GMWM Package - Part I
A Computationally Efficient Platform for Inertial Sensor Calibration

Stéphane Guerrier

Department of Statistics
University of Illinois at Urbana-Champaign

joint work with
James Balamuta (UIUC)
Roberto Molinari (U. Geneva)
Jan Skaloud (EPFL)
Xinyu Zhang (CAS)

September 16, 2015

Outline

Part I

1. **Introduction**
   1. Latent time series models
   2. GMWM estimator
   3. GMWM-based model selection
   4. Why use the GMWM?
2. **GMWM package**
   1. Software architecture
   2. Computational efficiency
   3. List of features

Part II

3. **GMWM Demo**
   1. Loading data
   2. Computing WV and AV
   3. Visualization
   4. Estimation and inference
   5. Model selection
4. **Conclusions**
   1. Advantages of the GMWM
   2. Upcoming features

IMU Calibration

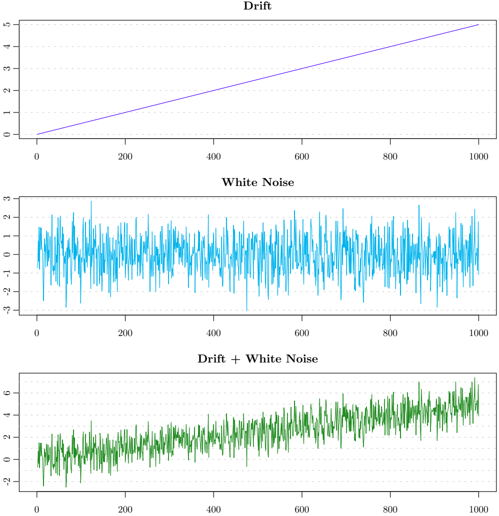
Errors in Inertial Sensors

- Possible causes:
  - Non-orthogonalities of the sensor axes
  - Environmental conditions (e.g. temperature)
  - Electronics
  - Dynamics
  - Others
- Error types:
  - Deterministic (calibration models, physical models, ...)
  - **Random error components** (typically latent time series models,...)

Correct stochastic sensor error modeling implies:

- Correct stochastic assumptions for inference
- **Better navigation or post-processing performance**

An easy latent time series model



Remarks:

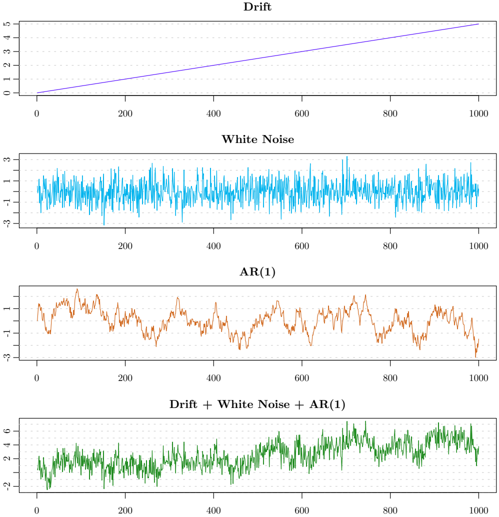
- Simple linear regression model:

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

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

- 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 GMWM estimator

Definition:

The GMWM estimator is the solution of the following optimization problem

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

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

Theoretical results (under technical conditions)

- $\boldsymbol{\nu}(\boldsymbol{\theta})$ is one-to-one in $\boldsymbol{\Theta}$.
- $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed.
- $\boldsymbol{\Omega} = \mathbf{V}^{-1}$ (where $\mathbf{V} \equiv \text{cov}(\hat{\boldsymbol{\nu}})$) is asymptotically optimal.
- $(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}))^T \boldsymbol{\Omega}^\star (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}))$ is asymptotically χ^2.

GMWM model selection criterion

The Wavelet Variance Information Criterion (WVIC)

We define the WVIC as:

$$\text{WVIC} = \mathbb{E}\left[\mathbb{E}_0\left[\left(\hat{\boldsymbol{\nu}}^0 - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)^T \boldsymbol{\Omega}\left(\hat{\boldsymbol{\nu}}^0 - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)\right]\right].$$

This model selection criterion is specifically developed for the GMWM estimation technical (as the C_p for LS or AIC with MLE).

Theoretical results (under technical conditions)

- *Guerrier et al. 2015* proposed an unbiased estimator of WVIC using bootstrap method (WIC*).
- *Zhang et al. 2015* proposed an **asymptotically** unbiased estimator of WVIC that does not require the use of simulation (WIC). They also proved that this method is loss efficient.

Why use the GMWM?

Summary

1. **Reliable, efficient and flexible** estimation method that can estimate the parameters of state space models for which other methods (e.g. MLE) fails. Generally good finite sample performance.
2. **Well defined statistical properties:** consistent and asymptotically normally distributed for large class of models. This is (in general) not the case for Allan variance based methods.
3. **Correct inference** for the parameters and allows for goodness-of-fit test(s).
4. **Automatic model selection** based on criterion (equivalent of AIC for MLE). This selection procedure is loss efficient.
5. **Computationally inexpensive:** can estimate complex latent with millions of observation in less ks a second on a typical computer.

Design of the GMWM R Package

Design Principles

1. **Fast**
   - Estimate large time series (10^6 and more) quickly.
   - Rapidly iterate over different models.
2. **Simple**
   - Accessible to everyone.
   - Straightforward commands.
3. **Portable**
   - Reusable across different computational platforms.
   - Possibly usable in embedded computer systems.

Software Architecture

Backend of GMWM Package

- The internals of the package are based solely on the matrix functions afforded by the **Armadillo** templated C++ Matrix Algebra Library .
- This enables the code to be **portable** to any computational platform that supports calling C++ functions.
- However, there is a **dependency** on R's graphics and optimization functions.

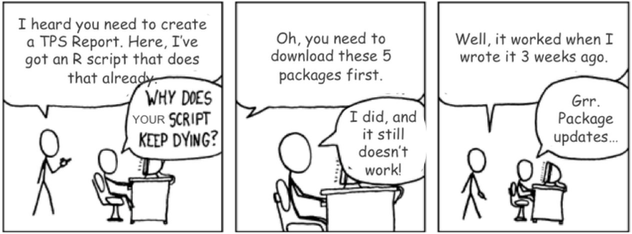
Frontend of GMWM Package

- The user interacts with GMWM backend through **R's GUI**.
- To facilitate this, we use a package called **RcppArmadillo**, which facilitates the transference of R objects into C++.
- The graphics generated **require** R-specific packages ggplot2, scales, grid, and gridExtra.

Why use R?

Benefits of R

- It's a **FREE** mature open source platform
- Similar to MATLab rapid prototyping from interpreter-architecture
- Able to integrate Fortran, C, C++, and Python code via wrappers
- Easy install and distribution of **latest statistical methods**
- Support from Fortune 500 companies (e.g. Oracle and Microsoft)
- But...



High-Performance Computing

Optimizing for Speed

- The GMWM package is very computationally efficient since a lot of work was put forth to decrease the amount of time required for each task required to compute the GMWM estimator.
- So, not only is the code for the GMWM estimation fast, but also for other parts of the package such as bootstrapping, inference, robustness, et cetera is also very quick.

Result:

The GMWM is able to estimate complex time series models with sample sizes of about 10^6 observations in less than a second.

Timings on Decomposition

Function	N		
	100	1,000	10,000
AV MO (gmwm R package)	0.0000	0.0005	**0.0261**
AV MO (avar R package)	0.0074	0.1053	**4.0189**
AV TO (gmwm R package)	0.0002	0.0001	**0.0003**
AV TO (avar R package)	0.0009	0.0104	**0.1053**

Table: Computational time in seconds performed on an iMac Late 2013 (3.5ghz i5 with 8 GB 1600 Mhz)

Note:

The AV gmwm-based implementation is (at least!) 100 times faster.

Package Features

- **User-friendly syntax** for time series model specification and simulation.
- **Efficient computation** of AV and WV (DWT or MODWT).
- Diverse error signal **visualization and model-validation** options.
- Tailor-made GMWM estimation for **IMU calibration models**.
- **Model Selection** via bootstrapped or asymptotic methods.
- **Robust alternatives** for all features against contaminated data.
- And many more..

Open Source + 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

Thank you very much for your attention!

A special thanks to

- Dr. Yannick Stebler (EPFL/u-blox)
- Prof. Maria-Pia Victoria-Feser (U. Geneva)
- Wenchao Yang (UIUC)

Any questions?

More info...

http://publish.illinois.edu/SMAC-group

stephane@illinois.edu