Another Look at Estimation for MA(1) Processes With a Unit Root

F. Jay Breidt

Richard A. Davis

Nan-Jung Hsu

Murray Rosenblatt

Colorado State University
Columbia University
National Tsing-Hua University
U. of California, San Diego

(http://www.stat.columbia.edu/~rdavis/lectures)

Program

Model: $Y_t = Z_t - \theta Z_{t-1}$, $\{Z_t\} \sim \text{IID}(0, \sigma^2)$

- Introduction
  - Asymptotics in nonstandard inference settings
  - The MA(1) unit root problem
  - Why study non-invertible MA(1)'s?
    - over-differencing
    - random walk + noise
- Gaussian Likelihood Estimation
  - Identifiability
  - Limit results
  - Extensions
    - non-zero mean
    - heavy tails
- Laplace Likelihood/LAD estimation
  - Joint and exact likelihood
  - Limit results
  - Limit distribution/simulation comparisons
  - Pile-up probabilities
    - joint likelihood
    - exact likelihood

Asymptotics in Non-standard Inference Settings

Subtitle: When a Taylor series expansion fails.

- objective function is not differentiable (LAD)
- remainder is of same order as last included term (heavy-tailed noise)

Applications to:

- LAD estimation
  (Pollard `91; Davis and Dunsmjuir `95)
- M-estimation with infinite variance
  (Pollard `91; Davis, Knight and Liu `92; Davis `95)
- Unit root problems (AR + MA)
  (Davis and Dunsmuir `95; Davis, Chen and Dunsmuir `95)

Illustration of the Paradigm

LAD Estimation:

1. Median. Let $\{Z_t\}$ ~ IID with median 0 and pdf f such that f(0) > 0.

Objective function:

$$S_n(\theta) = \sum_{t=1}^{n} |Z_t - \theta| - \sum_{t=1}^{n} |Z_t|, \qquad \hat{\theta} = \arg\min S(\theta) = \text{median}$$

Reparameterize: $\theta = u/n^{1/2}$

$$T_n(u) = \sum_{t=1}^{n} |Z_t - u/n^{1/2}| - \sum_{t=1}^{n} |Z_t| \to_d T(u) := -uN + u^2 f(0) \quad \text{on C(R)}, \ N \sim \text{N}(0,1)$$

It follows that

$$\begin{aligned} \hat{u}_n &= \arg\min T_n(u) = n^{1/2}(\hat{\theta} - 0) \\ &\to_d \hat{u} = \arg\min(-uN + u^2 f(0)) \\ &= -N/(2f(0)) \sim N(0, \frac{1}{4f^2(0)}) \end{aligned}$$

Illustration of the Paradigm

2. AR(1) Let $\{X_t\}$ be the AR(1) process

$$X_t = \phi_0 X_{t-1} + Z_t,$$

where $\{Z_t\}$ ~ IID with median 0 and pdf f such that f(0) > 0.

Objective function:

$$S_n(\theta) = \sum_{t=2}^{n} |X_t - \phi X_{t-1}| - \sum_{t=1}^{n} |Z_t|, \qquad \hat{\phi} = \arg\min S(\phi) = \text{LAD estimate}$$

Reparameterize: $\phi = \phi_0 + u/n^{1/2}$

$$T_n(u) = \sum_{t=1}^{n} |Z_t - u/n^{1/2} X_{t-1}| - \sum_{t=1}^{n} |Z_t| \to_d T(u) := -u\nu N + u^2\nu^2 f(0), \;\; N \sim \mathrm{N}(0,1)$$

It follows that

$$\hat{u}_n = \arg\min T_n(u) = n^{1/2}(\hat{\phi} - \phi_0)$$

$$\to_d \hat{u} = \arg\min(-u\nu N + u^2\nu^2 f(0))$$

$$= -N/(2\nu f(0)) \sim N(0, \frac{1}{4\nu^2 f^2(0)}), \qquad \nu^2 = \mathrm{Var}(X_t).$$

MA(1) unit root problem

MA(1): (world's simplest time series model!)

$$Y_t = Z_t - \theta\, Z_{t-1}\,,\quad \{Z_t\} \sim \text{IID}\,(0, \sigma^2)$$

Properties:

- $|\theta| < 1 \quad \Rightarrow \quad Z_t = \sum_{j=0}^{\infty} \theta^{j} Y_{t-j}$ (invertible)
- $|\theta| > 1 \quad \Rightarrow \quad Z_t = -\sum_{j=1}^{\infty} \theta^{-j} Y_{t+j}$ (non-invertible)
- $|\theta| = 1 \quad \Rightarrow \quad Z_t \in \text{sp}\{Y_t, Y_{t-1}, \ldots,\}$ and $Z_t \in \text{sp}\{Y_{t+1}, Y_{t+2}, \ldots,\}$

  $\Rightarrow \quad P_{\text{sp}\{Y_s,\, s \neq 0\}} Y_0 = Y_0$ (perfect interpolation)

- $|\theta| < 1 \quad \Rightarrow \quad \hat{\theta}_{MLE}$ is $\text{AN}(\theta, (1-\theta^2)/n)$

MLE = maximum (Gaussian) likelihood, n = sample size

What if $\theta = 1$?

Why study MA(1) with a unit root?

a) differencing (to remove non-stationarity)

- linear trend model: $X_t = a + bt + Z_t$.

  $Y_t = X_t - X_{t\text{-}1} = b + Z_t - Z_{t\text{-}1}$ ~ MA(1) with $\theta = 1$.

- seasonal model: $X_t = s_t + Z_t$, s_t seasonal component w/ period 12.

  $Y_t = X_t - X_{t\text{-}12} = Z_t - Z_{t\text{-}12}$ ~ MA(12) with $\theta = 1$.

b) random walk + noise

$X_t = X_{t\text{-}1} + U_t$ (random walk signal)

$Y_t = X_t + V_t$ (random walk signal + noise)

Then

$$Y_t - Y_{t\text{-}1} = U_t + V_t - V_{t\text{-}1} \sim \text{MA}(1)$$

with θ=1 if and only if $\text{Var}(U_t) = 0$.

Identifiability and the Gaussian likelihood

Identifiability ($Y_t = Z_t - \theta Z_{t-1}$, $\{Z_t\} \sim$ IID $(0, \sigma^2)$)

- $|\theta| > 1 \quad \Rightarrow \quad Y_t = \varepsilon_t - \theta^{-1} \varepsilon_{t-1}$, where $\{\varepsilon_t\} \sim \mathrm{WN}(0,\theta^2\sigma^2)$.
- $\{\varepsilon_t\}$ is IID *if and only if* $\{Z_t\}$ is Gaussian (Breidt and Davis `91)
- $\{\varepsilon_t\}$ is a special case of an ***All-Pass Model*** (Breidt, Davis, Trindade `01, Andrews et al. `05a, `05b)

Gaussian Likelihood

$$L_G(\theta, \sigma^2) = L_G(1/\theta, \theta^2\sigma^2) \Rightarrow \theta \text{ is only identifiable for } |\theta| \leq 1.$$

Notes:

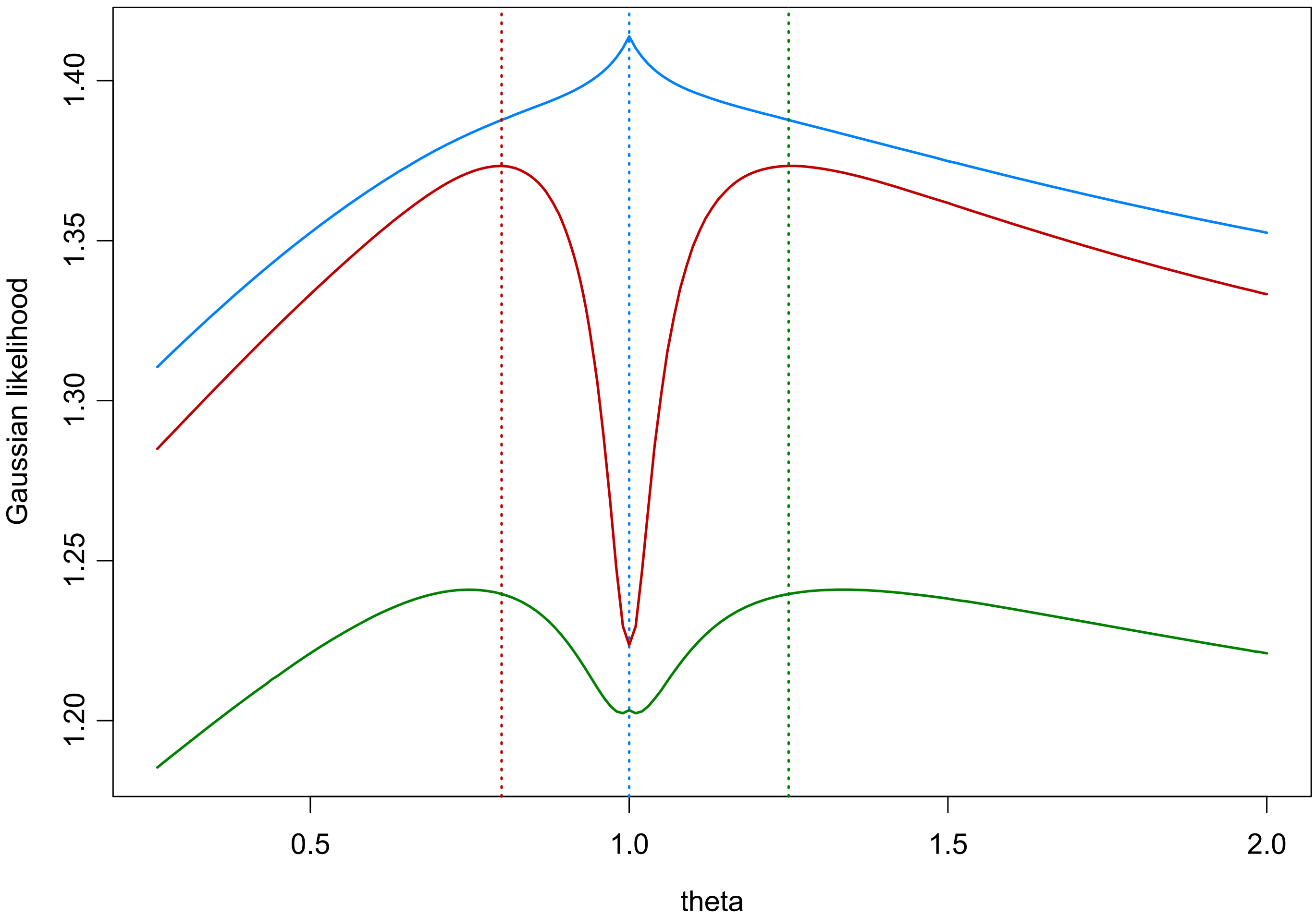
i) this implies $L_G(\theta) = L_G(1/\theta)$ for the profile likelihood and $\theta = 1$ is a critical point, $L'_G(1) = 0$.

ii) a *pile-up effect* ensues, i.e., $P(\hat{\theta} = 1) > 0$ even if $\theta < 1$.

Gaussian likelihood, non-Gaussian data

100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, $\{Z_t\} \sim$ IID, Laplace pdf

$\theta_0 = .8$ $\theta_0 = 1.0$ $\theta_0 = 1.25$



Gaussian MLE for near-unit roots

Idea: build parameter normalization into the likelihood function.

Model: $Y_t = Z_t - (1-\beta/n)\, Z_{t-1}$, $t = 1,\ldots,n$.

$\beta = n(1-\theta)$, $\theta = 1- \beta/n$, $\theta_0 = 1- \gamma/n$

Gaussian Likelihood:

$L_n(\beta) = l_n\,(1- \beta/n) - l_n\,(1)$, $l_n\,(\,)$ = profile log-like.

Theorem (Davis and Dunsmuir `96): Under $\theta_0 = 1-\gamma / n$,

$$L_n(\beta) \to_d Z_\gamma(\beta) \quad \text{on } C[0,\infty).$$

Results:

- $n(1-\hat{\theta}_{mle}) \to \hat{\beta}_{mle} = \text{argmax}\, Z_\gamma(\beta)$
- $n(1-\hat{\theta}_L) \to \hat{\beta}_L = \text{arg}\mathit{local}\text{max}\, Z_\gamma(\beta)$
- $P(\hat{\theta}_L = 1) \to P(\hat{\beta}_L = 0) = .6518 \quad \text{if } \gamma = 0.$

Extensions of MLE (Gaussian likelihood)

i) non-zero mean (Chen and Davis `00): same type of limit, except pile-up is more excessive.

$$P(\hat{\theta}_{mle} = 1) \to .955$$

This makes hypothesis testing easy!

Reject H_0: $\theta = 1$ if $\hat{\theta}_{mle} < 1$ (size of test is .045)

ii) heavy tails (Davis and Mikosch `98): $\{Z_t\}$ symmetric alpha stable (SαS). Then the max Gaussian likelihood estimator has the same normalizing rate, i.e.,

$$n(1 - \hat{\theta}_L) \to_d \hat{\beta}_L$$

$$P(\hat{\theta}_L = 1) \to P(\hat{\beta}_L = 0)$$

The pile-up decreases with increasing tail heaviness.

Laplace likelihood/LAD estimation

If noise distribution is non-Gaussian, the MA(1) parameter θ is identifiable for all real values.

Q1. For MLE (non-Gaussian) does one have $1/n$ or $1/n^{1/2}$ asymptotics?

Q2. Is there a *pile-up* effect?

Look at this problem with *non-Gaussian likelihood*

- Specifically, consider *Laplace likelihood / Least Absolute Deviations* for unit root only (not near-unit root)
- *Some results are preliminary only!*

Non-Gaussian likelihood – Joint and Exact

Model. $Y_t = Z_t - \theta Z_{t-1}$, $\{Z_t\} \sim$ IID with median 0 and $EZ^4 < \infty$. Initial variable.

$$Z^{init} = \begin{cases} Z_0, & \text{if } |\theta| \le 1, \\ Z_n - \sum_{t=1}^{n} Y_t, & \text{otherwise.} \end{cases}$$

Joint density: Let $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, then

$$f(\mathbf{y}_n, z^{init}) = f(z_0, z_1, \ldots, z_n)\left(1_{\{|\theta| \le 1\}} + |\theta|^{-n} 1_{\{|\theta| > 1\}}\right),$$

where the z_t are solved

forward by: $z_t = Y_t + \theta z_{t-1}$, $t = 1, \ldots, n$ for $|\theta| \le 1$ with $z_0 = z^{init}$

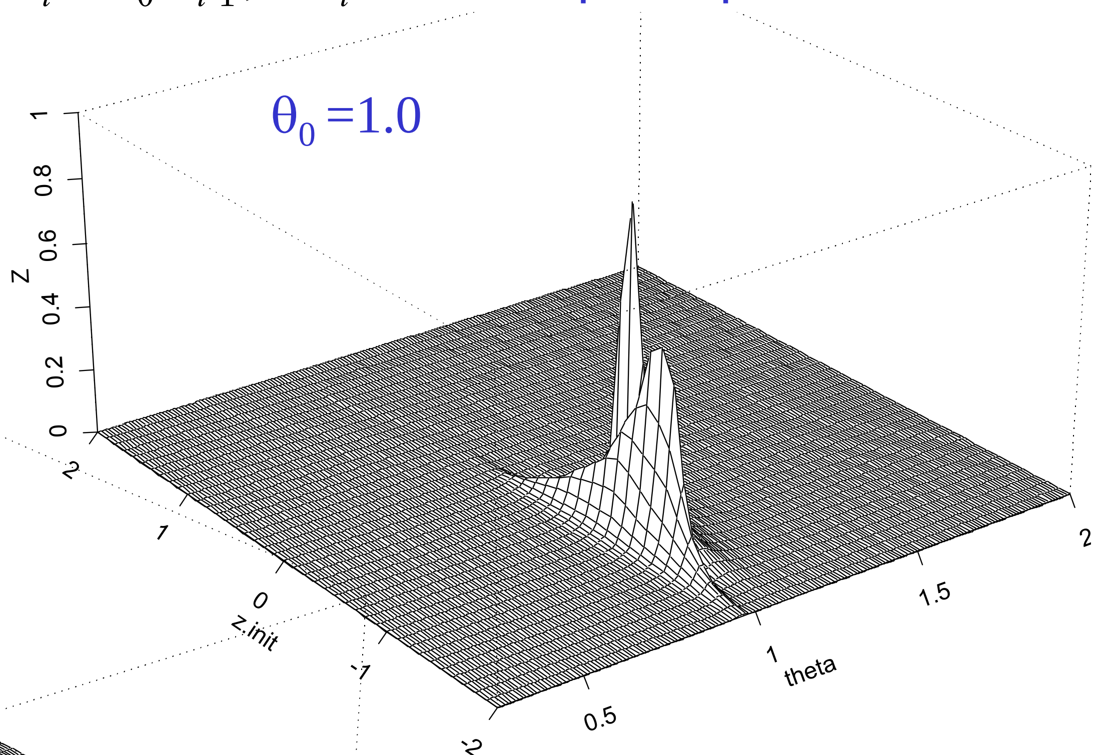
backward by: $z_{t-1} = \theta^{-1}(z_t - Y_t)$, $t = n, \ldots, 1$ for $|\theta| > 1$ with $z_n = z^{init} + Y_1 + \ldots + Y_n$

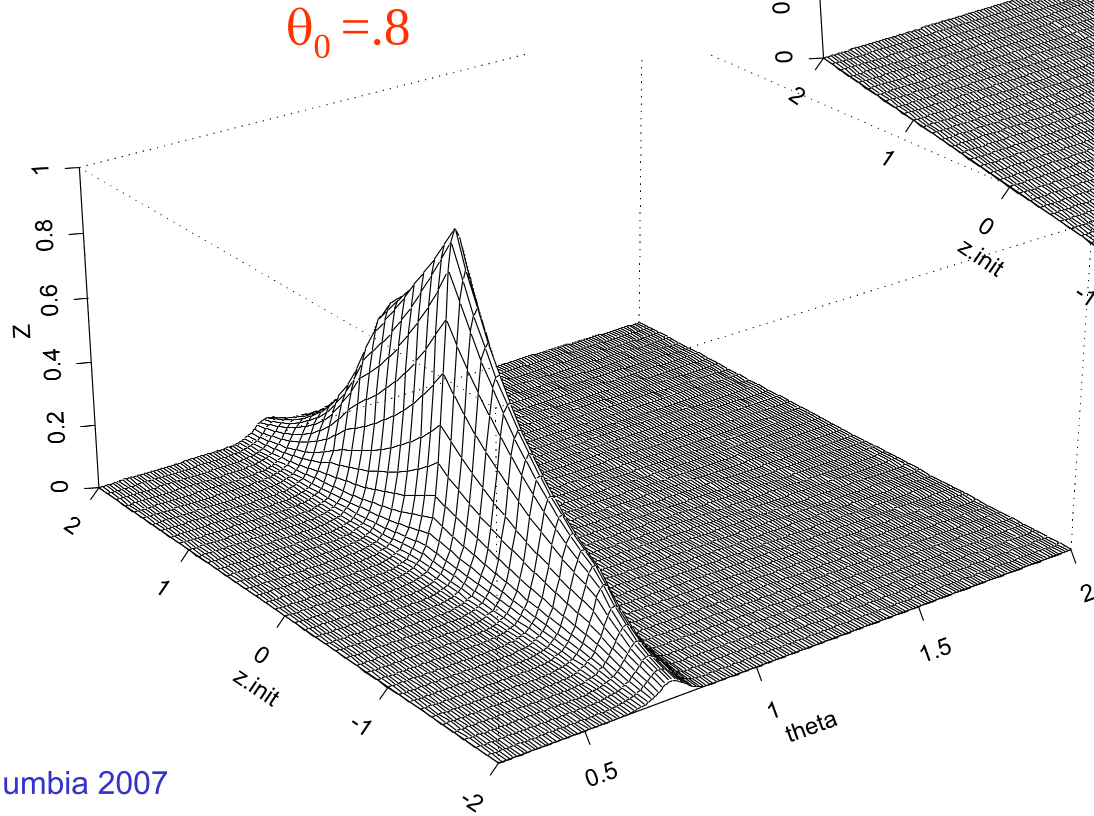
Note: integrate out z^{init} to get *Exact* likelihood.

$$f(\mathbf{y}_n) = \int_{-\infty}^{\infty} f(\mathbf{y}_n, z^{init})\, dz^{init}$$

Laplace likelihood examples

100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, $\{Z_t\} \sim$ IID Laplace pdf





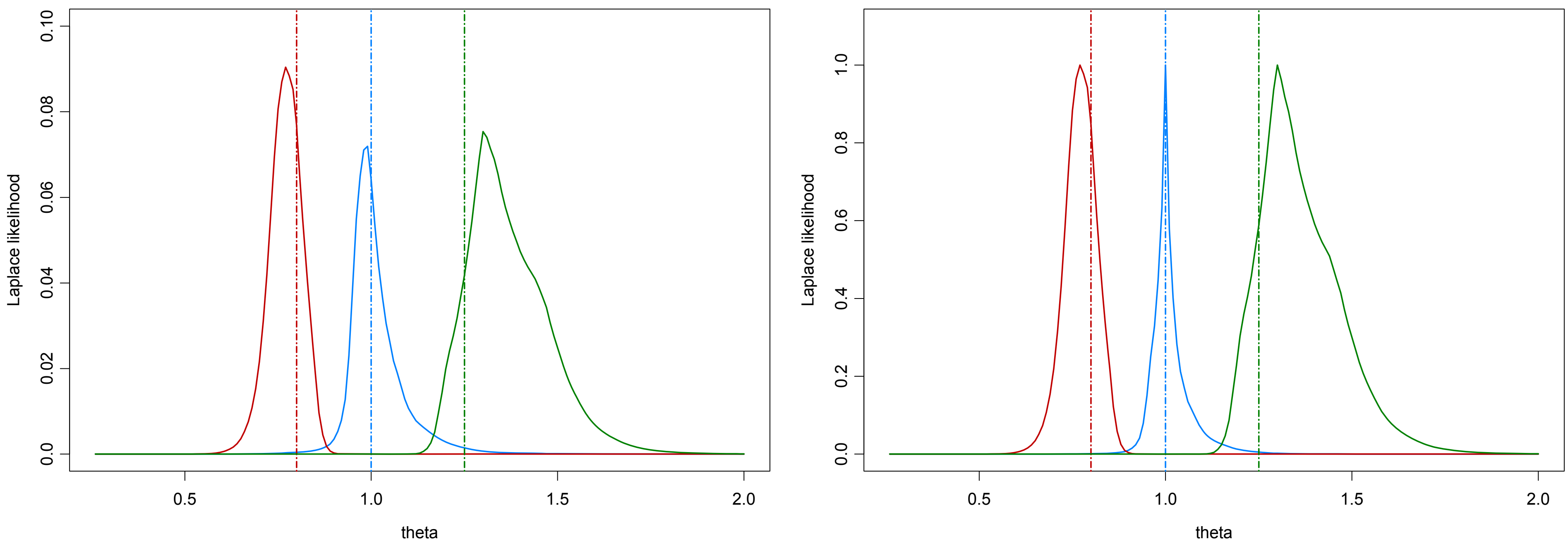
Laplace likelihood, Laplace noise

100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, $\{Z_t\} \sim$ IID Laplace pdf

$\theta_0 = .8$ $\theta_0 = 1.0$ $\theta_0 = 1.25$

Exact likelihood

Joint likelihood at $z_{max}(\theta)$



Laplace likelihood-LAD estimation

(Joint) Laplace log-likelihood. ($\sigma = \mathrm{E}|Z_0|$ is a scale parameter)

$$L(\theta, z^{init}, \sigma) = -(n+1)\log 2\sigma - \sigma^{-1}\sum_{t=0}^{n}|z_t| - n(\log|\theta|)1_{\{|\theta|>1\}}$$

Maximizing wrt σ, we obtain

$$\hat{\sigma} = \sum_{t=0}^{n}|z_t|/(n+1)$$

so that maximizing L is equivalent to minimizing

$$l_n(\theta, z^{init}) = \begin{cases} \sum_{t=0}^{n}|z_t|, & \text{if } |\theta| \le 1, \\ \sum_{t=0}^{n}|z_t||\theta|, & \text{otherwise.} \end{cases}$$

Joint Laplace likelihood — limit results

Result 1. Under the parameterizations,

$$\theta = 1 + \beta/n \quad \text{and} \quad z^{\text{init}} = Z_0 + \alpha\sigma/n^{1/2},$$

we have

$$U_n(\beta,\alpha) = \sigma^{-1}(l_n(\theta, z^{init}) - l_n(1, Z_0)) \to_d U(\beta,\alpha)$$

where

$$U(\beta,\alpha) = \int_0^1 \left(\beta \int_0^s e^{\beta(s-t)} dS(t) + \alpha e^{\beta s}\right) dW(s)$$

$$+ f(0)\int_0^1 \left(\beta \int_0^s e^{\beta(s-t)} dS(t) + \alpha e^{\beta s}\right)^2 ds$$

for $\beta \le 0$, and

$$U(\beta,\alpha) = \int_0^1 \left(-\beta \int_{s+}^1 e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)}\right) dW(s)$$

$$+ f(0)\int_0^1 \left(-\beta \int_s^1 e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)}\right)^2 ds$$

for $\beta > 0$.

Joint Laplace likelihood — limit results

The limits contain correlated Brownian Motions $S(t)$ and $W(t)$, obtained as the limits of the partial sum processes

$$S_n(t) = \frac{1}{\sigma\sqrt{n}}\sum_{i=0}^{[nt]} Z_i \to_d S(t), \qquad W_n(t) = \frac{1}{\sigma\sqrt{n}}\sum_{i=0}^{[nt]} \operatorname{sign}(Z_i) \to_d W(t).$$

From the limit,

$$U_n(\beta,\alpha) \to_d U(\beta,\alpha),$$

it suggests (from the continuous mapping theorem?) that

***limit(optimum(criterion))* = optimum(limit(criterion)).**

So for the optimizer of the **J**oint likelihood

$$\left(n\left(\hat{\theta}_J - 1\right), \sqrt{n}\sigma^{-1}\left(\hat{z}_J^{\text{init}} - Z_0\right)\right) \to_d \left(\hat{\beta}_J, \hat{\alpha}_J\right)$$

where

$$(\hat{\beta}_J, \hat{\alpha}_J) = \arg(\text{local}) \min U(\beta,\alpha).$$

Consistent estimation of noise?

Note that the previous results imply that

$$\hat{z}^{\text{init}} = Z_0 + \frac{\sigma}{\sqrt{n}}\hat{\alpha} = Z_0 + O_p\left(n^{-1/2}\right)$$

so that an unobserved random noise can be consistently estimated.

Does this make any sense?

Recall that in the unit root case,

$$Z_0 \in \overline{\text{sp}}\{Y_1, Y_2, \ldots, Y_n, \ldots\}$$

so that in fact, consistent estimation is possible.

Exact Laplace likelihood — limit results

Exact Laplace Likelihood:

$$L_n(\theta,\sigma) = \int_{-\infty}^{\infty} f(\mathbf{y}_n, z^{init}) dz^{init}$$

Result 2. For the local optimizer of the **E**xact likelihood,

$$n(\hat{\theta}_{\mathrm{E}} - 1) \to_d \hat{\beta}_{\mathrm{E}},$$

where

$$\hat{\beta}_{\mathrm{E}} = \arg\min U^*(\beta),$$

and $U^*(\beta)$ is a stochastic process defined in terms of $S(t)$ and $W(t)$.

Simulating from the limit process

Step 1. Simulate two indep sequences $(W_1, \ldots, W_m)$ and $(V_1, \ldots, V_m)$ of iid N(0,1) random variables with m=100000.

Step 2. Form $W(t)$ and $V(t)$ by the partial sum processes,

$$W(t) = \sum_{j=1}^{[100000\,t]} W_j / \sqrt{100000} \quad \text{and} \quad V(t) = \sum_{j=1}^{[100000\,t]} V_j / \sqrt{100000}.$$

Step 3. Set $S(t) = W(t) + c_1 V(t)$, where

$$c_1 = \sqrt{\operatorname{Var}(Z_t)/E^2|Z_0| - 1}.$$

Limit process depends only on c_1 and $f(0)$.

Step 4. Compute $U(\beta,\alpha)$ and $U^*(\beta)$ from the definition.

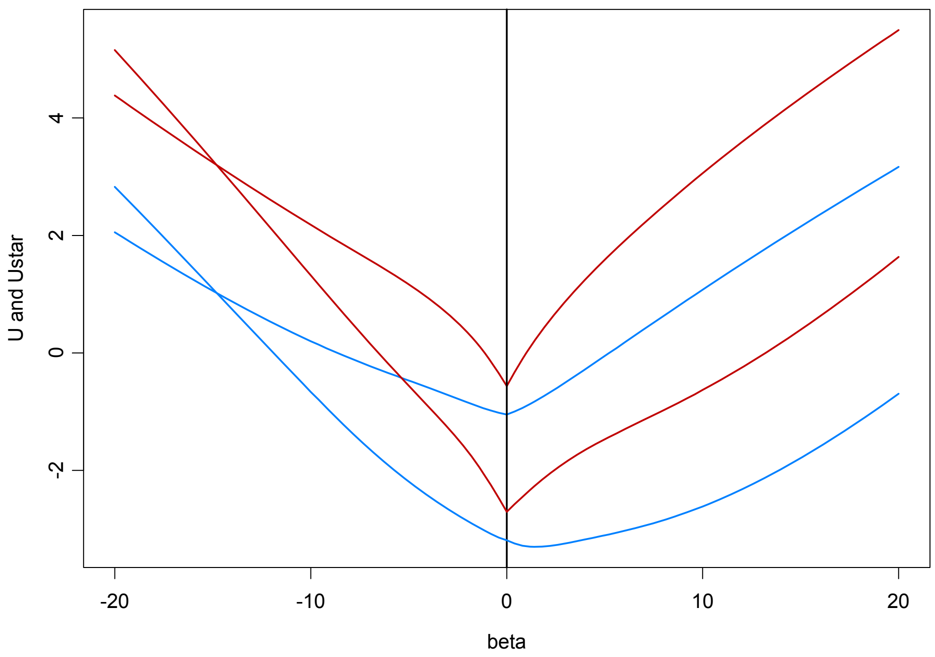
Step 5. Determine the respective **L**ocal and **G**lobal minimizers of **J**oint limit $U(\beta,\alpha)$ and **E**xact limit $U^*(\beta)$ numerically.

Simulated realizations of the limit processes

Simulate **Joint** and **Exact** limit processes, $U(\beta,\alpha(\beta))$, $U^*(\beta)$.

t(5) pdf

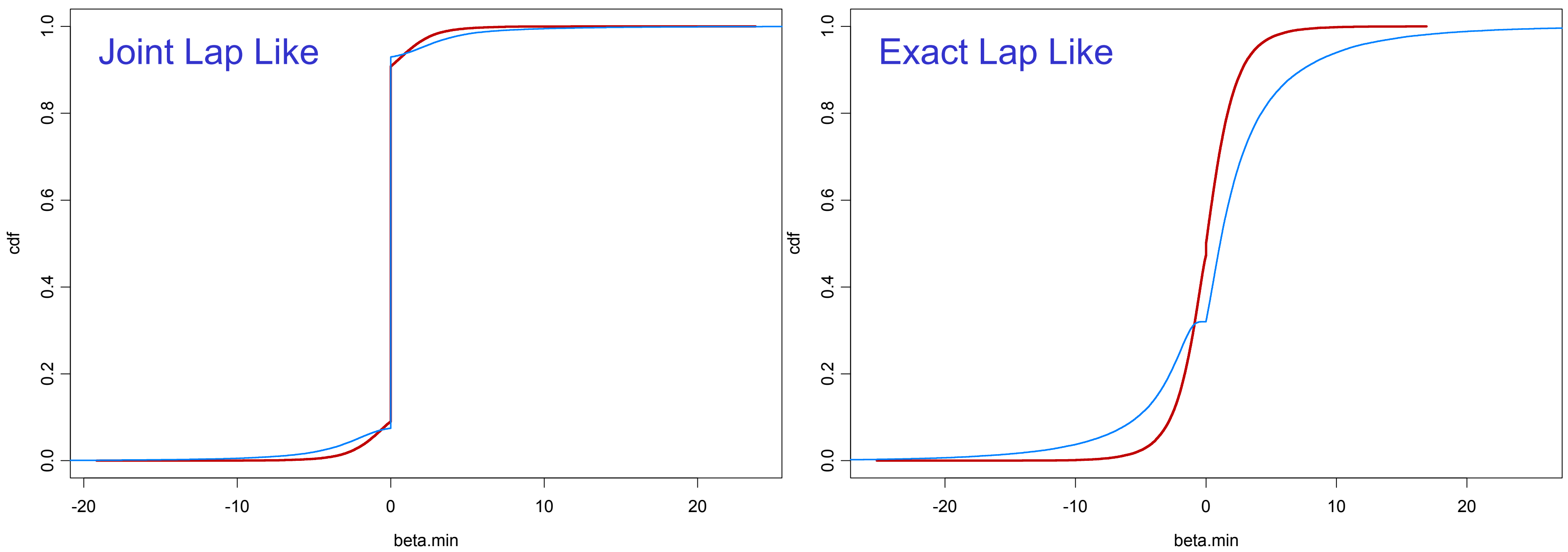
- Simulate realization of each limit process, joint and exact
- Compute local and global optima
- Repeat…
- Build up limit distribution functions



Limit cdf

red graph = Laplace pdf for Z_t

blue graph = Gaussian pdf for Z_t



Simulation results: Global Exact and Global Joint

Exact = MLE

Joint = maximize over θ and z_{init}

Laplace noise

$\theta = 1, \ \sigma = 1$

1000 reps

Note: **Joint** dominates **Exact** (rmse is half the size)

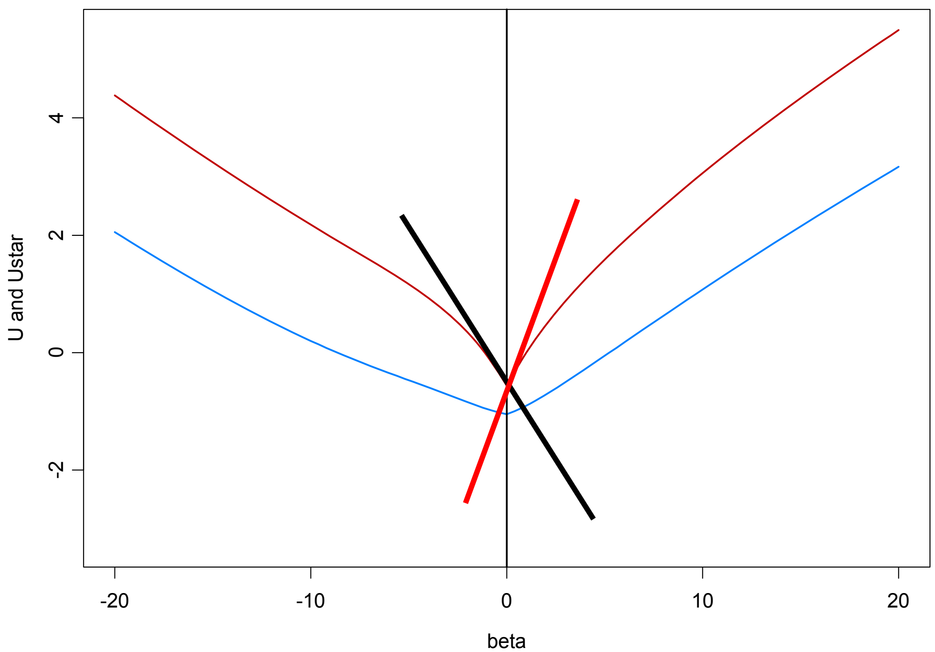
n		Exact $\hat{\theta}_E$	Joint $\hat{\theta}_J$
$n = 20$	bias	-.047	-.003
	rmse	**.224**	**.144**
$n = 50$	bias	-.013	.000
	rmse	**.096**	**.057**
$n = 100$	bias	.003	.000
	rmse	**.051**	**.011**
$n = 200$	bias	.000	.000
	rmse	**.028**	**.006**

Analysis of pile-up probabilities

Look back at realizations of the limit processes, $U(\beta,\alpha(\beta))$, $U^*(\beta)$.

- When is there a local optimum at $\theta = 1$?
- Check derivatives
- Negative derivative from the left
- Positive derivative from the right
- Local optimum at $\theta = 1$

t(5) pdf



Pile-up probabilities (Joint)

Result 3. (**J**oint Laplace likelihood)

$$\mathrm{P}(\hat{\theta}_{\mathrm{J}} = 1) \to \mathrm{P}(-1 < Y < 0),$$

where

$$Y = \int_0^1 S(s)dW(s) - W(1)\int_0^1 S(s)ds + \frac{W(1)}{2f(0)}\left(\int_0^1 W(s)ds - W(1)/2\right)$$

Idea: look at derivatives

$$\mathrm{P}(\hat{\theta}_{\mathrm{J}} = 1) = \mathrm{P}\left(\lim_{\beta\uparrow 0} \frac{\partial}{\partial\beta} U_n(\beta, \hat{\alpha}(\beta)) < 0 \text{ and } \lim_{\beta\downarrow 0} \frac{\partial}{\partial\beta} U_n(\beta, \hat{\alpha}(\beta)) > 0\right)$$

$$\to \mathrm{P}\left(\lim_{\beta\uparrow 0} \frac{\partial}{\partial\beta} U(\beta, \hat{\alpha}(\beta)) < 0 \text{ and } \lim_{\beta\downarrow 0} \frac{\partial}{\partial\beta} U(\beta, \hat{\alpha}(\beta)) > 0\right)$$

Now,

$$\lim_{\beta\downarrow 0} \frac{\partial}{\partial\beta} U(\beta, \hat{\alpha}(\beta)) = Y + 1$$

$$\lim_{\beta\uparrow 0} \frac{\partial}{\partial\beta} U(\beta, \hat{\alpha}(\beta)) = Y$$

and the result follows.

Pile-up probabilities (Exact)

Result 4. (**E**xact Laplace likelihood)

$$P(\hat{\theta}_E = 1) \to P\left[-\frac{1}{2} < Y < -\frac{1}{2}\right] = 0$$

The *pile-up probability* is always *zero* for the **E**xact, and always positive for the **J**oint (see Result 3).

Remark. (Laplace pile-up)

If Z_t has a Laplace density $f(z) = \frac{1}{2\sigma} e^{-|z|/\sigma}$, then

$$Y = \int_0^1 \left[W(1)s - W(s)\right] dV(s) + \frac{1}{2}.$$

where $W(s)$ and $V(s)$ are independent standard Brownian motions.

Laplace pile-up probabilities (cont)

It follows that the **J**oint estimator has pile-up probability

$$P(\hat{\theta}_J = 1) \to P(-1 < Y < 0)$$

$$= P(-.5 < \int_0^1 [W(1)s - W(s)]\, dV(s) < .5)$$

$$= E\left[P(-.5 < \int_0^1 [W(1)s - W(s)]\, dV(s) < .5 \mid W(t),\ t \in [0,1]) \right]$$

$$= E\left[2\Phi\left(.5 \left\{ \int_0^1 [W(1)s - W(s)]^2 ds \right\}^{-1/2} \right) - 1 \right]$$

$$\approx 0.820$$

But *no* pile-up probability for **L**ocal **E**xact:

Remark: if **L**ocal *does not* pile up, **G**lobal does not pile up

if **L**ocal *does* pile up, **G**lobal probably does as well

Simulation results – pile-up probabilities

Pile-up probabilities for **J**oint: $P(\hat{\theta}_J = 1)$

n	Gau	Lap	Unif	t(5)
20	.827	.796	.831	.796
50	.859	.806	.864	.823
100	.873	.819	.864	.817
200	.844	.819	.843	.831
500	.855	.809	.841	.846
∞	.858	.820	.836	.827

(No pile-up probabilities for **E**xact.)

Summary and Future Work

- Reviewed MA(1) unit root and near-unit root with Gaussian likelihood
    - 1/n asymptotics, pile-up even if $\theta<1$
- New results for MA(1) unit root with Least Absolute Deviations
    - 1/n asymptotics for **J**oint or **E**xact
    - **J**oint beats **E**xact;
    - **J**oint has pile-up and **E**xact does not
- Further work:
    - Nail down preliminary results, conduct further simulations
    - Other non-Gaussian criterion functions (MLE)?
    - Non-zero mean?
    - Near-unit root? $(1-\gamma/n)$
    - Performance of **J**oint with Gaussian likelihood?