Generalized Autoregressive Method of Moments

Drew Creal [1] Siem Jan Koopman [2] André Lucas [2]
Marcin Zamojski [2]

[1] University of Chicago, Booth School of Business [2] VU University Amsterdam and Tinbergen Institute

8th Annual SoFiE Conference 2015
Aarhus, June 23–26, 2015

Paper @ zmks.co/gamm Slides @ zmks.co/gamm/slides gasmodel.com

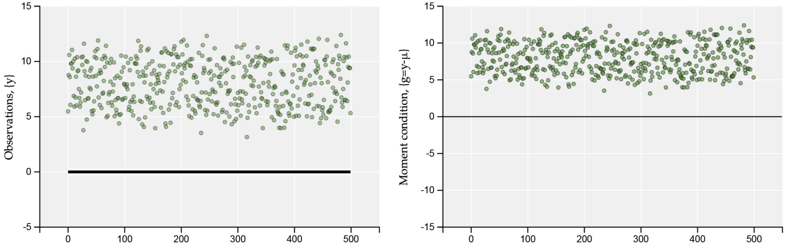
What is GaMM about?

- Part of the answer to the `million dollar' question of how to estimate time-varying parameters.
- Easily implementable and flexible method to be used when MLE is not feasible/desirable.

Outline

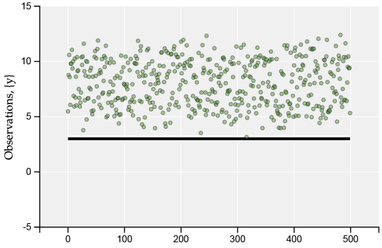
- Intuition
- Equations
- Application
- No time ☹
- Conclusions

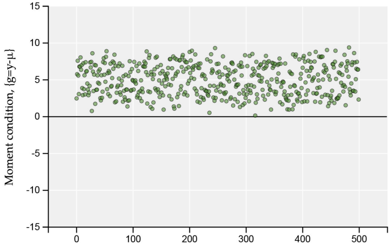
GMM



Given moment condition, $g_t = y_t - \mu$; minimize $\sum g_t^2$.

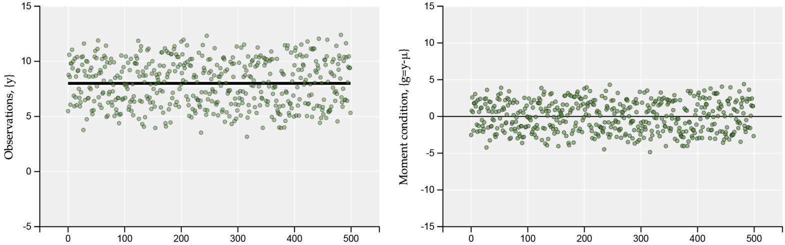
GMM





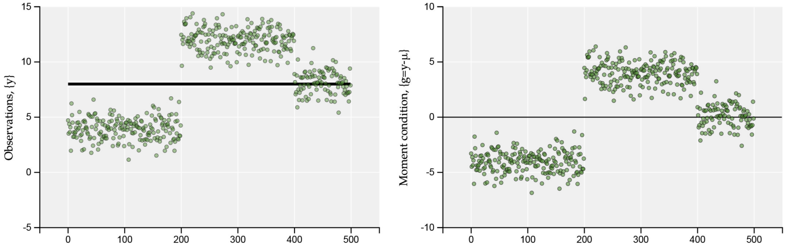
For a given initial guess for μ, optimizer looks at the *global misfit* of the model.

GMM



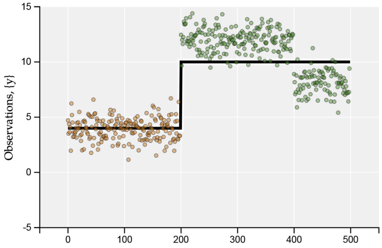
For a given initial guess for μ, optimizer looks at the *global misfit* of the model.

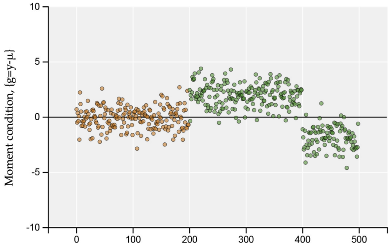
GMM with structural breaks



When data exhibits structural breaks, fitting the model with the full sample will not yield good results.

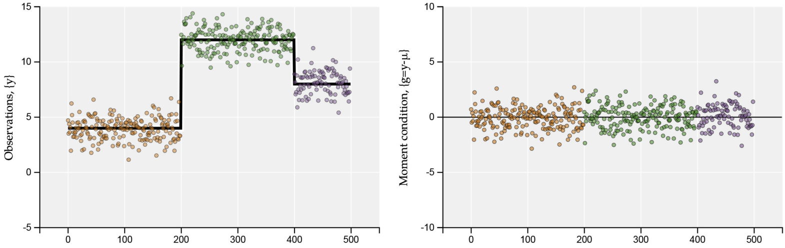
GMM with structural breaks



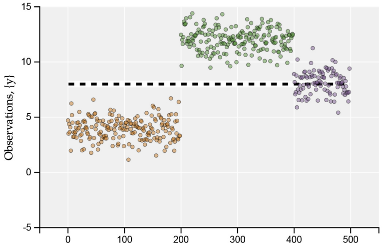


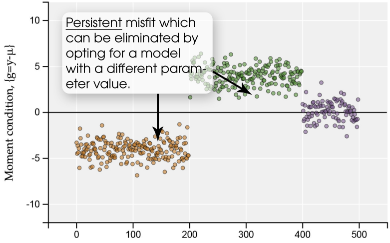
Instead, you would divide the sample into sub-samples and test whether **locally minimising misfit** of the model yields improvement over full sample, see Chow (1960), Quandt (1960), Andrews (1993), Bai (1997), Bai and Perron (1998), etc.

GMM with structural breaks



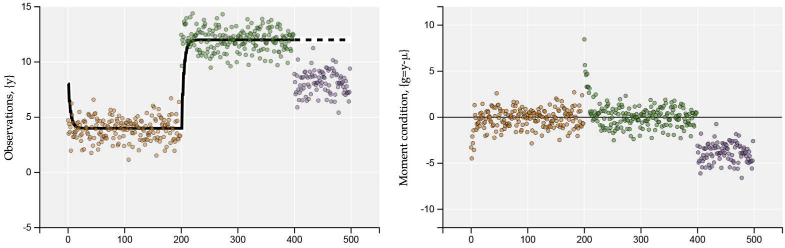
Our approach, GaMM





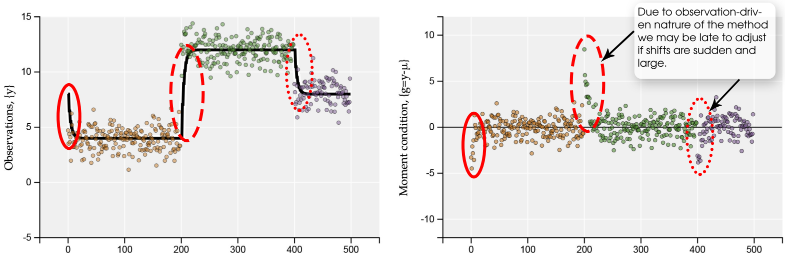
Observation driven method, so: $\hat{\mu}_{t+1} = \omega\,(I - B)\,B\hat{\mu}_t + Af\,(g_t)$

Our approach, GaMM



Observation driven method, so: $\hat{\mu}_{t+1} = B\hat{\mu}_t + Af(g_t)$

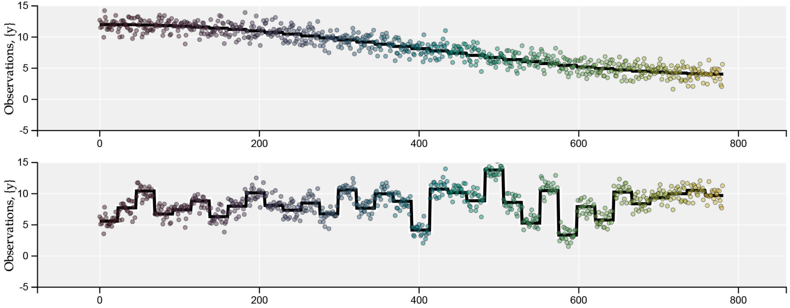
Our approach, GaMM



Observation driven method, so: $\hat{\mu}_{t+1} = B\hat{\mu}_t + Af(g_t)$

Contribution

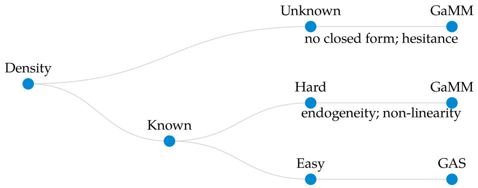
Observation driven GMM which allows any parameter dynamics.



Related literature

GaMM eats GAS

Creal, Koopman, and Lucas (2013)



Other methods

Parameter-driven: Kim and Nelson (1999), Durbin and Koopman (2012), McFadden (1989), Gouriéroux et al. (1993), Gallant and Tauchen (1996), Gallant et al. (2014).

GaMM(1,1)

$$\mathrm{E}_{t-1}\left[g_t\left(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta}\right)\right] = 0 \tag{1}$$

$$\{\boldsymbol{w}_t\} = \{y_t, \boldsymbol{x}_t, \boldsymbol{z}_t\} \tag{2}$$

GMM objective function at time t:

$$\mathrm{E}_{t-1}\left[g_t\left(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta}\right)\right]^\intercal \boldsymbol{\Omega}_t \, \mathrm{E}_{t-1}\left[g_t\left(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta}\right)\right] \tag{3}$$

We use the notion of Fréchet derivative to derive first-order conditions:

$$\boldsymbol{G}_t^{\epsilon\,\intercal} \boldsymbol{\Omega}_t \, \mathrm{E}_{t-1}^{\epsilon}\left[g_t\left(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta}\right)\right] = 0 \tag{4}$$

$$\boldsymbol{G}_t^{\epsilon} = \mathrm{E}_{t-1}^{\epsilon}\left[\frac{\partial \boldsymbol{g}_t\left(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta}\right)}{\partial \boldsymbol{f}_t^\intercal}\right] \tag{5}$$

Where: $\boldsymbol{\Omega}_t = \left\{\mathrm{E}_{t-1}\left[\boldsymbol{g}_t \boldsymbol{g}_t^\intercal\right]\right\}^\star$ (efficient) and $H^\star$ is the Moore-Penrose pseudo inverse. In the paper, we use $\boldsymbol{\Omega}_t = \boldsymbol{I}$ with good results. The nature of the problem determines whether and how $\boldsymbol{G}_t$ needs to be estimated.

Transition dynamics for $f_t \to f_{t+1}$

General updating step:

$$s_t = -\left(G_t^{\top}\Omega_t G\right)^{\star} G_t^{\top}\Omega_t g_t\left(w_t; f_t, \theta\right) \tag{6}$$

Time-varying parameter recursion:

$$f_{t+1} = B\left(f_t - \omega\right) + \omega + A s_t \tag{7}$$

Local optimality of dynamics

- Blasques et al. (2015) show that score updates (for GAS models) improve the local Kullback-Leibler divergence between the true data density and the model density.
- We extend these results from a fully parametric setting to the semi-parametric setting.
- GaMM dynamics efficiently process new observations and update f_t to f_{t+1}.
- Results hold whether or not the statistical model is correctly specified.

Estimation

GaMM instruments:

$$\bar{\boldsymbol{g}}_t = \boldsymbol{g}_t\left(\boldsymbol{w}_t; [\boldsymbol{\theta}, \boldsymbol{f}_t]\right) \otimes \left[\begin{array}{ccc} 1 & \boldsymbol{s}_{t-1} & \boldsymbol{f}_{t-1} \end{array}\right]^{\mathsf{T}} \tag{8}$$

$$\bar{\boldsymbol{g}} = \frac{1}{T}\sum_{t=1}^{T} \bar{\boldsymbol{g}}_t \approx \mathrm{E}\left[\bar{g}_t\left(\cdot\right)\right] = 0 \tag{9}$$

Criterion function, where $\boldsymbol{\theta}$ corresponds to all static parameters, including $\boldsymbol{\omega}$, $\boldsymbol{A}$, $\boldsymbol{B}$:

$$\min_{\boldsymbol{\theta} \in \Theta} \bar{\boldsymbol{g}}^{\mathsf{T}} \bar{\Omega} \bar{\boldsymbol{g}} \tag{10}$$

Estimation

GaMM instruments:

$$\bar{\boldsymbol{g}}_t = \boldsymbol{g}_t\left(\boldsymbol{w}_t; [\boldsymbol{\theta}, \boldsymbol{f}_t]\right) \otimes \begin{bmatrix} 1 & \boldsymbol{s}_{t-1} & \boldsymbol{f}_{t-1} \end{bmatrix}^{\top} \tag{8}$$

$$\bar{\boldsymbol{g}} = \frac{1}{T}\sum_{t=1}^{T} \bar{\boldsymbol{g}}_t \approx \mathrm{E}\left[\bar{g}_t(\cdot)\right] = 0 \tag{9}$$

Criterion function, where $\boldsymbol{\theta}$ corresponds to all static parameters, including $\boldsymbol{\omega}$, $\boldsymbol{A}$, $\boldsymbol{B}$.

$$\min_{\boldsymbol{\theta} \in \Theta} \bar{\boldsymbol{g}}^{\top} \bar{\Omega} \bar{\boldsymbol{g}} \tag{10}$$

Asymptotic theory

Consistency and asymptotic normality results follow from Hansen (1982), including an expression for the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ under standard high-level regularity conditions.

$$n^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{H}^{-1}\bar{\boldsymbol{D}}\boldsymbol{H}^{-1}\right), \qquad \bar{\boldsymbol{D}} = \bar{\boldsymbol{G}}^{\top}\bar{\Omega}\bar{\boldsymbol{V}}\bar{\Omega}\bar{\boldsymbol{G}}. \tag{11}$$

The efficient weighting matrix is $\bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{V}}^{-1}$, in which case the asymptotic covariance matrix collapses to $(\bar{\boldsymbol{G}}^{\top}\, \bar{\Omega}\, \bar{\boldsymbol{G}})^{-1}$.

Optimality of instruments

- $\begin{bmatrix} 1 & \boldsymbol{s}_{t-1} & \boldsymbol{f}_{t-1} \end{bmatrix}$ are not optimal.
- Optimal instruments would be of the form:

$$\boldsymbol{W}_t^\top = \boldsymbol{\Omega}_t \, \mathrm{E}_{t-1}\left[\frac{\partial \boldsymbol{g}_t(\boldsymbol{w}_t; \boldsymbol{f}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right] + \boldsymbol{\Omega}_t \, \boldsymbol{G}_t \, \frac{\mathrm{d}\, \boldsymbol{f}_t(\boldsymbol{w}_{t-1}; \boldsymbol{f}_{t-1}, \boldsymbol{\theta})}{\mathrm{d}\, \boldsymbol{\theta}^\top} \tag{12}$$

$$\frac{\mathrm{d}\, \boldsymbol{f}_{t+1}}{\mathrm{d}\, \boldsymbol{\theta}^\top} = \left(0 \;\vdots\; \mathbf{I} - \boldsymbol{B} \;\vdots\; \boldsymbol{f}_t^\top \otimes \mathbf{I} \;\vdots\; \boldsymbol{s}_t^\top \otimes \mathbf{I} \right) + \boldsymbol{B} \frac{\mathrm{d}\, \boldsymbol{f}_t}{\mathrm{d}\, \boldsymbol{\theta}^\top} + \boldsymbol{A} \frac{\partial \boldsymbol{s}_t}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{A} \frac{\partial \boldsymbol{s}_t}{\partial \boldsymbol{f}_t^\top} \frac{\mathrm{d}\, \boldsymbol{f}_t}{\mathrm{d}\, \boldsymbol{\theta}^\top} \tag{13}$$

- Optimal instruments are a smoothed version of $\begin{bmatrix} 1 & \boldsymbol{s}_{t-1} & \boldsymbol{f}_{t-1} \end{bmatrix}$.
- What we propose accounts for the dominant source of variation in the optimal instruments, while avoiding additional recursions during the estimation stage.

Flexibility

The joint GaMM dynamics for the vector $\tilde{\boldsymbol{f}}_{t+1} = (\boldsymbol{f}_t{}^\top, \boldsymbol{\theta}_{\boldsymbol{c}}^\top)^\top$ are

$$\tilde{\boldsymbol{f}}_{t+1} = \begin{bmatrix} \boldsymbol{f}_t \\ \hline \boldsymbol{\theta}_{\boldsymbol{c}} \end{bmatrix} = \tilde{\boldsymbol{\omega}} + \sum_{j=1}^{p} \tilde{\boldsymbol{B}}_j \left(\tilde{\boldsymbol{f}}_{t-j+1} - \tilde{\boldsymbol{\omega}} \right) + \sum_{i=1}^{q} \tilde{\boldsymbol{A}}_i \tilde{\boldsymbol{s}}_{t-i+1}, \tag{14}$$

$$\tilde{\boldsymbol{B}}_j = \begin{bmatrix} \boldsymbol{B}_j & \boldsymbol{C}_j^B \\ \hline \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \tilde{\boldsymbol{\omega}} = \begin{bmatrix} \boldsymbol{\omega} \\ \hline \boldsymbol{\theta} \end{bmatrix}, \quad \tilde{\boldsymbol{A}}_i = \begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{C}_i^A \\ \hline \boldsymbol{0} & \boldsymbol{0} \end{bmatrix},$$

The matrices $\boldsymbol{C}_i^A$ and $\boldsymbol{C}_j^B$ allow the time-varying parameter $\boldsymbol{f}_t$ to also react to the score of the static parameters $\boldsymbol{\theta}_{\boldsymbol{c}}$.

Flexibility

Consider a model with time-varying variance σ_t^2 and constant mean μ, i.e., $y_t = \mu + \varepsilon_t$, with $\varepsilon_t = \sigma_t z_t$, and $z_t \sim D(0,1)$ for some distribution D with zero mean and variance one. We obtain that

$$\boldsymbol{g}_t = \begin{bmatrix} y_t - \mu \\ (y_t - \mu)^2 - \sigma_t^2 \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t^2 - \sigma_t^2 \end{bmatrix}, \qquad \boldsymbol{G}_t = -\mathbf{I}. \tag{15}$$

Fixing $\boldsymbol{C}^B = 0$, the GaMM(1,1) recursion with $\boldsymbol{A}_1 = \boldsymbol{A}$ and $\boldsymbol{B}_1 = \boldsymbol{B}$ for $\boldsymbol{f}_t = \sigma_t^2$ is

$$\sigma_{t+1}^2 = \underbrace{\boldsymbol{\omega}\,(1 - \boldsymbol{B}) + A\varepsilon_t^2 + (\boldsymbol{B} - \boldsymbol{A})\sigma_t^2}_{\text{GARCH(1,1)}} + \underbrace{\boldsymbol{C}^A \varepsilon_t}_{\text{Leverage}}.$$

Which coincides with the familiar GARCH(1,1) model of Engle (1982) and Bollerslev (1986) with an additional leverage effect $\boldsymbol{C}^A \varepsilon_t$. The leverage effect is similar in form and spirit to the optimal leverage effect of GARCH filters in a mis-specified model setting as laid out in Nelson and Foster (1994).

Application: time-varying risk aversion in CCAPM

The standard power utility

$$1 = \mathrm{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R^x_{t+1} \right] \tag{16}$$

Where, as in Hansen and Singleton (1982), C_t is aggregate real consumption measure and R^x_t is real return from a test asset.

Facts:

- discount factor is consistently estimated close to 1.0; Ghysels and Hall (1990); Hansen and Singleton (1982)
- risk-aversion is sensitive to initial values, sample selection, and when estimated it tends to be an order of magnitude too high; Hansen and Singleton (1982), Savov (2011), Mehra and Prescott (1985)

Application: time-varying risk aversion in CCAPM

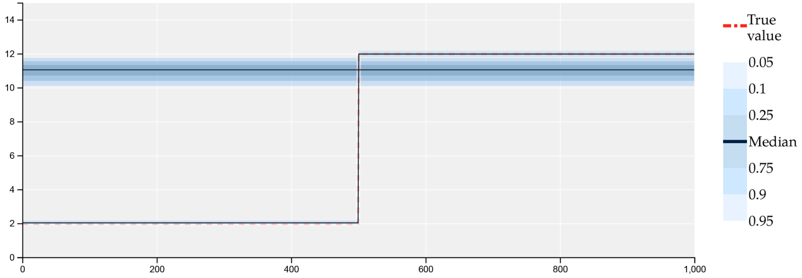
$$1 = \mathsf{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_t} R^x_{t+1} \right] \tag{17}$$

We assume that shocks to risk-aversion are exogenous and that agents are myopic in the sense that they consider γ_t to remain fixed forever when making their decision at time t.

We can directly employ the GaMM framework by taking $f_t = \gamma_t$.

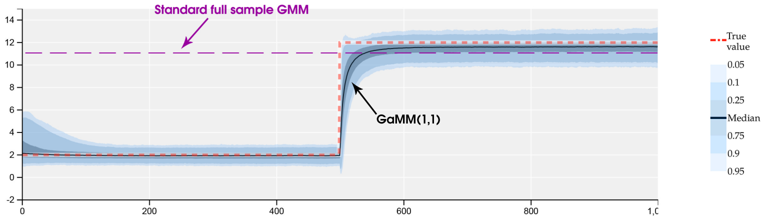
Simulations: structural breaks

GMM



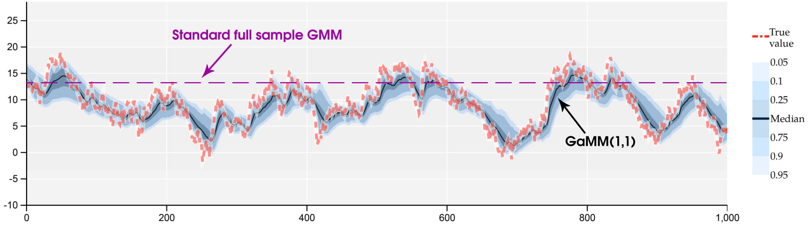
Simulations: structural breaks

GaMM



Simulations: AR(1)

GaMM



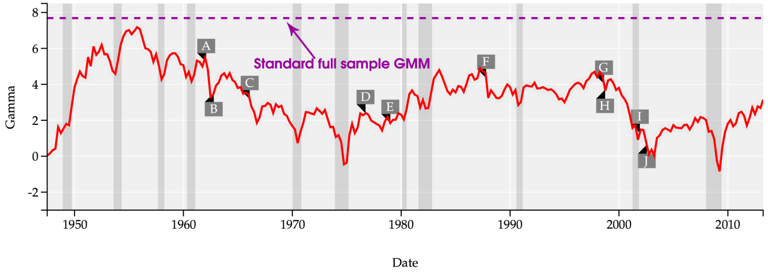
U.S. quaterly data (1947–2013)

- quaterly data from 1947–2013
- current price returns from Shiller (2001)
- current price consumption expenditures on non-durables and services from BEA
- deflators from Shiller (2001) and updated based on BEA's NIPA tables
- we fix $\beta = 0.995$ in estimation to bring focus to the risk-aversion parameter; Hansen, et. al. (2008); Savov (2011)

U.S. quaterly data (1947–2013)

Period	Real returns [%]	Real cons. growth [%]	Risk-aversion
1950–1960	7.36	0.49	5.53
1990–2000	4.66	0.6	3.76
2000–2010	1.43	0.41	1.5

U.S. quaterly data (1947–2013)



A U.S. announces embargo agains Cuba

B Cuban missile crisis

C 1966 Bear Market and Credit Crunch

D 1976 Bear Market (stagflation)

E Second Oil Crisis

F Black Monday

G Russia defaults on its domestic debt

H LTCM bailout

I Enron files for bankruptcy

J WorldCom files for bankruptcy

Also in the paper...

- Two additional applications:
  - Time-varying scale in stable distributions.
    GaMM estimates the time-varying scale and remaining static parameters without bias.
  - Endogenous regressor problem.
    We benchmark GaMM against Kalman Filter, the Kalman Filter results are biased but with low sampling variability, whereas paths obtained with GaMM are unbiased on average, but at the cost of a higher sampling variance.
- Penalized criterion function which further reduces RMSE.

Conclusions

- GaMM extends Generalized Method of Moments (GMM) to a setting where a **subset of the parameters are expected to vary over time** with unknown dynamics.
- You only need to specify a set of conditional moment conditions and a set of parameters which are assumed to vary over time.
- **We approximate the dynamics by an autoregressive process driven by the score** of the local GMM criterion function to filter out the dynamic path of the time-varying parameter.
- Our approach is **completely observation driven**, such that estimation and inference are entirely straightforward.
- GaMM is flexible, easily implementable, and allows for any parameter dynamics, e.g., structural breaks, sinusoidal waves, AR(p), etc.
- GaMM generalizes GAS (Creal et al., 2013) to settings where densities are unknown or hard to derive.