Estimation of Possibly Non-Stationary Long Memory Processes via Adaptive Overdifferencing

Maryclare Griffin

University of Massachusetts Amherst
Department of Mathematics and Statistics

June 4, 2025

Joint work with David S. Matteson, Gennady Samorodnitsky (Cornell)

TECHNOMETRICS
2025, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/00401706.2025.2453207

Likelihood Inference for Possibly Nonstationary Processes via Adaptive Overdifferencing

Maryclare Griffin[a], Gennady Samorodnitsky[b], and David S. Matteson[c]

[a]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA; [b]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY; [c]Department of Statistics and Data Science, Cornell University, Ithaca, NY

ABSTRACT
We make an observation that facilitates exact likelihood-based inference for the parameters of the popular ARFIMA model without requiring stationarity by allowing the upper bound $\bar{d}$ for the memory parameter d to exceed 0.5: estimating the parameters of a single nonstationary ARFIMA model is equivalent to estimating the parameters of a sequence of stationary ARFIMA models. This allows for the use of existing methods for evaluating the likelihood for an invertible and stationary ARFIMA model. This enables improved inference because many standard methods perform poorly when estimates are close to the boundary of the parameter space. It also allows us to leverage the wealth of likelihood approximations that have been introduced for estimating the parameters of a stationary process. We explore how estimation of the memory parameter d depends on the upper bound $\bar{d}$ and introduce adaptive procedures for choosing $\bar{d}$. We show via simulation how our adaptive procedures estimate the memory parameter well, relative to existing alternatives, when the true value is as large as 2.5.

ARTICLE HISTORY
Received April 2024
Accepted January 2025

KEYWORDS
ARFIMA; FARIMA; long memory

Outline

Outline

- Basics

Outline

- Basics
    - Time Series Data

Outline

- Basics
  - Time Series Data
  - Long-Memory

Outline

- Basics
    - Time Series Data
    - Long-Memory
    - Autoregressive Moving Average (ARMA) Models

Outline

- Basics
    - Time Series Data
    - Long-Memory
    - Autoregressive Moving Average (ARMA) Models
    - Fractional Integration (FI)

Outline

- Basics
    - Time Series Data
    - Long-Memory
    - Autoregressive Moving Average (ARMA) Models
    - Fractional Integration (FI)
    - Autoregressive FI Moving Average (ARFIMA) Models

Outline

- Basics
  - Time Series Data
  - Long-Memory
  - Autoregressive Moving Average (ARMA) Models
  - Fractional Integration (FI)
  - Autoregressive FI Moving Average (ARFIMA) Models
- Statement of the problem with existing methods

Outline

- Basics
  - Time Series Data
  - Long-Memory
  - Autoregressive Moving Average (ARMA) Models
  - Fractional Integration (FI)
  - Autoregressive FI Moving Average (ARFIMA) Models
- Statement of the problem with existing methods
- Review of possible solutions and limitations

Outline

- Basics
    - Time Series Data
    - Long-Memory
    - Autoregressive Moving Average (ARMA) Models
    - Fractional Integration (FI)
    - Autoregressive FI Moving Average (ARFIMA) Models
- Statement of the problem with existing methods
- Review of possible solutions and limitations
- Proposal of a solution

Outline

- Basics
  - Time Series Data
  - Long-Memory
  - Autoregressive Moving Average (ARMA) Models
  - Fractional Integration (FI)
  - Autoregressive FI Moving Average (ARFIMA) Models
- Statement of the problem with existing methods
- Review of possible solutions and limitations
- Proposal of a solution
- Demonstration of success

Time Series Data

Time Series Data

Consists of n consecutive values

Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

For this talk, we're going to assume the observations $\boldsymbol{y}$ are:

Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

For this talk, we're going to assume the observations $\boldsymbol{y}$ are:

- Equally spaced in time

Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

For this talk, we're going to assume the observations $\boldsymbol{y}$ are:

- Equally spaced in time
- Univariate

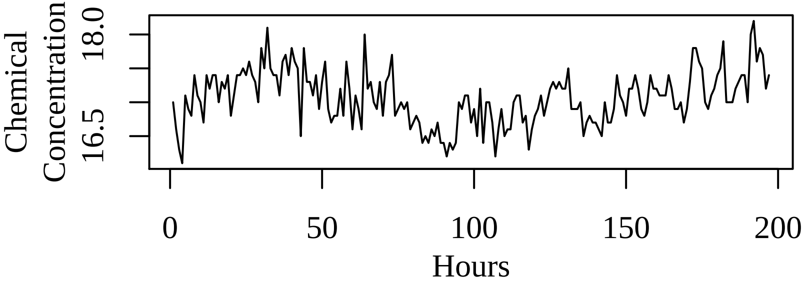
Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

For this talk, we're going to assume the observations $\boldsymbol{y}$ are:

- Equally spaced in time
- Univariate



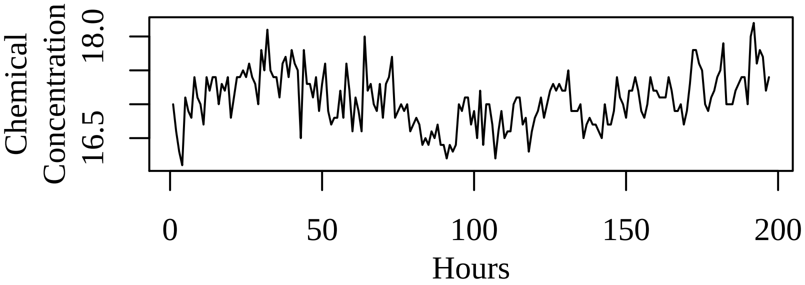
Time Series Data

Consists of n consecutive values

$$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_{n-2}, y_{n-1}, y_n)$$

For this talk, we're going to assume the observations $\boldsymbol{y}$ are:

- Equally spaced in time
- Univariate



Box and Jenkins (1970)

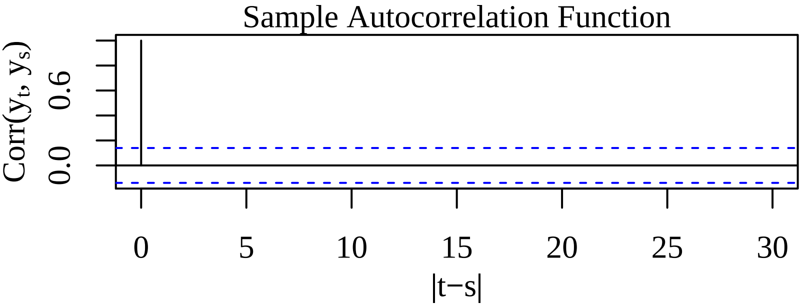
Long Memory

Long Memory

$\operatorname{Corr}(y_t, y_s)$ decreases very slowly as $|t - s|$ increases

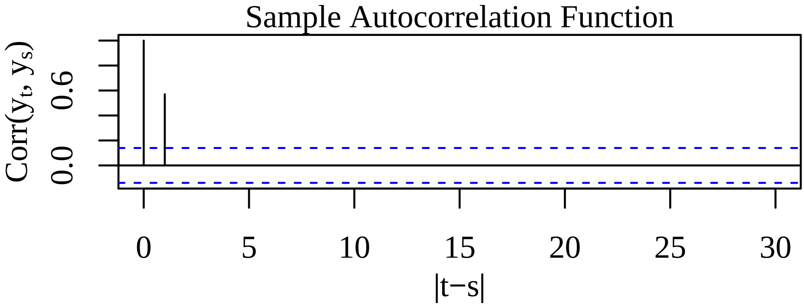
Long Memory

Corr (y_t, y_s) decreases very slowly as $|t - s|$ increases



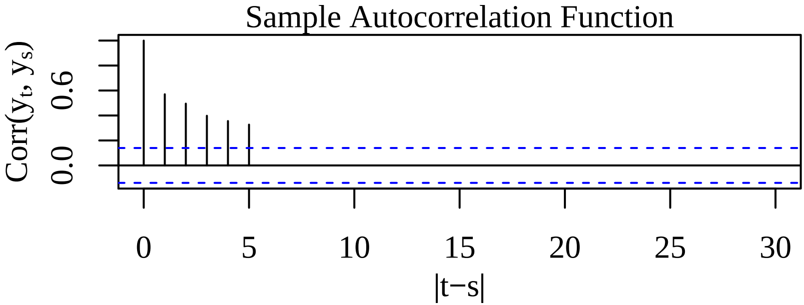
Long Memory

Corr (y_t, y_s) decreases very slowly as $|t - s|$ increases



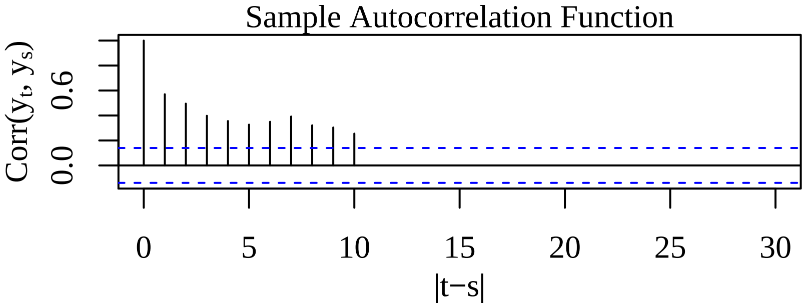
Long Memory

$\text{Corr}(y_t, y_s)$ decreases very slowly as $|t-s|$ increases



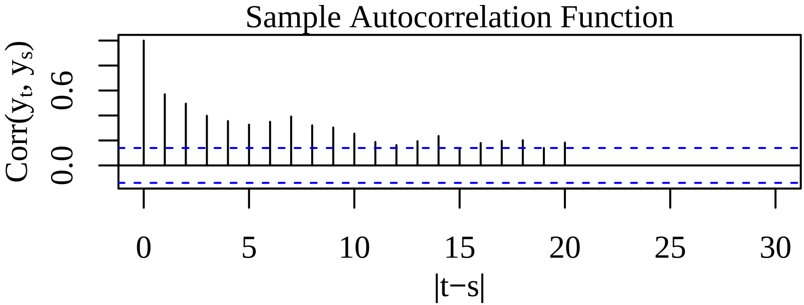
Long Memory

$\text{Corr}(y_t, y_s)$ decreases very slowly as $|t - s|$ increases



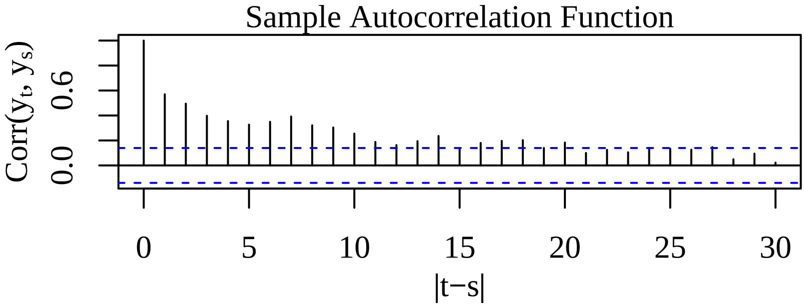
Long Memory

$\text{Corr}(y_t, y_s)$ decreases very slowly as $|t - s|$ increases



Long Memory

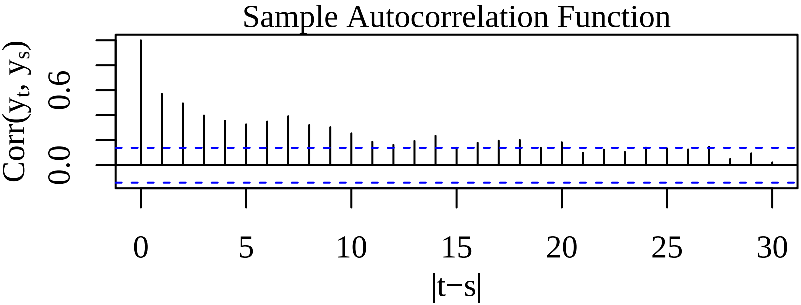
Corr (y_t, y_s) decreases very slowly as $|t - s|$ increases



Laboissiere et al. (2015); Barassi et al. (2018); Lo et al. (1993)

Long Memory

Corr (y_t, y_s) decreases very slowly as $|t - s|$ increases

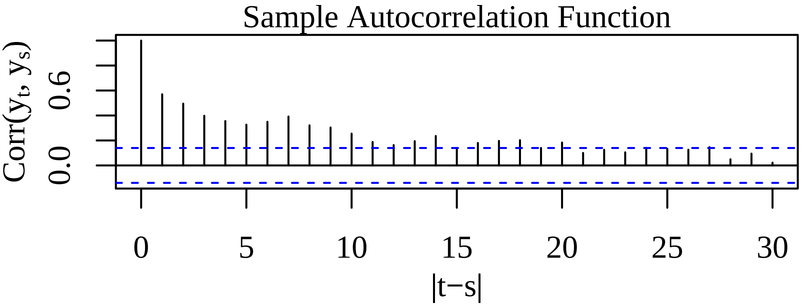


Long memory is observed in many types of data

Laboissiere et al. (2015); Barassi et al. (2018); Lo et al. (1993)

Long Memory

Corr (y_t, y_s) decreases very slowly as $|t-s|$ increases



Long memory is observed in many types of data

- Stream flows, earth surface temperature, CO_2 emissions
- EEG, MRI, measurements of movement
- Behavior of cells in culture

Laboissiere et al. (2015); Barassi et al. (2018); Lo et al. (1993)

Autoregressive Moving Average (ARMA) Model

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

Autoregressive Moving Average (ARMA) Model

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t}^{\text{ARMA}(p,q)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Box and Jenkins (1970)

Autoregressive Moving Average (ARMA) Model

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t}^{\text{ARMA}(p,q)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

- AR ϕ_j gives a previous deviation $y_{t-j} - \mu_{t-j}$'s contribution
- MA θ_l gives a previous error z_{t-l}'s contribution

Box and Jenkins (1970)

Autoregressive Moving Average (ARMA) Model

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t}^{\text{ARMA}(p,q)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

- AR ϕ_j gives a previous deviation $y_{t-j} - \mu_{t-j}$'s contribution
- MA θ_l gives a previous error z_{t-l}'s contribution
- Mean constant $\mu_t = \mu$ or a linear in covariates $\mu_t = \boldsymbol{x}_t' \boldsymbol{\beta}$

Box and Jenkins (1970)

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Often rewritten in terms of the lag operator $B^\ell y_t = y_{t-\ell}$

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Often rewritten in terms of the lag operator $B^\ell y_t = y_{t-\ell}$

- $1 - \sum_{j=1}^{p} \phi_j B^j = \phi\left(B\right)$

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Often rewritten in terms of the lag operator $B^{\ell} y_t = y_{t-\ell}$

- $1 - \sum_{j=1}^{p} \phi_j B^j = \phi\left(B\right)$
- $1 + \sum_{l=1}^{q} \theta_l B^l = \theta\left(B\right)$

Autoregressive Moving Average (ARMA) Model

$$\left(y_t - \mu_t\right) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Often rewritten in terms of the lag operator $B^\ell y_t = y_{t-\ell}$

- $1 - \sum_{j=1}^{p} \phi_j B^j = \phi\left(B\right)$
- $1 + \sum_{l=1}^{q} \theta_l B^l = \theta\left(B\right)$

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Often rewritten in terms of the lag operator $B^\ell y_t = y_{t-\ell}$

- $1 - \sum_{j=1}^{p} \phi_j B^j = \phi\left(B\right)$
- $1 + \sum_{l=1}^{q} \theta_l B^l = \theta\left(B\right)$

$$\implies \phi\left(B\right)\left(y_t - \mu_t\right) = \theta\left(B\right) z_t$$

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Constrained to be

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Constrained to be

- **Stationary:** All roots of $\phi(x)$ are outside the unit circle
  $\rightarrow$ Variance is finite, we can compute the likelihood

Autoregressive Moving Average (ARMA) Model

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Constrained to be

- **Stationary:** All roots of $\phi(x)$ are outside the unit circle
  $\rightarrow$ Variance is finite, we can compute the likelihood
- **Invertible:** All roots of $\theta(x)$ are outside the unit circle
  $\rightarrow$ Unique values of $\theta_1, \ldots, \theta_q$ give unique likelihood

ARMA and Long Memory

ARMA and Long Memory

First - let's simplify things

ARMA and Long Memory

First - let's simplify things

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

ARMA and Long Memory

First - let's simplify things

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + \sum_{l=1}^{q} \theta_l z_{t-l} + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

MA parameters θ_l contribute up to l-th autocorrelation

$\rightarrow$ Can't help us model long memory

AR and Long Memory

AR and Long Memory

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

AR and Long Memory

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Long memory $\implies$ $|\phi_j|$ not "small" for large p, effectively

AR and Long Memory

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Long memory $\implies$ $|\phi_j|$ not "small" for large p, effectively

$$(y_t - \mu_t) = \sum_{j=1}^{\infty} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

AR and Long Memory

$$(y_t - \mu_t) = \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

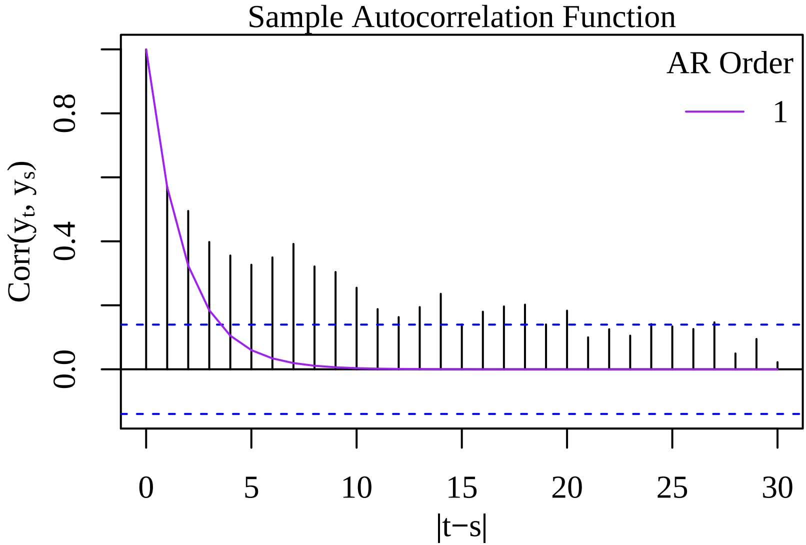
Long memory $\implies$ $|\phi_j|$ not "small" for large p, effectively

$$(y_t - \mu_t) = \sum_{j=1}^{\infty} \phi_j \left(y_{t-j} - \mu_{t-j}\right) + z_t$$

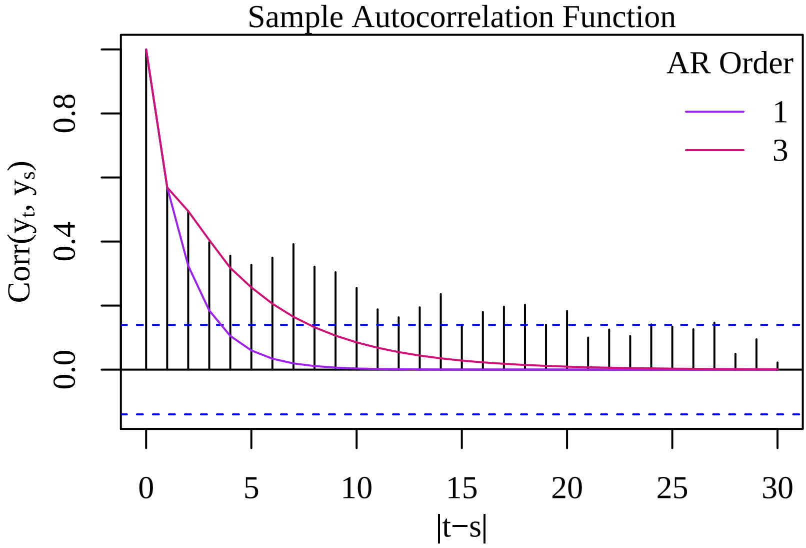
$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

In practice, long memory requires $p \to \infty$ as $n \to \infty$

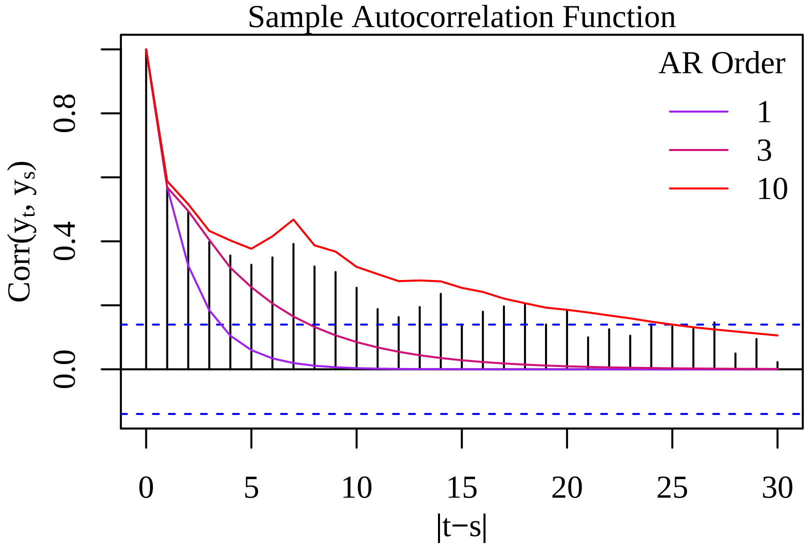
AR and Long Memory



AR and Long Memory



AR and Long Memory



Fractional Integration (FI)

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^{j-1} (y_{t-j} - \mu_{t-j}) + z_t}^{\text{FI}(d)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Granger and Joyeux (1980); Hosking (1981)

Fractional Integration (FI)

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^{j-1} (y_{t-j} - \mu_{t-j}) + z_t}^{\mathrm{FI}(d)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Can be written $(1 - B)^d (y_t - \mu_t) = z_t$

Granger and Joyeux (1980); Hosking (1981)

Fractional Integration (FI)

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^{j-1} (y_{t-j} - \mu_{t-j}) + z_t}^{\mathrm{FI}(d)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Can be written $(1 - B)^d (y_t - \mu_t) = z_t$

Stationary if $d < 0.5$, "invertible" for $d > -0.5$

Granger and Joyeux (1980); Hosking (1981)

Fractional Integration (FI)

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^{j-1} (y_{t-j} - \mu_{t-j}) + z_t}^{\mathrm{FI}(d)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Can be written $(1 - B)^d (y_t - \mu_t) = z_t$

Stationary if $d < 0.5$, "invertible" for $d > -0.5$

Invertible means something different here...

Granger and Joyeux (1980); Hosking (1981)

Fractional Integration (FI)

$$\overbrace{(y_t - \mu_t) = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^{j-1} (y_{t-j} - \mu_{t-j}) + z_t}^{\text{FI}(d)}$$

$$z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Can be written $(1 - B)^d (y_t - \mu_t) = z_t$

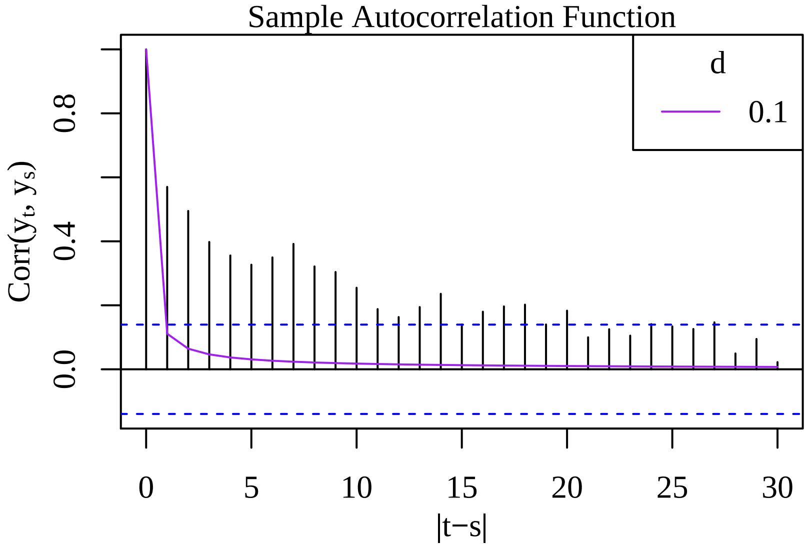
Stationary if $d < 0.5$, "invertible" for $d > -0.5$

Invertible means something different here...

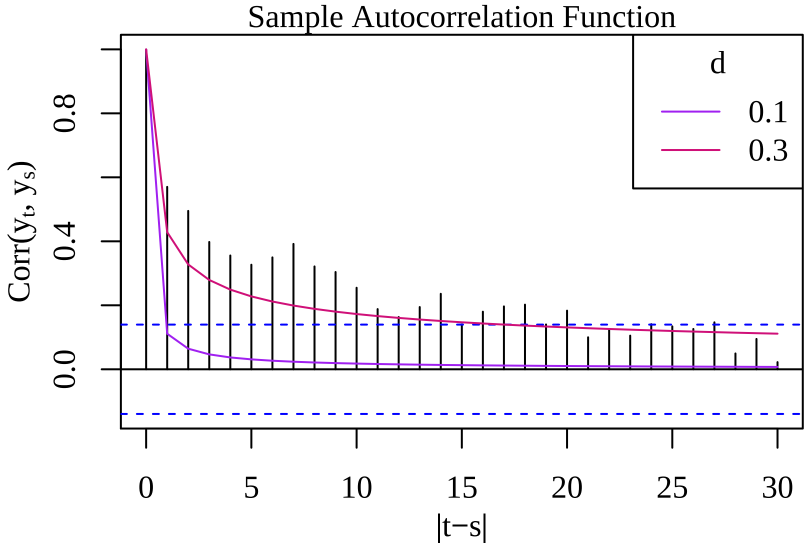
Unique values of $d \leq -0.5$ give unique likelihood

Granger and Joyeux (1980); Hosking (1981)

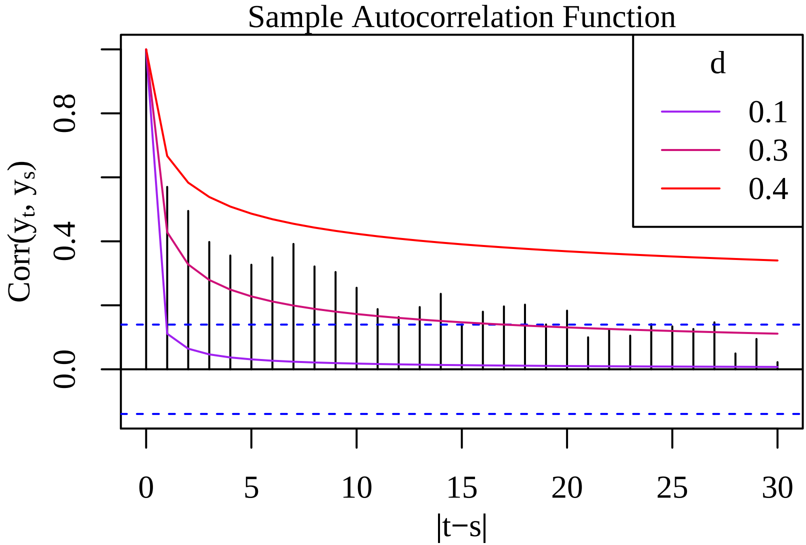
Fractional Integration (FI)



Fractional Integration (FI)



Fractional Integration (FI)



Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi\left(B\right)\left(1-B\right)^{d}\left(y_t-\mu_t\right)=\theta\left(B\right)z_t, \quad z_t \stackrel{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi(B)(1-B)^d(y_t - \mu_t) = \theta(B)\, z_t, \quad z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Constrained to be

Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi(B)(1-B)^d(y_t-\mu_t)=\theta(B)z_t, \quad z_t \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Constrained to be

- **Stationary:** Roots of $\phi(x)$ outside unit circle, $d < 0.5$

Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi\left(B\right)\left(1-B\right)^{d}\left(y_t-\mu_t\right)=\theta\left(B\right)z_t, \quad z_t \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Constrained to be

- **Stationary:** Roots of $\phi\left(x\right)$ outside unit circle, $d < 0.5$
- **Invertible:** Roots of $\theta\left(x\right)$ outside unit circle, $d \geq -0.5$

Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi\left(B\right)\left(1-B\right)^{d}\left(y_t-\mu_t\right)=\theta\left(B\right)z_t, \quad z_t \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Constrained to be

- **Stationary:** Roots of $\phi\left(x\right)$ outside unit circle, $d < 0.5$
- **Invertible:** Roots of $\theta\left(x\right)$ outside unit circle, $d \geq -0.5$

This talk: Focus on FI(d) =ARFIMA($0, d, 0$)

Autoregressive FI Moving Average (ARFIMA) Model

$$\overbrace{\phi\left(B\right)\left(1-B\right)^{d}\left(y_t-\mu_t\right)=\theta\left(B\right)z_t, \quad z_t \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)}^{\text{ARFIMA}(p,d,q)}$$

Constrained to be

- **Stationary:** Roots of $\phi\left(x\right)$ outside unit circle, $d < 0.5$
- **Invertible:** Roots of $\theta\left(x\right)$ outside unit circle, $d \geq -0.5$

This talk: Focus on FI(d) =ARFIMA$(0, d, 0)$

(Content applies to ARFIMA(p, d, q) as well)

Estimation of ARFIMA(p, d, q) Processes

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d
- Likelihood

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d
- Likelihood
  - ☹ Requires assumptions on range of d

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d
- Likelihood
  - ☹ Requires assumptions on range of d
  - ☹ Needs more computational tricks

Baillie (1996)

Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d
- Likelihood
  - ☹ Requires assumptions on range of d
  - ☹ Needs more computational tricks
  - ☺ Theory for standard errors, confidence intervals applies

Baillie (1996)

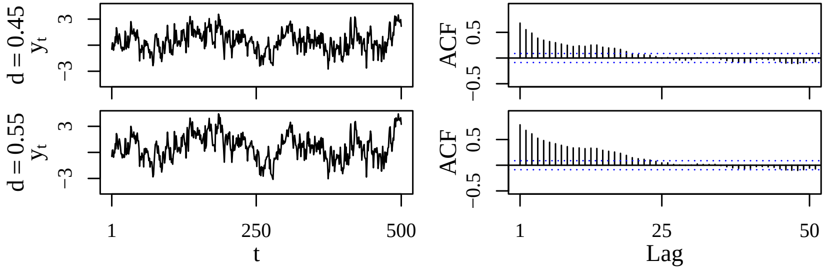
Estimation of ARFIMA(p, d, q) Processes

Lots of methods, based on:

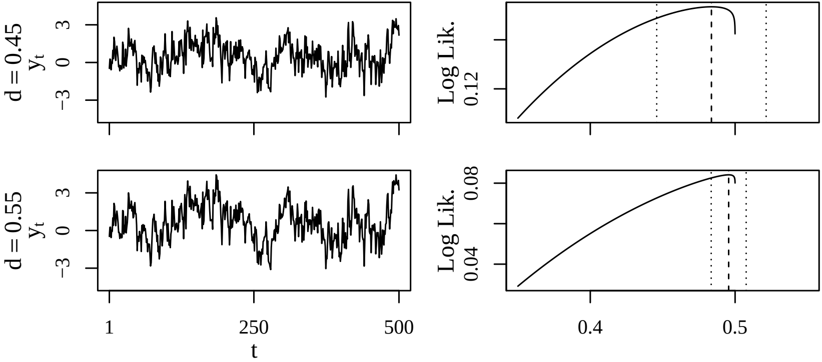
- Regression
  - ☺ Doesn't have to make assumptions on range of d
  - ☹ "Throws out" data, may not give estimates of all parameters
  - ☹ Often introduces additional tuning parameters
  - ☹ Difficult to get standard errors, confidence intervals for d
- Likelihood
  - ☹ **Requires assumptions on range of d**
  - ☹ Needs more computational tricks
  - ☺ Theory for standard errors, confidence intervals applies

Baillie (1996)

Why do we care about assumptions on d?



Why do we care about assumptions on d?



Can we get around this?

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

$$(1 - B)^{d-m} (1 - B)^{m} (y_t - \mu_t) = z_t$$

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

$$(1 - B)^{d-m} (1 - B)^m (y_t - \mu_t) = z_t$$

- If $0.5 \le d < 1.5$, choose $m = 1$ to get a stationary process

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

$$(1-B)^{d-m}(1-B)^{m}(y_t-\mu_t)=z_t$$

- If $0.5 \leq d < 1.5$, choose $m=1$ to get a stationary process

$$(1-B)(y_t-\mu_t)=y_t-y_{t-1}-(\mu_t-\mu_{t-1})$$

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

$$(1-B)^{d-m}(1-B)^{m}(y_t-\mu_t)=z_t$$

- If $0.5 \leq d < 1.5$, choose $m=1$ to get a stationary process

$$(1-B)(y_t-\mu_t)=y_t-y_{t-1}-(\mu_t-\mu_{t-1})$$

- If $1.5 \leq d < 2.5$, choose $m=2$ to get a stationary process

Can we get around this?

Yes - because polynomials in B are just like regular polynomials

$$(1 - B)^{d-m} (1 - B)^{m} (y_t - \mu_t) = z_t$$

- If $0.5 \leq d < 1.5$, choose $m = 1$ to get a stationary process

$$(1 - B)(y_t - \mu_t) = y_t - y_{t-1} - (\mu_t - \mu_{t-1})$$

- If $1.5 \leq d < 2.5$, choose $m = 2$ to get a stationary process

$$(1 - B)^2 (y_t - \mu_t) = y_t - 2y_{t-1} + y_{t-2} - (\mu_t - 2\mu_{t-1} + \mu_{t-2})$$

Can we get around this?

Suggests the following procedure

Can we get around this?

Suggests the following procedure

1. For $d < 0.5$, fit an FI(d) model to y_t

Can we get around this?

Suggests the following procedure

1. For $d < 0.5$, fit an FI(d) model to y_t
2. For $0.5 \leq d < 1.5$, fit an FI($d-1$) model to $(1-B)\,y_t$

Can we get around this?

Suggests the following procedure

1. For $d < 0.5$, fit an FI(d) model to y_t
2. For $0.5 \leq d < 1.5$, fit an FI($d-1$) model to $(1-B)\,y_t$
3. For $1.5 \leq d < 2.5$, fit an FI($d-2$) model to $(1-B)^2\,y_t$

Can we get around this?

Suggests the following procedure

1. For $d < 0.5$, fit an FI(d) model to y_t
2. For $0.5 \leq d < 1.5$, fit an FI($d-1$) model to $(1-B)\,y_t$
3. For $1.5 \leq d < 2.5$, fit an FI($d-2$) model to $(1-B)^2\,y_t$

Select $d \leq 2.5$ that maximizes likelihood

Can we get around this?

Suggests the following procedure

1. For $d < 0.5$, fit an FI(d) model to y_t
2. For $0.5 \le d < 1.5$, fit an FI($d-1$) model to $(1-B)\,y_t$
3. For $1.5 \le d < 2.5$, fit an FI($d-2$) model to $(1-B)^2\,y_t$

Select $d \le 2.5$ that maximizes likelihood

Problem: Likelihoods in each step based on different data!

Can we get around this?

Suggests the following procedure

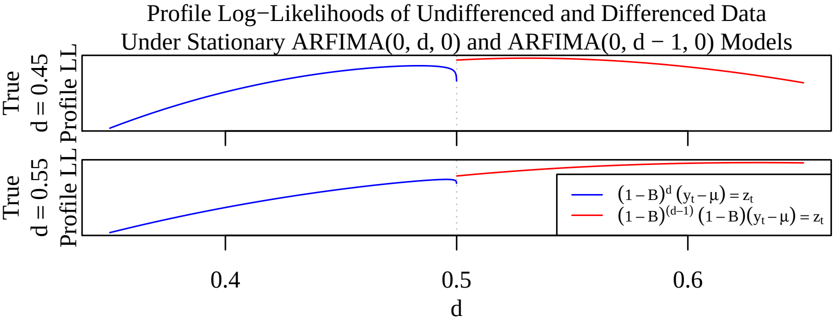
1. For $d < 0.5$, fit an FI(d) model to y_t
2. For $0.5 \leq d < 1.5$, fit an FI($d - 1$) model to $(1 - B)\, y_t$
3. For $1.5 \leq d < 2.5$, fit an FI($d - 2$) model to $(1 - B)^2\, y_t$

Select $d \leq 2.5$ that maximizes likelihood

Problem: Likelihoods in each step based on different data!

Beran (1995); Velasco and Robinson (2000); Mayoral (2007)

What's wrong with this procedure?



Is there anything else we can do?

Is there anything else we can do?

- **Problem:** Data changes as d changes

Is there anything else we can do?

- **Problem:** Data changes as d changes
- **Solution:** Fix data in advance to get stationarity for $d < \bar{d}$

Key Idea

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1 - B)^m \, y_t$ is stationary $\mathrm{FI}(d - m)$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1 - B)^m y_t$ is stationary $\mathrm{FI}(d - m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1 - B)^m y_t$ is stationary FI$(d - m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

  - Corresponds to a <u>non-invertible</u> FI$(d - m)$ model

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1-B)^m y_t$ is stationary FI$(d-m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

  - Corresponds to a non-invertible FI$(d-m)$ model

$$(1-B)^{d-m}(1-B)^m(y_t - \mu_t) = z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1 - B)^m y_t$ is stationary FI$(d - m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

- Corresponds to a non-invertible FI$(d - m)$ model

$$(1 - B)^{d-m} (1 - B)^m (y_t - \mu_t) = z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

☺ Can find integer r such that $d - m + r > -0.5$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1 - B)^m y_t$ is stationary FI$(d - m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

  - Corresponds to a non-invertible FI$(d - m)$ model

$$(1 - B)^{d-m} (1 - B)^m (y_t - \mu_t) = z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

☺ Can find integer r such that $d - m + r > -0.5$

$$(1 - B)^{d-m+r} (1 - B)^m (y_t - \mu_t) = (1 - B)^r z_t$$

Odaki (1993); Doornik and Ooms (2003)

Key Idea

- Given $\bar{d}$, we can find m such that $\bar{d} - m < 0.5$
- As long as $d < \bar{d}$, $(1-B)^m y_t$ is stationary FI$(d-m)$

☹ Most software won't give likelihood if $d - m \leq -0.5$

  - Corresponds to a <u>non-invertible</u> FI$(d-m)$ model

$$(1-B)^{d-m}(1-B)^m(y_t - \mu_t) = z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

☺ Can find integer r such that $d - m + r > -0.5$

$$(1-B)^{d-m+r}(1-B)^m(y_t - \mu_t) = (1-B)^r z_t$$

$(1-B)^m(y_t - \mu_t)$ stationary ARFIMA$(0, d-m+r, r)$

Odaki (1993); Doornik and Ooms (2003)

New Procedure

New Procedure

- Pick $\bar{d}$, maximum possible d

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$
- Evaluate sequence of stationary ARFIMA likelihoods

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$
- Evaluate sequence of stationary ARFIMA likelihoods
    1. For $-0.5 \leq d - m < 0.5$, $w_t \sim$ARFIMA$(0, d - m, 0)$

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$
- Evaluate sequence of stationary ARFIMA likelihoods
    1. For $-0.5 \leq d - m < 0.5$, $w_t \sim$ARFIMA$(0, d - m, 0)$
    2. For $-1.5 \leq d - m < -0.5$, $w_t \sim$ARFIMA$(0, d - m + 1, 1)$

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$
- Evaluate sequence of stationary ARFIMA likelihoods
    1. For $-0.5 \leq d - m < 0.5$, $w_t \sim$ARFIMA$(0, d - m, 0)$
    2. For $-1.5 \leq d - m < -0.5$, $w_t \sim$ARFIMA$(0, d - m + 1, 1)$
    3. For $-2.5 \leq d - m < -1.5$, $w_t \sim$ARFIMA$(0, d - m + 2, 2)$

New Procedure

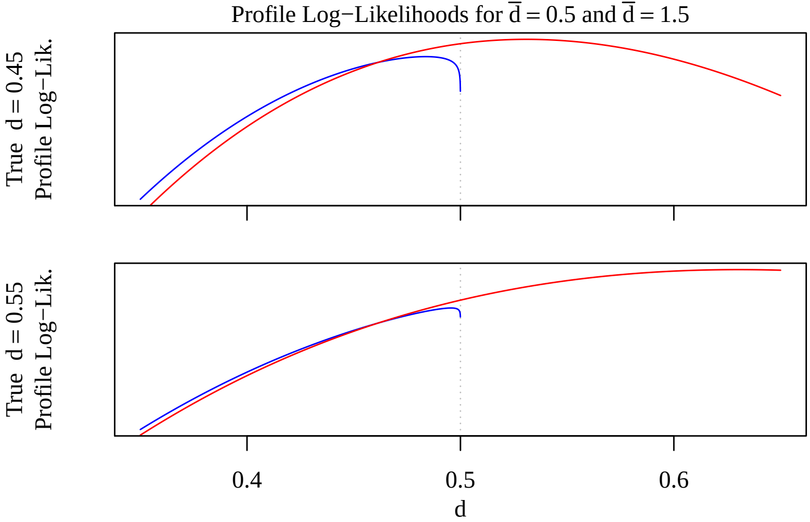
- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \leq \bar{d} - m \leq 0.5$
- Compute $w_t = (1 - B)^m \, y_t$
- Evaluate sequence of stationary ARFIMA likelihoods
    1. For $-0.5 \leq d - m < 0.5$, $w_t \sim$ARFIMA$(0, d - m, 0)$
    2. For $-1.5 \leq d - m < -0.5$, $w_t \sim$ARFIMA$(0, d - m + 1, 1)$
    3. For $-2.5 \leq d - m < -1.5$, $w_t \sim$ARFIMA$(0, d - m + 2, 2)$
- Select $-2.5 + m \leq d \leq 0.5 + m$ that maximizes likelihood

New Procedure

- Pick $\bar{d}$, maximum possible d
- Find integer m such that $-0.5 \le \bar{d} - m \le 0.5$
- Compute $w_t = (1 - B)^m y_t$
- Evaluate sequence of stationary ARFIMA likelihoods
  1. For $-0.5 \le d - m < 0.5$, $w_t \sim$ARFIMA$(0, d - m, 0)$
  2. For $-1.5 \le d - m < -0.5$, $w_t \sim$ARFIMA$(0, d - m + 1, 1)$
  3. For $-2.5 \le d - m < -1.5$, $w_t \sim$ARFIMA$(0, d - m + 2, 2)$
- Select $-2.5 + m \le d \le 0.5 + m$ that maximizes likelihood

This gives a continuous likelihood in d!

New Procedure



Simulations

Simulations

The new procedure gives different estimates, hopefully better?

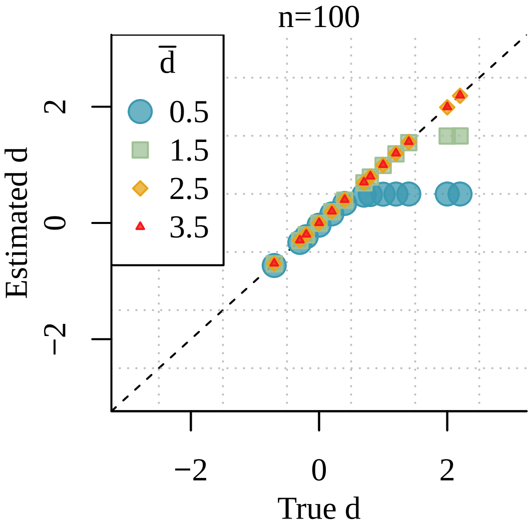
Simulations

The new procedure gives different estimates, hopefully better?

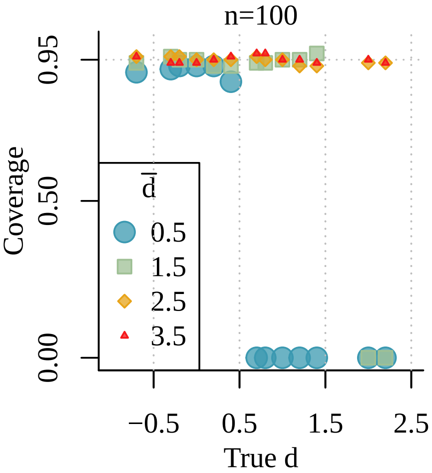
Simulate $10,000$ time series $\boldsymbol{y}$ for each:

- $d \in \{0, 0.1, \ldots, 2.2\}$
- $n \in \{100, 500\}$

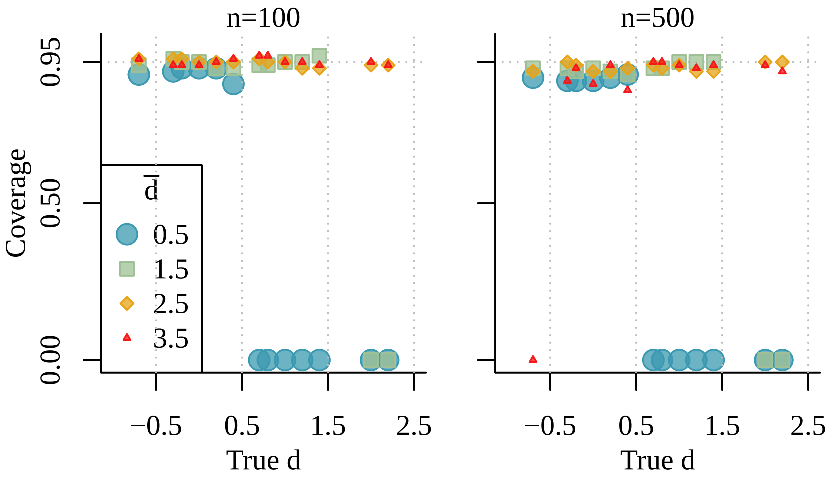
Simulations



Simulations



Simulations



What's left?!

What's left?!

Need a principled/optimal way to choose $\bar{d}$

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

$\Pr(\hat{d}_{\bar{d}} > \bar{d}) \approx \epsilon$, small ϵ facilitates:

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

$\Pr(\hat{d}_{\bar{d}} > \bar{d}) \approx \epsilon$, small ϵ facilitates:

- Construction of CI's for d with nominal coverage

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

$\Pr(\hat{d}_{\bar{d}} > \bar{d}) \approx \epsilon$, small ϵ facilitates:

- Construction of CI's for d with nominal coverage
- Implementation of tests of $d = c$ with nominal level

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

$\Pr(\hat{d}_{\bar{d}} > \bar{d}) \approx \epsilon$, small ϵ facilitates:

- Construction of CI's for d with nominal coverage
- Implementation of tests of $d = c$ with nominal level

When $\epsilon = 0.5$ we'll call estimator **BND**

What's left?!

Need a principled/optimal way to choose $\bar{d}$

Choose smallest $\bar{d}$ for which:

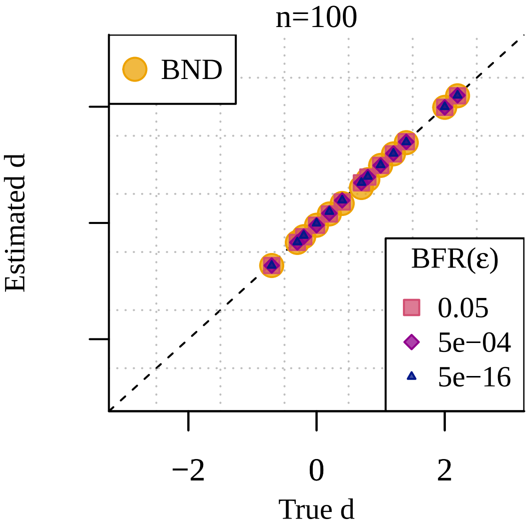
- Likelihood is decreasing in d near boundary $\bar{d}$
- Approximate $(1-\epsilon)\,\%$ quantile of $\hat{d}_{\bar{d}}$ is $< \bar{d}$

$\Pr(\hat{d}_{\bar{d}} > \bar{d}) \approx \epsilon$, small ϵ facilitates:

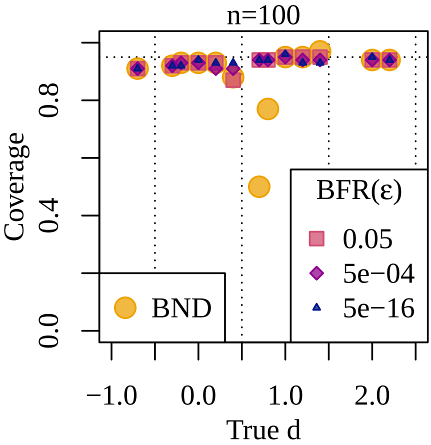
- Construction of CI's for d with nominal coverage
- Implementation of tests of $d = c$ with nominal level

When $\epsilon = 0.5$ we'll call estimator **BND**
When $\epsilon < 0.5$, we'll call estimator $\mathbf{BFR}(\epsilon)$

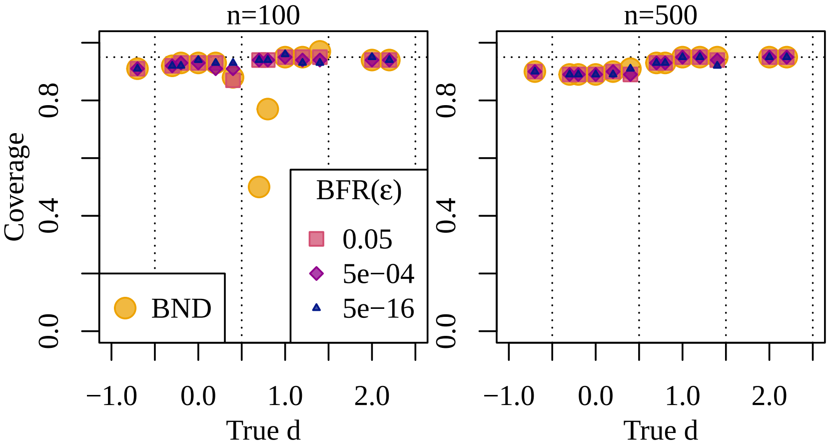
Adaptive Procedure Simulations



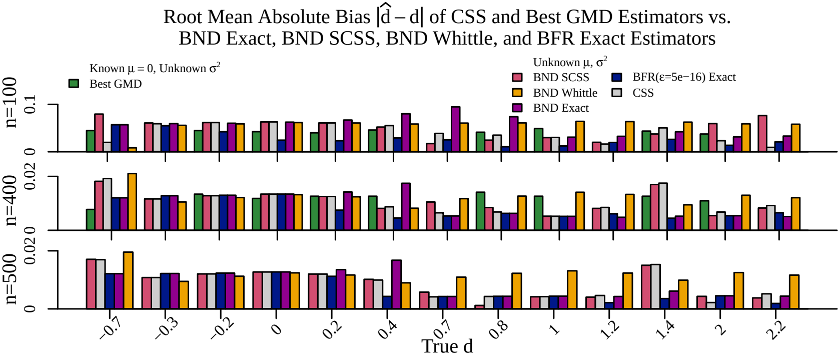
Adaptive Procedure Simulations



Adaptive Procedure Simulations

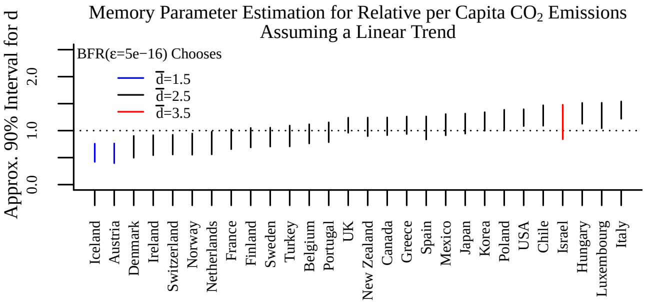


Adaptive Procedure vs. Competition



Application to CO^2 Emission Convergence

$$(1-B)^{d_c}\left(y_{tc}-\mu_c-\beta_c t\right)=z_{tc}, \quad z_{tc} \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma_c^2\right),$$



Barassi et al. (2018)

Conclusions and Future Work

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Maximizing over possibly non-stationary FI models
$\approx$ Maximizing a sequence of stationary ARFIMA models

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Maximizing over possibly non-stationary FI models
$\approx$ Maximizing a sequence of stationary ARFIMA models

Everything we have done:

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Maximizing over possibly non-stationary FI models
$\approx$ Maximizing a sequence of stationary ARFIMA models

Everything we have done:

- Extends to ARFIMA(p, d, q)

$$\phi(B)(1-B)^{d}(y_t - \mu_t) = \theta(B) z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

  - ⋆ Optimization is harder
  - ⋆ Likelihood may be multimodal

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Maximizing over possibly non-stationary FI models
$\approx$ Maximizing a sequence of stationary ARFIMA models

Everything we have done:

- Extends to ARFIMA(p, d, q)

$$\phi(B)(1-B)^d(y_t - \mu_t) = \theta(B)z_t,\; z_t \overset{i.i.d.}{\sim} \text{normal}\left(0, \sigma^2\right)$$

  - ⋆ Optimization is harder
  - ⋆ Likelihood may be multimodal
- Can be made approximate

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Barassi, M. R., N. Spagnolo, and Y. Zhao (2018). Fractional Integration Versus Structural Change: Testing the Convergence of CO_2 Emissions. Environmental and Resource Economics 71(4), 923–968.

Beran, J. (1995). Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models. Journal of the Royal Statistical Society. Series B (Methodological) 57(4), 659–672.

Box, G. E. P. and G. M. Jenkins (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden Day.

Doornik, J. A. and M. Ooms (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. Computational Statistics and Data Analysis 42(3), 333–348.

Granger, C. W. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis 1(1), 15–29.

Hosking, J. R. M. (1981). Fractional Differencing. Biometrika 68(1), 165–176.

Laboissiere, R., J. C. Letievant, E. Ionescu, P. A. Barraud, M. Mazzuca, and C. Cian (2015). Relationship between spectral characteristics of spontaneous postural sway and motion sickness susceptibility. PLoS ONE 10(12), 1–15.

Lo, C.-M., C. R. Keese, and I. Giaever (1993). Monitoring Motion of Confluent Cells in Tissue Culture. Experimental Cell Research 204, 102–109.

Mayoral, L. (2007). Minimum distance estimation of stationary and non-stationary ARFIMA processes. Econometrics Journal 10(1), 124–148.

Odaki, M. (1993). On the Invertibility of Fractionally Differenced ARIMA Processes. Biometrika 80(3), 703–709.

Velasco, C. and P. M. Robinson (2000). Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series. Journal of the American Statistical Association 95(452), 1229–1243.

Acknowledgements

Thank you to:

- Robert Gramacy for inviting me!

Acknowledgements

Thank you to:

- Robert Gramacy for inviting me!
- My coauthors, David and Gena

Acknowledgements

Thank you to:

- Robert Gramacy for inviting me!
- My coauthors, David and Gena
- Grant support:
  - Xerox PARC Faculty Research Award
  - National Science Foundation Awards 1455172, 1934985, 1940124, and 1940276
  - USAID
  - Cornell University Atkinson Center for a Sustainable Future

Acknowledgements

Thank you to:

- Robert Gramacy for inviting me!
- My coauthors, David and Gena
- Grant support:
  - Xerox PARC Faculty Research Award
  - National Science Foundation Awards 1455172, 1934985, 1940124, and 1940276
  - USAID
  - Cornell University Atkinson Center for a Sustainable Future

Conclusions and Future Work

We took a difficult problem and turned it into an easy problem!

Maximizing over possibly non-stationary FI models
$\approx$ Maximizing a sequence of stationary ARFIMA models

Everything we have done:

- Extends to ARFIMA(p, d, q)
  $$\phi(B)(1-B)^d(y_t-\mu_t)=\theta(B)z_t,\ z_t \overset{i.i.d.}{\sim} \text{normal}\left(0,\sigma^2\right)$$
  - ⋆ Optimization is harder
  - ⋆ Likelihood may be multimodal
- Can be made approximate