Aggregation and Long Memory

J. Eduardo Vera-Valdés
PhD Defence

Aarhus University and CREATES

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

November 15, 2016

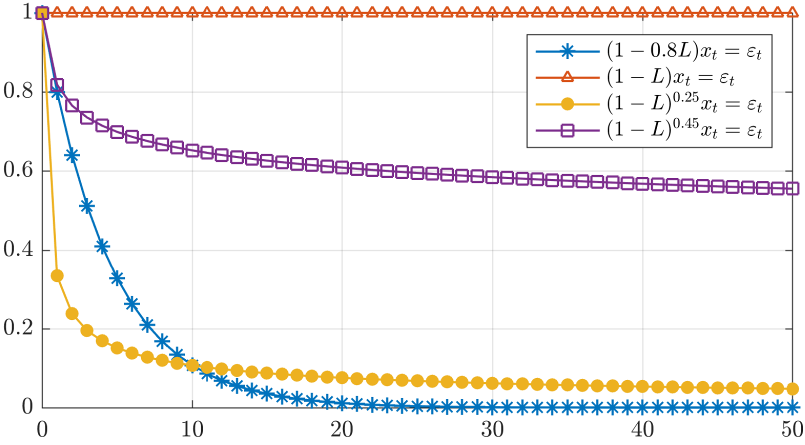
Outline

Outline

Long Memory

- Granger (1966) addressed the shape of the spectrum for economic variables. He found that their spectrum shows a pole at the origin.
- This implies long lasting correlations in the form of an hyperbolic decay instead of the standard geometric one.
- Long memory has since been detected in many time series in Economics, Finance, and other disciplines: GDP, inflation, volatility series, and river flows, to name a few.
- Its presence has implications for estimation and forecasting.

Autocorrelation Function.



Long Memory

The dissertation contributes to three branches of analysis:

- First chapter, coauthored with Niels Haldrup, deals with theoretical reasonings behind the presence of long memory. In particular, cross-sectional aggregation of heterogeneous micro units leading to long memory.
- The second chapter analyzes the forecasting performance of *ARFIMA* models when dealing with long memory governed by processes other than *ARFIMA* processes.
- The third chapter, coauthored with Daniela Osterrieder and Daniel Ventosa-Santaulària, assesses estimation of unbalanced regressions with long memory.

Outline

Cross-Sectional Aggregation

- A common motivation behind the presence of long memory in the time series data is cross-sectional aggregation, Granger (1980).

- Granger showed that if the series is the result of the cross-sectional aggregation of $AR(1)$ processes with random coefficients of a particular type, then the aggregated series would show long memory.

Cross-Sectional Aggregation

First Chapter:

- We show that the cross-sectional aggregation result extends to other definitions of long memory.
- Granger's result is asymptotic, we conduct a Monte Carlo study to analyze its finite sample properties.
- We study the limiting properties of a fractionally differenced long memory process that is generated by cross-sectional aggregation.

Definitions of Long Memory

Let x_t be a stationary time series with autocovariance function $\gamma_x(k)$ and spectral density function $f_x(\lambda)$, let $d \in (0, 1/2)$, then:

(i) x_t has long memory in the **covariance sense** if $\gamma_x(k) \approx Ck^{2d-1}$ as $k \to \infty$.

(ii) x_t has long memory in the **spectral sense** if $f_x(\lambda) \approx C\lambda^{-2d}$ as $\lambda \to 0$.

(iii) x_t has long memory in the **rate of the partial sum sense** if $\mathrm{Var}(\sum_t^T x_t) \approx CT^{1+2d}$ as $T \to \infty$.

(iv) x_t has long memory in the **self-similar sense** if $m^{1-2d}\mathrm{Cov}(x_t^{(m)}, x_{t+k}^{(m)}) \approx Ck^{2d-1}$ as $k, m \to \infty$ where $x_t^{(m)} = \frac{1}{m}(x_{tm-m+1} + \cdots + x_{tm})$ with $m \in \mathbb{N}$.

(v) x_t has long memory in the **distribution sense** if scaled partial sums converge to fractional Brownian motion.

Random Coefficient $AR(1)$ model

Consider a process defined as

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} \quad i = 1;$$

where $\varepsilon_{i,t}$ is a white noise process independent of α_i with $\mathrm{E}[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2$, $\forall t \in \mathbb{Z}$, and $\alpha_i^2 \sim \mathcal{B}(\alpha; p, q)$ with $\mathcal{B}(\alpha; p, q)$ the Beta distribution:

$$\mathcal{B}(\alpha; p, q) = \frac{1}{B(p, q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0, 1),$$

where $p, q > 1$ and $B(\cdot, \cdot)$ is the Beta function.

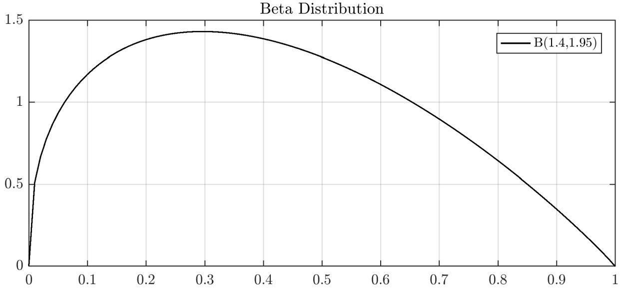
Random Coefficient $AR(1)$ model

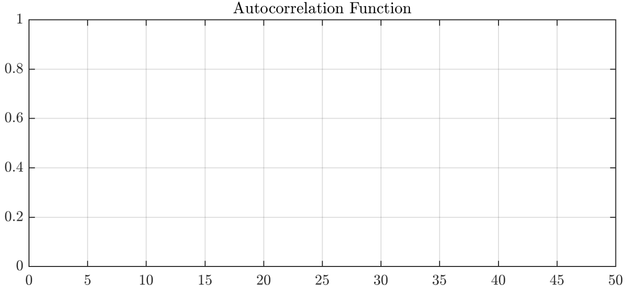
Robinson (1978) computed the second moments of this process.
Let $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}$ as before, then for $k \in \mathbb{N}$:

$$\begin{aligned}
\gamma_{x_i}(k) &:= E[x_{i,t}x_{i,t+k}] = E\left[E[x_{i,t}x_{i,t+k}|\alpha_i]\right] = \sigma_\varepsilon^2 E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right] \\
&= \sigma_\varepsilon^2 \int_0^1 \frac{x^{k/2}}{1-x}\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}dx \\
&= \sigma_\varepsilon^2 \frac{\Gamma(q-1)}{B(p,q-1)}\frac{\Gamma(p+k/2)}{\Gamma(p+k/2+q-1)}.
\end{aligned}$$

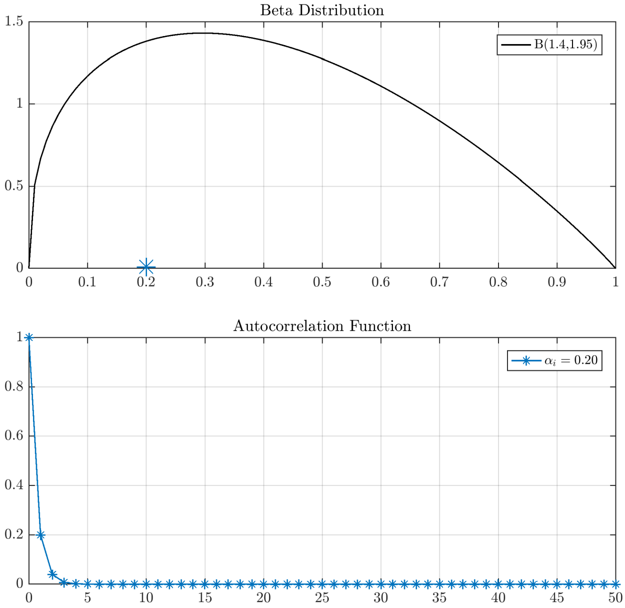
- Using Stirling's approximation, $\gamma_{x_i}(k) = C\Gamma(p+k/2)/\Gamma(p+k/2+q-1) \approx Ck^{1-q}$; thus it shows hyperbolic decaying autocorrelations.
- Nonetheless, it is not an ergodic process.

Random Coefficient $AR(1)$ model

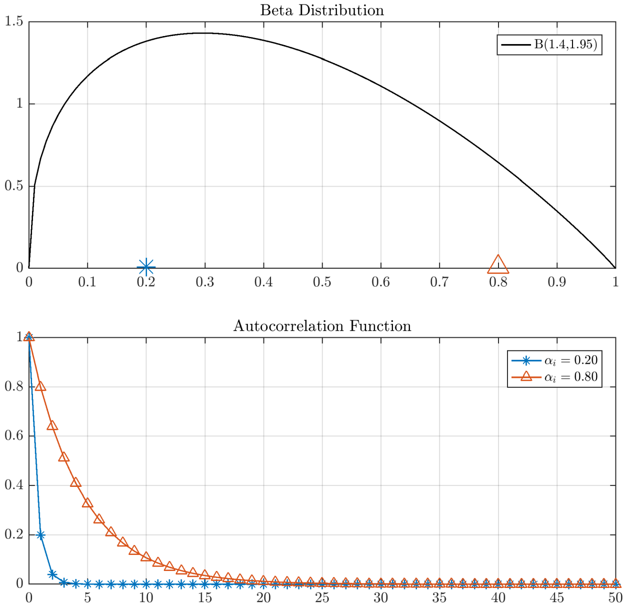




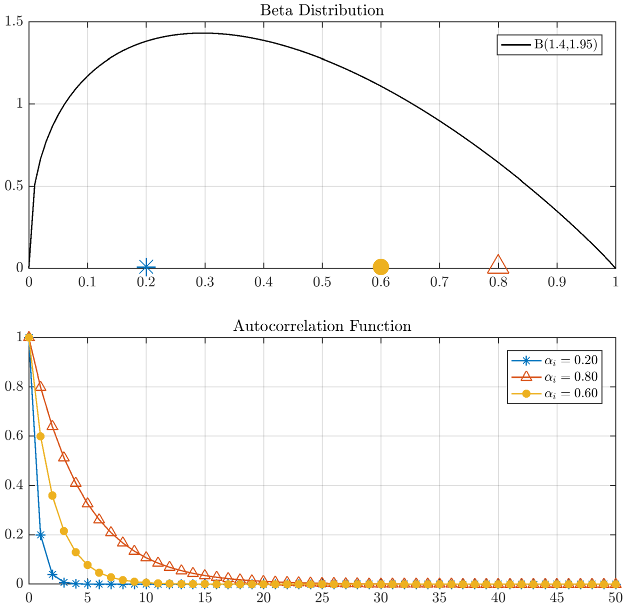
Random Coefficient $AR(1)$ model



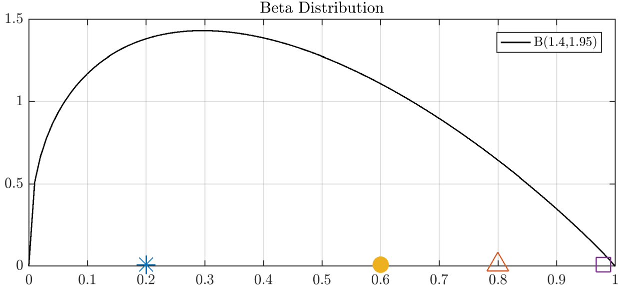
Random Coefficient $AR(1)$ model

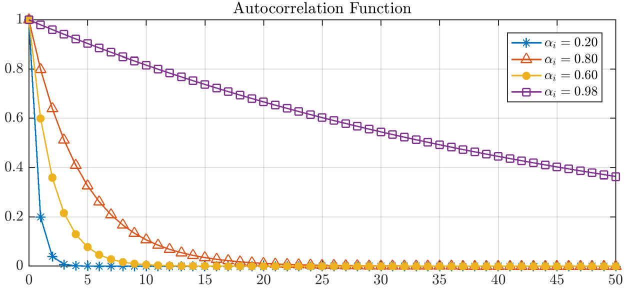


Random Coefficient $AR(1)$ model



Random Coefficient $AR(1)$ model





Cross-Sectional Aggregation

Granger considered N of those series defined as

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} \quad i = 1, 2, \cdots, N;$$

where $\varepsilon_{i,t}$ is a white noise process independent of α_i $\forall i$ with $\mathrm{E}[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2$ $\forall i \in \{1, 2, \cdots, N\}$, $\forall t \in \mathbb{Z}$, and $\alpha_i^2 \sim \mathcal{B}(\alpha; p, q)$ with $\mathcal{B}(\alpha; p, q)$ the Beta distribution:

$$\mathcal{B}(\alpha; p, q) = \frac{1}{B(p, q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0, 1),$$

where $p, q > 1$ and $B(\cdot, \cdot)$ is the Beta function.

- Furthermore, define the cross-sectional aggregated series as:

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}. \tag{1}$$

Cross-Sectional Aggregation

Granger showed that it generates long memory in the covariance sense, definition (i).
Let x_t be the cross-sectional aggregated process as before, then for $k \in \mathbb{N}$:

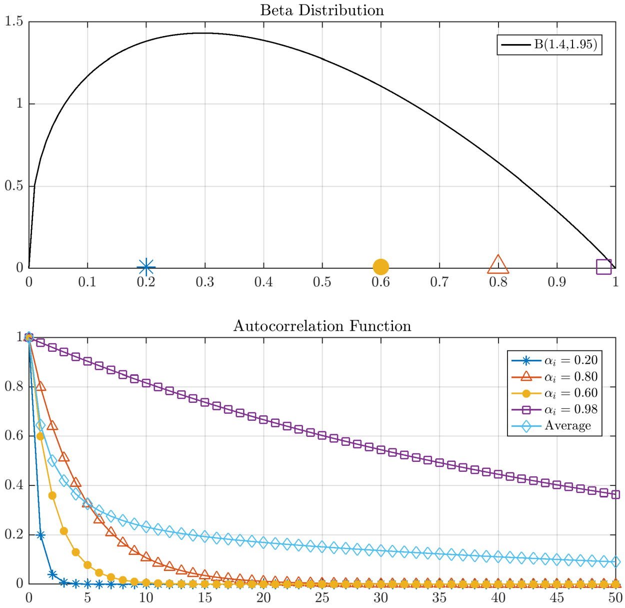
$$\begin{aligned}\gamma_x(k) &:= E[x_t x_{t+k}] = E\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_{i,t}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_{i,t+k}\right)\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} E\left[E[x_{i,t}x_{i,t+k}|\alpha_i]\right] = \frac{\sigma_\varepsilon^2}{N}\sum_{i=1}^{N} E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right].\end{aligned}$$

Now, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\alpha_i^k}{1-\alpha_i^2} \approx E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right].$$

- Considering the cross-sectional aggregated process achieves ergodicity.

Cross-Sectional Aggregation



Cross-Sectional Aggregation

We extend Granger's results to definitions (*ii*) through (*v*).

Theorem

Let x_t be defined as in (1) then, as $N \to \infty$, x_t has long memory with parameter $d = 1 - q/2$ in the sense of definitions (i) through (iv). Furthermore, if $\varepsilon_{i,t}$ is an $i.i.d.$ process, then x_t has long memory in the sense of definition (v).

Proof: In the dissertation.

Outline

Finite Sample Simulation

- Granger's result is asymptotic. We now perform a Monte Carlo simulation to examine its finite sample properties.
- We generate series under different settings with focus on three dimensions:
  - The distribution of the autocorrelation coefficient near one in the context of the implied long memory $d = 1 - q/2$.
  - The cross-sectional dimension.
  - The sample size.
- For comparison we also simulate an $FI(d)$ processes using the exact algorithm suggested by Jensen and Nielsen (2014).

Finite Sample Simulation

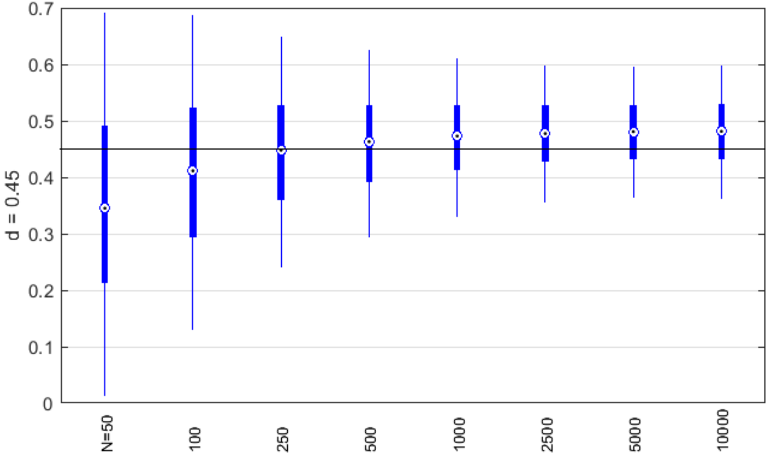
Table: Mean and standard deviation in parentheses of the estimated long memory parameter. $T = N = 10,000$; $R = 1,000$.

Theo.	Cross-sectional aggregated			$FI(d)$		
d	*GPH*	*LW*	*MLE*	*GPH*	*LW*	*MLE*
0.45	0.493	0.490	0.475	0.454	0.450	0.449
	(0.072)	(0.060)	(0.035)	(0.071)	(0.058)	(0.007)
0.35	0.393	0.393	0.420	0.347	0.349	0.349
	(0.071)	(0.058)	(0.069)	(0.072)	(0.057)	(0.008)
0.25	0.320	0.320	0.362	0.252	0.250	0.248
	(0.073)	(0.058)	(0.072)	(0.070)	(0.052)	(0.008)
0.15	0.262	0.259	0.294	0.149	0.141	0.148
	(0.074)	(0.063)	(0.087)	(0.074)	(0.059)	(0.008)

Note. The estimators considered are Geweke and Porter-Hudak (1983), *GPH*; the local Whittle estimator of Robinson (1995) and Kunsch (1986), *LW*; and the Maximum Likelihood Estimator of Sowell (1992), *MLE*; respectively.

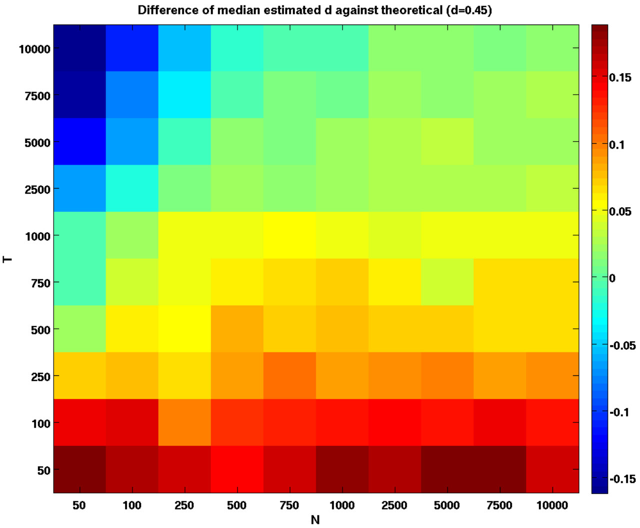
Finite Sample Simulation

Box-plots of $\widehat{d}_{GPH}$, $d = 0.45$, $T = R = 10,000$,
$N \in \{50, 100, 250, 500, 1000, 2500, 5000, 10000\}$.



Finite Sample Simulation

Heat-maps of $(\widehat{d}_{GPH} - d)$, $d = 0.45$, $R = 1,000$,
$T, N \in \{50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000\}$.



Finite Sample Simulation

- For small cross-sectional dimension and large sample size, the median is below the theoretical value in all cases.
- For smaller sample sizes, the memory appears to be exaggerated.
- This suggests that if we use aggregation to simulate long memory we need to increase the cross-sectional dimension as well as the sample size.
- In summary, it shows that the aggregation scheme to generate long memory is not as precise as fractional differencing, particularly for small degrees of memory, while being more computationally demanding.

Outline

Aggregation and *ARFIMA* processes

We are interested in assessing if fractionally differencing a cross-sectional aggregated process removes the long memory.

Let $y_t = (1 - L)^d x_t$ and note that:

$$y_t = (1 - L)^d \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (1 - L)^d x_{i,t}.$$

- In the next Theorem we obtain its autocorrelation function.

Aggregation and *ARFIMA* processes

Theorem

Let $y_t = (1-L)^d x_t$ *where* x_t *is defined as in (1) and* $\gamma_y(k) = E[y_t y_{t-k}]\ \forall k \in \mathbb{N}$ *then, as* $N \to \infty$*,*

$$\gamma_y(k) = \frac{\gamma^*(k)}{B(p,q)}\left[B(p,q-1)\left(F_1(k)-1\right) + B(p+\frac{1}{2},q-1)F_2(k)\right],$$

where $F_1(k)$ *and* $F_2(k)$ *are generalized hypergeometric functions,* $\gamma^*(k)$ *is the autocorrelation function of an* $I(-d)$ *process, and* $B(\cdot,\cdot)$ *is the Beta function.*

Proof: In the dissertation

- $$\gamma_y(k) = E[y_t y_{t+k}] = \frac{1}{N}E[E[(1-L)^d x_{i,t}(1-L)^d x_{i,t+k}|\alpha_i]],$$

  where we substitute the autocovariance of an *ARFIMA*$(1,-d,0)$ obtained by Sowell.

Aggregation and *ARFIMA* processes

Corollary

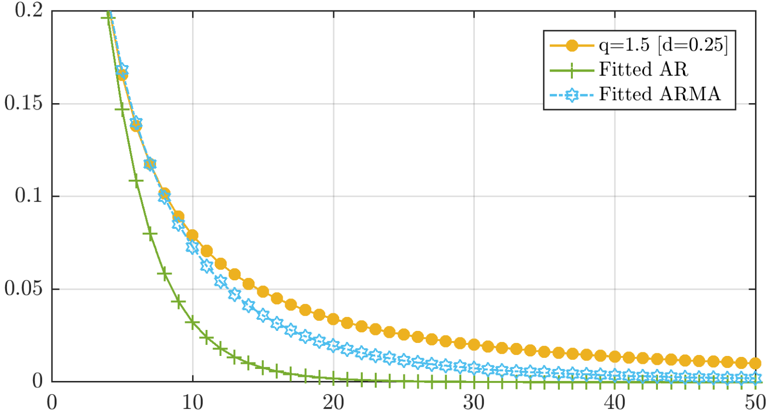
As $k \to \infty$, $\gamma_y(k) \approx \tau(k)k^{-1-2d}$, where $\tau(k)$ is a slowly-varying function in the sense that, for $c > 0$, $\lim_{k\to\infty} \tau(ck)/\tau(k) = 1$. Moreover, the autocorrelations are absolutely summable, that is, $\sum_{i=0}^{\infty} |\rho_y(k)| = \sum_{i=0}^{\infty} |\gamma_y(k)/\gamma_y(0)| < \infty$.

Proof: In the dissertation

- In particular, it has hyperbolic decaying autocorrelations; yet, it satisfies Davidson's definition of an $I(0)$ process.

Aggregation and *ARFIMA* processes

Autocovariance function for the fractionally differenced cross-sectional aggregated series and fitted short memory models.



Note: Lags were selected using the Bayesian Information Criteria given the results of Beran(1998).

Aggregation and *ARFIMA* processes

- Taking a fractional difference with the true long memory parameter helps in controlling the long memory behavior.

- Nonetheless, caution must be taken when using parametric estimation methods since the resulting process after fractional differencing is not an *ARMA* process.

Outline

Forecasting Long Memory

- As shown previously, long memory may not necessarily come from fractional differencing.

- Given the popularity of the *ARFIMA* model, we may still be interested in evaluating its forecasting performance.

Forecasting Long Memory

Chapter 2:

- Evaluate the forecasting performance of *ARFIMA* models on long memory generated by sources other than the *ARFIMA*.
- We consider the cross-sectional aggregated process and the error duration model as sources of long memory.
- We compute the Root Mean Square Error (*RMSE*) for each model and determine the Model Confidence Set (*MCS*).
- We consider forecasts horizons $h = 5, 10, 30, 50, 100, 300$.

Table: Competing Models

$FI(d)$	$HAR(3)$	$AR(22)$	$HAR(4)$	$AR(50)$	$I(1)$

- Following Ray (1993), we include high-order *AR* processes.
- The $HAR(3)$ model of Corsi(2009) is a constrained $AR(22)$ given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + \epsilon_t,$$

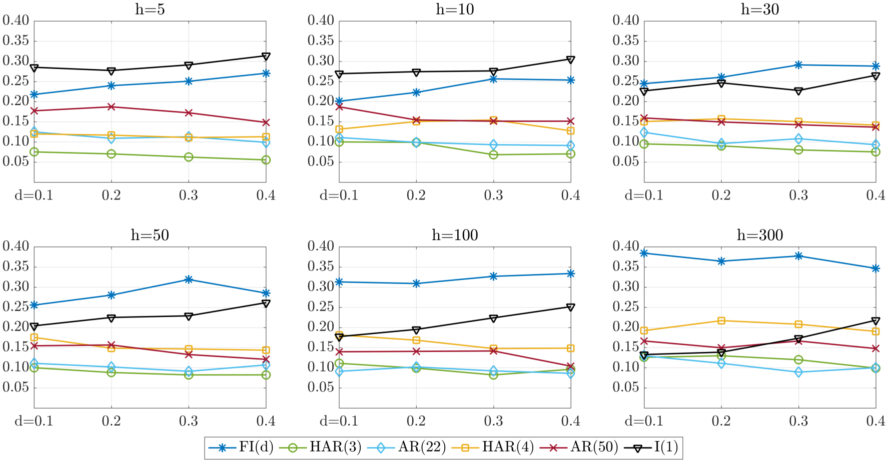
  where $x_{t-1}^{(f)} = x_{t-1}$, $x_{t-1}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i}$ and, $x_{t-1}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} x_{t-i}$.
- Analogously, define the $HAR(4)$, a constrained $AR(50)$, as

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + a_4 x_{t-1}^{(b)} + \epsilon_t,$$

  where $x_{t-1}^{(b)} = \frac{1}{50}\sum_{i=1}^{50} x_{t-i}$.

Forecasting Long Memory

Percentage number of times the model is contained in the *MCS*.



Forecasting Long Memory

- The figure shows that $FI(d)$ models are well suited for long horizon forecasts of long memory generated by cross-sectional aggregation.
- The results for the $HAR(3)$ and $AR(22)$, and $HAR(4)$ and $AR(50)$ models point to a bias-variance trade-off.
- *HAR* models are a compromise between the rigid $FI(d)$ and flexible high-order *AR* models. They incorporate high-order autoregressive specifications while greatly restricting the number of parameters to be estimated.
- This arrangement provides better long horizon forecasts than pure *AR* specifications.

Outline

Conclusions

- We show that the cross-sectional aggregation result extends to other definitions of long memory.
- In finite samples, cross-sectional aggregation is not as precise as fractional differencing, while being more computationally demanding.
- We study the limiting properties of a fractionally differenced long memory process that is generated by cross-sectional aggregation.
- We find that *ARFIMA* models are well suited for long horizon forecasts of long memory generated by processes other than the *ARFIMA*.