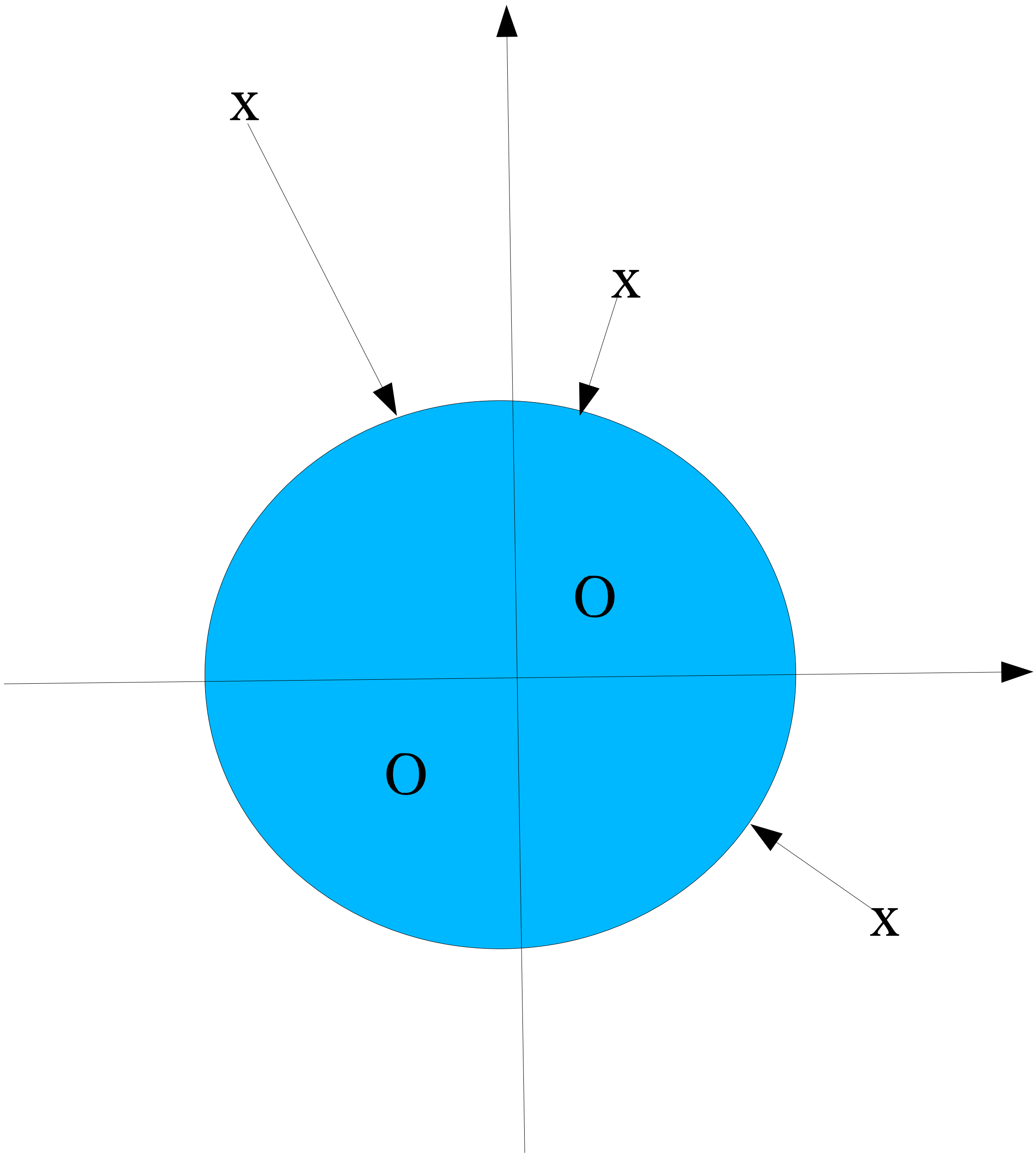
$$\begin{aligned} h_c^{A,\tau}(x,a) &:= \mathcal{H}_c(A[h(x;a)-\tau]) \qquad (2) \\ &= A[h(x;a)-\tau] w_c(A[h(x;a)-\tau]) \end{aligned}$$

where the nonsingular matrix $A \in R^{H\times H}$ and the vector $\tau \in R^H$ are determined by the implicit equations:

$$E_{\theta_0} h_c^{A,\tau}(X_1, a(P_{\theta_0})) = 0 \quad , \qquad (3)$$

and

$$E_{\theta_0} h_c^{A,\tau}(X_1, a(P_{\theta_0})) h_c^{\prime A,\tau}(X_1, a(P_{\theta_0})) = I. \; (4)$$



Sketch of the Huber function

- $h_c^{A,\tau}$ can be interpreted as a *truncated version* of h. Because of the truncation, h must be shifted by τ in order to satisfy the orthogonality condition (3). Moreover, (4) ensures that c is an upper bound on the self-standardized influence of the corresponding GMME, because – by construction – the selfstandardized norm of $h_c^{A,\tau}$ is equal to its euclidean norm which itself is bounded by c.

- The GGME $\tilde{a}_c^{A,\tau}$ associated to the modified orthogonality function $h_c^{A,\tau}$ is a consistent estimator for $a(P_{\theta_0})$ that is asymptotically best and robust.

- Whereas the original moment conditions h are usually dictated by economic theory, the truncated version $h_c^{A,\tau}$ takes into account the *realistic case* that only the "majority of the data" can reasonably fit the original moment conditions. The weights $w_c(A[h(x;a)-\tau])$ assigned to each observation x can be used to detect outlying points.

- The bound imposed on the self-standardized influence of *any* GMME cannot be chosen arbitrarily small. Indeed, $c \geq \sqrt{H}$.

- No further model assumptions are needed in order to do this construction.

♦ Robust Inference with GMM Estimators

Hansen's test: For the case $r := H - k > 0$ the GMM specification test is a test of the overidentifying restrictions implied by the null hypothesis (1). It is constructed with a functional $\xi^G(P_{\theta_n}) := U'^G(P_{\theta_n})U^G(P_{\theta_n})$, where

$$U^G(P_{\theta_n}) := W_0^{\frac{1}{2}} E_{\theta_n} h(X_1; \tilde{a}(P_{\theta_n})) \quad .$$

Can also consider GMM versions of the classical ML-tests (Wald, score, likelihood-ratio type) to test a null hypothesis

$$\{a \in \mathcal{A} \quad | \quad g(a) = 0\}$$

defined using a smooth function $g : \mathcal{A} \to R^r$.

Theorem: *Let $\tilde{a}$ be a GMME induced by an orthogonality function h and denote by $\alpha : dom(\alpha) \rightarrow R$ the level functional of the test.*

Let further $(P_{\epsilon,n,G})_{n\in N}$ be a sequence of ϵ, n, G–contaminations of the underlying null distribution P_{θ_0}, each of them belonging to a corresponding neighborhood

$$\mathcal{U}_{\epsilon,n}(P_{\theta_0}) := \left\{ P_{\epsilon,n,G} := (1 - \frac{\epsilon}{\sqrt{n}})P_{\theta_0} + \frac{\epsilon}{\sqrt{n}}G \right\}.$$

The following two statements then hold:

- *The bias of the level functional α under $P_{\epsilon,n,\delta_x}-$ contaminations is bounded if and only if h is bounded.*

- *If h is bounded, the bias of α is uniformly bounded by the inequality:*

$$\lim_{n\to\infty} \quad |\alpha(P_{\epsilon,n,G}) - \alpha_0| \leq \epsilon^2 \cdot \mu \sup_x (||h(x, a(P_{\theta_0}))||_{W_0})^2 + o(\epsilon^2) \quad ,$$

where

$$\begin{aligned} \mu &= -\frac{\partial}{\partial\beta} H_r(\eta_{1-\alpha_0}; \beta)|_{\beta=0} \\ &= \frac{(1-\alpha_0)}{2} - \frac{1}{2} H_{r+2}(\eta_{1-\alpha_0}; 0), \end{aligned}$$

$H_r(\cdot;\beta)$ is the cumulative distribution function of a noncentral $\mathcal{X}^2(r;\beta)$ distribution with r degrees of freedom and noncentrality parameter $\beta \geq 0$, and $\eta_{1-\alpha_0}$ is the $1-\alpha_0$ quantile of a $\mathcal{X}^2(r;0)$.

Maximal asymptotic bias of the level of tests that are derived from the robust GMME can be bounded by

$$\lim_{n\to\infty} |\alpha(P_{\epsilon,n,G}) - \alpha_0| \leq \mu \cdot (\epsilon c)^2 + o(\epsilon^2) \quad .$$

Can choose c in dependence of both the maximal amount of contamination (ϵ) expected by the researcher – given some prior information on the nature of the data – and the maximal bias for the level ($maxbias$), he (she) is going to accept from his (her) research strategy, i.e. $c = \frac{1}{\epsilon}\sqrt{maxbias/\mu}$.

For $\alpha_0 = 5\%$, this can be approximated by $c \approx \frac{3}{\epsilon} r^{0.3} (maxbias)^{1/2}$.

For instance, for $\epsilon = 5\%$ and $maxbias = 0.5\%$, we obtain values of c which vary from 4 for $r = 1$ to 6 for $r = 5$.

Similar result for the power.

Robust Analysis ($c = 6$)

- Robust GMM specifications tests *reject practically all constrained models at a 5% significance level*.

- Some observations are identified as potentially influential (e.g. March 1980).

This analysis shows that it is very difficult to distinguish – by means of the classical Hansen's test – the specification properties of the models, even when only a *single* observation is changed in the data.

A robust Hansen's test should be used if one is interested in obtaining decisions that are not primarily determined by a few observations in the sample.

Models with $\gamma \geq 1$ cannot be motivated, when using *robust* model selection strategies.

Influential Observations

- Some observations are identified as potentially influential.

- A clustering of influential observations is visible in the 1979-1982 period.

- Other influential observations are found in the 1973-1975 (the first OPEC crisis) and after the stock market crash in 1987.

- The 1979-1982 period is well-known to coincide with a temporary change in the monetary policy of the Federal Reserve.

- The influential observations coincide with the typical regimes found in regime switching studies, e.g. Gray(1996) who identified similar regimes at exactly those points where the infuential observations are found.

♦ Extensions of the CKLS Models

Misspecification of the CKLS models
Need sophisticated multi-factor models or more complex single factor models?

- Regime-switching models (e.g. as proposed in Cai (1994), Gray (1996), Ang and Bekaert (2000b))

- Models allowing for nonlinearities in the drift and diffusion term (e.g. Ait Sahalia (1996), Stanton (1997), Jiang (1998) and Ahn and Gao (1999))

- Models adding GARCH and similar features (e.g. Brenner, Harjes and Kroner (1996), Koedijk, Nissen, Schotman and Wolff (1997) and Ball and Torous (1999))

We re-analyze the Ahn and Gao (1999) model which adds a quadratic drift term. Dataset spanning the period Dec. 1942 - February 1991.

Model	α_1	α_2	α_3	σ	ξ	$d.f$
Cl. AG	-0.070	0.065	-0.007	0.037	4.5	2
	(-1.08)	(1.73)	(-1.80)	(19.93)	(0.103)	
Rob. AG	-0.028	0.043	-0.005	0.038	21.3	2
	(-0.60)	(1.50)	(-1.50)	(26.17)	(0.000)	

- The classical Hansen test does not reject the AG model, while the robust test does.

- The classical Hansen test is still highly sensitive with respect to a contamination of a single point.

- There are still influential observations clustering in the 1979-1982 period.

♦ Conclusion

It seems that adding a nonlinearity to the drift, does still not suffice to explain the 1979-1982 period. Similar conlusions are found for models allowing parameter shifts (e.g. Bliss and Smith(1999)).

Details can be found in :

Ronchetti, E. and Trojani, F. (2001)
"Robust Inference with GMM Estimators"
Journal of Econometrics, 101, 37-69.

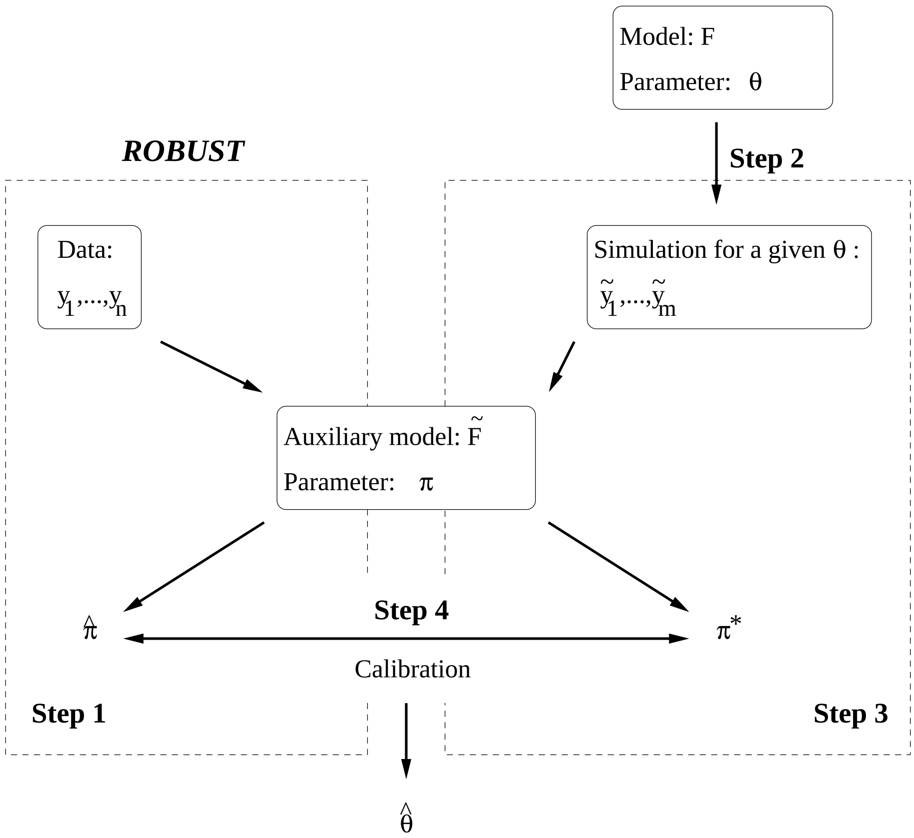
Dell'Aquila, R., Ronchetti, E., and Trojani, F. (2003)
"Robust GMM Analysis of Models for the Short Rate Process"
Journal of Empirical Finance, 10, 373-397.

♦ Indirect Inference

- Indirect inference to reduce the bias of estimators based on a crude discretization of the stoch. diff. eq.; cf. Gouriéroux and Monfort (1996)

- Study this property when the model does not hold exactly

- Geometric Brownian motion with drift; conclusions for more complex models



Schematic illustration of the robust indirect estimation algorithm

Special case:

$\pi = h(\theta) = E_\theta[T_n]$
for some estimator T_n.

"target estimator"
Cabrera & Fernholz(1999), *Ann. Stat.*

Has smaller MSE than T_n

♦ Geometric Brownian Motion With Drift

Assume that the price y_t of an asset satisfies the model

$$dy_t = \mu y_t dt + \sigma y_t dW_t, \tag{5}$$

where W_t is a Brownian motion and μ and σ are the drift and volatility parameters respectively.

Crude discretization of (5)

$$\begin{aligned} y_t &= y_{t-1} + \mu y_{t-1} + \sigma y_{t-1}\epsilon_t \\ &= (1+\mu)y_{t-1} + \sigma y_{t-1}\epsilon_t, \end{aligned} \tag{6}$$

where $\{\epsilon_t\}_{t=1,\ldots,n}$ are iid $N(0,1)$.

Log-likelihood function:

$$\tilde{l}_n = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\ \sigma^2 - \sum_{t=1}^{n}\log\ y_{t-1}$$
$$- \frac{1}{2}\sum_{t=1}^{n}\frac{[y_t - (1+\mu)y_{t-1}]^2}{\sigma^2 y_{t-1}^2} \quad (7)$$

MLE for (μ, σ^2):

$$\tilde{\mu} = \bar{r}_t - 1 \quad (8)$$
$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r}_t)^2, \quad (9)$$

where $r_t = y_t/y_{t-1}$.

Estimators based on the **crude discretization are biased**:

$$\begin{aligned} bias(\tilde{\mu}) &= E\tilde{\mu} - \mu \\ &= e^{\mu} - (1 + \mu) = O(1) \end{aligned}$$

Use indirect inference to eliminate this bias.

(6) is used as auxiliary model, (8) and (9) as auxiliary estimators, and the simulation is carried out from a finer discretization of (5). The resulting estimators $(\widehat{\mu}_I, \widehat{\sigma}_I^2)$ are then unbiased up to order $O(n^{-1})$.

Next proposition: investigate the bias of the auxiliary and indirect inference estimators **when the underlying model does not hold exactly**.

We still consider (5) as an ideal underlying model but we take into account that the observed prices are only **approximately** normally distributed.

Proposition: Assume model (6), where $\{\epsilon_t\}$ are iid variables with distribution
$(1-\varepsilon)N(0,1)+\varepsilon N(0,\tau^2)$,
$0 \leq \varepsilon \leq 1$ and $\tau \geq 1$.
Let $\tilde{\mu}$ be the auxiliary estimator for μ defined by (8), $\hat{\mu}_I$ the corresponding indirect estimator, and s the number of simulation runs.

$$
\begin{aligned}
bias(\tilde{\mu}, \varepsilon) &= E\tilde{\mu} - \mu \\
&= e^{\mu} - (1+\mu) + \varepsilon e^{\mu}(e^{\sigma^2(\tau^2-1)/2} - 1) \\
and & \\
bias(\widehat{\mu}_I, \varepsilon) &= E\widehat{\mu}_I - \mu \\
&= -\frac{1}{n}\frac{1}{2}(1+\frac{1}{s})(e^{\sigma^2} - 1 \\
&\quad +\varepsilon(e^{\sigma^2(2\tau^2-1)} - 2e^{\sigma^2(\tau^2+1)/2} + e^{\sigma^2})) \\
&\quad +O(\varepsilon^2) + O(n^{-2}).
\end{aligned}
$$

Two components of bias for $\tilde{\mu}$ and $\widehat{\mu}_I$. For $\tilde{\mu}$ the **bias due to the discretization** is

$$e^{\mu} - (1 + \mu)$$

and the **bias due to deviations** from the model is

$$\varepsilon e^{\mu}(e^{\sigma^2(\tau^2-1)/2} - 1)$$

Both components are positive.

The bias of the indirect estimator is negative and of order $O(n^{-1})$.

Fig. 2: Bias as a function of tau for the auxiliary estimator (...) and for the indirect inference estimator (-); mu=0.2,sigma=0.5,n=100,s=10

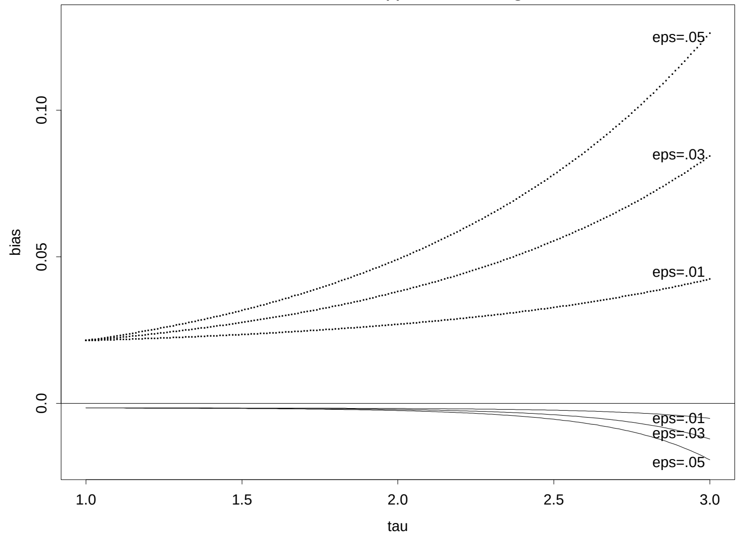
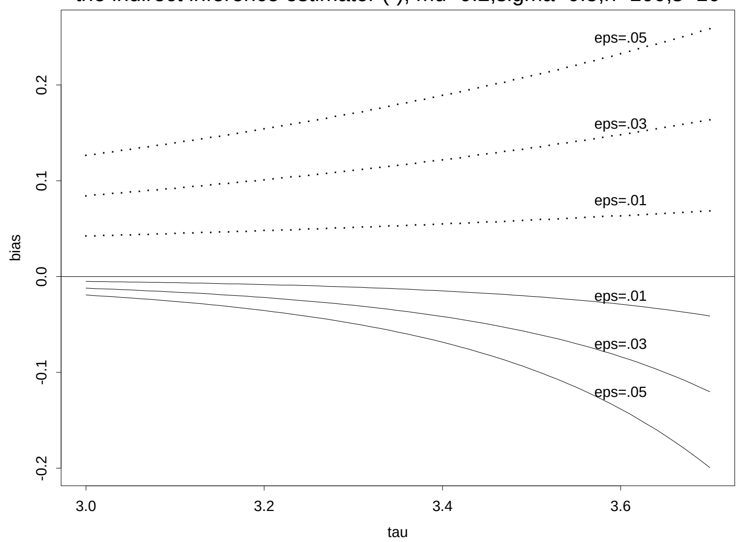


Fig. 3: Bias as a function of tau for the auxiliary estimator (...) and for the indirect inference estimator (-); mu=0.2,sigma=0.5,n=100,s=10



Therefore, the level of contamination is very low and it would be difficult to detect such a deviation from the model on a real sample.

Small deviations from the stochastic structure of the model can wipe out the bias improvement obtained with indirect inference.

This conclusion will carry out in more complex models.

$\Rightarrow$ Robust Indirect Estimators

♦ Simulation Study

- Estimators

(a) Exact discretization $(\widehat{\mu}, \widehat{\sigma}^2)$

(b) Exact discretization robust $(\widehat{\mu}_R, \widehat{\sigma}_R^2)$

(c) Crude discretization $(\tilde{\mu}, \tilde{\sigma}^2)$

(d) Crude discretization robust $(\tilde{\mu}_R, \tilde{\sigma}_R^2)$

(e) Indirect $(\widehat{\mu}_I, \widehat{\sigma}_I^2)$

(f) Indirect robust $(\widehat{\mu}_{RI}, \widehat{\sigma}_{RI}^2)$

- Distributions for ϵ_t: $N(0,1)$ and $0.95N(0,1)+0.05N(0,5^2)$

- Generate realizations of the process $\{y_t\}$ with a fine discretization ($\mu = 0.2$,$\sigma^2 = 0.25$) and draw samples of size $n = 100$

- The number of simulations s in the indirect procedure is chosen to be 10

- Tuning constant for the robust estimators: $c = 1.345$ (95% efficiency at model)

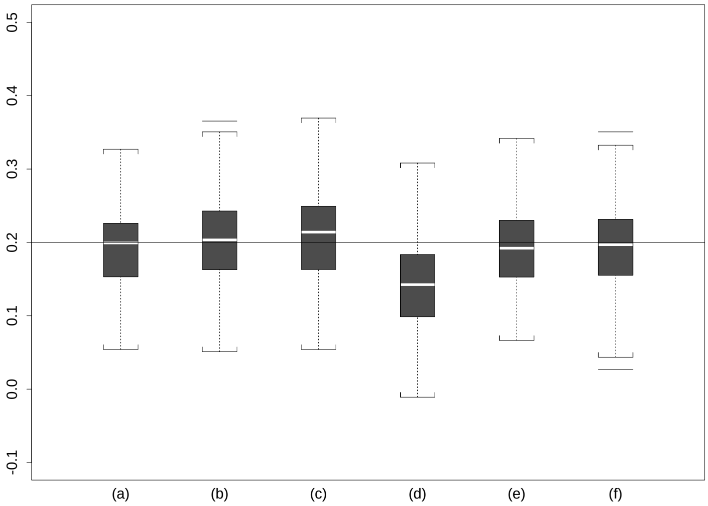


Fig. 4: Estimation of μ without contamination

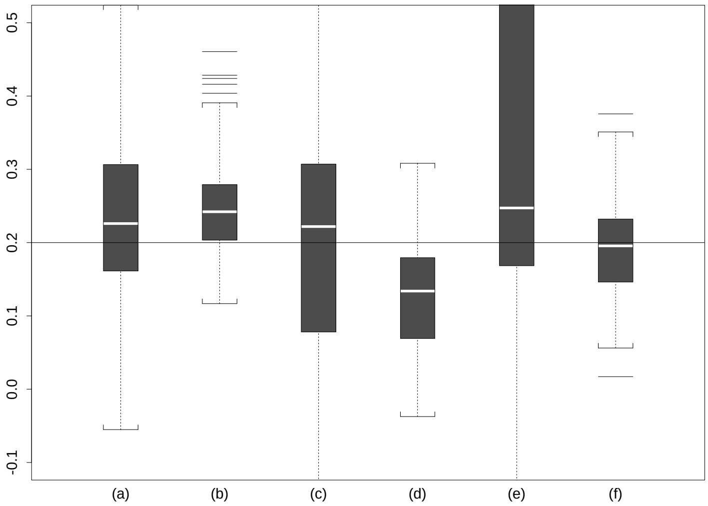


Fig. 5: Estimation of μ with 5% additive outlier contamination

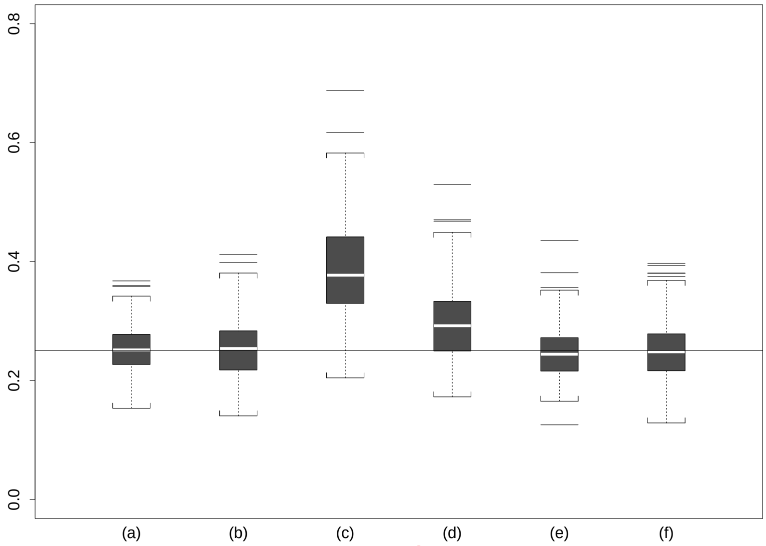


Fig. 6: Estimation of σ^2 without contamination

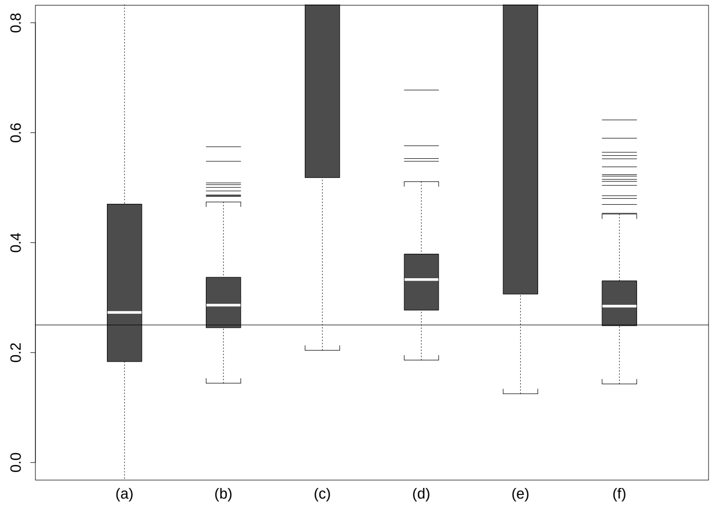


Fig. 7: Estimation of σ^2 with 5% additive outlier contamination

♦ Discussion

Figure 4: estimation of μ under the model

- Crude estimators (c) and (d) are biased under the model

- Bias corrected by their indirect estimators (e) and (f) (cf. Figure 1 at $\tau = 1$)

- Indirect estimators behave like the estimators obtained by the "exact discretization"

- Robust estimators are a little more variable than the classical ones

Figure 6: estimation of σ^2 under the model

- Overall picture is the same, but larger gains in bias reduction

Figure 5: estimation of μ under the contaminated model

- Crude estimators (c) and (d) are again biased and especially (c) exhibits large variability

- Classical indirect estimator (e) cannot correct the bias and reduce the variability

- Robust indirect estimator (f) shows overall the best performance

- Under contamination the estimators obtained by exact discretization are worse in terms of bias and variance than the robust indirect estimators

Figure 7: estimation of σ^2 under the contaminated model

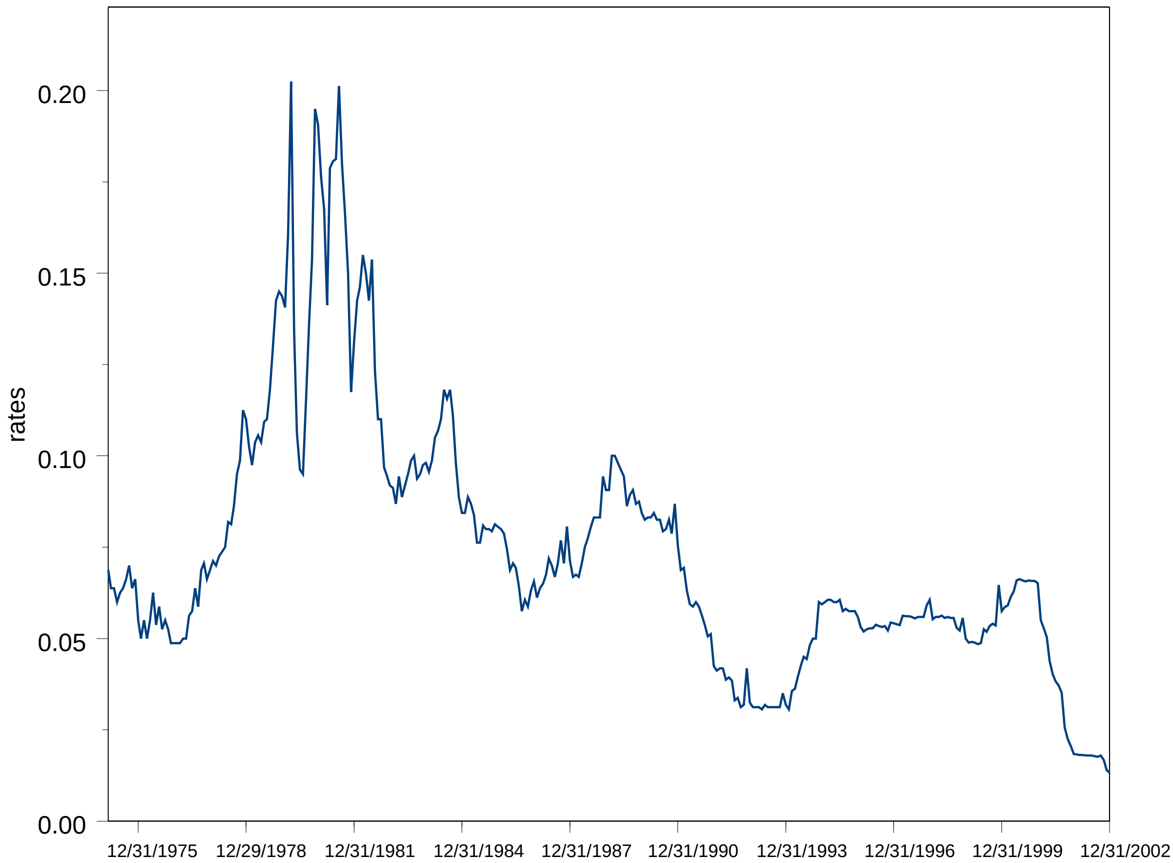
- Effects much larger than in the case of μ and same conclusions

To summarize:

The robust indirect estimators for μ and σ^2 exhibit a good performance in terms of bias and variance **even when the model is not exact** and the data are generated by a distribution in a neighborhood of the model.

Moreover, **indirect inference is a general procedure** that can be carried out for general models, for instance when an exact discretization is not available.

♦ Forecasting

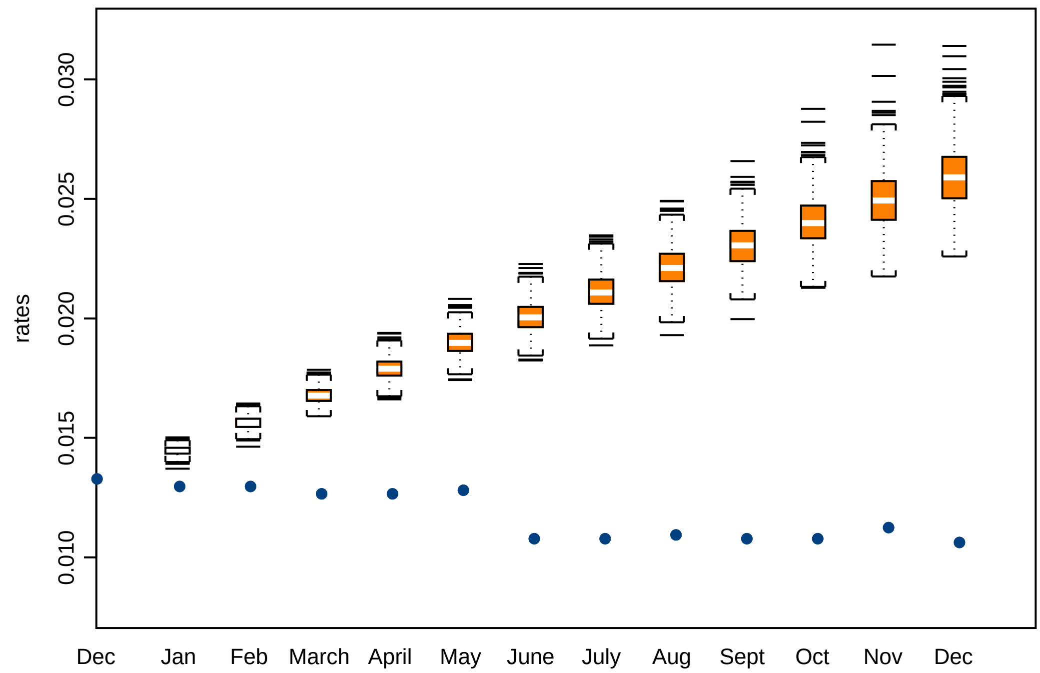


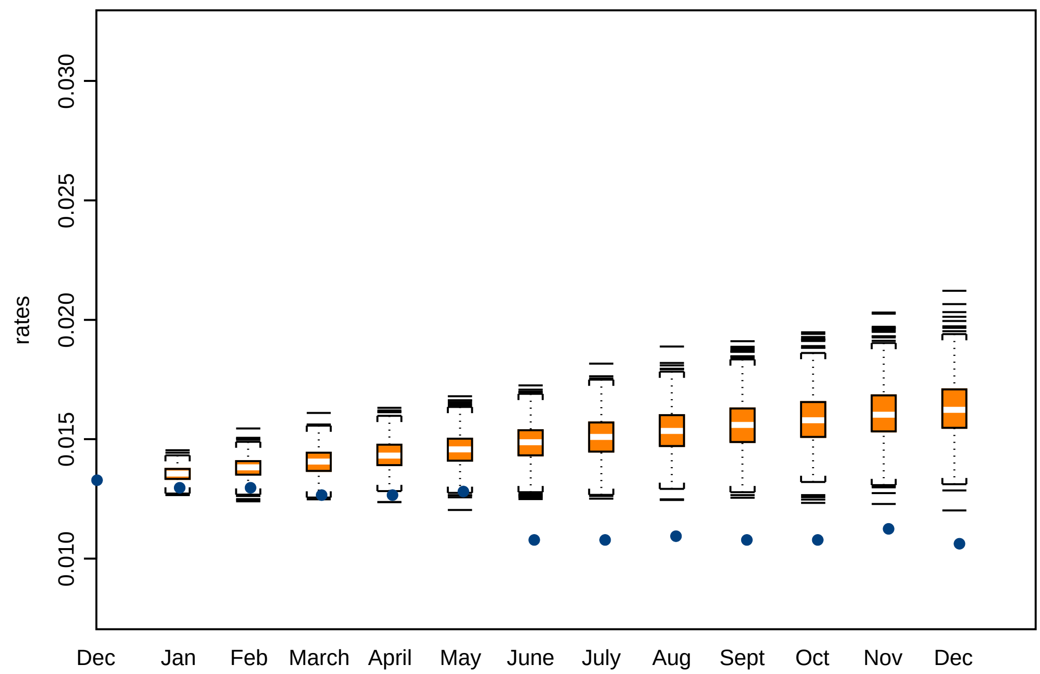
US Eurocurrency rates from
February 28, 1975 to December 31, 2002

Simulate from
model with crude discretization with parameter estimates
and
errors ϵ_t generated from a standard normal distribution.

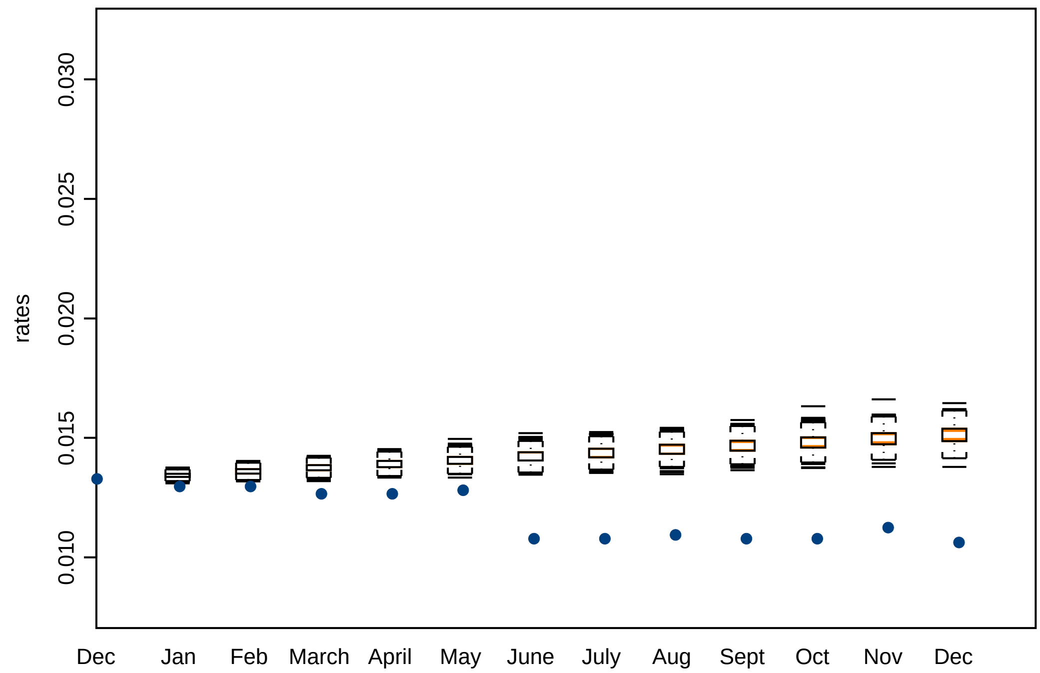
This produces a path of simulated future monthly values in 2003 starting from the last observed interest rate in December 2002.

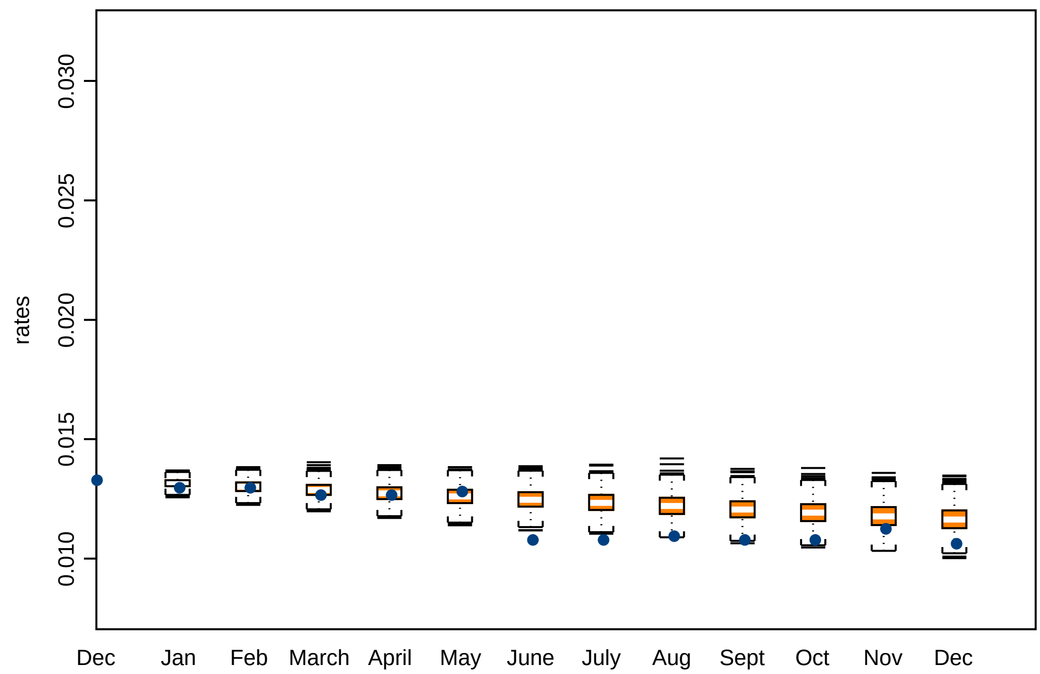
Simulate 1000 such paths.
The next figures present the simulated future values by means of boxplots.
Actual values that occurred in 2003 are indicated by circles.
In each figure the parameters are estimated by GMM, RGMM, IGMM, IRGMM, respectively.





Forecast for US Eurocurrency rates for 2003 using GMM and RGMM.





Forecast for US Eurocurrency rates for 2003 using IGMM and IRGMM.

Compare numerically the goodness-of-fit of different forecasting techniques by the following measures:

$$\widehat{RAMSE} = \Big(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{m}\sum_{s=1}^{m}(y_t^{(s)} - y_t)^2\Big)^{\frac{1}{2}}, \qquad (10)$$

$$\widehat{AMAD} = \frac{1}{n}\sum_{t=1}^{n}\operatorname*{median}_{s}\Big(|y_t^{(s)} - \operatorname*{median}_{l}(y_t^{(l)})|\Big), \qquad (11)$$

$$\widehat{AMEDBIAS} = \frac{1}{n}\sum_{t=1}^{n}\Big(|\operatorname*{median}_{s}(y_t^{(s)}) - y_t|\Big), \qquad (12)$$

$\{y_t^{(s)}\}_{t=1,\ldots,n}$, $s = 1, \ldots, m$: m simulated values for n periods,
$\{y_t\}_{t=1,\ldots,n}$: real observations.

$\widehat{AMEDBIAS}$: bias of the the forecast,
$\widehat{AMAD}$: variability of the forecast,
$\widehat{RAMSE}$: the root mean squared error of the forecast, a combination of bias and variability.

	α	β	σ	γ
GMM	0.0015 (0.80)	-0.0225 (-0.75)	0.7333 (1.50)	1.9305 (6.73)
IGMM	0.0002 (0.11)	-0.0035 (-0.11)	0.8278 (0.91)	2.0993 (4.31)
RGMM	0.0003 (0.29)	-0.0058 (-0.30)	0.3167 (1.99)	1.6111 (7.79)
IRGMM	-0.0002 (-0.10)	0.0023 (0.08)	0.5240 (1.64)	1.8398 (7.25)

Parameter estimates for US Eurocurrency rates

	$\widehat{RAMSE}$	$\widehat{AMAD}$	$\widehat{A.M.B.}$	$\cdot 10^{-3}$
GMM	9.28	0.42	7.92	
IGMM	2.87	0.16	2.39	
RGMM	3.60	0.49	2.93	
IRGMM	1.02	0.26	0.70	

Prediction quality for US Eurocurrency rates

Similar results for

- forecast with 5-years periods of weekly data

- Eurocurrency rates for Switzerland, UK, and Japan.

Details can be found in:

Genton, M.G. and Ronchetti, E. (2003)
"Robust Indirect Inference"
Journal of the American Statistical Association, 98, 67-76.

Czellar, V., Karolyi, G. A., and Ronchetti, E. (2007)
"Indirect Robust Estimation of the Short-term Interest Rate Process"
Journal of Empirical Finance, 14, 546-563.

♦ Nonparametric Estimation of Drift and Volatility

First order approximations for μ and σ:

$$\mu(r_t) = \frac{1}{\Delta} E[r_{t+\Delta} - r_t \mid r_t] + o(\Delta) \quad ,$$

and

$$\sigma^2(r_t) = \frac{1}{\Delta} E[(r_{t+\Delta} - r_t)^2 \mid r_t] + o(\Delta) \quad .$$

Regress

$r_{t+\Delta} - r_t \quad vs \quad r_t$

and

$(r_{t+\Delta} - r_t)^2 \quad vs \quad r_t$.

Kernel estimators

Smoothing splines

♦ Robust Estimation and Inference for Dynamic Location-Scale Models

♦ Dynamic Location-Scale Models

$\{y_t\}_{t\in\mathbb{Z}}$ a real valued strictly stationary random sequence

y_t has a conditionally Gaussian distribution
$y_t|\mathcal{F}_{t-1} \sim \mathcal{N}(\mu_t(\theta), \sigma_t^2(\theta))$ i.e.

$$\begin{aligned} y_t &= \mu_t(\theta) + \varepsilon_t(\theta), \\ \varepsilon_t^2(\theta) &= \sigma_t^2(\theta) + \nu_t(\theta), \end{aligned}$$

where
$\mu_t(\theta) = E[y_t|\mathcal{F}_{t-1}]$ and $\sigma_t^2(\theta) = var[y_t|\mathcal{F}_{t-1}]$ parameterize the conditional mean and the conditional variance of y_t given the information $\mathcal{F}_{t-1}$ up to time $t-1$.

- ARMA Box & Jenkins(1975)

$$\begin{aligned}\mu_t(\theta) &= \rho_0 + \rho_1 y_{t-1}\\ \sigma_t^2(\theta) &= \sigma^2\end{aligned}$$

$\rho_0 \in \mathbb{R},\ |\rho_1| < 1.$

- ARCH Engle(1982)

$$\begin{aligned}\mu_t(\theta) &= \rho_0 + \rho_1 y_{t-1}\\ \sigma_t^2(\theta) &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2(\theta)\\ &= \alpha_0 + \alpha_1 (y_{t-1} - \rho_0 - \rho_1 y_{t-2})^2\end{aligned}$$

$\rho_0 \in \mathbb{R},\ |\rho_1| < 1,\ \alpha_0 > 0,\ 0 \le \alpha_1 < 1.$

- GARCH Bollerslev(1986)

$$\begin{aligned}\mu_t(\theta) &= 0 \quad (y_t = \varepsilon_t)\\ \sigma_t^2(\theta) &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \delta_1 \sigma_{t-1}^2(\theta)\\ &= \alpha_0/(1-\delta_1) + \alpha_1 \sum_{j=0}^{\infty} \delta_1^j y_{t-1-j}^2\end{aligned}$$

$\alpha_0, \alpha_1, \delta_1 > 0,\ \ \alpha_1 + \delta_1 < 1.$
GARCH model is an ARCH model with an infinite number of lagged y variables.

- Double threshold ARCH

  Glosten, Jagannathan, Runkle(1993)

  Li & Li(1996)

$$\begin{aligned}
\mu_t(\theta) &= \rho_0 + (\rho_1 + \rho_2\, d_{1,t-1})\, y_{t-1} \\
\sigma_t^2(\theta) &= \alpha_0 + (\alpha_1 + \alpha_2\, d_{2,t-1}) \\
&\quad (y_{t-1} - \rho_0 - (\rho_1 + \rho_2\, d_{1,t-2})\, y_{t-2})^2 \\
&\quad + \alpha_3\, d_{1,t-1}
\end{aligned}$$

  with the dummy variable

  $d_{1,t-1} = 1$ if $\rho_0 + \rho_1\, y_{t-1} > 0$ and 0 othw.

  $d_{2,t-1} = 1$ if $\varepsilon_{t-1}(\theta) < 0$ and 0 othw.

Estimation

Conditional moment condition:
$E_\theta[\psi(y_1, ..., y_m; a(\mathbb{P}_\theta^m)|\mathcal{F}_{m-1}] = 0$.

For example, $\psi = s$, the conditionally Gaussian score function

$$\begin{aligned} s(y_1, ..., y_m; \theta) &= -k_{1,m} + k_{2,m} u_m(\theta) \\ &\quad + k_{1,m} u_m(\theta)^2, \end{aligned}$$

$u_m(\theta) = \varepsilon_m(\theta)\sigma_m(\theta)^{-1}$,
$k_{1,m} := \frac{1}{2\sigma_m^2(\theta)}\frac{\partial\sigma_m^2(\theta)}{\partial\theta}$,
$k_{2,m} := \frac{1}{\sigma_m(\theta)}\frac{\partial\mu_m(\theta)}{\partial\theta}$
defines a conditionally unbiased estimator of θ.

But non-robust
$\Rightarrow$ More general ψ functions (conditional M-estimators);

Mancini, Ronchetti, Trojani (2005)
J. Am. Stat. Ass. (MATLAB)

♦ RGMM Testing For Conditional Heteroscedasticity

ARCH(1,1)

$$
\begin{aligned}
y_t &= \rho_0 + \rho_1 y_{t-1} + \sigma_t u_t \quad , \\
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2
\end{aligned}
$$

where $(u_t)_{t\in\mathbb{N}}$ is a standardized i.i.d sequence of r.v.

Orthogonality conditions for a GMM estimation of the parameters $(\alpha_0, \alpha_1, \beta_0, \beta_1)$:

$$
\begin{array}{rr}
E[\epsilon_t] = 0 \quad , & E[\epsilon_t y_{t-1}] = 0 \quad , \\
E[\eta_t - h_t] = 0 \quad , & E[\eta_t \eta_{t-1}] = 0 \quad ,
\end{array}
$$

where $\epsilon_t = y_t - \rho_0 - \rho_1 y_{t-1} \quad , \quad \eta_t = \epsilon_t^2 \quad .$

Orthogonality conditions **unbounded**!

Test $\alpha_1 = 0$ vs $\alpha_1 > 0$

Compare level and power of classical and robust GMM tests under the following distributions of $\{u_t\}$

- Standard normal $N(0, 1)$
- Contaminated normal $CN(\epsilon, K^2)$
  $\epsilon = 0.05, K = 10$
- Student t_ν , $\nu = 5, 9$
- Double exponential

- $(\rho_0, \rho_1, \alpha_0) = (0.4, 0.3, 0.25)$

- $\alpha_1 \quad : 0 - 0.3$

- $T = 250, 500, 1000$

- 1000 simulations

- Tuning constant $c = 2.09$
  ($\epsilon = 10\%$, max bias level $= +/- 0.5\%$)

Table 2: GMM and RGMM Simulation Results under $u_t \sim \mathcal{N}(0,1)$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the $\mathcal{X}^2$ test. The constant c for the RGMM test was set to $c = 2.09$.

	GMM			RGMM		
α_1; T	250	500	1000	250	500	1000
0.00	0.08	0.08	0.05	0.02	0.02	0.02
0.05	0.05	0.09	0.19	0.02	0.06	0.07
0.10	0.09	0.28	0.62	0.06	0.14	0.29
0.15	0.20	0.52	0.90	0.12	0.31	0.62
0.20	0.32	0.74	0.97	0.21	0.51	0.87
0.25	0.45	0.84	0.98	0.35	0.71	0.95
0.30	0.56	0.89	0.98	0.49	0.86	0.99

Table 3: GMM and RGMM Simulation Results under $u_t \sim DE$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the $\mathcal{X}^2$ test. The constant c for the RGMM test was set to $c = 2.09$.

	GMM			RGMM		
α_1; T	250	500	1000	250	500	1000
0.00	0.11	0.10	0.09	0.03	0.03	0.03
0.05	0.04	0.04	0.09	0.03	0.06	0.12
0.10	0.04	0.12	0.31	0.07	0.14	0.32
0.15	0.06	0.23	0.54	0.11	0.26	0.58
0.20	0.10	0.33	0.71	0.18	0.41	0.78
0.25	0.16	0.43	0.78	0.26	0.57	0.91
0.30	0.21	0.50	0.79	0.34	0.70	0.96

Table 4: GMM and RGMM Simulation Results under $u_t \sim t_9$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the $\mathcal{X}^2$ test. The constant c for the RGMM test was set to $c = 2.09$.

	GMM			RGMM		
α_1; T	250	500	1000	250	500	1000
0.00	0.09	0.09	0.07	0.02	0.02	0.02
0.05	0.05	0.05	0.11	0.04	0.06	0.10
0.10	0.05	0.16	0.42	0.09	0.14	0.30
0.15	0.12	0.35	0.69	0.13	0.31	0.62
0.20	0.21	0.54	0.83	0.23	0.50	0.84
0.25	0.30	0.65	0.87	0.35	0.83	0.95
0.30	0.38	0.73	0.88	0.46	0.83	0.99

Table 5: GMM and RGMM Simulation Results under $u_t \sim t_5$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the $\mathcal{X}^2$ test. The constant c for the RGMM test was set to $c = 2.09$.

	GMM			**RGMM**		
α_1; T	250	500	1000	250	500	1000
0.00	0.10	0.11	0.11	0.02	0.02	0.03
0.05	0.05	0.05	0.06	0.03	0.07	0.11
0.10	0.06	0.10	0.24	0.05	0.14	0.33
0.15	0.11	0.18	0.43	0.11	0.28	0.61
0.20	0.15	0.29	0.59	0.17	0.46	0.82
0.25	0.21	0.40	0.67	0.29	0.64	0.93
0.30	0.27	0.48	0.71	0.40	0.78	0.97

Table 6: GMM and RGMM Simulation Results under $u_t \sim CN(0.05, 100)$

Each entry in the Table corresponds to the empirical rejection frequency of the hypothesis $\alpha_1 = 0$ obtained using 5% critical values for the $\mathcal{X}^2$ test. The constant c for the RGMM test was set to $c = 2.09$.

	GMM			RGMM		
α_1; T	250	500	1000	250	500	1000
0.00	0.35	0.51	0.48	0.02	0.01	0.02
0.05	0.16	0.19	0.17	0.02	0.03	0.06
0.10	0.09	0.08	0.05	0.03	0.06	0.14
0.15	0.06	0.04	0.02	0.06	0.11	0.24
0.20	0.04	0.03	0.03	0.07	0.16	0.36
0.25	0.04	0.03	0.06	0.10	0.22	0.48
0.30	0.04	0.04	0.11	0.13	0.28	0.60

- Robust test yields **stable level and power** across distributions

- Classical test shows a **drastic liberal behavior**

- The **power advantage** of the classical test under normality is **lost** for **very small deviations from normality**

Empirical Application

Apply classical and robust Wald tests for ARCH structure to weekly exchange rate returns of the Swedish krona against US dollar Period Nov. 29$^{\text{th}}$, 1993 - Nov. 17$^{\text{th}}$, 2003; 522 observations (from Datastream).

'Regular' time series with no clear outlier.

The first ten sample autocorrelations of squared and absolute returns are not significantly different from zero.

Moreover, the Jarque-Bera test has a p-value of 0.47 not rejecting normality.

Estimates for the parameters
ρ_0, ρ_1, α_0 and α_1 of an AR(1)-ARCH(1) model and (Wald test p-values for the hypothesis that the corresponding parameter is zero):

Classical
0.02 (0.73), -0.030 (0.53), 1.86 (0), 0.06 (**0.22**)

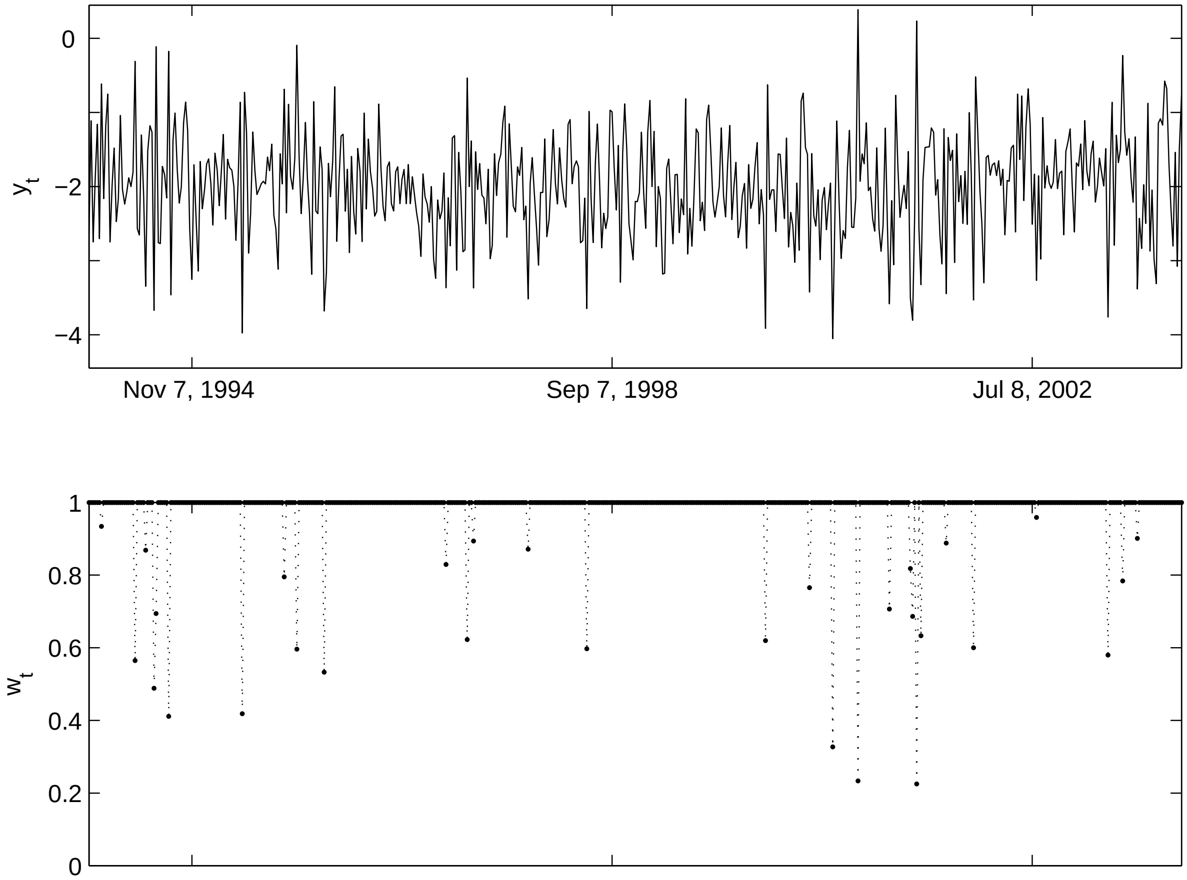
Robust ($c = 4$)
0.01 (0.88), 0.014 (0.75), 1.64 (0), 0.47 (**0**).

As in typical financial return series, the conditional mean parameters are not significantly different from zero.

Moreover, the classical estimate of the ARCH parameter α_1 is also not significant. Hence, the classical Wald test does not reject the homoscedasticity hypothesis.

By contrast, the robust estimate of this ARCH parameter is highly significant, showing that ARCH effects in the data are possibly obscured by some outlying observations detected by the robust weights.

It is interesting to notice that one would expect outliers to *enhance* the ARCH structure. Instead, because the estimation of the volatility by classical techniques is inflated, the potential ARCH structure is hidden by the presence of a few outlying observations.



Weekly exchange rate returns of the Swedish krona vs. US dollar (Nov 29, 1993 - Nov 17,2003 (top panel) and weights implied by the rob. est. of the AR(1)-ARCH(1) model with $c = 4$ (bottom panel).

Mancini, Ronchetti, Trojani (2005)
J. Am. Stat. Ass.

♦ **Other Topics and Outlook**

♦ Robust estimation of beta

- Sharpe (1978) (L_1)
- Chan&Lakonishok (1992), *J. of Fin. and Quant. Analysis*
- Knez & Ready (1997), *Journal of Finance*
- Martin & Simin (2003), *Financial Analysts Journal*

- ♦ Modelling the tails of the distribution (VaR etc.)

  Use:

  - Dupuis (1997), *J. of Hydrology*
  - Dupuis & Field (1998), *Can. Journ. of Stat.*
  - Dupuis(1998), *Extremes*

- ♦ Robust Dynamic Conditional Correlations

  - Extend Engle (2002), *J. Business & Econ. Stat.* (open)

ROBUST STATISTICS IN FINANCE
References
E. Ronchetti
2006

General Books on Statistics in Finance

Carmona, R. A. (2004), *Statistical Analysis of Financial Data in S-PLUS*, New York: Springer.

McNeil, A. J., Frey, R., and Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press.

Zivot, E., and Wang, J. (2003), *Modeling Financial Time Series with S-PLUS* New York: Springer.

Books on Robust Statistics

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

Dynamic Location-Scale Models

Mancini, L., Ronchetti, E., and Trojani, F. (2005), Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models, *Journal of the American Statistical Association*, **100**, 628-641.

Single Factor Model

Ait-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review Financial Studies*, **12**, 721-762.

Chan, K.G.G., Karolyi, G.A., Longstaff, F.A., and Sanders, A.B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Rate," *Journal of Finance*, **47**, 1209-1227.

Ahn, D.H., and Gao, B. (1999), "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, **12**, 721-762.

GMM Estimation

Gourieroux, C., and Monfort, A. (1995), *Statistics and Econometric Models*, Cambridge University Press.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, **50**, 1029-1054.

Robust GMM Estimation

Dell'Aquila, R., Ronchetti, E., and Trojani, F. (2003), "Robust GMM Analysis of Models for the Short Rate Process," *Journal of Empirical Finance*, **10**, 373-397.

Ronchetti, E., and Trojani, F. (2001), "Robust Inference with GMM Estimators," *Journal of Econometrics*, **101**, 37-69.

Indirect Inference

Czellar, V., Karolyi, G.A., and Ronchetti, E. (2007), "Indirect Robust Estimation of the Short-term Interest Rate Process," *Journal of Empirical Finance*, **14**, 546-563.

Genton, M.G., and Ronchetti, E. (2003), "Robust Indirect Inference," *Journal of the American Statistical Association*, **98**, 67-76.

Gourieroux, C., and Monfort, A. (1996), *Simulation Based Econometric Methods*, Oxford University Press.

Gourieroux, C., Monfort, A., and Renault, A. E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, **8**, 85-118.

Heggland, K., and Frigessi, A. (2004), "Estimating Functions in Indirect Inference," *Journal of the Royal Statistical Society, Series B*, **66**, 447-462.

Nonparametric Techniques

Ait-Sahalia, Y., Lo, A.W. (1998), "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, **53**, 499-547.

Green, P. J., and Silverman, B. W.(1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, New York: Chapman and Hall.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, New York: Chapman and Hall.

Stanton, R. (1997), "A Nonparametric Model of the Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, **52**, 1973-2002.