Robust Inference for Time Series Models: a Wavelet-Based Framework

Stéphane Guerrier

Department of Statistics
University of Illinois at Urbana-Champaign

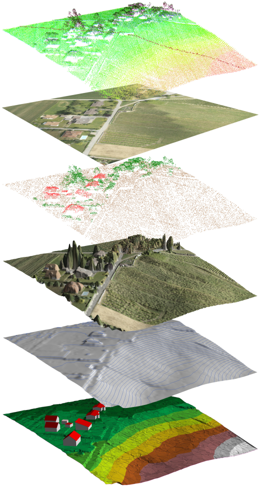
joint work with
R. Molinari (U. of Geneva)

October 29, 2015

Helicopter-based mobile mapping

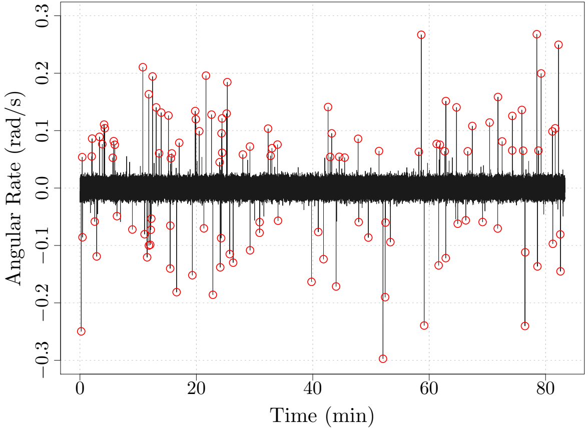
Ecole Polytechnique Fédérale de Lausanne (EPFL)

scan2map™
HANDHELD AIRBORNE MAPPING SYSTEM



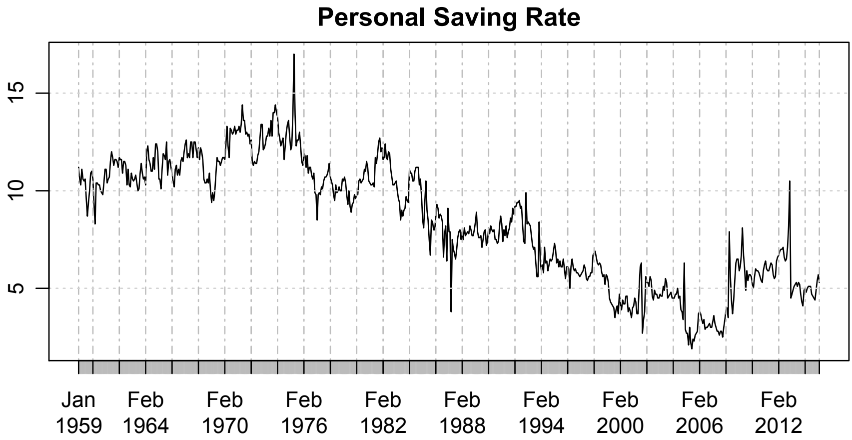
Virtual Reality: Ted?

Gyroscope signal

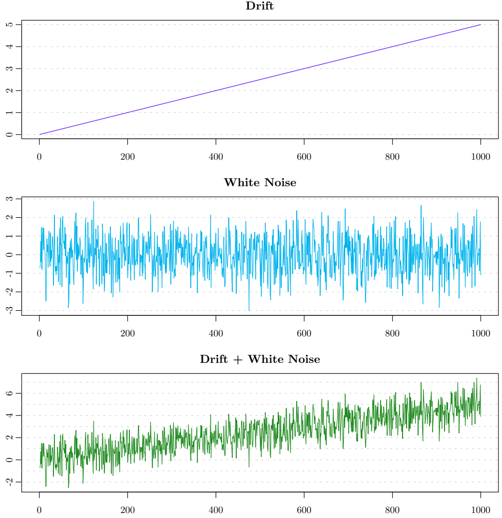


Stebler et al., 2014 suggest a latent model made by the sum of two AR(1) models and a white noise process for this kind of signals.

Personal US Saving Rate



An easy latent time series model



Remarks:

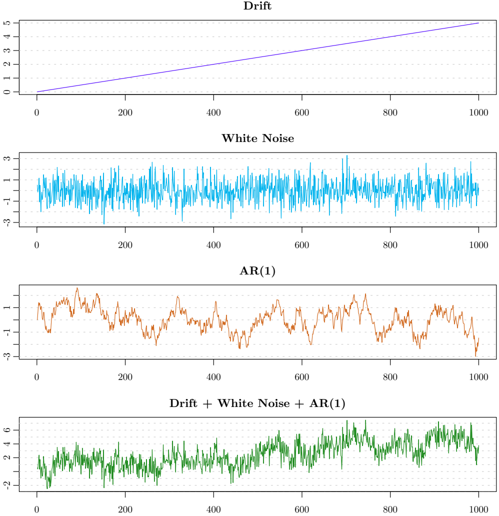
- Simple linear regression model:

$$y_t = \omega t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

- MLE is perfectly fine.
- **What if we add an AR(1) process?**

Adding an autoregressive process



Remarks:

- Not a linear regression model but a **state space model**.
- Computing the likelihood is not an easy task (Kalman filter).
- **MLE (in fact EM-KF) fails**.

The proposed approach

Robust Inference for Time Series Models:

We present a new framework for the **robust estimation of time series models**. This approach provides estimators which are

- consistent and asymptotically normally distributed,
- applicable to a broad spectrum of time series models:
  - ARMA
  - State-space models
  - Latent models used in engineering and other fields
- straightforward to implement,
- computationally efficient.

The framework is based on the recently developed **Generalized Method of Wavelet Moments** (GMWM) (see *Guerrier et al, 2013*) and a **new robust estimator of the Wavelet Variance** (WV).

Outline

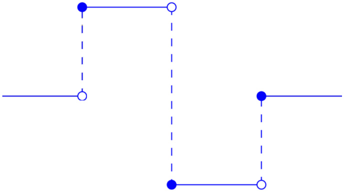
1. Introduction
   - Spectrum (SDF/PSD) and wavelet approximations
2. Wavelet Variance (WV)
   - Classical estimators
   - Robust estimators
   - (Simulations)
3. Generalized Method of Wavelet Moments (GMWM)
   - Classical approach
   - Robust extension (RGMWM)
   - Simulations
   - Examples
     - Precipitation data
     - US personal saving rate
     - Inertial data
4. Conclusions and outlook

Wavelet Filters

But what is the WV exactly? Let us start by defining the wavelet coefficients as follows:

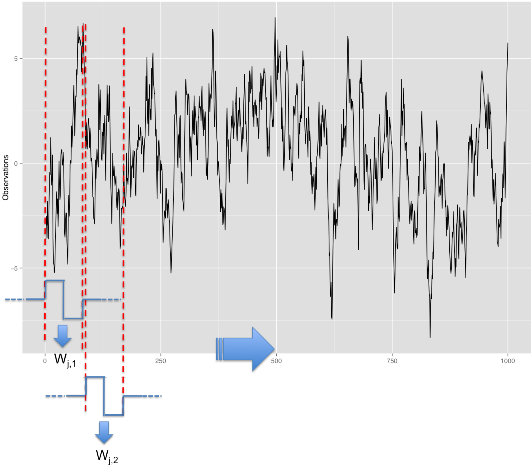
$$W_{j,t} = \sum_{l=1}^{L_j} h_{j,l} X_{t-l}$$

where $W_{j,t}$ represents the wavelet coefficients for scales j and $h_{j,l}$ are so called wavelet filters issued from a mother wavelet function (for example the Haar wavelet) and L_j is the length of the filter



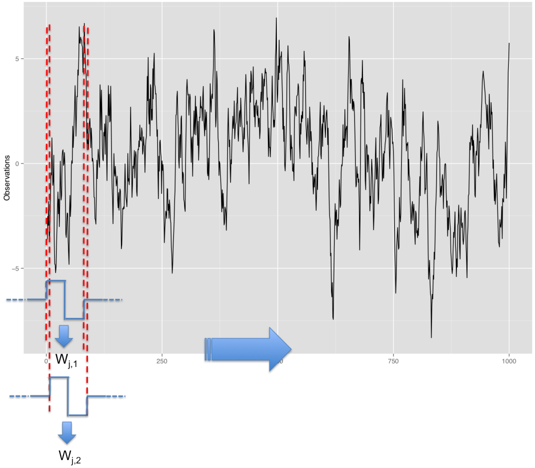
Discrete Wavelet Transform (DWT)

The wavelet filters can be applied in different ways but there are two main transforms used for time series:



Maximum Overlap Discrete Wavelet Transform (MODWT)

The wavelet filters can be applied in different ways but there are two main transforms used for time series:



MODWT WV Estimator

Definition

Once we have defined the wavelet coefficients, we can now define the WV as being the variance of the wavelet coefficients

$$\nu_j^2 \equiv var[W_{j,t}]$$

Estimation

An unbiased estimator for the WV is given by the **MODWT estimator** *(see Percival, 1995)*

$$\tilde{\nu}_j^2 = \frac{1}{M_j(T)} \sum_{t=L_j}^{T} W_{j,t}^2$$

where $M_j(T) = T - L_j + 1$.

SDF vs WV of an AR(1) process

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$$

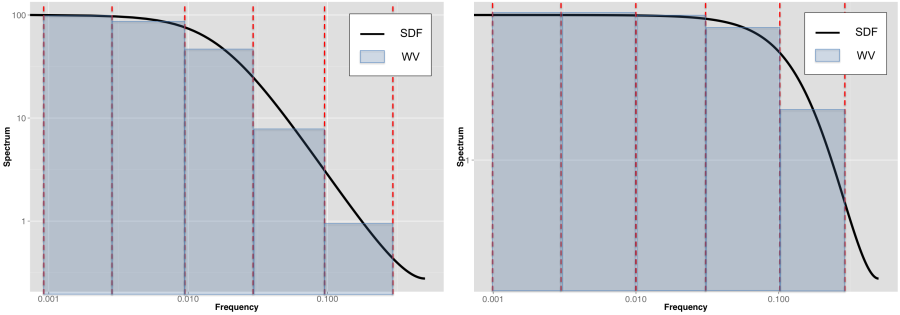


Figure: Left: AR(1) with $\phi = 0.9$; Right: AR(1) with $\phi = 0.2$

Robustness of WV

The contamination model

Suppose we are interested in the model $F_{\boldsymbol{\theta}}$ but what we observe is

$$F_{\epsilon} = (1-\epsilon)F_{\boldsymbol{\theta}} + \epsilon H$$

with small $\epsilon > 0$, $F_{\boldsymbol{\theta}}$ being the distribution function of the wavelet coefficients ($\boldsymbol{\theta}$ being the parameter vector) and H a generic contaminating distribution different from $F_{\boldsymbol{\theta}}$.

Robustness of WV

In a recent paper *Mondal and Percival, 2012* underline how the classic estimator of WV $\tilde{\nu}_j^2$ can become **highly biased when in the presence of small contaminations**.

An approximate solution

Since the IF of $\tilde{\nu}_j^2$ is unbounded, a robust WV estimator would be required to **bound the bias** of the estimator when there is contamination.

A robust proposal

For this purpose *Mondal and Percival, 2012* propose a robust estimator of WV ($\bar{\nu}_j^2$) for when (X_t) is a **stationary Gaussian process** using the following approach

- make a log-transformation of $W_{j,t}^2$ to turn the estimation of a scale parameter into a location parameter
- use a bounded $\psi(\cdot)$ function to estimate the location parameter
- obtain an approximately unbiased estimator of ν_j^2 by correcting and inversely transforming the previous estimate

Robust WV

Huber's Proposal 2

The exact estimator takes the form of an M-estimator expressed as a *weighted maximum likelihood estimator* and is the implicit solution in ν_j^2 of

$$\sum_{t=1}^{M_j(T)} w^2(r_{j,t}, \nu_j^2, c) r_{j,t}^2 - a(\nu_j^2, c) = 0$$

where

- $w(\cdot)$ is a weight function
- $r_{j,t} = {}^{W_{j,t}}/_{\nu_j}$ with $\nu_j = \sqrt{\nu_j^2}$
- c is a tuning constant which regulates the trade-off between "robustness" and efficiency of the estimator given $w(\cdot)$

Robust WV

Definition

The **correction term** $a(\nu_j^2, c)$ ensures Fisher consistency of the estimator and is defined as

$$a(\nu_j^2, c) = \int w^2(r_{j,t}, \nu_j^2, c) r_{j,t}^2 dF_{\theta}$$

Weight function

The chosen weight function $w(\cdot)$ is the **Tukey biweight function** (Biw) which gives weights

$$w(r_{j,t}; \nu_j^2, c) = \begin{cases} \left(\left(\frac{r_{j,t}}{c}\right)^2 - 1\right)^2 & \text{if } \; |r_{j,t}| \leq c \\ 0 & \text{if } \; |r_{j,t}| > c \end{cases}$$

Robust WV Estimator

The proposed estimator

Based on the (robust) weight function $w(\cdot)$, the proposed estimator can be expressed as

$$\hat{\nu}_j^2 = \underset{\nu_j^2 \in \mathbb{R}^+}{argmin} \left\{ \left[\frac{1}{M_j(T)} \sum_{t=1}^{M_j(T)} w^2(r_{j,t}; \nu_j^2, c) r_{j,t}^2 - a_\psi(c) \right]^2 \right\}$$

where $a_\psi(c)$ is the equivalent of $a(\nu_j^2, c)$ assuming that the wavelet coefficients $W_{j,t}$ follow a **stationary Gaussian process**.

Global identifiability

Although it is often hard to prove and is assumed for simplicity (e.g. see *Mondal and Percival (2010)*), we prove the **global identifiability** of $\hat{\nu}_j^2$ based on the Biw function (as well as on the Hub function).

Result

Using the Biw or Hub functions and if ν_0^2 represents the true WV at a given scale j, then

$$\mathbb{E}[\psi(W_{j,t}, \nu_j^2)] = 0$$

if and only if $\nu_j^2 = \nu_0^2$.

Robust WV

Assumptions

The process $(W_{j,t})$ is a stationary and ergodic Gaussian process with finite variance $0 < \nu_j^2 < \infty$

Asymptotic results

Under this condition and for the considered weight functions we have that:

$$||\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}||_2 = o_p\left(\sqrt{\frac{\log_2(T)}{T}}\right)$$

$$\sqrt{T}\boldsymbol{s}^T\boldsymbol{\Sigma}^{-1/2}\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}\left(0, 1\right)$$

where $\boldsymbol{s} \in \mathbb{R}^J$ is such that $||\boldsymbol{s}|| = 1$ and, $\boldsymbol{\Sigma} = \mathbf{M}^{-T}\mathbf{S}_{\psi}(\mathbf{0})\mathbf{M}^{-1}$ (with $\mathbf{S}_{\psi}(\mathbf{0})$ the SDF of $\psi(\mathbf{W}_t, \boldsymbol{\nu})$ and $\mathbf{M} = \mathbb{E}[-\psi'(\mathbf{W}_t, \boldsymbol{\nu})]$)

WV implied by $F_{\boldsymbol{\theta}}$

WV as a function of $\boldsymbol{\theta}$

Given a model $F_{\boldsymbol{\theta}}$, the WV has an implicit link with the parameters $\boldsymbol{\theta}$ given by

$$\nu_j^2(\boldsymbol{\theta}) = \int_{-1/2}^{1/2} |H_j(f)|^2 S_{F_{\boldsymbol{\theta}}}(f) df$$

where

- $S_{F_{\boldsymbol{\theta}}}$ is the SDF associated to the model $F_{\boldsymbol{\theta}}$
- H_j is the Fourier transform of the wavelet filters with $H_1(f) = \sum_{l=0}^{L_1-1} h_{1,l} e^{-i2\pi fl}$
- $|\cdot|$ denotes the modulus

GMWM

The idea

Using the properties of the WV estimator $\tilde{\nu}_j^2$ and the implicit link of the WV with a process' parameters $\boldsymbol{\theta}$, *Guerrier et al., 2013* propose to match the estimated WV with the theoretical WV through the **Generalized Method of Wavelet Moments** (GMWM).

The GMWM estimator

The GMWM estimator is the solution of the following optimization problem

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \; \left(\tilde{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^T \boldsymbol{\Omega} \left(\tilde{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)$$

where $\boldsymbol{\Omega}$ is a positive definite weighting matrix which is chosen in a suitable manner such that the quadratic form is convex.

Identifiability

Result

The function $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\nu}(\boldsymbol{\theta}) - \boldsymbol{\nu}(\boldsymbol{\theta}_0) = 0$ iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for the following models:

- Causal and invertible ARMA models
- Latent time series models (e.g. $X_t = \sum \text{AR}(1) + \sum \text{MA}(1) + \text{WN} + \text{QN}$, commonly used in many engineering applications)
- Several other stationary and non-stationary state-space models (e.g. $X_t = \text{QN} + \text{WN} + \text{AR}(1) + \text{RW} + \text{DR}$, commonly used in natural sciences)

GMWM

Theorem: Asymptotic Normality

We showed that (under some regularity conditions) $\tilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ and

$$\sqrt{T}\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}\left(0, \mathbf{Q}\right)$$

where $\mathbf{Q} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$, $\boldsymbol{\Gamma} = \left(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}\right)^{-1} \mathbf{D}^T \boldsymbol{\Omega}$, $\mathbf{D} = \partial \boldsymbol{\nu}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T$ and $\boldsymbol{\Lambda}$ is the asymptotic covariance matrix of $\tilde{\boldsymbol{\nu}}^2$.

Important condition:

Asymptotic normality for $\tilde{\boldsymbol{\theta}}$ is a consequence of asymptotic normality for $\tilde{\boldsymbol{\nu}}$.

A Robust GMWM

Idea

The GMWM takes the form of an **indirect estimator** with auxiliary parameter being the estimator of WV. *Genton and Ronchetti, 2003* underline that the IF of the *indirect estimator is bounded only if the auxiliary estimator is bounded*.

The RGMWM

Using the proposed estimator $\hat{\nu}_j^2$ we can therefore deliver a **Robust GMWM** (RGMWM):

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \; \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^T \boldsymbol{\Omega} \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)$$

Simulations

Simulations Setting

- $B = 500$ replications for each simulation
- $T = 1000$
- H is a zero-mean Gaussian distribution
    - large σ_ϵ^2 (9 or 100)
    - *additive type outliers*

Performance measure

We define a relative version of the **root mean squared error**

$$\text{RMSE*} = \sqrt{\mathbb{E}\left(\frac{\hat{\mu} - \mu_0}{\mu_0}\right)^2 + var\left(\frac{\hat{\mu}}{\mu_0}\right)}$$

with μ_0 being the true value of a parameter and $\hat{\mu}$ being the estimation of μ_0.

Simulations

Simulated processes

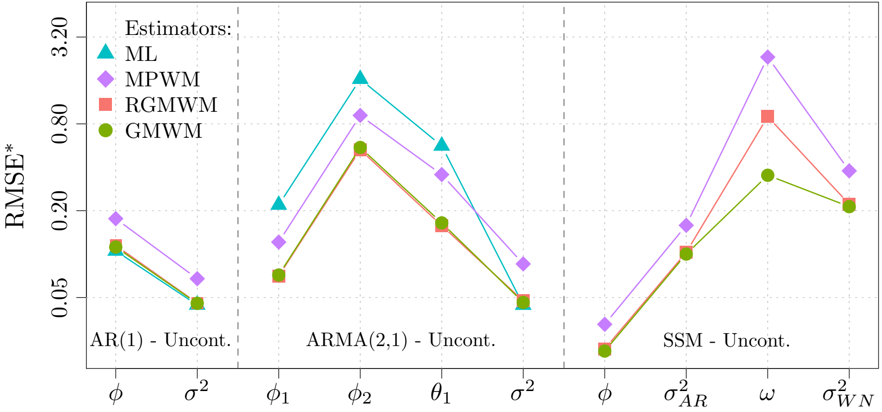
Three processes were used to compare the maximum likelihood estimator (ML), the GMWM, the GMWM based on $\bar{\nu}_j^2$ (MPWM) and the RGMWM.

- AR(1) with $\phi = 0.3$ and $\sigma^2 = 5$
- ARMA(2,1) with $\phi_1 = 0.8$, $\phi_2 = 0.1$, $\theta_1 = 0.3$ and $\sigma^2 = 5$
- SSM defined as

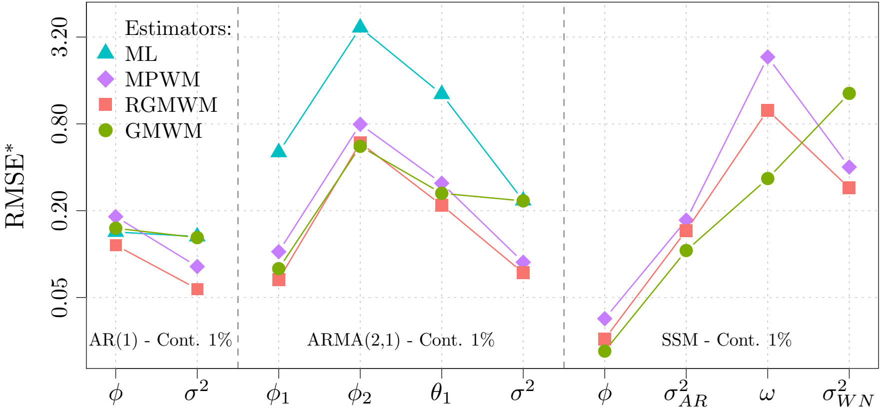
$$
\begin{aligned}
Y_t =& \phi Y_{t-1} + W_t, \;\; W_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{AR}) \\
X_t =& Y_t + Z_t + U_t, \;\; Z_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{WN}).
\end{aligned}
$$

with U_t being a drift with parameter $\omega = 0.05$, $\phi = 0.95$, $\sigma^2_{AR} = 16$ and $\sigma^2_{WN} = 4$.

Uncontaminated processes



Contaminated processes



Applications: the Environmental System Model

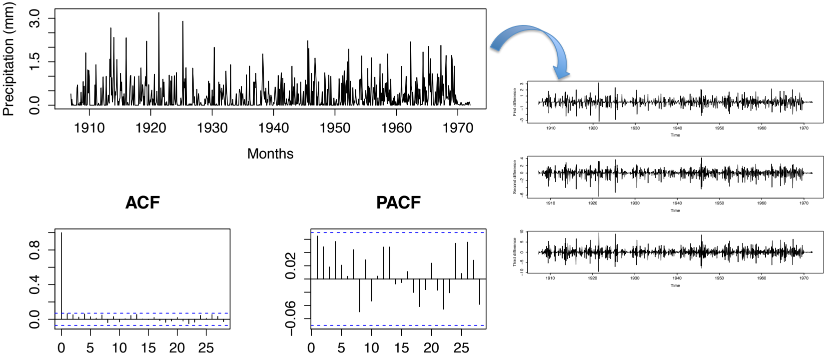


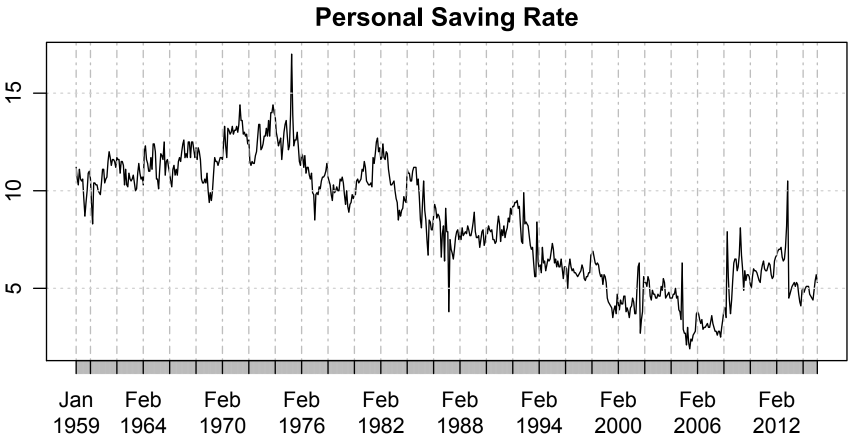
Figure: Monthly precipitation data from 1907 to 1972 taken form *Hipel and McLeod, Elsevier, 1994*

Applications: the Environmental System Model

For the precipitation phase, models which are usually considered at the white noise or AR(1). The ACF and PACF would suggest the first but let's test the AR(1)...

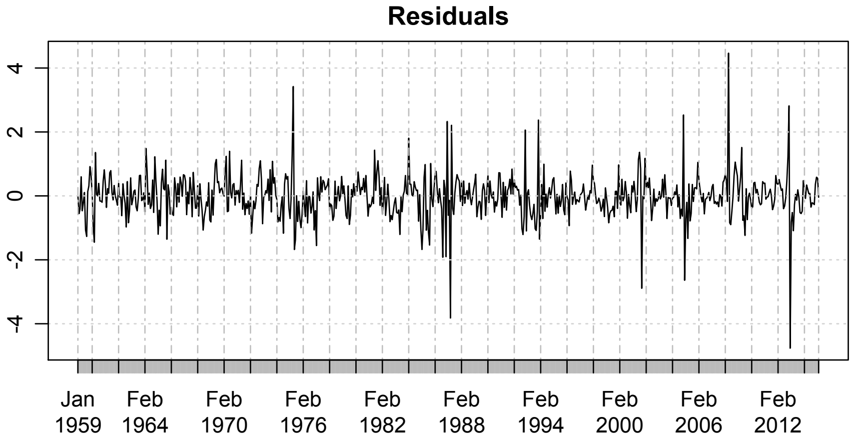
	ϕ	σ^2
MLE	$6.463 \cdot 10^{-2}$	$2.222 \cdot 10^{-1}$
CI	$[-5.702 \cdot 10^{-3}, 1.255 \cdot 10^{-1}]$	$[2.014 \cdot 10^{-1}, 2.413 \cdot 10^{-1}]$
GMWM	$5.384 \cdot 10^{-2}$	$2.205 \cdot 10^{-1}$
CI	$[-1.758 \cdot 10^{-2}, 1.255 \cdot 10^{-1}]$	$[1.984 \cdot 10^{-1}, 2.439 \cdot 10^{-1}]$
RGMWM	$3.892 \cdot 10^{-1}$	$1.016 \cdot 10^{-1}$
CI	$[3.008 \cdot 10^{-1}, 4.813 \cdot 10^{-1}]$	$[8.943 \cdot 10^{-2}, 1.133 \cdot 10^{-1}]$

Application: Personal Saving Rate



Application: Personal Saving Rate

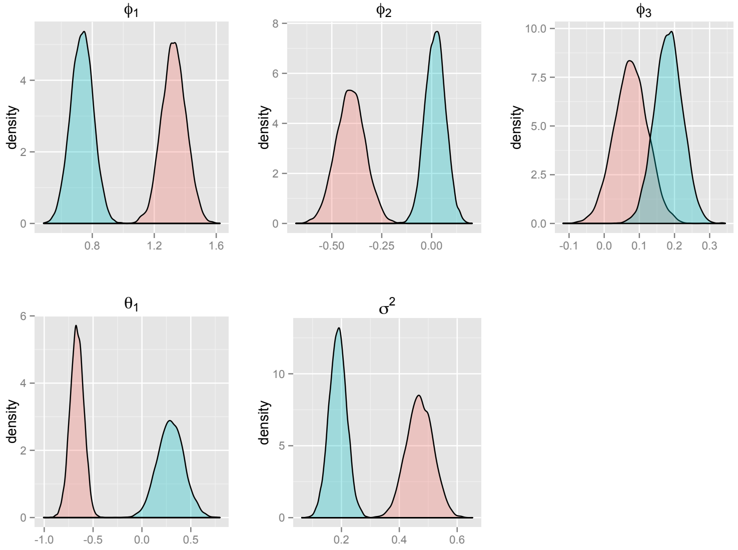
Classical model selection procedures (e.g. AIC, BIC, HQ) all suggests an ARMA(3,1) when selecting p, d, q of an assumed ARIMA(p, d, q).



Application: Personal Saving Rate

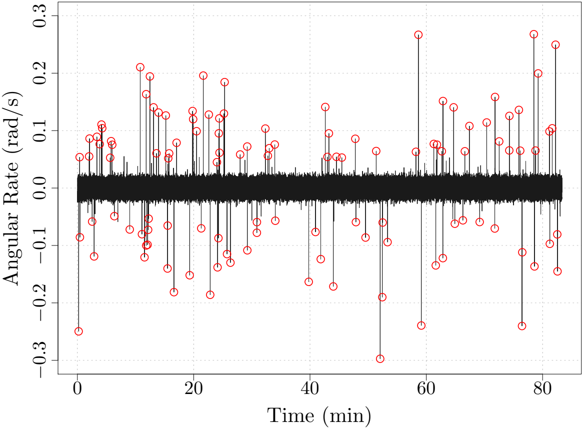
	MLE	RGMWM
ϕ_1	1.327 (0.079)	0.734 (0.091)
ϕ_2	-0.407 (0.073)	0.018 (0.051)
ϕ_3	0.075 (0.048)	0.182 (0.074)
θ_1	-0.665 (0.071)	0.295 (0.134)
σ^2	0.471 (0.060)	0.188 (0.022)

Application: Personal Saving Rate



Applications: Ted?

Applications: Gyroscope signal



Stebler et al., 2014 suggest a latent model made by the sum of two AR(1) models and a white noise process for this kind of signals.

Applications: Gyroscope signal

		Estimate	CI
σ^2	GMWM	$5.868 \cdot 10^{-5}$	$[5.843 \cdot 10^{-5}, 5.891 \cdot 10^{-5}]$
	RGMWM	$5.498 \cdot 10^{-5}$	$[5.473 \cdot 10^{-5}, 5.526 \cdot 10^{-5}]$
ϕ_1	GMWM	$9.969 \cdot 10^{-1}$	$[9.960 \cdot 10^{-1}, 9.975 \cdot 10^{-1}]$
	RGMWM	$9.965 \cdot 10^{-1}$	$[9.955 \cdot 10^{-1}, 9.973 \cdot 10^{-1}]$
σ_1^2	GMWM	$1.723 \cdot 10^{-9}$	$[1.256 \cdot 10^{-9}, 2.376 \cdot 10^{-9}]$
	RGMWM	$2.082 \cdot 10^{-9}$	$[1.452 \cdot 10^{-9}, 2.885 \cdot 10^{-9}]$
ϕ_2	GMWM	$9.048374 \cdot 10^{-1}$	$[9.048362 \cdot 10^{-1}, 9048378 \cdot 10^{-1}]$
	RGMWM	$9.048372 \cdot 10^{-1}$	$[9.048314 \cdot 10^{-1}, 9.048389 \cdot 10^{-1}]$
σ_2^2	GMWM	$1.493 \cdot 10^{-8}$	$[5.542 \cdot 10^{-17}, 3.142 \cdot 10^{-8}]$
	RGMWM	$2.948 \cdot 10^{-8}$	$[1.207 \cdot 10^{-8}, 4.679 \cdot 10^{-8}]$

Conclusions

- We hope that the particular use of the WV and the principle of the GMWM have unlocked a potential field of statistical research
- The RGMWM is a development of the GMWM providing a general framework for inference on time series models which is
    - robust (bounded IF)
    - consistent (identifiable) and asymptotically normally distributed
    - easy to implement and computationally efficient
    - applicable to a large class of time series models
- "GMWM" package is no available on CRAN.

R package & Collaboration

Open Source License

The GMWM project has been released using the **Q Public License**.

Collaborate with us on GitHub!

The GMWM is currently hosted on GitHub underneath the **SMAC-group @ UIUC** user account. To contribute, fork a copy and submit a pull request with any new features or bug fixed!

`https://github.com/SMAC-group/GMWM`

Outlook

Extensions

A great deal of work in progress based on the principle of the GMWM:

- Multivariate discrete process (Olympia Hadjiliadis)
- Spatial modeling (Bo Li)
- Non-stationary model estimation with time-varying parameters and covariates (Yanyuan Ma and Roberto Molinari)
- Model selection (time series and mixed-effect models) (Xinyu Zhang, Roberto Molinari and Haotian Xu)
- Further development of the GMWM R package (James Balamuta and Roberto Molinari)

Thank you very much for your attention!

A special thanks to

- James Balamuta & Wenchao Yang (UIUC)
- Dr. Jan Skaloud & Dr. Yannick Stebler (EPFL)
- Prof. Maria-Pia Victoria-Feser (U. Geneva)

Any questions?

More info...

SMAC-group.com
github.com/SMAC-Group
stephane@illinois.edu

Main References

GMWM:

- Guerrier, S., Stebler, Y., Skaloud, J., Victoria-Feser, M.-P.: Wavelet variance based estimation for composite stochastic processes. JASA, 108(503) (2013)
- Guerrier, S., Molinari, R.: Robust Inference for Time Series Models: a Wavelet-based Framework. Submitted working paper (2015)
- Guerrier, S., Molinari, R., Stebler, Y., Victoria-Feser, M.P.: Study of Allan Variance-based Techniques for Time Series Model Estimation and Wavelet Variance-based Improvements. Submitted working paper (2015)

Engineering applications:

- Stebler, Y., Guerrier, S., Skaloud, J.: An Approach for Observing and Modeling Errors in MEMS-based Inertial Sensors under Vehicle Dynamic, IEEE Transactions on Aerospace and Electronic Systems (2015, to appear).
- Stebler, Y., Guerrier, S., Skaloud, J., Victoria-Feser, M.P.: The Generalized Method of Wavelet Moments for Inertial Navigation Filter Design, IEEE Transactions on Aerospace and Electronic Systems (2014).
- Stebler, Y., Guerrier, S., Skaloud, J., Victoria-Feser, M.P.: A Framework for Inertial Sensor Calibration Using Complex Stochastic Error Models. Position Location and Navigation Symposium (PLANS), IEEE/ION (2012)

Wavelet Variance:

- Percival, D.P.: On estimation of the wavelet variance. Biometrika 82(3) (1995)
- Serroukh, A., Walden, A.T., Percival, D.B.: Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. Journal of the American Statistical Association 95 (2000)
- Zhang, N.F.: Allan variance of time series models for measurement data. Metrologia 45(5) (2008)
- Mondal, D. Percival, D.B.: M-estimation of Wavelet Variance. Annals of the Institute of Statistical Mathematics 64, 27-53 (2010)

Robust Time Series:

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