Forecasting Time Series using GAS models

Francisco Blasques

ISF2014 Rotterdam Workshop

VRIJE
UNIVERSITEIT
AMSTERDAM

www.gasmodel.com

Today's Schedule

09:00 - 10:15	Lecture
10:15 - 10:30	Coffee Break
10:30 - 11:30	Lecture
11:30 - 12:00	Informal Q&A
12:00 - 13:00	Lunch
13:00 - 14:15	Lecture
14:15 - 14:30	Coffee Break
14:30 - 15:30	Lecture
15:30 - 16:00	Informal Q&A

Today's Summary

Morning: Theory

1. Introduction to GAS models
2. Why use GAS models?
3. Stochastic properties of GAS models
4. Estimating the parameters of GAS models

Afternoon: Applications

- GAS Forecasting: Comparison with State Space Models
- Dynamic Factor Models with Mixed Measurements and Mixed Frequencies
- Modeling Dynamic Volatilities and Correlations with GAS

Morning Summary

1. Introduction to GAS models
2. Why use GAS models?
   - The GAS update is optimal
3. Stochastic properties of GAS models
   - Strict stationarity and ergodicity
   - Existence of unconditional moments
4. Estimating the parameters of GAS models
   - Estimation in mis-specified models
   - Estimation in well-specified models

Introduction to GAS Models

Creal, Koopman and Lucas (2014)
"A General Framework for Observation Driven Time-Varying Parameter Models"

www.gasmodel.com

What are GAS models?

GAS models are state space models :

i.e. Behavior of $\{y_t\}$ explained using a **tv parameter** $\{f_t\}$:

$$y_t = f_t + u_t \quad \text{where} \quad \{u_t\} \text{ is iid}$$

What are GAS models?

GAS models are state space models :

i.e. Behavior of $\{\boldsymbol{y_t}\}$ explained using a **tv parameter** $\{\boldsymbol{f_t}\}$:

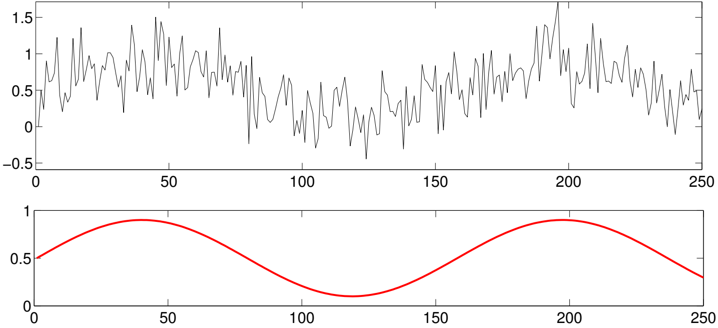
$$y_t = f_t + u_t \quad \text{where} \quad \{u_t\} \;\text{ is iid}$$

Note: This is a postulated structure! DGP can be different!

GAS models are state space models :

i.e. Behavior of $\{y_t\}$ explained using a **tv parameter** $\{f_t\}$:

$$y_t = f_t + u_t \quad \text{where} \quad \{u_t\} \text{ is iid}$$

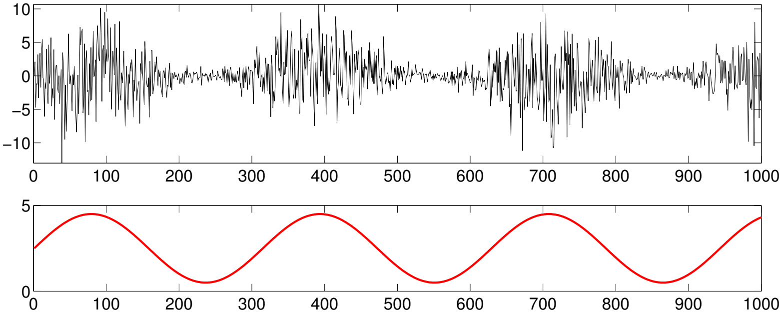


What are GAS models?

GAS models are state space models :

i.e. Behavior of $\{y_t\}$ explained using a **tv parameter** $\{f_t\}$:

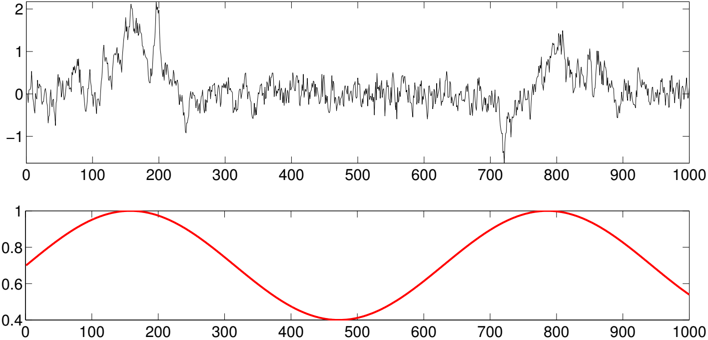
$$y_t = f_t \cdot u_t \quad \text{where} \quad \{u_t\} \text{ is iid}$$



GAS models are state space models :

i.e. Behavior of $\{y_t\}$ explained using a **tv parameter** $\{f_t\}$:

$$y_t = f_t \cdot y_{t-1} + u_t \quad \text{where} \quad \{u_t\} \text{ is iid}$$



What are GAS models?

GAS models are observation-driven models:

i.e. Update of $\boldsymbol{f_{t+1}}$ depends on observed $\boldsymbol{y_t}$

$$y_t|f_t \sim p(y_t|f_t) \quad , \quad f_{t+1} = \phi(y_t, f_t)$$

What are GAS models?

GAS models are observation-driven models:

i.e. Update of $\boldsymbol{f_{t+1}}$ depends on observed $\boldsymbol{y_t}$

$$y_t|f_t \sim p(y_t|f_t) \quad , \quad f_{t+1} = \phi(y_t, f_t)$$

Alternative: Parameter-driven models:

i.e. Update of $\boldsymbol{f_t}$ depends on exogenous innovation $\boldsymbol{u_t}$

$$y_t|f_t \sim p(y_t|f_t) \quad , \quad f_{t+1} = \phi(u_t, f_t)$$

What are GAS models?

GAS models are observation-driven models:

i.e. Update of $\boldsymbol{f_{t+1}}$ depends on observed $\boldsymbol{y_t}$

$$y_t | f_t \sim p(y_t | f_t) \quad , \quad f_{t+1} = \phi(y_t, f_t)$$

So what is special about the GAS?

- Computationally simple to use (observation driven)
- Substitute ad-hoc and problem-specific parameter update function ϕ by an internally consistent and general update ϕ
- The GAS update is optimal!

Consider the following tv mean $\{y_t\}$ and estimated $\{f_t\}$ ...

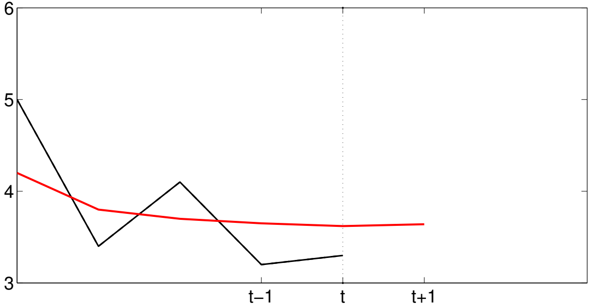


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

Consider the new observation of y_{t+1}...

Question: How should f_{t+2} be updated?

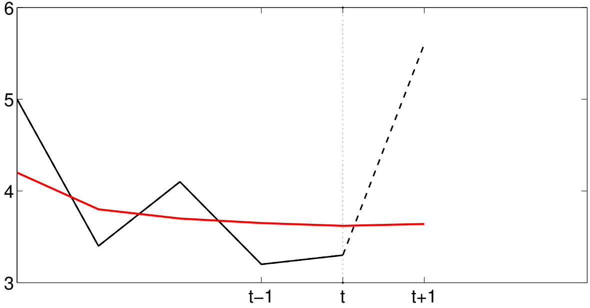


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

The idea... time-varying mean example

A possible answer: $f_{t+2} = \phi(y_{t+1}, f_{t+1}) = \omega + \alpha y_{t+1} + \beta f_{t+1}$

GAS answer: it depends! what does $p(y_{t+1}|f_{t+1})$ look like?

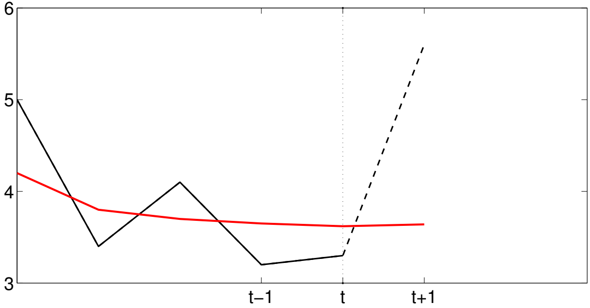


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

The idea... time-varying volatility example

Consider the following tv volatilty $\{y_t\}$ and estimated $\{f_t\}$ ...

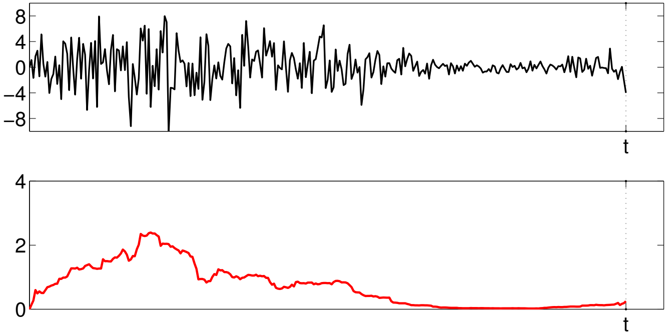


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

Consider the new observation of y_{t+1}...

Question: How should f_{t+2} be updated?

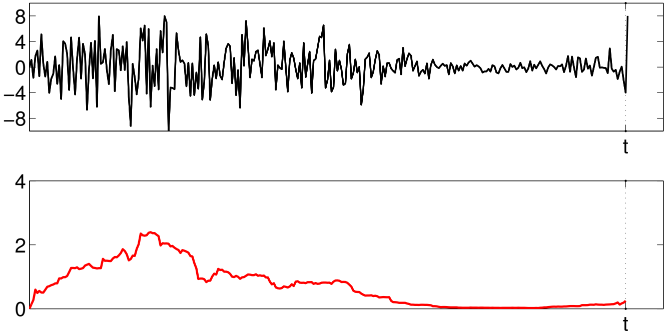


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

A possible answer: $f_{t+2} = \phi(y_{t+1}, f_{t+1}) = \omega + \alpha y_{t+1}^2 + \beta f_{t+1}$

GAS answer: it depends! how is $p(y_{t+1}|f_{t+1})$?

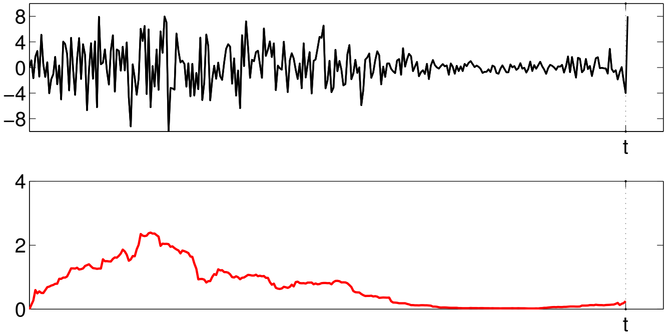


Figure : $\{y_t\}$ (black line) , $\{f_t\}$ (red line)

The GAS Model

$$y_t = g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u$$

$$f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t.$$

- ω, α, β and λ are **time-invariant** scalar parameters.
- $s(y_t, f_t) = \nabla(y_t, f_t) \cdot S(f_t)$ is a scaled score:

$$\nabla(y_t, f_t) := \partial \ln p(y_t | f_t) / \partial f$$

- $s(y_t, f_t)$ can be **linear, nonlinear or invariant in** f_t.

Distinctive feature of GAS: The use of $s(y_t, f_t)$

$$f_{t+1} \;=\; \phi(y_t, f_t) \;=\; \omega + \alpha s(y_t, f_t) + \beta f_t$$

$$s(y_t, f_t) \;:=\; \nabla(y_t, f_t) \cdot S(f_t) \;=\; \frac{\partial \ln p(y_t|f_t)}{\partial f_t} \cdot S(f_t),$$

Intuition: update parameter f_t to more likely value f_{t+1} by taking steepest ascent step in direction $s(y_t, f_t)$.

Scaling: use local curvature to improve step; e.g. a power $(a = 0, 1, \frac{1}{2})$ of the inverse information matrix

$$S(f_t) = E_{t-1}\Big[\nabla^2(y_t, f_t)\Big]^{-a}.$$

GAS Models: Generality

Question: What kind of updates do we get?

Answer: Well... many!

GAS encompasses well-known observation driven time-varying parameter models as characterized by Cox (1981):

- **GARCH:** Engle (1982) and Bollerslev (1986),
- **EGARCH:** Nelson (1991) ,
- **ACD:** Engle & Russell (1998),
- **MEM:** Engle (2002),
- **ACM:** Rydberg & Shephard (2003).

Example: Time-Varying Mean with GAS Dynamics

$$y_t = f_t + \sigma u_t \quad \text{where} \quad f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t,$$

$$\{u_t\} \;\; \text{iid} \;\; \text{with pdf } p_u \quad , \quad (\omega, \alpha, \beta, \sigma) = \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^4$$

If : $u_t \sim N(0,1)$ and $S(f) = 1$

Then : $p(y_t|f_t) \sim N(f_t, \sigma^2)$, $p(y_t|f_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y_t - f_t)^2}{2\sigma^2})$

Hence : $s(y_t, f_t) = \frac{\partial \log p_u(y_t - f_t)}{\partial f} = \frac{y_t - f_t}{\sigma^2}$

So the GAS update is: $f_{t+1} = \omega + \alpha(y_t - f_t)/\sigma^2 + \beta f_t$

Example: Time-Varying Mean with GAS Dynamics

$$y_t = f_t + \sigma u_t \quad \text{where} \quad f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t,$$

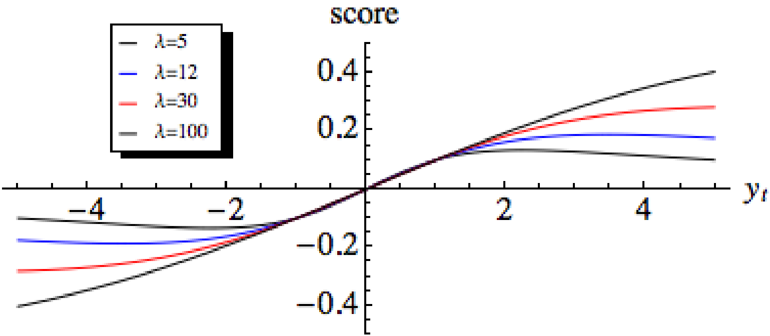
$$\{u_t\} \ \text{ iid } \ \text{with pdf } p_u(\lambda) \quad , \quad (\omega, \alpha, \beta, \sigma, \lambda) = \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^5$$

If : $u_t \sim t(\lambda) \quad \text{and} \quad S(f) = 1$

Then : $$s(y_t, f_t) = \frac{(\lambda+1)(y_t - f_t)/\sigma^2}{1 + (y_t - f_t)^2/\sigma^2 + \lambda}$$

Hence : $$f_{t+1} = \omega + \alpha \frac{(\lambda+1)(y_t - f_t)/\sigma^2}{1 + (y_t - f_t)^2/\sigma^2 + \lambda} + \beta f_t$$

Example: Time-Varying Mean with GAS Dynamics



Example: Time-Varying Mean with GAS Dynamics

$$y_t = h(f_t) + u_t \quad \text{where} \quad f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t,$$

$$\{u_t\} \ \text{iid} \ \text{with pdf } p_u \quad , \quad (\omega, \alpha, \beta, \sigma) = \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^4$$

In this model : $s(y_t, f_t) = -S(f_t)\nabla p_u(y_t - h(f_t))h'(f_t)$

If : $u_t \sim N(0,1)$ and $S(f) = 1$

Then : $s(y_t, f_t) = h'(f_t)(y_t - h(f_t))$

Important : $p_u,\ S,\ h \quad \Rightarrow \quad p(y_t|f_t) \quad \Rightarrow \quad s(y_t, f_t)$

Example: Dynamic Volatility Model

$$y_t = h(f_t)u_t \quad \text{where} \quad f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t,$$

$$\{u_t\} \ \text{ iid } \ \text{with pdf } p_u \quad , \quad (\omega, \alpha, \beta) = \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^3$$

If : $u_t \sim N(0,1) \quad , \quad h(f_t) = \sqrt{f_t} \quad \text{and} \quad S(f_t) = \mathcal{I}^{-1}$

Then : $s(y_t, f_t) = y_t^2 - f_t \quad \text{(GARCH)}$

If : $u_t \sim t(\lambda) \quad , \quad h(f_t) = \sqrt{f_t} \quad \text{and} \quad S(f_t) = \mathcal{I}^{-1}$

Then : $s(y_t, f_t) = (1 + 3\lambda^{-1})\left(\dfrac{y_t^2(1 + \lambda^{-1})}{1 + \lambda^{-1} y_t^2 / f_t} - f_t\right)$

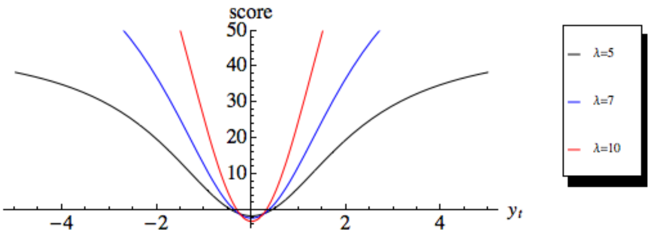
Example: Time-Varying Volatility with GAS Dynamics

Volatility Example: $y_t = \sqrt{\tilde{f}_t} u_t \quad , \quad u_t \sim \mathrm{t}(\lambda)$

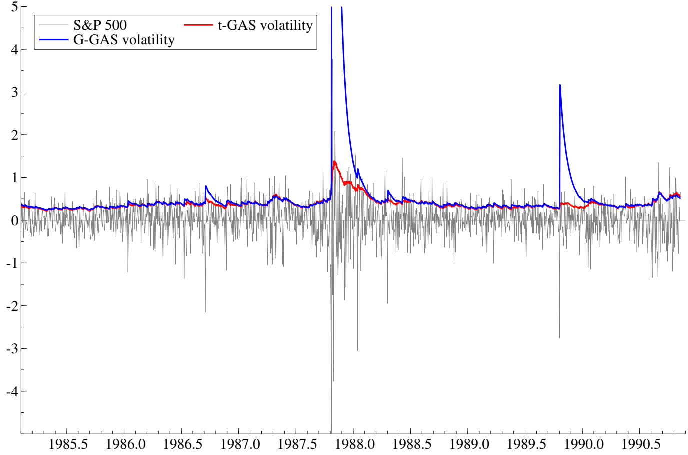
t-GARCH update: $\tilde{f}_{t+1} = \omega + \alpha y_t^2 + \beta \tilde{f}_t$

t-GAS update: $\tilde{f}_{t+1} = \omega + \alpha(1 + 3\lambda^{-1})\left(\frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda \tilde{f}_t)} - \tilde{f}_t\right) + \beta \tilde{f}_t$

Main idea: score suggests conservative update for small λ



Application: Dynamic Volatility Model



Optimality of GAS Models

Blasques, Koopman and Lucas (2012)
"Information Theoretic Optimality of Observation Driven Time Series Models"

www.gasmodel.com

First Motivation

Motivation 1:

1. GAS update is very intuitive!
2. Using the score seems optimal in some sense!
3. Likelihood and KL divergence are closely related

Question: Is the GAS update optimal in some KL sense?

Challenge 1: GAS update is surely not always correct!

Challenge 2: Comparing different observation-driven models is only interesting under misspecification with very general DGP!

Optimal Observation-Driven Update

Question: Is there an optimal updating equation?

Answer: This depends on the notion of optimality!

Result 1: The GAS update is optimal *in KL-variation*

Result 2: Only a GAS-type update is optimal *in KL-variation*

Note: Results hold for very general DGP!

Motivation 2: Great variety of update functions in time-varying parameter models.

Example: Volatility models: $y_t = \sqrt{f_t} u_t$

GARCH Engle (1992), Bollerslev (1986)

$$f_{t+1} = \omega + \alpha y_t^2 + \beta f_t$$

AV-GARCH Taylor (1986), Nelson & Foster (1994)

$$f_{t+1} = \omega + \alpha |y_t| + \beta f_t$$

t-GAS Creal at al. (2011,2013), Lucas et al. (2014)

$$f_{t+1} = \omega + \alpha(1 + 3\lambda^{-1})\left(\frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda f_t)} - f_t\right) + \beta f_t$$

Question: which one is best?

Question: which one is best?

Answer: depends on notion of best

Answer: depends on the distribution of the data (DGP)

Refined question: Allowing for a very general DGP, which update is optimal in the Kullback-Leibler (KL) sense?

Short answer: The GAS update!

Claim 1: The GAS update is optimal in 'KL variation'

Claim 2: Only a GAS update can be optimal in 'KL variation'

Claim 3: Optimality in KL variation matters in practice!

DGP and Observation-Driven Model

True sequence of conditional densities:

$$\left\{ \, p(y_t|f_t) \, \right\} , \quad \textit{true} \text{ time-varying parameter } \{f_t\}$$

Conditional densities *postulated by probabilistic model:*

$$\left\{ \, \tilde{p}\big(y_t|\tilde{f}_t;\boldsymbol{\theta}\big) \, \right\} , \quad \textit{filtered} \text{ time-varying parameter } \{\tilde{f}_t\}$$

$\tilde{p}\big(y_t|\tilde{f}_t;\boldsymbol{\theta}\big)$ is defined *implicitly by observation equation:*

$$y_t = g\big(\tilde{f}_t \, , \, u_t \, ; \, \boldsymbol{\theta}\big), \qquad u_t \sim p_u(\boldsymbol{\theta}),$$

Observation-driven parameter update:

$$\tilde{f}_{t+1} = \phi\big(y_t \, , \, \tilde{f}_t \, ; \, \boldsymbol{\theta}\big), \qquad \forall \; t \in \mathbb{N}, \qquad \tilde{f}_1 \in \mathcal{F} \subseteq \mathbb{R},$$

Optimal Observation-Driven Update

Question: Is there an optimal form for the update equation?

$$\tilde{f}_{t+1} = \phi\big(y_t\,,\,\tilde{f}_t\,;\,\boldsymbol{\theta}\big), \qquad \forall\, t \in \mathbb{N}, \qquad \tilde{f}_1 \in \mathcal{F} \subseteq \mathbb{R},$$

Answer: Yes! The GAS update which takes the form

$$\tilde{f}_{t+1} = \omega + \alpha s(y_t, \tilde{f}_t) + \beta \tilde{f}_t, \qquad \forall\, t \in \mathbb{N},$$

where $s(y_t, \tilde{f}_t)$ is the **scaled score**

$$s(y_t, \tilde{f}_t) = S(\tilde{f}_t) \cdot \frac{\partial \log \tilde{p}(y_t | \tilde{f}_t; \boldsymbol{\theta})}{\partial f}$$

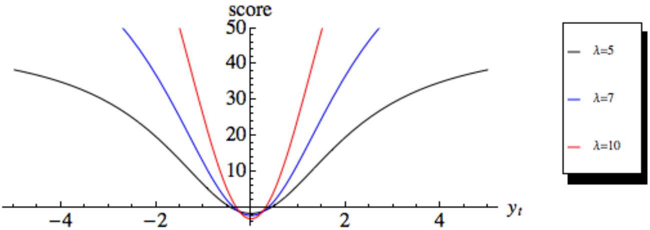
Main Idea: update parameter like in a Newton algorithm!

Volatility Example: $y_t = \sqrt{\tilde{f}_t} u_t \quad , \quad u_t \sim \mathrm{t}(\lambda)$

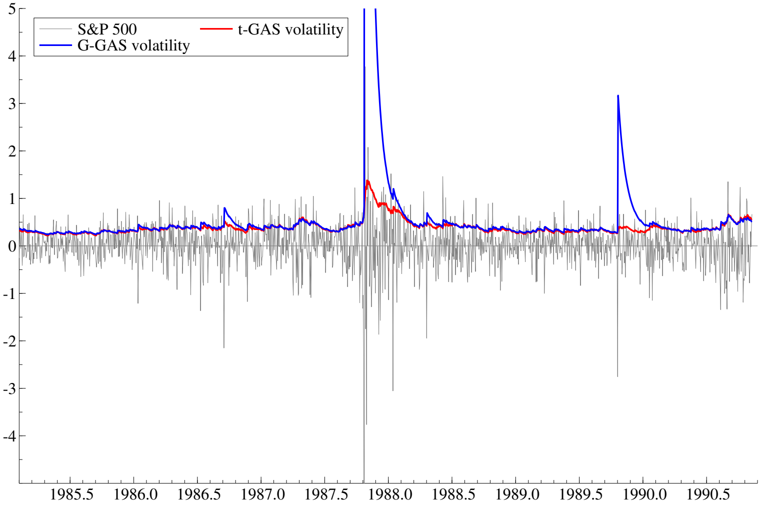
t-GARCH update: $\tilde{f}_{t+1} = \omega + \alpha y_t^2 + \beta \tilde{f}_t$

t-GAS update: $\tilde{f}_{t+1} = \omega + \alpha(1 + 3\lambda^{-1})\left(\frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda \tilde{f}_t)} - \tilde{f}_t\right) + \beta \tilde{f}_t$

Main idea: score suggests conservative update for small λ



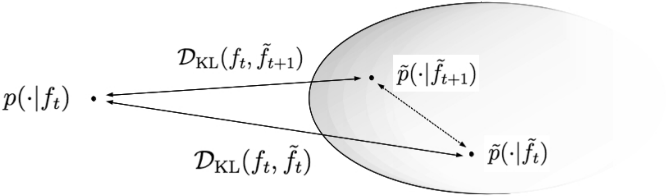
GAS Model: Volatility Example



Definition: The *KL variation* Δ of a parameter update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ is defined as

$$\Delta = \mathcal{D}_{\text{KL}}(f_t, \tilde{f}_{t+1}) - \mathcal{D}_{\text{KL}}(f_t, \tilde{f}_t)$$

where $$\mathcal{D}_{\text{KL}}(f_t, \tilde{f}_t) \equiv \mathcal{D}_{\text{KL}}\Big(p(\cdot|f_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_t;\boldsymbol{\theta})\Big).$$



Definition: A parameter update is *KLV-optimal* if $\Delta < 0$.

The First Result: KLV-Optimality

Proposition

LET $s(y_t, \tilde{f}_t; \boldsymbol{\theta}) = S(\tilde{f}_t) \cdot \partial \log \tilde{p}(y_t | \tilde{f}_t; \boldsymbol{\theta}) / \partial f$ *with*

- $S(\tilde{f}) > 0 \; \forall \, \tilde{f} \in \tilde{\mathcal{F}}$ *and*
- $\partial \log \tilde{p}(y_t | \tilde{f}_t; \boldsymbol{\theta}) / \partial f$ *is continuously differentiable.*

THEN the GAS update is locally KLV-optimal uniformly in p, $f_t \in \mathcal{F}$, *and* $\tilde{f}_t \in \tilde{\mathcal{F}}$.

Local optimality: $\tilde{f}_{t+1}$ is in the neighborhood of $\tilde{f}_t$

In paper: non-local results, expected and realized optimality

Definition: A parameter update $\tilde{f}_{t+1} = \phi(y_t, \tilde{f}_t; \boldsymbol{\theta})$ is said to be *score-equivalent* if

$$\text{sign}(\nabla\phi(f, y; \boldsymbol{\theta})) \; = \; \text{sign}(s(f, y; \boldsymbol{\theta})) \quad \text{for almost every } (y, f)$$

PROPOSITION

LET the assumptions above hold.

THEN an update function ϕ is locally KLV-optimal if and only if it is score-equivalent.

Application: Volatility Model

Data Generating Process:

$$y_t = \sqrt{f_t} u_t, \qquad u_t \sim \tau(\lambda),$$
$$\log f_t = a + b \log f_{t-1} + u_t, \qquad u_t \sim \text{NID}(0, \sigma_u^2),$$

Parameter Values: $a = 0$, $b = 0.98$, $\sigma_u = 0.065$, $\lambda \in [2, 8]$.

Compared Models: GARCH, t-GARCH and t-GAS.

$$y_t = \sqrt{f_t} u_t$$

(GARCH) $\quad f_t = \omega + \alpha y_{t-1}^2 + \beta f_{t-1} \quad , \quad u_t \sim N(0, \sigma^2)$

(t-GARCH) $\quad f_t = \omega + \alpha y_{t-1}^2 + \beta f_{t-1} \quad , \quad u_t \sim \tau(\nu)$

(t-GAS) $\quad f_t = \omega + \alpha s(y_{t-1}, f_{t-1}) + \beta f_{t-1} \quad , \quad u_t \sim \tau(\nu)$

Data Generating Process:

$$y_t = \sqrt{f_t} u_t, \qquad u_t \sim \tau(\lambda),$$
$$\log f_t = a + b \log f_{t-1} + u_t, \qquad u_t \sim \text{NID}(0, \sigma_u^2),$$

Parameter Values: $a = 0$, $b = 0.98$, $\sigma_u = 0.065$, $\lambda \in [2, 8]$.

Note: Comparison at pseudo-true parameters $\boldsymbol{\theta}_0^* = \arg\min KL$

Sample size T: Large enough for ML estimator to converge to pseudo-true parameter (at 3rd decimal place).

Asymptotic Theory: Convergence of ML estimator to pseudo-true parameter in misspecified GAS is ensured!

KLV Optimality Regions

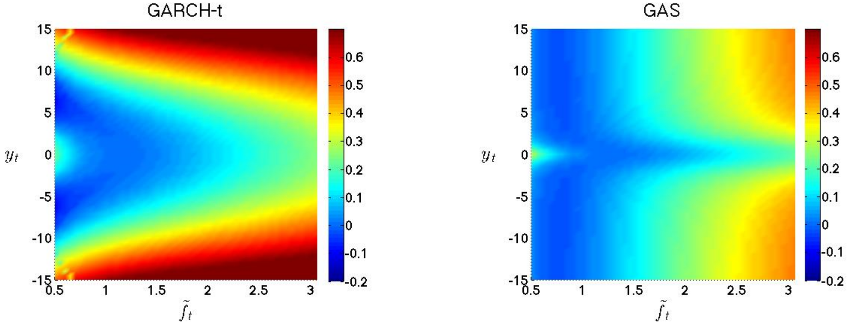


Figure : KLV regions for a conditional density of y_t given f_t used in these pictures is the standard Student's t with $\lambda = 3$ degrees of freedom. The regions are plotted for a given true $f_t \approx 1.2$.

Relative KL and Relative RMSE

Relative KL Divergence:

$$\text{RKL}\Big(\text{GAS}\,,\;\text{Amod}\Big) = 1 - \frac{\text{KL}\Big(\text{DGP}\,,\;\text{GAS}\Big)}{\text{KL}\Big(\text{DGP}\,,\;\text{Amod}\Big)}$$

Relative RMSE:

$$\text{RRMSE}\Big(\text{GAS}\,,\;\text{Amod}\Big) = 1 - \frac{\text{RMSE}\Big(\text{DGP}\,,\;\text{GAS}\Big)}{\text{RMSE}\Big(\text{DGP}\,,\;\text{Amod}\Big)}$$

Note: GAS is better when RKL > 0, RRMSE > 0

Note: GAS is infinitely better when RKL $\to 1$, RRMSE $\to 1$

Relative KL Divergence and Relative RMSE

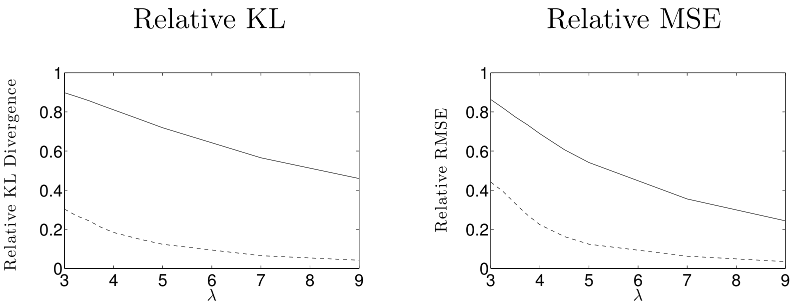


Figure : **Left:** Relative KL divergences and RMSE for t-GAS relative to GARCH (solid curve) and to t-GARCH (dashed curve). $T = 35,000$.

Note: GAS is better when RKL > 0, RRMSE > 0

Note: GAS is infinitely better when RKL $\to 1$, RRMSE $\to 1$

Summary: Optimality

1. The local GAS update is optimal in KL variation under very mild smoothness conditions
2. The optimality in KL variation results in considerably smaller KL divergence between the model and the DGP
3. Non-local results and other extensions are also available in the paper.

Stochastic Properties of GAS models

Blasques, Koopman and Lucas (2012)
"Stationarity and Ergodicity of Univariate Generalized Autoregressive Score Processes"

Blasques, Koopman and Lucas (2014)
"Maximum Likelihood Estimation for GAS Models: Feedback Effects and Contraction Conditions"

www.gasmodel.com

Stochastic properties of simple models:

What you know for sure...

Model Linear AR(1): $x_t = \alpha + \beta x_{t-1} + u_t$ with iid $\{u_t\}$

Then $\{x_t\}$ is strictly stationary and ergodic if $|\beta| < 1$

Then $\{x_t\}$ is weakly stationary if $|\beta| < 1$ and $\mathbb{E}|u_t|^2 < \infty$.

What you might not know...

This is only a special case of a much more general theory!

Lyapunov (1882), The General Problem of Stability of Motion

Nonlinear stochastic AR models of great complexity!

The more general theory:

Model: Non-linear AR(1): $x_t = \phi(x_{t-1}, u_t)$ with iid $\{u_t\}$

IMPORTANT: $\{x_t\}$ is strictly stationary and ergodic **if**

1. $\phi(x, u_t)$ has a log moment for some x:

$$\mathbb{E} \log^+ \phi(x, u_t) < \infty$$

2. $\phi(x, u_t)$ is contracting on average:

$$\mathbb{E} \log \sup_x \left| \frac{\partial \phi(x, u_t)}{\partial x} \right| < 0 \quad \Leftarrow \quad \mathbb{E} \sup_x \left| \frac{\partial \phi(x, u_t)}{\partial x} \right| < 1$$

Linear AR: $x_t = \alpha + \beta x_{t-1} + u_t$ we have:

$$\mathbb{E} \sup_x \left| \frac{\partial \phi(x, u_t)}{\partial x} \right| = \mathbb{E} \sup_x |\beta| < 1 = |\beta| < 1$$

Quadratic AR: $x_t = \alpha + \beta x_{t-1}^2 + u_t$ we have:

$$\mathbb{E} \sup_x \left| \frac{\partial \phi(x, u_t)}{\partial x} \right| = \mathbb{E} \sup_x |2\beta x| = \infty \qquad \Rightarrow \text{unstable unless } \beta = 0.$$

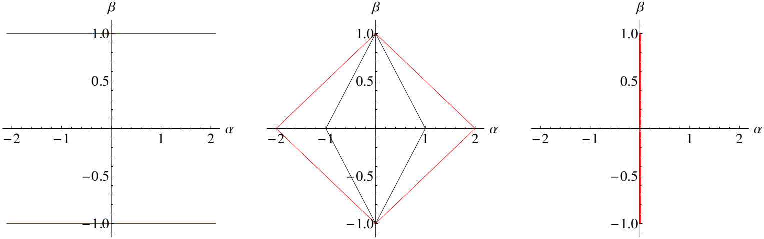
Random-Coefficient AR: $x_t = \alpha + \beta_t x_{t-1} + u_t$ we have:

$$\mathbb{E} \sup_x \left| \frac{\partial \phi(x, u_t)}{\partial x} \right| = \mathbb{E} \sup_x |\beta_t| = \mathbb{E}|\beta_t| < 1$$

Stationarity and Ergodicity

Model: $y_t = g(f_t, u_t) \quad , \quad f_{t+1} = \omega + \alpha s(f_t, y_t) + \beta f_t$

Stationarity and Ergodicity of filter $\quad \mathrm{E} \sup_{f^*} \left| \dfrac{\partial s(f^*, y_t)}{\partial f} \right| < \dfrac{1 - |\beta|}{|\alpha|}$

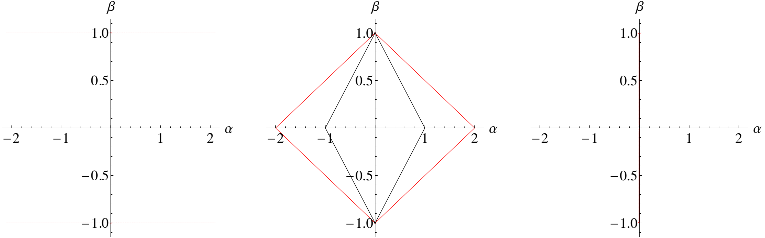


(a) Maximal (b) Non-Degenerate (c) Degenerate

Model: $y_t = g(f_t, u_t) \quad , \quad f_{t+1} = \omega + \alpha s(f_t, g(f_t, u_t)) + \beta f_t$

Stationarity and Ergodicity of filter

$$\mathrm{E} \sup_{f^*} \left| \frac{\partial s(f^*, g(f^*, u_t))}{\partial f} \right| < \frac{1 - |\beta|}{|\alpha|}$$



(d) Maximal (e) Non-Degenerate (f) Degenerate

Reformulation of GAS Model

$$
\begin{aligned}
y_t =& g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda) \\
f_{t+1} =& \omega + \alpha s(y_t, f_t) + \beta f_t.
\end{aligned}
$$

Reformulation of GAS Model

$$\begin{aligned} y_t =& g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda) \\ f_{t+1} =& \omega + \alpha s(y_t, f_t) + \beta f_t. \end{aligned}$$

- IMPORTANT:

  The dynamic properties of the filtered f_t depend on $\{y_t\}$

  $$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s(y_t, f_t)$$

  The dynamic properties of the true f_t depend on $\{u_t\}$

  $$f_{t+1} = \omega + \beta f_t + \alpha s\Big(g(f_t, u_t)\,,\, f_t\Big)$$

Reformulation of GAS Model

$$y_t = g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda)$$
$$f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t.$$

- IMPORTANT:

  The dynamic properties of the filtered f_t depend on $\{y_t\}$

  $$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s(y_t, f_t)$$

  The dynamic properties of the true f_t depend on $\{u_t\}$

  $$f_{t+1} = \omega + \beta f_t + \alpha s\Big(g(f_t, u_t)\,,\ f_t\Big)$$

- IMPORTANT: These two sequences are very different!

Reformulation of GAS Model: GARCH example

$$\begin{aligned} y_t =& g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda) \\ f_{t+1} =& \omega + \alpha s(y_t, f_t) + \beta f_t . \end{aligned}$$

- IMPORTANT:

  The dynamic properties of the filtered f_t depend on $\{y_t\}$

  $$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s(y_t, f_t)$$

  The dynamic properties of the true f_t depend on $\{u_t\}$

  $$f_{t+1} = \omega + \beta f_t + \alpha s\Big(g(f_t, u_t)\, ,\, f_t\Big)$$

- IMPORTANT: These two sequences are very different!

Reformulation of GAS Model: GARCH example

$$y_t = g(f_t, u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda)$$
$$f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t.$$

- IMPORTANT:

  The dynamic properties of the filtered f_t depend on $\{y_t\}$

  $$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha \left(y_t^2 - \tilde{f}_t \right)$$

  The dynamic properties of the true f_t depend on $\{u_t\}$

  $$f_{t+1} = \omega + \beta f_t + \alpha \left(u_t^2 - 1 \right) \cdot f_t$$

- IMPORTANT: These two sequences are very different!

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t; \lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t;\lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of filter: (difficult!)

$$\mathbb{E} \log \sup_f \left| \beta + \alpha \partial s(f_t, y_t; \lambda)/\partial f_t \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha(1 + 3\lambda^{-1})\Big(\frac{(1+\lambda)(y_t^2/(\lambda f_t))^2}{(1 + y_t^2/(\lambda f_t))^2} - 1 \Big) \right| < 0$$

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t; \lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of filter: (difficult!)

$$\mathbb{E} \log \sup_f \left| \beta + \alpha \partial s(f_t, y_t; \lambda)/\partial f_t \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha(1 + 3\lambda^{-1}) \Big(\frac{(1+\lambda)(y_t^2/(\lambda f_t))^2}{(1 + y_t^2/(\lambda f_t))^2} - 1 \Big) \right| < 0$$

Note: Supremum obtained as $f \downarrow 0 = \beta + (\lambda + 3) \cdot \alpha$

Note: (α, β)-SE regions become smaller for larger λ

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t; \lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of filter: (difficult!)

$$\mathbb{E} \log \sup_f \left| \beta + \alpha \partial s(f_t, y_t; \lambda) / \partial f_t \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha (1 + 3\lambda^{-1}) \Big(\frac{(1+\lambda)(y_t^2/(\lambda f_t))^2}{(1 + y_t^2/(\lambda f_t))^2} - 1 \Big) \right| < 0$$

Note: Supremum obtained as $f \downarrow 0 = \beta + (\lambda + 3) \cdot \alpha$

Note: (α, β)-SE regions become smaller for larger λ

Note: Discontinuity at infinity towards the normal

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t; \lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of DGP: (easy!)

$$\mathbb{E} \sup_f \left| \beta + \alpha \partial s_u(f, u_t; \lambda) / \partial f \right| < 1$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1 + u_t^2/(\lambda))^2} - 1 \right) \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1 + u_t^2/\lambda)^2} - 1 \right) \right| < 0$$

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t; \lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of DGP: (easy!)

$$\mathbb{E} \sup_f \left| \beta + \alpha \partial s_u(f, u_t; \lambda)/\partial f \right| < 1$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1 + u_t^2/(\lambda))^2} - 1 \right) \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1 + u_t^2/\lambda)^2} - 1 \right) \right| < 0$$

Note: Contraction does not depend on f_t

Example: t-GAS Volatility model

$$y_t = g(f_t, u_t) = \sqrt{f_t} \cdot u_t, \quad \text{with } p_u(u_t;\lambda) = t(\lambda)$$

Creal et al. (2011, 2013)

Contraction of DGP: (easy!)

$$\mathbb{E} \sup_f \left| \beta + \alpha \partial s_u(f, u_t; \lambda)/\partial f \right| < 1$$

$$\Leftrightarrow \mathbb{E} \log \sup_f \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1+u_t^2/(\lambda))^2} - 1 \right) \right| < 0$$

$$\Leftrightarrow \mathbb{E} \log \left| \beta + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1+\lambda)(u_t^2/\lambda)^2}{(1+u_t^2/\lambda)^2} - 1 \right) \right| < 0$$

Note: Contraction does not depend on f_t

Note: Easy to find the combinations of α and β that satisfy this condition

Note: When $\phi(f_t, y_t)$ is contracting we say that the filter is invertible. This is because

$$f_{t+1} = \phi(f_t, y_t) = \xi\big(y_t, y_{t-1}, y_{t-2}, ...\big)$$

Note: When $\phi(f_t, u_t)$ is contracting then the true $\{f_t\}$ is SE

Note: If the DGP of $\{f_t\}$ is SE and g is continuous, then the data $\{y_t\}$ generated by the model $y_t = g(f_t, u_t)$ is also SE.

Note: Invertibility of the model is always needed for estimation because the filtered $\{f_t\}$ enters the likelihood function

Note: Contraction of $\phi(f_t, u_t)$ is needed under correct specification to show that the data $\{y_t\}$ is SE.

Asymptotic Properties of MLE for GAS

Blasques, Koopman and Lucas (2014)

"Maximum Likelihood Estimation for Generalized Autoregressive Score Models"

www.gasmodel.com

Estimation in Practice

IMPORTANT: GAS models are very easy to estimate!

Note: The log likelihood function is just the average of the log conditional densities:

$$\ell(\boldsymbol{\theta}; y_1, ..., y_T) := \frac{1}{T}\sum_{t=1}^{T} \log p(y_t | f_t(\boldsymbol{\theta}), \lambda)$$

Step-by-step:

1. for a starting value $(\omega, \alpha, \beta, \lambda) = \boldsymbol{\theta} \in \Theta$
2. initialize the filter at some $f_1(\boldsymbol{\theta})$
3. use the data $y_1, ..., y_T$ to calculate $f_2(\boldsymbol{\theta}), ..., f_T(\boldsymbol{\theta})$
4. calculate the log likelihood $\ell(\boldsymbol{\theta}; y_1, ..., y_T)$
5. iterate over $\boldsymbol{\theta}$ to find the maximum of $\ell(\boldsymbol{\theta}; y_1, ..., y_T)$.

Estimation in Practice: Gaussian local level

Example: $y_t = f_t + u_t$ with $u_t \sim N(0,1)$

$$\ell(\boldsymbol{\theta}; y_1, ..., y_T) = \frac{1}{T}\sum_{t=1}^{T} \log\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y_t - f_t)^2}{2}\right)\right]$$
$$\propto \frac{1}{T}\sum_{t=1}^{T}(y_t - f_t(\boldsymbol{\theta}))^2$$

Step-by-step:

1. for a starting value $(\omega, \alpha, \beta) = \boldsymbol{\theta} \in \Theta$
2. initialize the filter at some $f_1(\boldsymbol{\theta})$
3. use the data $y_1, ..., y_T$ to calculate $f_2(\boldsymbol{\theta}), ..., f_T(\boldsymbol{\theta})$
4. calculate the log likelihood $\ell(\boldsymbol{\theta}; y_1, ..., y_T)$
5. iterate over $\boldsymbol{\theta}$ to find the maximum of $\ell(\boldsymbol{\theta}; y_1, ..., y_T)$.

Estimation in Practice: Gaussian local level

The theory of ML estimation is a bit more complicated...

... but still a lot of fun!

Asymptotic Properties of Estimators

What you surely know!

In simple models we can...

1. derive an expression for the estimator
2. show that LLNs and CLTs and apply
3. obtain CAN results for correctly specified models

Linear AR(1): $x_t = \beta x_{t-1} + u_t$

Estimator of β: $\hat{\beta}_T = \dfrac{\sum_{t=1}^{T} x_t x_{t-1}}{\sum_{t=1}^{T} x_{t-1}^2} = \dfrac{\frac{1}{T}\sum_{t=1}^{T} x_t x_{t-1}}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2}$

- If $|\beta| < 1$ then LLNs and CLTs apply and $\hat{\beta}_T$ is CAN!

What you might not know...

1. We don't need the estimator to be tractable!
2. We don't need the model to be well specified!

A very brief history of estimation:

LLN: Cardano (1500's) stated LLN without proof. First proof in Jacob Bernoulli's (1713) Ars Conjectandi. Simpler proof by Pafnuty Chebyshev (1874) using (unproved) inequality. Markov's (1884) thesis contains proof.

OLS: Discovered by Gauss (1821) and republished by Markov (1900).

General: Doob (1934), Cramer (1946), **Wald (1949)**, Le Cam (1949), **Jennrich (1969)**, Malinvaud (1970).

Modern CAN proofs:

Analytically intractable estimators:

using theory of Wald (1949) and Jennrich (1969)!

Mis-specified models:

CAN w.r.t. pseudo-true parameter $\boldsymbol{\theta}_0^*$

- $\boldsymbol{\theta}_0^*$ minimizes KL div in ML and weighted L2 norm in LS
- $\boldsymbol{\theta}_0^*$ of other estimators minimizes other distances
- we can compare $\boldsymbol{\theta}_0^*$ of different estimators
- if model is well-specified then $\boldsymbol{\theta}_0^* = \boldsymbol{\theta}_0$

The Univariate GAS Model

Real-valued stochastic sequence $\{y_t\}$ with **tv** $p(y_t|f_t)$,

$$y_t = g(h(f_t), u_t) \quad , \quad \{u_t\} \text{ iid } \sim p_u(\lambda)$$
$$f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t.$$

- f_t is a real-valued **tv** parameter,
- $h : \mathbb{R} \to \mathbb{R}$ is parameterization function,
- ω, α, β and λ are **time-invariant** scalar parameters,
- $s(y_t, f_t) = \nabla(y_t, f_t) \cdot S(f_t)$ is a scaled score:

$$\nabla(y_t, f_t) := \partial \ln p(y_t|h(f_t))/\partial f_t$$

- $s(y_t, f_t)$ can be **linear, nonlinear or invariant in** f_t.

Why use the score?

- Likelihood optimization: steepest ascent update
- Optimal in KL divergence

Why use a scaling function for the score?

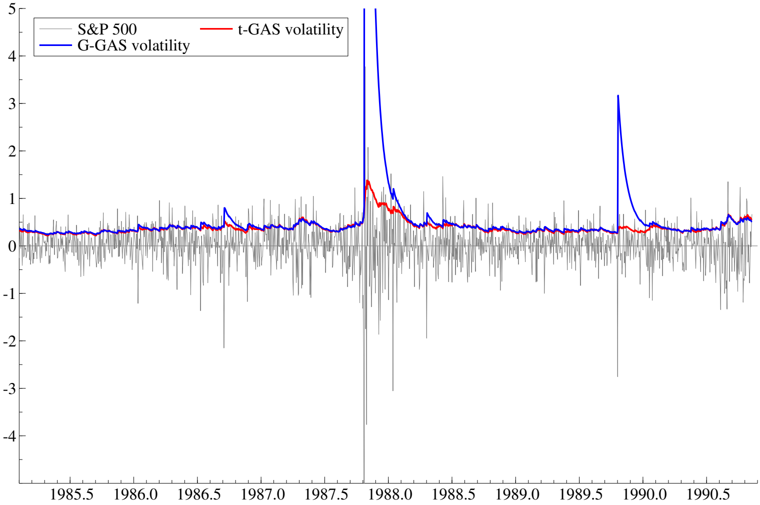
- Control step size
- Inverse information scaling: Newton-Rapshon update

How general is this?

e.g. GARCH, EGARCH, MEM, ACD, ACM and more!

GARCH:
$$y_t = g\big(h(f_t), u_t\big) = \sqrt{f_t}u_t$$
$$f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t = \omega + a y_t^2 + b f_t$$

Application: Dynamic Volatility Model



ML Estimation of Time-Invariant Parameters

Existing results on ML estimation of $\boldsymbol{\theta} := (\omega, \alpha, \beta, \lambda)$

- **Special cases:**
  - GARCH, EGARCH, MEM, ACD, ACM
- **Some general classes:**

  e.g. Straumman-Mikosch (2006) conditional volatility models:

  $$y_t = g\big(h(f_t), u_t\big) = \sqrt{f_t} u_t \quad \text{and} \quad f_{t+1} = F(f_t, y_t; \boldsymbol{\theta}).$$
- **Score driven models:**
  - Harvey (2010, 2013):
    - Conditional volatility GAS with exponential link function
    - Linear dynamics + Correct specification + Local results

ML Estimation of Time-Invariant Parameters

$$y_t = g(h(f_t), u_t) \quad , \quad f_{t+1} = \omega + \alpha s(y_t, f_t) + \beta f_t$$

Objective: Find simple conditions on g, h, p_u, S and Θ that ensure consistency and asymptotic normality of MLE:

- allowing for nonlinear dynamics in $\{f_t\}$.
- allowing for local and global results
- allowing for correct and incorrect model specification

Problem: Generality comes at a cost... a given condition might fit some choice of g, h, p_u and S, but not others.

Solution: Balance between generality and easy applicability!

Consistency Conditions:

Consistency Conditions:

IF Uniform convergence of likelihood function:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L_\infty(\boldsymbol{\theta})| \stackrel{a.s.}{\to} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

ML Consistency

Consistency Conditions:

IF Uniform convergence of likelihood function:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta\in\Theta\ :\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \ \ \forall\ u > 0$$

Consistency Conditions:

IF Uniform convergence of likelihood function:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta\in\Theta\ :\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF Uniform convergence of likelihood function:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta\in\Theta\ :\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta \ : \ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \ \ \forall \ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta\in\Theta\ :\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0)\ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)| < \infty\ \forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0)\ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)| < \infty\ \forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)$ is smooth in $\boldsymbol{\theta}$

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0)\ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)| < \infty\ \forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)$ is smooth in $\boldsymbol{\theta}$

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0)\ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)\}$ is **SE** $\forall\,(\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)| < \infty \; \forall\,(\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $L_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T)$ is smooth in $\boldsymbol{\theta}$

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta \,:\, \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \quad \forall\, u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta \;:\; \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \;\; \forall \; u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \quad \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0)\ \ \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta\ :\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \quad \forall\ u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\, (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Identifiable uniqueness of $\boldsymbol{\theta}_0$:

$$\sup_{\theta \in \Theta \;:\; \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > u} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \;\; \forall\, u > 0$$

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Model parameter identification (local or global)

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\, (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Model parameter identification (local or global)

IF Correct specification <u>or</u> unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Model parameter identification (local or global)

IF Correct specification or unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Correct specification or unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

ML Consistency

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Invertibility conditions on g, h, S and p_u

IF Correct specification <u>or</u> unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

IF $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Invertibility conditions on g, h, S and p_u

IF Restrictions on Θ and $\mathbb{V}\text{ar}(s_t(f_t)) > 0$

IF Correct specification or unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Consistency Conditions:

- **IF** $\{(y_t, f_t(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\, (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.
- **IF** $\mathbb{E}|y_t|^m < \infty$ and $\mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty$

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ

IF Invertibility conditions on g, h, S and p_u

IF Restrictions on Θ and $\mathbb{V}\text{ar}(s_t(f_t)) > 0$

IF Correct specification <u>or</u> unique pseudo-true parameter

THEN $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.

Asymptotic Normality Conditions:

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T} L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

ML Asymptotic Normality

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T}L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Uniform convergence of second derivative:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L''_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L''_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T}L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Uniform convergence of second derivative:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L''_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L''_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

ML Asymptotic Normality

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T}L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Uniform convergence of second derivative:

$$\sup_{\boldsymbol{\theta}\in\Theta} |L''_T(\boldsymbol{\theta}, f_1|\mathbf{y}_T) - L''_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T} L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Uniform convergence of second derivative:

$$\sup_{\boldsymbol{\theta} \in \Theta} |L''_T(\boldsymbol{\theta}, f_1 | \mathbf{y}_T) - L''_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF CLT for the score:

$$\sqrt{T} L'_T(\boldsymbol{\theta}_0, f_1) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)) \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF Uniform convergence of second derivative:

$$\sup_{\boldsymbol{\theta} \in \Theta} |L''_T(\boldsymbol{\theta}, f_1 | \mathbf{y}_T) - L''_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall\ f_1 \in \mathcal{F} \ \text{ as } T \to \infty.$$

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \stackrel{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \stackrel{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\left\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\right\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ.

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\left\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\right\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ.

IF $\mathbb{I}(\boldsymbol{\theta}_0)$ is invertible.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

IF $\{(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1))\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.

IF $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ.

IF Previous identification conditions on Θ and g, h, S and p_u.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Asymptotic Normality Conditions:

- **IF** $\left\{\left(y_t,\ f_t(\boldsymbol{\theta}, f_1),\ f_t'(\boldsymbol{\theta}, f_1),\ f_t''(\boldsymbol{\theta}, f_1)\right)\right\}$ is **SE** $\forall\ (\boldsymbol{\theta}, f_1) \in \Theta \times \mathcal{F}$.
- **IF** $\mathbb{E}|y_t|^m < \infty,\ \mathbb{E}|f_t(\boldsymbol{\theta}, f_1))|^m < \infty\ ,\ \mathbb{E}|f_t'(\boldsymbol{\theta}, f_1))|^m < \infty$ and $\mathbb{E}|f_t''(\boldsymbol{\theta}, f_1))|^m < \infty$.

IF g, h, S and p_u are smooth with bounded n^{th} derivative

IF Compact Θ.

IF Previous identification conditions on Θ and g, h, S and p_u.

IF $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ as $T \to \infty$.

THEN $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} N(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1})$ as $T \to \infty$.

Stationarity, Ergodicity and Moments

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: Contracting update $f_{t+1} = \omega + \alpha s(y_t, f_t; \lambda) + \beta f_t$

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: Contracting update $f_{t+1} = \omega + \alpha s(y_t, f_t; \lambda) + \beta f_t$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Stationarity, Ergodicity and Moments

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: Contracting update $f_{t+1} = \omega + \alpha s(y_t, f_t; \lambda) + \beta f_t$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Stationarity, Ergodicity and Moments

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: $\mathrm{E}\sup_f \left|\alpha + \beta\partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

- IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

  IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Known DGP: (Correct Specification)

DGP: $y_t = g(h(f_t), u_t) \quad , \quad f_{t+1} = \omega + \alpha s(y_t, f_t; \lambda) + \beta f_t$

Stationarity, Ergodicity and Moments

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

- IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Known DGP: (Correct Specification)

DGP: $y_t = g(h(f_t), u_t)$, $f_{t+1} = \omega + \alpha s(y_t, f_t; \lambda) + \beta f_t$

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

- IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

  IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Known DGP: (Correct Specification)

DGP: $y_t = g(h(f_t), u_t) \quad , \quad f_{t+1} = \omega + \alpha s_u(u_t, f_t; \lambda) + \beta f_t$

Stationarity, Ergodicity and Moments

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

- IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Known DGP: (Correct Specification)

DGP: $y_t = g(h(f_t), u_t) \quad , \quad f_{t+1} = \omega + \alpha s_u(u_t, f_t; \lambda) + \beta f_t$

IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s_u(u_t, f; \lambda)/\partial f\right| < 1$

What we want: SE and Moments for $\{(y_t, f_t^d(\boldsymbol{\theta}, f_1))\}$

Unknown DGP: (Incorrect Specification)

- IF: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$ (unknown DGP, test data)

IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s(y_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{f_t^{(d)}(\boldsymbol{\theta}, f_1)\}$ is SE and $\mathbb{E}|f_t^{(d)}(\boldsymbol{\theta}, f_1)|^{n_f} < \infty$.

Known DGP: (Correct Specification)

DGP: $y_t = g(h(f_t), u_t) \quad , \quad f_{t+1} = \omega + \alpha s_u(u_t, f_t; \lambda) + \beta f_t$

IF: $\mathrm{E}\sup_f \left|\alpha + \beta \partial s_u(u_t, f; \lambda)/\partial f\right| < 1$

THEN: $\{y_t\}$ is SE and $\mathbb{E}|y_t|^{n_y} < \infty$.

Example: Time-Varying Volatility with GAS Dynamics

Let $y_t = h(f_t)u_t \ , \quad u_t \sim p_u(\lambda) \ , \quad f_t = \omega + \alpha s_t(f_t;\lambda) + \beta f_{t-1}.$

Then $s_t(f_t;\lambda) = -S(f_t) \cdot \nabla h(f_t) \cdot \Big(\nabla p_u(u_t;\lambda) u_t + 1 \Big)$

where $\nabla h(f_t) = \frac{\partial \log h(f_t)}{\partial f_t}$ and $\nabla p_u(u_t) = \frac{\partial \log p_u(u_t)}{\partial u_t}$

Example: Time-Varying Volatility with GAS Dynamics

Let $y_t = h(f_t)u_t \ , \quad u_t \sim p_u(\lambda) \ , \quad f_t = \omega + \alpha s_t(f_t;\lambda) + \beta f_{t-1}.$

Then $s_t(f_t;\lambda) = -S(f_t) \cdot \nabla h(f_t) \cdot \Big(\nabla p_u(u_t;\lambda) u_t + 1 \Big)$

where $\nabla h(f_t) = \frac{\partial \log h(f_t)}{\partial f_t}$ and $\nabla p_u(u_t) = \frac{\partial \log p_u(u_t)}{\partial u_t}$

ex.1 If: $S(f_t) = \mathcal{I}_t^{-1/2}(f_t) = \nabla h(f_t)^{-1}$

Then: CAN conditions are simple

Example: Time-Varying Volatility with GAS Dynamics

Let $y_t = h(f_t)u_t \ , \quad u_t \sim p_u(\lambda) \ , \quad f_t = \omega + \alpha s_t(f_t;\lambda) + \beta f_{t-1}.$

Then $s_t(f_t;\lambda) = -S(f_t) \cdot \nabla h(f_t) \cdot \Big(\nabla p_u(u_t;\lambda)u_t + 1 \Big)$

where $\nabla h(f_t) = \frac{\partial \log h(f_t)}{\partial f_t}$ and $\nabla p_u(u_t) = \frac{\partial \log p_u(u_t)}{\partial u_t}$

ex.1 If: $S(f_t) = \mathcal{I}_t^{-1/2}(f_t) = \nabla h(f_t)^{-1}$

Then: CAN conditions are simple

ex.2 If: $h(f_t) = \exp(f_t)$ and $S(f_t) = 1$

Then: CAN conditions are simple

Example: Time-Varying Volatility with GAS Dynamics

Let $y_t = h(f_t)u_t \ , \quad u_t \sim p_u(\lambda) \ , \quad f_t = \omega + \alpha s_t(f_t;\lambda) + \beta f_{t-1}.$

Then $s_t(f_t;\lambda) = -S(f_t) \cdot \nabla h(f_t) \cdot \Big(\nabla p_u(u_t;\lambda)u_t + 1 \Big)$

where $\nabla h(f_t) = \frac{\partial \log h(f_t)}{\partial f_t}$ and $\nabla p_u(u_t) = \frac{\partial \log p_u(u_t)}{\partial u_t}$

ex.1 If: $S(f_t) = \mathcal{I}_t^{-1/2}(f_t) = \nabla h(f_t)^{-1}$

Then: CAN conditions are simple

ex.2 If: $h(f_t) = \exp(f_t)$ and $S(f_t) = 1$

Then: CAN conditions are simple

ex.3 If: Nonlinear h and general $S(f_t)$

Then: CAN conditions are more complicated

CAN: Gaussian GAS

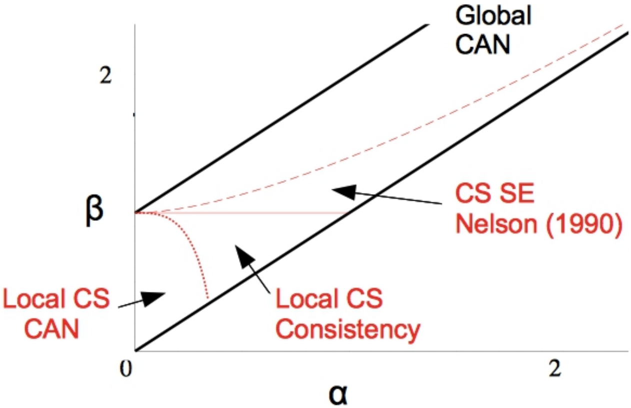


Figure : CAN regions for Gaussian Volatility GAS

$$y_t = \sqrt{f_t} u_t \quad , \quad u_t \sim N(0, \sigma^2) \quad , \quad \text{Inverse Info Scaling}$$

Diagonal restriction $\beta > 1 + \alpha$ ensures $f_t > 0$.

Example: Time-Varying Volatility with GAS Dynamics

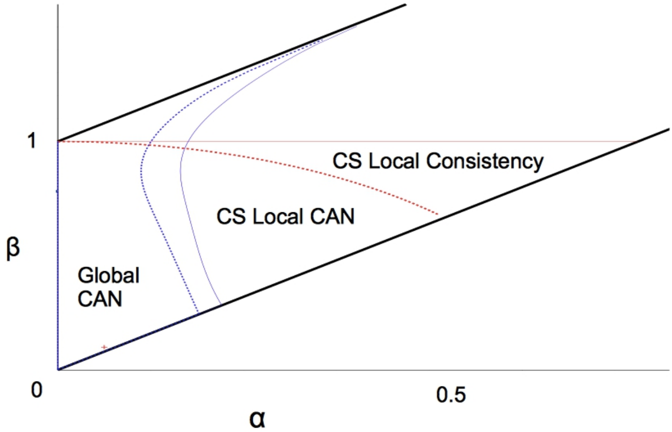


Figure : CAN regions for t-Volatility GAS in Creal at al. (2011)

$$y_t = \sqrt{f_t} u_t \quad , \quad u_t \sim \tau(\lambda) \quad , \quad \text{Inverse Info Scaling}$$

Diagonal restriction: $\beta > (1 + 3\lambda^{-1})\alpha$ ensures $f_t \geq 0$.

Summary: Consistency and Asymptotic Normality

1. Estimation is really simple because the likelihood function is immediately available!

2. We showed that the MLE is CAN under appropriate regularity conditions

3. We have seen that the size and shape of regions on which we can establish CAN depend essentially on:
    1. ensuring strict stationarity and ergodicity
    2. ensuring invertibility
    3. ensuring enough moments exist
    4. ensuring identification

Summary: GAS Theory

1. Introduced GAS model and discussed how intuitive it is!
2. We have seen how the GAS update adapts so well to the distribution of the data
3. Showed that the GAS update is optimal in KL variation
4. Showed that optimality in KL variation may result in substantially better model fit
5. Derived conditions for the invertibility and strict stationarity and ergodicity of the GAS filter
6. Derived conditions for strict stationarity and ergodicity of the GAS as a DGP
7. Showed that MLE is consistent and asymptotically normal under appropriate conditions

Today's Schedule

09:00 - 10:15	Lecture
10:15 - 10:30	Coffee Break
10:30 - 11:30	Lecture
11:30 - 12:00	Informal Q&A
12:00 - 13:00	Lunch
13:00 - 14:15	Lecture
14:15 - 14:30	Coffee Break
14:30 - 15:30	Lecture
15:30 - 16:00	Informal Q&A