Signal Extraction in Long Memory Stochastic Volatility

Josu Arteche
Universidad del País Vasco UPV/EHU
III_t **Workshop in Time Series Econometrics**

Zaragoza, Abril 2013

Structure of the talk

- Introduction.
  - Motivation.
  - LMSV models.
- Signal extraction for volatility estimation in LMSV.
  - Kalman filter.
  - Time domain methods (Harvey, 1998).
  - Frequency domain methods.
- Finite sample behaviour.
- Application to Dow Jones Industrial index.

Structure of the talk

- Introduction.
  - Motivation.
  - LMSV models.
- Signal extraction for volatility estimation in LMSV.
  - Kalman filter.
  - Time domain methods (Harvey, 1998).
  - Frequency domain methods.
- Finite sample behaviour.
- Application to Dow Jones Industrial index.

Structure of the talk

- Introduction.
  - Motivation.
  - LMSV models.
- Signal extraction for volatility estimation in LMSV.
  - Kalman filter.
  - Time domain methods (Harvey, 1998).
  - Frequency domain methods.
- Finite sample behaviour.
- Application to Dow Jones Industrial index.

Structure of the talk

- Introduction.
  - Motivation.
  - LMSV models.
- Signal extraction for volatility estimation in LMSV.
  - Kalman filter.
  - Time domain methods (Harvey, 1998).
  - Frequency domain methods.
- Finite sample behaviour.
- Application to Dow Jones Industrial index.

- Why Long Memory in Stochastic Volatility?

- Why Long Memory in Stochastic Volatility?
  - Persistent autocorrelation in proxys of the volatility of financial time series (squares or other powers of absolute values)$\Rightarrow$ Long memory in volatility is a stylized fact.

- Why Long Memory in Stochastic Volatility?
  - Persistent autocorrelation in proxys of the volatility of financial time series (squares or other powers of absolute values)$\Rightarrow$ Long memory in volatility is a stylized fact.
  - SV models more flexible than ARCH based models (for example, contrary to ARCH extensions, covariance stationarity and long memory in squares is possible).

- Why ARCH based models more popular among empirical researchers?

- Why ARCH based models more popular among empirical researchers?
  - **Estimation**: MLE much easier in ARCH $\Rightarrow$ Recent advances in LMSV using Whittle QMLE (Breidt et al, 1998, Zaffaroni, 2009, JoE).

- Why ARCH based models more popular among empirical researchers?
  - **Estimation**: MLE much easier in ARCH $\Rightarrow$ Recent advances in LMSV using Whittle QMLE (Breidt et al, 1998, Zaffaroni, 2009, JoE).
  - **Volatility estimation**: Conditional variances (volatility) are very easily obtained in ARCH based models because they are deterministic functions of the past.

- Why ARCH based models more popular among empirical researchers?
  - **Estimation**: MLE much easier in ARCH $\Rightarrow$ Recent advances in LMSV using Whittle QMLE (Breidt et al, 1998, Zaffaroni, 2009, JoE).
  - **Volatility estimation**: Conditional variances (volatility) are very easily obtained in ARCH based models because they are deterministic functions of the past.

Goal of this paper: propose a simple to implement, general and robust technique of volatility extraction in LMSV

The series (returns) is defined as

$$z_t = \sigma \sigma_t \epsilon_t$$

- $\sigma > 0$ a scale factor,
- $\epsilon_t \sim iid(0, 1)$,
- $\sigma_t = \exp(x_t/2)$ for x_t a long memory process with a spectral density (pseudo spectral density in the nonstationary case) function

$$f_x(\lambda) = \lambda^{-2d} g_x(\lambda) \; , \quad 0 < \lambda \leq \pi$$

  - $0 < d < 1$ (stationary or nonstationary but mean reverting x_t),
  - $g_x(\lambda)$ positive, finite, symmetric around the origin and twice continuously differentiable.

Denoting $y_t = \log z_t^2$

$$\boxed{y_t = \mu + x_t + u_t}$$

where

- $\mu = \log \sigma^2 + E \log \epsilon_t^2$
- $u_t = \log \epsilon_t^2 - E \log \epsilon_t^2$ is a mean zero white noise process with finite variance σ_u^2.

Denoting $y_t = \log z_t^2$

$$\boxed{y_t = \mu + x_t + u_t}$$

where

- $\mu = \log \sigma^2 + E \log \epsilon_t^2$
- $u_t = \log \epsilon_t^2 - E \log \epsilon_t^2$ is a mean zero white noise process with finite variance σ_u^2.

The estimation of the volatility component x_t is then just a particular case of signal extraction in a long memory signal plus white noise process.

A.1: If $0 < d_0 < 1/2$ then $x_t = v_t$ and for $1/2 \leq d_0 < 1$ then $x_t = x_0 + \sum_{s=1}^{t} v_s$ for x_0 a random variable not depending on t and $v_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$, $\sum_{j=0}^{\infty} b_j^2 < \infty$ where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$, $E(\varepsilon_t^3) < \infty$, $E(\varepsilon_t^4) < \infty$.

A.2: The spectral density of v_t is

$$f_v(\lambda) = \lambda^{-2d_v} g_v(\lambda) \ , \quad 0 < \lambda \leq \pi \ ,$$

$0 < d_v = d_0 < 1/2$ (if x_t stationary) and $-1/2 \leq d_v = d_0 - 1 < 0$ (nonstationary x_t).

A.3: u_t is zero i.i.d. with finite fourth moment.

A.4: u_t and ε_s are uncorrelated at all leads and lags (correlation between ε and ϵ allowed).

A.5: $cum(\varepsilon_t, \varepsilon_s, u_l, u_m) = k < \infty$ if $t = s = l = m$ and 0 otherwise.

Kalman Filter

- A parametric specification of x_t is needed.
- Huge dimension of the state space representation in long memory series (Chan and Palma, 1998).

Kalman Filter

- A parametric specification of x_t is needed.
- Huge dimension of the state space representation in long memory series (Chan and Palma, 1998).
- **Alternatively**: Use a truncated AR (better than MA because AR coefficients converge faster to 0) and use smoothing for volatility estimation.

Kalman Filter

- A parametric specification of x_t is needed.
- Huge dimension of the state space representation in long memory series (Chan and Palma, 1998).
- **Alternatively**: Use a truncated AR (better than MA because AR coefficients converge faster to 0) and use smoothing for volatility estimation.

  Problems:

  - The number of parameters to be estimated increases with the truncation.
  - Large dimension of the state space model (a large truncation is needed), large number of parameters to be estimated and large sample sizes $\Rightarrow$ KF quite computationally demanding and subject to numerical inaccuracies.

Wiener-Kolmogorov filter in the time domain

- Harvey (1998), for the stationary case, proposed a linear estimator of x_t based on a Wiener-Kolmogorov filter that minimizes the mean square error

$$\tilde{x} = (I - \sigma_u^2 \Sigma_y^{-1})(y - \mu)$$

where Σ_y is the variance covariance matrix of y and σ_u^2 is the variance of the noise.

Wiener-Kolmogorov filter in the time domain

● Harvey (1998), for the stationary case, proposed a linear estimator of x_t based on a Wiener-Kolmogorov filter that minimizes the mean square error

$$\tilde{x} = (I - \sigma_u^2 \Sigma_y^{-1})(y - \mu)$$

where Σ_y is the variance covariance matrix of y and σ_u^2 is the variance of the noise. Problems:

- Requires inversion of Σ_y, which can be rather computationally demanding if n is large.
- Variances and covariances have to be estimated and the quality of the estimates significantly affects the signal extraction.
- Only valid under stationary long memory.

Wiener-Kolmogorov filter in the frequency domain

- We consider a frequency domain minimum MSE linear estimator of x_t defined as

$$x_{t|\infty} = \sum_{j=-\infty}^{\infty} \psi_j (y_{t-j} - \mu)$$

where

$$\psi_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_{xy}(\lambda)}{f_y(\lambda)} e^{ij\lambda} \mathrm{d}\lambda = \mathbf{1}_{j=0} - \frac{1}{\pi} \int_0^{\pi} \frac{\theta}{f_y(\lambda)} \cos(j\lambda) \mathrm{d}\lambda$$

due to uncorrelation of signal and noise, $f_y(\lambda)$ and $f_{yx}(\lambda)$ are the (pseudo) sdf of y_t and cross sdf ofy_t and x_t, $\theta = \sigma_u^2/2\pi$ is the (constant) sdf of the noise and $\mathbf{1}_{j=0} = 1$ if $j = 0$ and $\mathbf{1}_{j=0} = 0$ otherwise.

- The gain of this filter is $f_x(\lambda)/f_y(\lambda)$.

Feasible WK filter in the frequency domain

- A feasible plug-in version of ψ_j is $\hat{\psi}_j = \hat{\psi}_j(\hat{\theta}, \hat{f}_y)$ where

$$\hat{\psi}_j(\hat{\theta}, \hat{f}_y) = \mathbf{1}_{j=0} - \frac{1}{n^*} \sum_{p=1}^{n^*} \frac{\hat{\theta}}{\hat{f}_y(\lambda_p)} \cos(j\lambda_p)$$

for $n^* = [n/2]$, $[]$ denoting “the integer part of”, n is the sample size, $\lambda_p = 2\pi p/n$ are Fourier frequencies.

Feasible WK filter in the frequency domain

- A feasible plug-in version of ψ_j is $\hat{\psi}_j = \hat{\psi}_j(\hat{\theta}, \hat{f}_y)$ where

$$\hat{\psi}_j(\hat{\theta}, \hat{f}_y) = \mathbf{1}_{j=0} - \frac{1}{n^*} \sum_{p=1}^{n^*} \frac{\hat{\theta}}{\hat{f}_y(\lambda_p)} \cos(j\lambda_p)$$

for $n^* = [n/2]$, $[]$ denoting “the integer part of”, n is the sample size, $\lambda_p = 2\pi p/n$ are Fourier frequencies.

- For consistency of $\hat{\psi}_j$ we need:
  - A consistent estimator of $\theta \Longrightarrow$ Local Whittle (Hurvich et al. 2005).
  - An estimator of $f_y(\lambda_p)$ consistent uniformly over $p = 1, ..., n^*$, that is consistent for constant frequencies ($p = O(n)$) and for Fourier frequencies collapsing to zero ($p = o(n)$), no matter how far from or close to the spectral pole.

Feasible WK filter in the frequency domain

- Uniformly consistent estimator of $f_y(\lambda)$ (Hidalgo and Yajima, 2002):

$$\boxed{\hat{f}_y(\lambda_v) = \frac{|\lambda_v|^{-2\hat{d}}}{2m^*+1} \sum_{j=-m^*}^{m^*} |\lambda_v + \lambda_j|^{2\hat{d}} I_y(\lambda_v + \lambda_j)}$$

for $\hat{d}$ an estimator of d_0 such that $(\hat{d} - d_0) = o_p(\log^{-1} m^*)$ (for example the local Whitle), $\lambda_v = 2\pi v/n$, $v = 1, ..., n^*$ and m^* satisfying $\frac{1}{m^*} + \frac{m^*}{n} \to 0$ as $n \to \infty$.

Feasible WK filter in the frequency domain

- Uniformly consistent estimator of $f_y(\lambda)$ (Hidalgo and Yajima, 2002):

$$\boxed{\hat{f}_y(\lambda_v) = \frac{|\lambda_v|^{-2\hat{d}}}{2m^*+1} \sum_{j=-m^*}^{m^*} |\lambda_v + \lambda_j|^{2\hat{d}} I_y(\lambda_v + \lambda_j)}$$

for $\hat{d}$ an estimator of d_0 such that $(\hat{d} - d_0) = o_p(\log^{-1} m^*)$ (for example the local Whitle), $\lambda_v = 2\pi v/n$, $v = 1, ..., n^*$ and m^* satisfying $\frac{1}{m^*} + \frac{m^*}{n} \to 0$ as $n \to \infty$.

Theorem (consistency for $0 < d < 1$): *Under linearity (in a martingale difference) of the (stationary part of) the signal and finite fourth moments of the noise and innovations of the signal, as $n \to \infty$, uniformly over $v = 1, ..., n^*$*

$$\boxed{\hat{f}_y(\lambda_v) = f_y(\lambda_v)\,(1 + o_p(1)))}$$

Feasible WK filter in the frequency domain

Theorem : *Let $\hat{\theta}$ be a consistent estimator of θ and $\hat{f}_y(\lambda_p)$ estimate consistently $f_y(\lambda_p)$ uniformly over $p = 1, ..., n^*$. Then as $n \to \infty$*

$$\hat{\psi}_j = \psi_j(1 + o_p(1))$$

for j satisfying $|j|^{2d+1}/n \to 0$ as $n \to \infty$.

Feasible WK filter in the frequency domain

- The plug in estimator of the volatility component x_t is then

$$\boxed{\hat{x}_{t|n} = \sum_{j=\max(t-n,-M)}^{\min(t-1,M)} \hat{\psi}_j (y_{t-j} - \hat{\mu})}$$

for M satisfying $M^{-1} + n^{-1}M^{2d+1} \to 0$ as $n \to \infty$.

Feasible WK filter in the frequency domain

- The plug in estimator of the volatility component x_t is then

$$\boxed{\hat{x}_{t|n} = \sum_{j=\max(t-n,-M)}^{\min(t-1,M)} \hat{\psi}_j (y_{t-j} - \hat{\mu})}$$

for M satisfying $M^{-1} + n^{-1} M^{2d+1} \to 0$ as $n \to \infty$.

- We restrict the lags of the observable to a maximum of M:
  - Consistency of $\hat{\psi}_j$ is guaranteed.
  - Since $\psi_j = O(j^{-1-2d})$, weights for larger lags are virtually zero.
  - Relatively low values of M allow a symmetric feasible filter to be applied for a large number of intermediate observations, avoiding in that way undesirable phase shifts.

Feasible WK filter in the frequency domain

An estimate of the constant μ is needed for feasible signal extraction.

- Under stationarity ($d < 1/2$) the sample mean $\hat{\mu} = \bar{y}$ is consistent.
- If $d \geq 1/2$ the sample mean is not consistent. Following Shimotsu (2010) $\hat{\mu} = y_1$ or any mean of a fixed number of observations, which are $O_p(1)$ such that the (nonstationary) signal asymptotically dominates the constant μ.

Feasible WK filter in the frequency domain

Step 1: Get $\hat{d}$, $\hat{\theta}$ by local Whittle. The bandwidth m is selected within a stable region of estimates (Taqqu and Teverovsky, 1996).

Step 2: Get $\hat{f}_y(\lambda_p)$ for $p = 1, 2, ..., n^*$ with m^* selected based on the smoothness of the spectral density at frequencies far from the origin. The higher the smoothness the larger m^*.

Step 3: Construct the weights $\hat{\psi}_j$. Chose M as the lowest value such that for $j > M$ then $|\hat{\psi}_j| < \eta$ for a prespecified $\eta > 0$.

Step 4: With this M get $\hat{x}_{t|n}$.

Step 5 (Validation): Check that the standardized residuals $\hat{\varepsilon}_{t|n} = z_t \exp(-\hat{x}_{t|n}/2)$ are *i.i.d.* (e.g. portmanteau test on $\hat{\varepsilon}^2_{t|n}$).

$$y_t = x_t + u_t \quad t = 1, 2, ..., n,$$

for $x_t = \kappa x_t^*$, $(1-L)^{d_0} x_t^* = w_t$ and:

Model 1: $d_0 = 0.4$, $w_t = w_t^*$ and $u_t = \log \epsilon_t^2$ with

$$\begin{pmatrix} \varepsilon_t \\ w_{t-1}^* \end{pmatrix} \sim NID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

Model 2: As Model 1 but with $(1-0.8L)w_t = w_t^*$.

Model 3: As Model 1 but with $(1-0.2L+0.8L^2)w_t = w_t^*$.

Model 4: As Model 1 but with $d_0 = 0.8$.

Model 5: *Higher order dependence*. $d_0 = 0.4$ and $(w_t, u_t)' = H_t^{1/2}\eta_t$ for $\eta_t \sim N(0, I_2)$, $H_t = diag(a_1 h_{1t}, a_2 h_{2t})$, $h_{it} = \alpha_0 + \alpha_1 w_{t-1}^2 + \alpha_2 u_{t-1}^2$ for $i = 1, 2$, and a_1, a_2 are constants chosen to maintain the unconditional variances of signal and noise as in Model 1.

- $\alpha = (\alpha_0, \alpha_1, \alpha_2) = (0.0001, 0.25, 0.04)$ (as in Wong and Li, 1997, *Biometrika*).
- $\rho = 0, -0.8$, the latter corresponding to a large negative correlation (leverage).
- κ is chosen to satisfy $f_u(0)/\kappa^2 f_w(0) = \pi^2, 5\pi^2$, which are two long run NSR close to those empirically found in financial time series (Breidt et al. 1998, Pérez and Ruiz 2001).
- The constant μ is estimated in Models 1, 2, 3 and 5 by the sample mean and by the average of the first 10 observations in Model 4 (Shimotsu, 2010, ET)).
- $n = 2048$ comparable with the size of many financial series.
- 1000 replications.

Monte Carlo

1. $\hat{x}_{t|n}^{(1)}$ with $M = 100$, $(m^*, m) = (80, 1000)$ for Models 1, 2, 4 and 5 and $(60, 300)$ in Model 3.

2. $\hat{x}_{t|n}^{(2)}$ with $M = 100$ and true f_y and θ.

3. $\hat{x}_{t|n}^{(3)}$ is $\tilde{x}$ with true variances and covariances. Following Harvey (1998) we consider weights for a sample size $n = 256$. Prior first differencing and posterior integration back in the nonstationary case.

4. $\hat{x}_{t|n}^{(4)}$ is as before but with the covariances of y_t estimated by their sample counterparts and σ_u^2 by local Whittle with m as in $\hat{x}_{t|n}^{(1)}$.

5. $\hat{x}_{t|n}^{(5)}$ is obtained by the Kalman filter applied to an AR(10).

6. The naive $\hat{x}_{t|n}^{(6)} = y_t - \hat{\mu}$, often used as proxy of the volatility.

Two criteria for global comparison:

- The average over the 1000 replications of the sample correlation between x_t and $\hat{x}_{t|n}^{(i)}$, $i = 1, 2, ..., 6$.
- The global typified Monte Carlo Mean Square Error defined as

$$MCMSE(i) = \frac{1}{\sigma^2}\frac{1}{2048}\sum_{t=1}^{2048}\frac{1}{1000}\sum_{l=1}^{1000}(\hat{x}_{t,l}^{(i)} - x_{t,l})^2$$

  where $\sigma^2 = \sigma_x^2$ in Models 1, 2, 3 and 5, $\sigma^2 = \sigma_v^2$ in Model 4, t, l indicates observation t in Monte Carlo replication l.

- Number of times that the Ljung-Box statistic does not reject the hypothesis that the first 100 autocorrelations of the squared standardized residuals $\hat{\varepsilon}_{t|n}^{2(i)} = \exp(y_t - \hat{x}_{t|n}^{(i)})$ are null at 5% significance level.

Sensitivity analysis of $\hat{x}^{(1)}_{t|n}$

Table 1: Sensitivity to the choice of m and m^*. Model 2

	$NSR = \pi^2$					
	m=40	m=100	m=300	m=600	m=800	m=1000
$m^* = 40$	5.659	3.634	1.367	0.882	0.793	0.771
	(0.317)	(0.438)	(0.575)	(0.633)	(0.647)	(0.655)
$m^* = 60$	5.603	3.619	1.351	0.862	0.773	0.749
	(0.330)	(0.451)	(0.591)	(0.653)	(0.668)	(0.677)
$m^* = 80$	5.563	3.614	1.348	0.857	0.767	0.744
	(0.339)	(0.458)	(0.599)	(0.663)	(0.678)	(0.687)
$m^* = 100$	5.530	3.614	1.351	0.860	0.770	0.746
	(0.347)	(0.463)	(0.604)	(0.668)	(0.684)	(0.693)
	$NSR = 5\pi^2$					
$m^* = 40$	18.500	15.389	10.408	7.177	6.428	5.987
	(0.193)	(0.236)	(0.281)	(0.302)	(0.309)	(0.310)
$m^* = 60$	18.446	15.372	10.381	7.142	6.391	5.949
	(0.205)	(0.252)	(0.302)	(0.326)	(0.334)	(0.335)
$m^* = 80$	18.452	15.405	10.414	7.168	6.416	5.972
	(0.211)	(0.259)	(0.310)	(0.336)	(0.344)	(0.345)
$m^* = 100$	18.488	15.457	10.475	7.222	6.471	6.024
	(0.213)	(0.262)	(0.313)	(0.339)	(0.347)	(0.348)

MCMSE and correlation with true signal (between round brackets) of $\hat{x}^{(1)}_{t|n}$ with different m and m^*.

Table 2: Sensitivity to the choice of m and m^*. Model 3.

	$NSR = \pi^2$					
	m=40	m=100	m=300	m=600	m=800	m=1000
$m^* = 40$	1.088	0.767	0.512	0.723	0.876	1.041
	(0.590)	(0.692)	(0.740)	(0.626)	(0.505)	(0.555)
$m^* = 60$	1.072	0.763	0.508	0.707	0.844	1.019
	(0.595)	(0.695)	(0.743)	(0.630)	(0.518)	(0.567)
$m^* = 80$	1.060	0.763	0.509	0.698	0.821	1.004
	(0.598)	(0.695)	(0.742)	(0.629)	(0.525)	(0.574)
$m^* = 100$	1.053	0.765	0.514	0.693	0.805	0.994
	(0.598)	(0.693)	(0.739)	(0.625)	(0.528)	(0.579)
	$NSR = 5\pi^2$					
$m^* = 40$	3.269	2.620	1.890	1.297	1.438	2.325
	(0.353)	(0.398)	(0.431)	(0.293)	(0.363)	(0.384)
$m^* = 60$	3.264	2.622	1.891	1.280	1.430	2.321
	(0.353)	(0.397)	(0.431)	(0.291)	(0.364)	(0.385)
$m^* = 80$	3.271	2.635	1.905	1.279	1.437	2.330
	(0.347)	(0.390)	(0.423)	(0.282)	(0.358)	(0.379)
$m^* = 100$	3.284	2.653	1.925	1.286	1.451	2.343
	(0.340)	(0.380)	(0.411)	(0.269)	(0.348)	(0.370)

MCMSE and correlation with true signal (between round brackets) of $\hat{x}^{(1)}_{t|n}$ with different m and m^*.

Global Comparison

Table 3: Global MSE and correlation

	$\tilde{x}_{t\|\infty}$	$\hat{x}^{(1)}_{t\|n}$	$\hat{x}^{(2)}_{t\|n}$	$\hat{x}^{(3)}_{t\|n}$	$\hat{x}^{(4)}_{t\|n}$	$\hat{x}^{(5)}_{t\|n}$	$\hat{x}^{(6)}_{t\|n}$
			Model 1				
$NSR = \pi^2$	*0.523*	**0.788**	0.711	0.719	1.299	1.166	4.948
	(0.556)	**(0.545)**	(0.579)	(0.572)	(0.331)	(0.158)	(0.379)
		[842]	[923]	[916]	[422]	[28]	[377]
$NSR = 5\pi^2$	*0.662*	4.017	0.871	0.875	6.032	**2.718**	23.983
	(0.357)	**(0.289)**	(0.376)	(0.372)	(0.116)	(0.082)	(0.178)
		[678]	[923]	[918]	[164]	[153]	[375]
			Model 2				
$NSR = \pi^2$	*0.286*	**0.744**	0.638	0.656	1.620	1.616	8.528
	(0.713)	**(0.687)**	(0.734)	(0.716)	(0.389)	(0.083)	(0.271)
		[717]	[869]	[861]	[122]	[0]	[387]
$NSR = 5\pi^2$	*0.443*	5.972	0.798	0.803	9.453	**4.257**	41.468
	(0.488)	**(0.345)**	(0.504)	(0.498)	(0.115)	(0.038)	(0.123)
		[650]	[907]	[890]	[116]	[133]	[409]

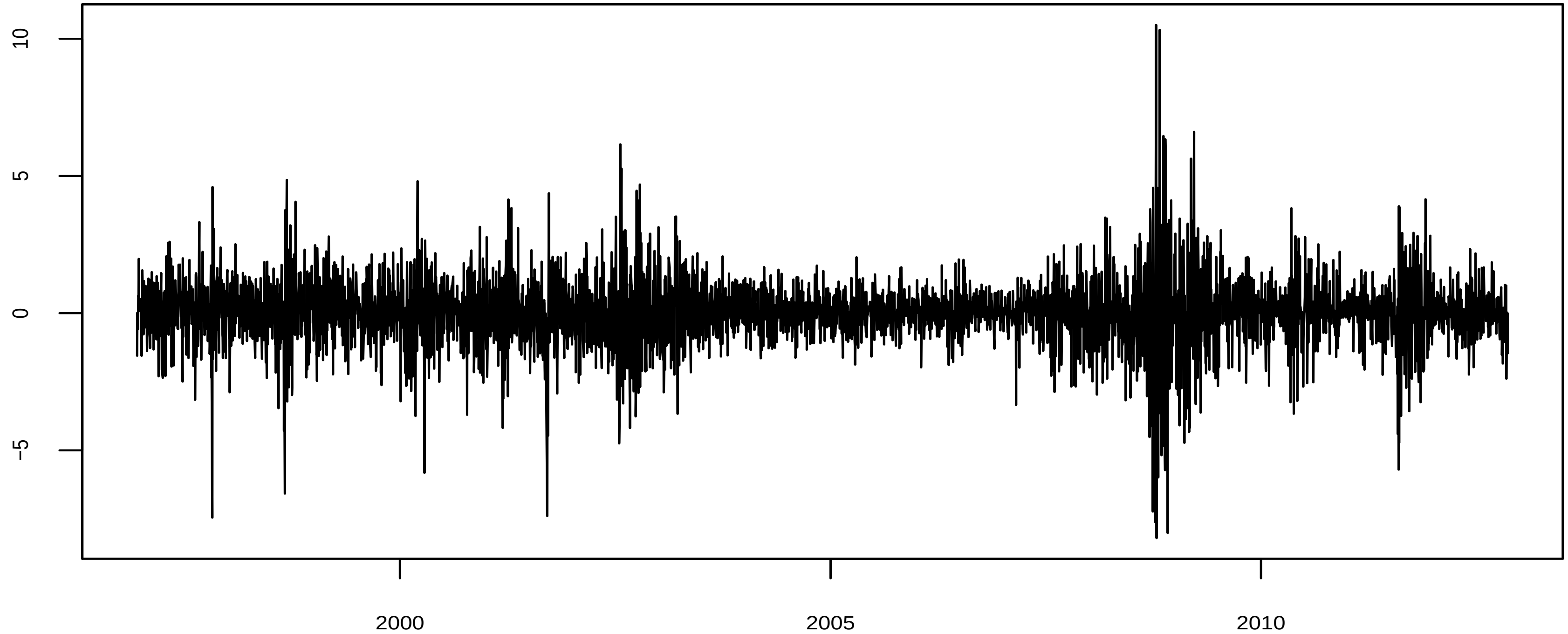
Note: MCMSE, global correlation between x_t and $\hat{x}^{(i)}_{t|n}$ (between round brackets) and nonrejections of no correlation in squared standardized residuals (between square brackets). Optimals in italic (benchmark) and best feasibles in bold.

Table 4: Global MSE and correlation

	$\tilde{x}_{t\|\infty}$	$\hat{x}^{(1)}_{t\|n}$	$\hat{x}^{(2)}_{t\|n}$	$\hat{x}^{(3)}_{t\|n}$	$\hat{x}^{(4)}_{t\|n}$	$\hat{x}^{(5)}_{t\|n}$	$\hat{x}^{(6)}_{t\|n}$
			Model 3				
$NSR = \pi^2$	*0.368* *(0.777)*	**0.508** **(0.743)** **[775]**	0.427 (0.779) [879]	0.428 (0.778) [872]	0.661 (0.652) [677]	0.865 (0.416) [299]	1.439 (0.636) [413]
$NSR = 5\pi^2$	*0.654* *(0.544)*	1.891 **(0.431)** **[676]**	0.714 (0.545) [925]	0.715 (0.544) [909]	2.444 (0.297) [351]	**1.340** (0.206) [47]	6.959 (0.345) [379]
			Model 4				
$NSR = \pi^2$		**86.088** **(0.968)** **[820]**	85.970 (0.970) [856]	85.509 (0.970) [899]	91.806 (0.926) [102]	109.608 (0.567) [8]	94.156 (0.847) [321]
$NSR = 5\pi^2$		**91.255** **(0.933)** **[692]**	90.235 (0.942) [786]	85.254 (0.943) [844]	124.873 (0.747) [33]	118.380 (0.288) [0]	135.317 (0.603) [323]
			Model 5				
$NSR = \pi^2$	*0.523* *(0.556)*	**0.811** **(0.545)**	0.709 (0.581)	0.709 (0.581)	1.317 (0.336)	1.188 (0.158)	4.944 (0.379)
$NSR = 5\pi^2$	*0.662* *(0.357)*	4.521 **(0.293)**	0.865 (0.382)	0.869 (0.382)	6.507 (0.119)	**2.960** (0.085)	24.010 (0.180)

Daily Dow Jones Industrial

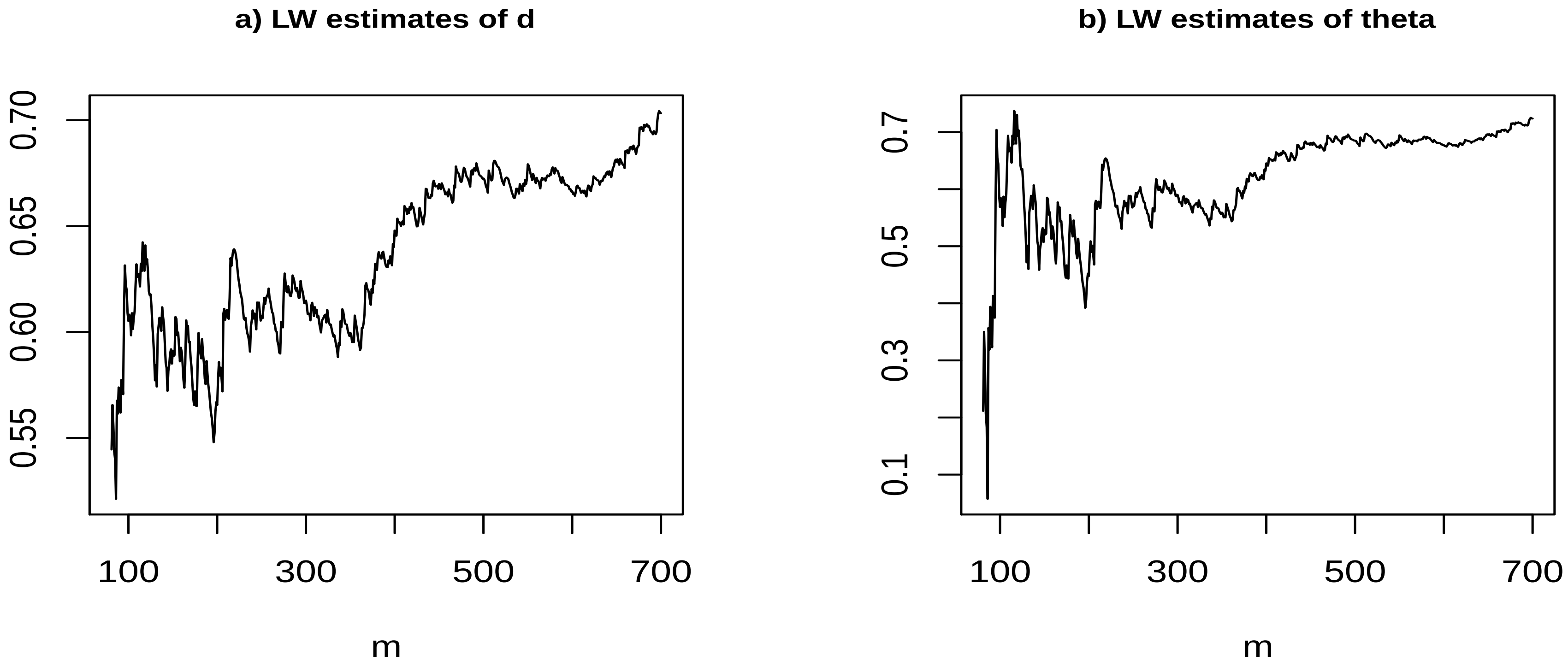
Figure 1: Dow Jones Industrial returns (12/12/1996-11/14/2012)



$p - value$ of robustified Box-Pierce statistic (100 lags) 0.164

- **Step 1:** Estimation of d and θ by Local Whittle (selection of m)

Figure 2: Local Whittle estimates



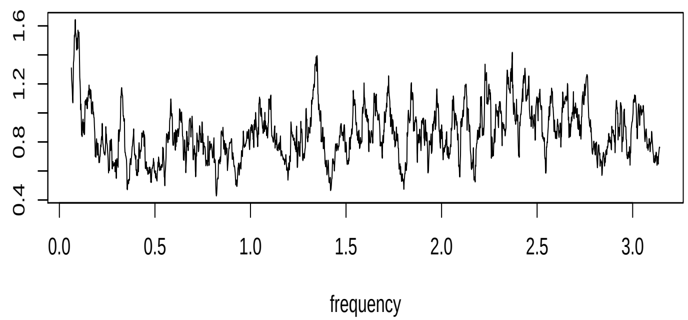
With $m = 500$, $\hat{d} = 0.67$ and $\hat{\theta} = 0.69 \Rightarrow \hat{\mu} = \sum_{t=1}^{10} y_t/10$

Daily Dow Jones Industrial: FD strategy

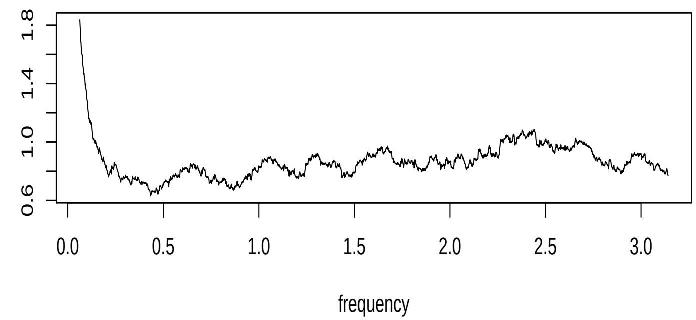
- **Step 2:** Estimate $f_y(\lambda_p)$ (select m^*)

Figure 3: Estimates of $f_y(\lambda_p)$ for $p \geq 40$

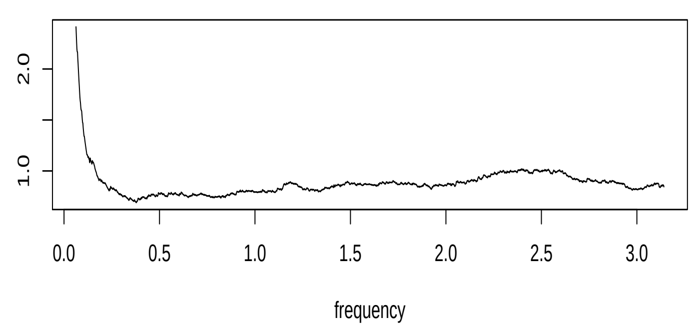
(a) $m^* = 10$



(b) $m^* = 60$



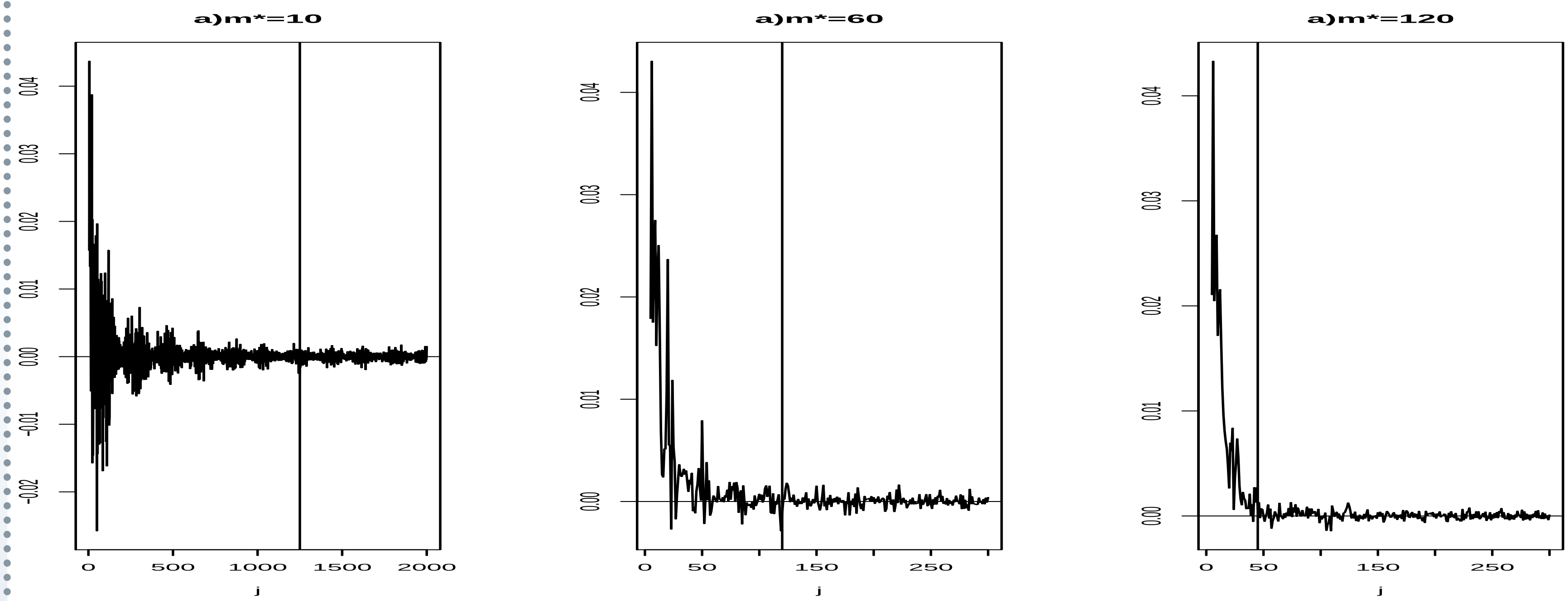
(c) $m^* = 120$



Daily Dow Jones Industrial: FD strategy

- **Step 3:** Construct the weights $\hat{\psi}_j$ (selection of M)

Figure 4: Weights $\hat{\psi}_j$ and M



M selected as the lowest value such that $|\hat{\psi}_j| \leq 0.002$, $\forall j > M$:
(M=1250 ($m^* = 10$), 120 ($m^* = 60$), 45 ($m^* = 120$)).

Daily Dow Jones Industrial: FD strategy

- **Step 4:** We estimate the variance of the returns conditional on the volatility component in a LMSV model as

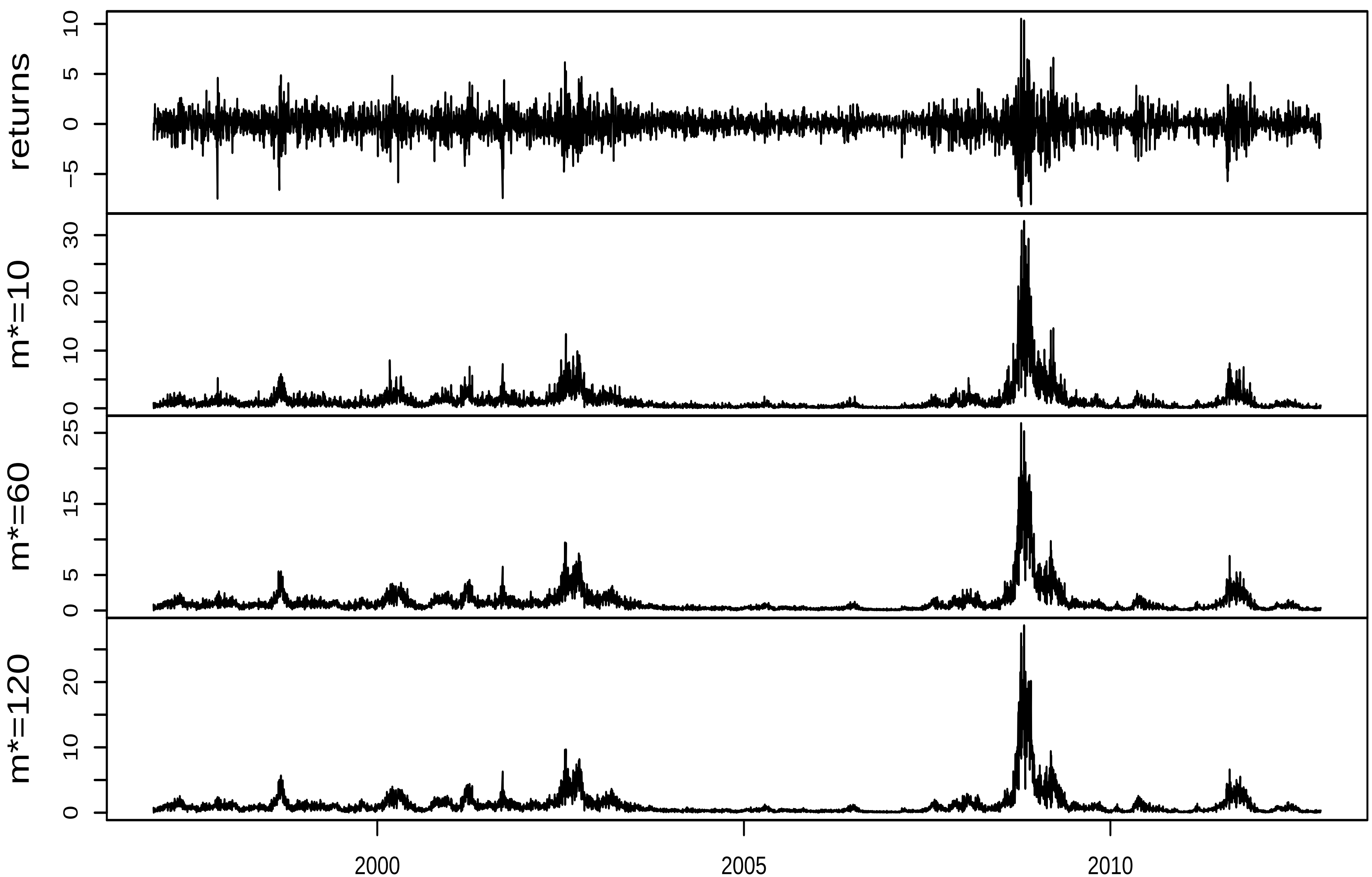
$$\hat{\sigma}_t^2 = \hat{\sigma}^2 \exp(\hat{x}_{t|n}^{(1)})$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} z_t^2 \exp(-\hat{x}_{t|n}^{(1)}).$$

Daily Dow Jones Industrial: FD strategy

Figure 5: Estimation of the conditional variance



Daily Dow Jones Industrial: Validation

- **Step 5:** Validation:
  - The standardized residuals, $\hat{\varepsilon}_{t|n} = z_t / \sqrt{\hat{\sigma}_t^2}$ should be close to $i.i.d$ (not necessarily Gaussian) and in particular their squares should be uncorrelated.

Table 5: Ljung-Box test for 100 ac (p-values)

	$m^* = 10$	$m^* = 60$	$m^* = 120$
$\hat{\varepsilon}^2_{t\|n}$	0.000	0.113	0.138

Conclusions

- We have proposed a robust, general and simple to implement Wiener-Kolmogorov signal extraction technique based on a local spectral specification of the signal (either stationary or nonstationary).

Conclusions

- We have proposed a robust, general and simple to implement Wiener-Kolmogorov signal extraction technique based on a local spectral specification of the signal (either stationary or nonstationary).

- For that, we propose a pre-whitened (in the frequency domain) sdf estimator and show its consistency in the whole band of Fourier frequencies for stationary and nonstationary signals.

Conclusions

- We have proposed a robust, general and simple to implement Wiener-Kolmogorov signal extraction technique based on a local spectral specification of the signal (either stationary or nonstationary).
- For that, we propose a pre-whitened (in the frequency domain) sdf estimator and show its consistency in the whole band of Fourier frequencies for stationary and nonstationary signals.
- It can be used for signal extraction in a general context: economic mechanisms with different factors for short run and long run behaviour, measurement errors, rational expectation models, Realized Volatility contaminated by market microstructure noise etc.

Conclusions

- We have proposed a robust, general and simple to implement Wiener-Kolmogorov signal extraction technique based on a local spectral specification of the signal (either stationary or nonstationary).
- For that, we propose a pre-whitened (in the frequency domain) sdf estimator and show its consistency in the whole band of Fourier frequencies for stationary and nonstationary signals.
- It can be used for signal extraction in a general context: economic mechanisms with different factors for short run and long run behaviour, measurement errors, rational expectation models, Realized Volatility contaminated by market microstructure noise etc.
- Next step: use the volatility series for asset pricing, risk management, forecasting, comparison with ARCH based methods....


Thank You