Long Memory, Fractional Integration, and Cross-Sectional Aggregation

Haldrup, Niels; Vera Valdés, J. Eduardo

Published in:
Journal of Econometrics

DOI (link to publication from Publisher):
10.1016/j.jeconom.2017.03.001

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Long Memory, Fractional Integration, and Cross-Sectional Aggregation*

Niels Haldrup[†] J. Eduardo Vera Valdés[‡]

March 13, 2017

Abstract

It is commonly argued that observed long memory in time series variables can result from cross-sectional aggregation of dynamic heterogenous micro units. In this paper we demonstrate that the aggregation argument is consistent with a range of different long memory definitions. A simulation study shows that the cross-section dimension needs to be rather large to reflect the theoretical memory when using commonly used methods to estimate the memory parameter, especially when the theoretical memory is not too high. We show that the aggregated process will converge to a generalized fractional process in the limit. The coefficients of the moving average representation of the series decay hyperbolically but they differ from the coefficients arising from inversion of the fractional difference filter. It appears that the fractionally differenced series will have an autocorrelation function that still exhibits hyperbolic decay, but at a rate that ensures summability. The fractionally differenced series is thus $I(0)$ but standard $ARFIMA$ modelling is invalid when the long memory is caused by aggregation. It is shown that standard methods for estimating and selecting $ARFIMA$ specifications fail in properly fitting the dynamics of the series.

Keywords: Long memory, Fractional Integration, Aggregation.

JEL classification: C2, C22.

*Author's accepted manuscript. DOI published journal article: https://doi.org/10.1016/j.jeconom.2017.03.001

[†]nhaldrup@creates.au.dk

[‡]eduardo@math.aau.dk

1 Introduction

Without specifically discussing long memory, the study of this concept in econometrics goes back to Granger (1966) in his article about the spectral shape near the origin for economic time series variables. He found that *long-term fluctuations, if decomposed into frequency components, are such that the amplitudes of the components decrease smoothly with decreasing period* (Granger, 1966, p. 155). This certainly applies for non-stationary $I(1)$ processes and more generally for the class of fractionally integrated processes as demonstrated by Granger and Joyeux (1980). Such processes have long lasting autocorrelations that decay hyperbolically instead of the standard geometric decay characterizing $ARMA$ processes.

This kind of behavior has led to several definitions of long memory. In this paper we consider five definitions of long memory.

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

*(i) in the **covariance sense** if $\gamma_x(k) \approx C_x k^{2d-1}$ as $k \to \infty$ with C_x a constant,*

*(ii) in the **spectral sense** if $f_x(\lambda) \approx C_f \lambda^{-2d}$ as $\lambda \to 0$ with C_f a constant,*

*(iii) in the **rate of the partial sum sense** if $\mathrm{Var}(\sum_{t=1}^{T} x_t) = O_p(T^{1+2d})$,*

*(iv) in the **self-similar sense** if $m^{1-2d}\mathrm{Cov}(x_t^{(m)}, x_{t+k}^{(m)}) \approx C_m k^{2d-1}$ as $k, m \to \infty$ where $x_t^{(m)} = \frac{1}{m}(x_{tm-m+1} + \cdots + x_{tm})$ with $m \in \mathbb{N}$, $m/k \to 0$, and C_m is a constant,*

*(v) in the **weak convergence sense** if $X_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{\lfloor n\xi \rfloor} x_t \Rightarrow B_H(\xi)$, where $\sigma_n^2 = \mathrm{E}[(\sum_{t=1}^{n} x_t)^2]$, $\xi \in [0, 1]$, $B_H(\xi)$ is a fractional Brownian motion, $H = d + 1/2$, and $\Rightarrow$ denotes weak convergence on $D[0, 1]$, the space of real-valued functions that are continuous from the right with finite left limits.*

Above, $g(x) \approx h(x)$ as $x \to x_0$ means that $g(x)/h(x)$ converges to 1 as x tends to x_0, $O_p(\cdot)$ denotes order in probability, and $\lfloor \cdot \rfloor$ denotes the integer value of its argument.

Definition (i) is concerned with the behavior of the autocorrelation function for long lags and was one of the motivations behind the $ARFIMA$ class of models due to Adenstedt (1974), Granger and Joyeux (1980), and Hosking (1981). Basically, they extended the $ARMA$ model to account for fractional differencing. That is, for a stationary fractional process

$$A(L)(1-L)^d x_t = B(L)\epsilon_t, \tag{1}$$

where ϵ_t is a white noise process, $d \in (-1/2, 1/2)$, and $A(L)$, $B(L)$ are polynomials in the lag operator with no common roots, all outside the unit circle. They used the standard binomial expansion to decompose $(1-L)^d$ in a series with coefficients $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$ for $j \in \mathbb{N}$. Using Stirling's approximation it can be shown that these coefficients decay at a hyperbolic rate ($\pi_j \approx j^{d-1}$ as $j \to \infty$), which translates to slowly decaying autocorrelations.

Definition (ii) is the feature considered by Granger (1966) in his study of the typical spectral shape for economic variables. The behavior of the spectrum near the origin is also used in the construction of one of the most popular estimators for long memory due to Geweke and Porter Hudak (1983) who proposed an estimation procedure based on semiparametric log periodogram regression near the zero frequency.

Diebold and Inoue (2001) based their work on spurious long memory on definition (iii). They showed that structural breaks or regime switching schemes can be confused with long memory of the fractional type by focusing on the stochastic order of the variance of partial sums. Their paper demonstrates that certain stochastic processes are long memory by one definition but not necessarily by other definitions.

Definitions (iv) and (v) are largely based on the work of Mandelbrot and Van Ness (1968) for fractals. They defined the self-similarity condition and showed that the fractional Brownian motion in particular has this property. Basically, self-similarity implies that the degree of memory is constant for different levels of temporal aggregation. Weak convergence to a fractional Brownian motion of an appropriately scaled partial sum is important for many parametric long memory models, but the class of processes is broader than often being considered as we shall later see.

It is well known that $ARFIMA$ processes are long memory by definitions (i) through (iii), and an analogous derivation as in the proof of Theorem 1 below shows that it is also long memory in the self-similar sense, definition (iv). Moreover, a scaled partial sum of an $ARFIMA$

process converges to fractional Brownian motion, see for instance Davydov (1970) and Davidson and de Jong (2000). Thus, in the time series literature the $ARFIMA$ model has become the canonical specification for modeling long memory.

Even though the $ARFIMA$ model seems to be an appropriate specification to study long memory, the source underlying its dynamic features is still not clear. Physical (turbulence, see for instance Kolmogorov (1941)), as well as psychological reasons (Pearson (1902) *personal equation*), have been used to explain the presence of long memory. More recently, Parke (1999) proposed the error-duration model which relies on a decomposition of the time series into the sum of a sequence of shocks of stochastic magnitude and duration. He shows that if only a small proportion of the errors survive for large periods of time then the resulting series shows long memory in the covariance sense, definition (i). Nonetheless, given the nature in which the data is collected, one of the main arguments often given in economics to why time series data seems to have long memory features is due to cross-sectional aggregation. It is also commonplace to see arguments for cross-sectional aggregation motivating the presence of fractional long memory in real data.

Granger (1980), in line with the results of Robinson (1978) on random $AR(1)$ models, showed that cross-sectional aggregation of $AR(1)$ processes with random coefficients could produce long memory. Assuming a Beta distribution for the generation of cross-sectional $AR(1)$ coefficients, he showed that, as the cross-sectional dimension goes to infinity, the autocovariance function exhibits a slow hyperbolic decay, rather than the standard geometric decay characterizing $ARMA$ processes. Thus, cross-sectional aggregation of dynamic micro units can produce long memory in the covariance sense under certain conditions.

In this paper we focus on some features of the aggregation argument leading to long memory. We address the particular specification considered by Granger because the Beta distribution is a rather flexible specification that allows closed-form solutions but the analysis can be extended to other aggregation schemes as well. Zaffaroni (2004) shows that Granger's result applies to a broader class of distributions to which the Beta distribution belongs. We demonstrate that this aggregation scheme implies that the aggregated series is long memory using all the definitions considered in this paper.

Since the aggregation result is an asymptotic property we conduct a Monte Carlo simulation study to quantify how aggregation can lead to long memory in finite samples. The theoretical degree of memory of the aggregated series is tied to a particular parameter of the Beta distribution which affects the density mass around one. The simulations show that the cross-sectional dimension has to be rather large for the theoretical degree of memory to apply, while the time series dimension needs to be large to obtain a precise estimator. Finite samples of the series will still exhibit long memory but the estimated memory parameter can be rather large compared to its theoretical value, especially when the memory is only of moderate degree.

In the third part of the paper, we focus on the extent to which the memory implied by aggregation can be removed by fractional differencing. In particular, we are interested in how $ARFIMA$ type of long memory models can be useful for practical model building for the class of processes considered. It occurs that fractionally differencing the series, using the theoretical degree of memory, does remove the long memory of the process. The resulting series has absolutely summable autocorrelations and thus it is $I(0)$ by the definition of Davidson (2009). However, the fractionally differenced series will still have autocorrelations that decay hyperbolically and hence will decay slower than what an $ARMA$ specification will be able to fit. This feature is most dominant when the degree of memory is moderate as opposed to being close to non-stationarity. Our findings have implications for the argument that is often given for estimating $ARFIMA$ models in applications, namely that the observed long memory of time series can occur due to cross-sectional aggregation. A simulation study shows that fitted $ARFIMA$ models will generally be inadequate to fit the dynamics of the underlying process.

The paper is structured as follows. In section 2, the Granger aggregation scheme is presented and the features of the aggregated series are examined using the different long memory definitions that we consider. Section 3 presents the simulation study, and finally section 4 derives the features of fractional differencing of cross-sectionally aggregated long memory processes. The final section concludes.

2 Long Memory and Cross-Sectional Aggregation

Consider the random $AR(1)$ process given by:

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}, \tag{2}$$

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 $p, q > 1$ and $\mathcal{B}(\alpha; p, q)$ is the Beta distribution with density:

$$\mathcal{B}(\alpha; p, q) = \frac{1}{B(p,q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0,1), \tag{3}$$

where $B(\cdot,\cdot)$ is the Beta function.

Robinson (1978) showed that the process given by (2) admits a variance-covariance stationary solution. Furthermore, the unconditional autocorrelation function of this process shows hyperbolic decay. However, the process is not ergodic in the sense that random samples will depend on the realization of α_i.

Granger (1980) proposed[1] to consider the cross-sectional aggregation of the series specified in (2) which we here define as:

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}, \tag{4}$$

where $\{\alpha_i\}_{i=1}^n$ are $i.i.d.$ independent of $\{\varepsilon_{i,t}\}_{i\in\{1,2,\cdots,N\}}^{t\in\mathbb{Z}}$.

Note that considering (4) instead of (2) solves the ergodicity violation by eliminating the dependence of the autocorrelation function on the particular realization of the autoregressive coefficient. Intuitively, if N is large enough, samples from (4) will have similar realizations of $\{\alpha_i\}_{i=1}^N$ and thus will have similar autocorrelation functions.

Granger showed that, as $N \to \infty$, the autocorrelations of x_t decay at a hyperbolic rate and hence generate long memory in the covariance sense according to definition (i) with parameter $d = 1 - q/2$. Taking $q \in (1,2)$, the long memory generated falls in the stationary range,

[1] Granger (1980) also considered the case with dependence across the series and allowing for different variances across the cross-sectional units. He demonstrated that as long as a possible common component across the micro units has a memory smaller than $q/2$, then the aggregation results will remain unaltered. For ease of exposition, we will thus focus on the scenario under independence and equal variances since generalizations will not affect our general conclusion.

$d \in (0, 1/2)$. We will focus on this range for the rest of the analysis.

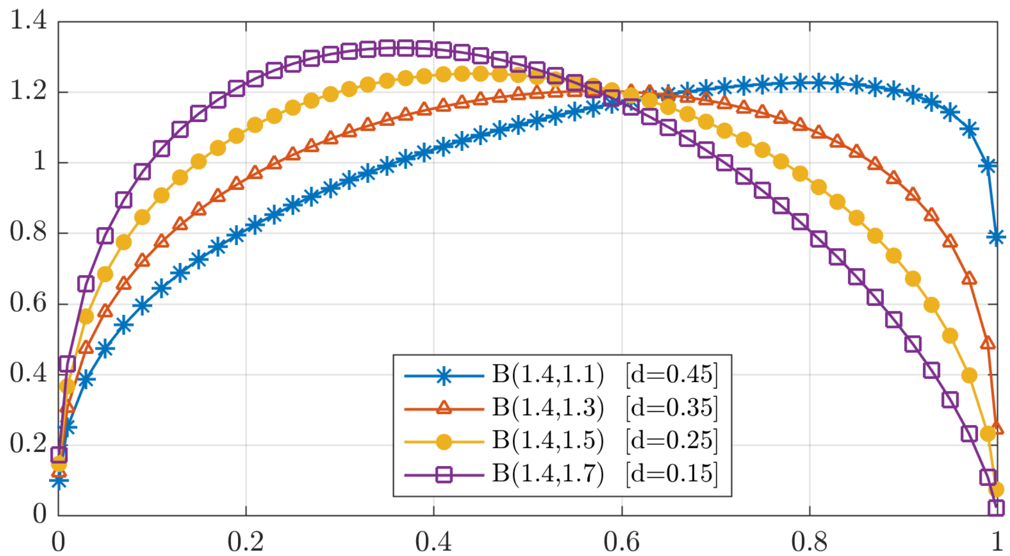
In Theorem 1, we extend the long memory result to definitions (ii) through (iv).

Theorem 1. *Let x_t be defined as in (4) then, as $N \to \infty$, x_t has long memory with parameter $d = 1 - q/2$ in the sense of definitions (i) through (iv).*

Proof: See appendix.

Theorem 1 shows that a cross-sectional aggregated series of infinitely many $AR(1)$ processes with squared autoregressive coefficients drawn from a Beta distribution has long memory with long memory parameter $d = 1 - q/2$. Note that the parameters p, q are shape parameters of the Beta distribution. In particular, q affects the density around one and thus the probability of aggregating over near unit-root $AR(1)$ processes. It appears that the value of p plays no role for this result as $N \to \infty$. As a consequence, Granger conjectured (and it was later confirmed by Zaffaroni (2004)) that asymptotically the memory only depends on the behavior of the distribution of the autoregressive coefficient near one. In Figure 1, we plot the beta distribution (3) for $p = 1.4$ and different values of q. As can be seen, the closer q is to one, the more density mass concentrates around one; which, as shown in Theorem 1, translates to a larger degree of memory in the cross-sectionally aggregated series, x_t.

Figure 1: The Beta distribution.



More generally, Zaffaroni (2004) showed that if the distribution of the autoregressive coefficient, α_i, belongs to a family of absolutely continuous distributions on $[0, 1)$, depending upon a real parameter $b \in (-1, \infty)$, with density

$$G(\alpha; b) \sim c_b(1-\alpha)^b \quad \text{as} \quad \alpha \to 1^-,$$

where $0 < c_b < \infty$, then the aggregated series, letting $N \to \infty$, will be long memory. Moreover, the more dense the distribution of α_i is around one, the greater the degree of long memory of the aggregate. Both the Uniform and Beta distributions are members of this family of distributions. Thus, the specific parametric assumption regarding the distribution of the autoregressive coefficient is not needed for the long memory result to apply, but as we will see below, the Beta distribution allows us to obtain closed-form expressions for one of the main results in the paper. Additionally, Zaffaroni (2004) extended the result for cross-sectional aggregation to general $ARMA$ processes of finite order without affecting the general conclusions.

Intuitively, the autocorrelations of the cross-sectional aggregated process can be seen as a weighted average of the autocorrelations of the $AR(1)$ processes, and, a higher proportion of high-persistent processes translates into a higher persistence in the aggregated series and the implied degree of memory.

In Theorem 1, we showed that cross-sectional aggregation satisfies long memory by definitions (i) through (iv). We now argue that under the additional condition that $\varepsilon_{i,t}$ is $i.i.d.$, the scaled partial sum of cross-sectional aggregated series converges weakly to fractional Brownian motion; that is, it has long memory in the sense of definition (v).

$ARFIMA$ processes are fractional differenced $ARMA$ processes after adopting the $(1-L)^d$ filter. The MA series resulting from inversion of the $(1-L)^d$ filter has hyperbolically decaying coefficients of the form $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$ for $j \in \mathbb{N}$ and this produces a series with hyperbolic decaying autocovariances according to Stirling's approximation $\Gamma(j+d)/\Gamma(j+1) \approx j^{d-1}$ for large j. We can generalize this construction to series that still show hyperbolic decaying coefficients, yet, the coefficients do not come from inversion of the fractional difference operator as defined above. We call these processes generalized fractional processes (see Davidson and de Jong (2000)).

We prove in Lemma 1 that under the stated conditions, the cross-sectional aggregated processes can be expressed as a generalized fractional process.

Lemma 1. *Let x_t be defined as in (4) and assume that $\varepsilon_{i,t}$ is an i.i.d. process, then, as $N \to \infty$, x_t can be expressed as*

$$x_t = \sum_{j=0}^{\infty} \phi_j \nu_{t-j},$$

where $\nu_j \sim N(0, \sigma_\varepsilon^2)$ are independent and $\phi_j = \left(B(p+j,q)/B(p,q)\right)^{1/2}$, $\forall j \in \mathbb{N}$.

Proof: See appendix.

Lemma 1 relies on the fact that when N goes to infinity the Central Limit Theorem can be applied. In this sense, it is in line with the work of Davidson and Sibbertsen (2005) who show that cross-sectional aggregated non-linear processes of appropriate form have linear representations in the sense of having $MA(\infty)$ representations. Note also that in Lemma 1 we could obtain a similar result if $\varepsilon_{i,t}$ is not *i.i.d.* but satisfies Lyapunov's condition. Furthermore, the resulting series inherits the uncorrelated property of $\varepsilon_{i,t}$ and, given normality, they are independent.

By Stirling's approximation the coefficients in the representation decay at a hyperbolic rate, $\phi_j \approx j^{-q/2} = j^{d-1}$ as $j \to \infty$ with $d = 1 - q/2$, but ϕ_j are not associated with the fractional differencing parameters, π_j, defined above. Thus, cross-sectional aggregated processes are generalized fractional processes. In Section 4, we will detail the study of the relationship between cross-sectional aggregated long memory processes and $ARFIMA$ processes.

Theorem 2 is a consequence of Theorem 4.6 in Beran et al. (2013).

Theorem 2. *Let x_t be defined as in (4) and assume that $\varepsilon_{i,t}$ is an i.i.d. process, then, as $N \to \infty$, x_t has long memory in the sense of definition (v) with parameter $d = 1 - q/2$.*

This result is in line with the findings of Zaffaroni (2004) when restricting the analysis to the Beta distribution which allows us to find closed-form solutions for the variance terms. This in turn translates into closed-form expressions for the coefficients of the generalized fractional process. Given this, Theorem 2 also follows directly from the developments of Davydov (1970) and Davidson and de Jong (2000).

In summary, Theorems 1 and 2 show that a cross-sectional aggregated series has long memory by all the definitions considered. However, although the coefficients of the MA representation decay hyperbolically, they are different from those arising from inversion of a fractional difference filter.

3 Finite Sample Study

In order to analyze the finite sample properties of Granger's aggregation result, which holds asymptotically, we conducted a Monte Carlo simulation experiment. Note that if we do not consider enough $AR(1)$ processes in the cross-sectional dimension, the resulting series may not have long memory as predicted theoretically. Granger (1990) proposed a division between cross-sectional aggregation in small scale, involving sums of a few time series variables, and large scale, involving the sums of very many variables. In particular, Chambers (1998) shows that when the number of variables is not large, the aggregation result cannot be obtained.

We generate x_t as in (4) under different parametric settings along three dimensions: the density of the autoregressive coefficient near one determined by the parameter q; the sample size T; and the cross-sectional dimension N.

The simulation proceeds as follows for R replications:

- Sample the N autoregressive coefficients from the density function, equation (3).
- Generate the individual $AR(1)$ series of size T, equation (2), using the sampled coefficients, with $\varepsilon_{i,t} \overset{i.i.d.}{\sim} N(0,1)$.
- Aggregate the individual series cross-sectionally according to equation (4).
- Estimate the long memory parameter by the GPH estimator of Geweke and Porter Hudak (1983); the local Whittle estimator of Robinson (1995) and Künsch (1986), LW; and the (Quasi)Maximum Likelihood Estimator of Sowell (1992), $QMLE$.

Note that both the GPH and LW estimators do not depend on a full parametric specification whereas $QMLE$ assumes an $ARFIMA$ specification. The importance of this will be made clearer in Section 4 when discussing the relationship between $ARFIMA$ processes and the processes under study.

Throughout, we use a bandwidth of $T^{0.5}$ when implementing the GPH and LW estimators as it is standard practice in the literature. As it is well known, the bandwidth affects the bias-precision trade-off. Results with different bandwidths are available upon request. Moreover, for reasons of space we present simulations for $p = 1.4$ throughout so that the density for the autoregressive coefficient takes the form shown in Figure 1. For robustness we have tried different values of p, available upon request, with similar qualitative results despite minor quantitative differences. Regarding the QML estimator, we select the $ARFIMA$ specification using the BIC-criterion given the results of Beran et al. (1998) on the validity of information criteria for long memory processes.

To analyze the importance of the density around one on the aggregation result, we report in Table 1 the results from the simulations for different values of q in (3) which is related to the degree of long memory $d = 1 - q/2$. As a reference point we also simulate $FI(d)$ series using the exact algorithm of Jensen and Nielsen (2014).

The table shows that for large degrees of memory, the estimates are close to their theoretical values but rather distant when the memory is low. In general, all memory estimates are biased upwards. Both the GPH and the LW estimators appear to provide rather similar estimates whereas the QML estimator performs worst which is also seen by the relatively poor coverage probabilities of this estimator. As we shall later see, the $ARFIMA$ model is misspecified under the given conditions. The conclusion is that the distribution of the autoregressive coefficient at the disaggregated level plays a key role for N and T as large as 10,000.[2] The simulations suggest that using cross-sectional aggregation as a way to simulate long memory works poorly when the memory index d is low. In contrast, Table 1 shows that when data is generated in accordance with a fractionally integrated process with a comparable memory parameter, then all estimators do an excellent job. In particular, MLE, which estimates a correctly specified model in this case, is performing especially well as one might expect, and with an overall larger coverage probability compared to the semi-parametric estimators.[3]

[2]Note that even though asymptotically we require both $T, N \to \infty$, the cross-sectional and time dimensions are not tied together. $N \to \infty$ is needed to determine the limiting memory degree and $T \to \infty$ is needed for the estimator to be consistent.

[3]Simulations not reported here show that we need a cross-sectional dimension N and a sample size T of more than $100,000$ to obtain results for the GPH and LW estimators that mimic those of data generated according to a $FI(d)$ process. This reflects that N needs to be large to reflect the theoretical memory. And T needs to be

Table 1: Mean and standard deviation (in parenthesis) of the estimated long memory parameter for cross-sectionally aggregated processes and $FI(d)$ processes with $T = N = 10,000$ based on $R = 10,000$ replications. In brackets the coverage probability at a 95% confidence level are reported.

Theoretical memory	Cross-sectional aggregated			$FI(d)$		
d	GPH	LW	$QMLE$	GPH	LW	MLE
0.45	0.4935	0.4899	0.4746	0.4544	0.4497	0.4487
	(0.0722)	(0.0595)	(0.0353)	(0.0706)	(0.0576)	(0.0073)
	[0.8756]	[0.8384]	[0.5083]	[0.9217]	[0.9227]	[0.9296]
0.35	0.3928	0.3925	0.4202	0.3470	0.3489	0.3495
	(0.0707)	(0.0577)	(0.0689)	(0.0723)	(0.0567)	(0.0079)
	[0.8482]	[0.8335]	[0.4055]	[0.9316]	[0.9019]	[0.9386]
0.25	0.3204	0.3199	0.3622	0.2515	0.2502	0.2484
	(0.0731)	(0.0578)	(0.0719)	(0.0703)	(0.0521)	(0.0080)
	[0.7483]	[0.6866]	[0.3262]	[0.9277]	[0.9346]	[0.9326]
0.15	0.2625	0.2593	0.2943	0.1494	0.1415	0.1482
	(0.0737)	(0.0628)	(0.0867)	(0.0740)	(0.0585)	(0.0077)
	[0.5632]	[0.4163]	[0.2243]	[0.9118]	[0.9068]	[0.9495]

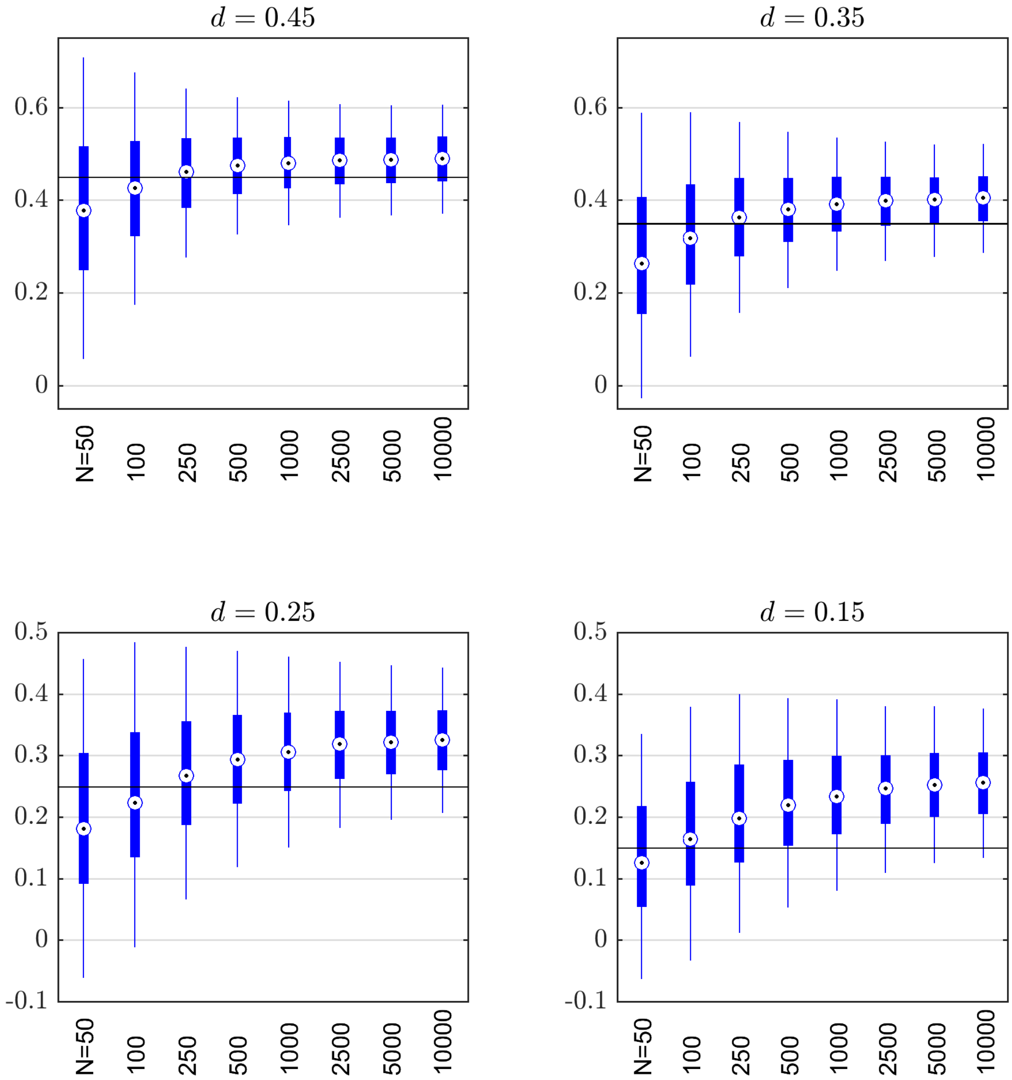
Note. The estimators considered are the semi-parametric estimator of Geweke and Porter Hudak (1983), GPH; the local Whittle estimator of Robinson (1995) and Künsch (1986), LW; and the (Quasi)Maximum Likelihood Estimator of Sowell (1992), $(Q)MLE$; respectively. Note that for data generated according to a $FI(d)$ process the likelihood is correctly specified.

Moving on to analyze the importance of the cross-sectional dimension, we present in Figure 2 box-plots from simulations with a sample size of $T = 10,000$ while varying the cross-sectional dimension N. For ease of exposition we only present results for the four theoretical degrees of long memory using the GPH estimator.

Figure 2 allows us to see how the long memory parameter evolves while increasing the cross-sectional dimension. It further shows the dependence of the result on the shape of the Beta distribution and the implied theoretical memory d. The larger the degree of memory (the denser the Beta distribution around one) the better we can approximate the asymptotic result. For small values of N the figures show that the median is below the theoretical value in all cases, which is line with the result by Chambers (1998) on small scale aggregation. It can also be seen that the memory parameter is generally imprecisely estimated when N is relatively small.

large for estimators to be precise and consistent.

Figure 2: Box-plot of the GPH estimator for different levels of aggregation. $T = 10,000$ observations and $R = 10,000$ replications. In each box the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the 95% coverage assuming symmetry.

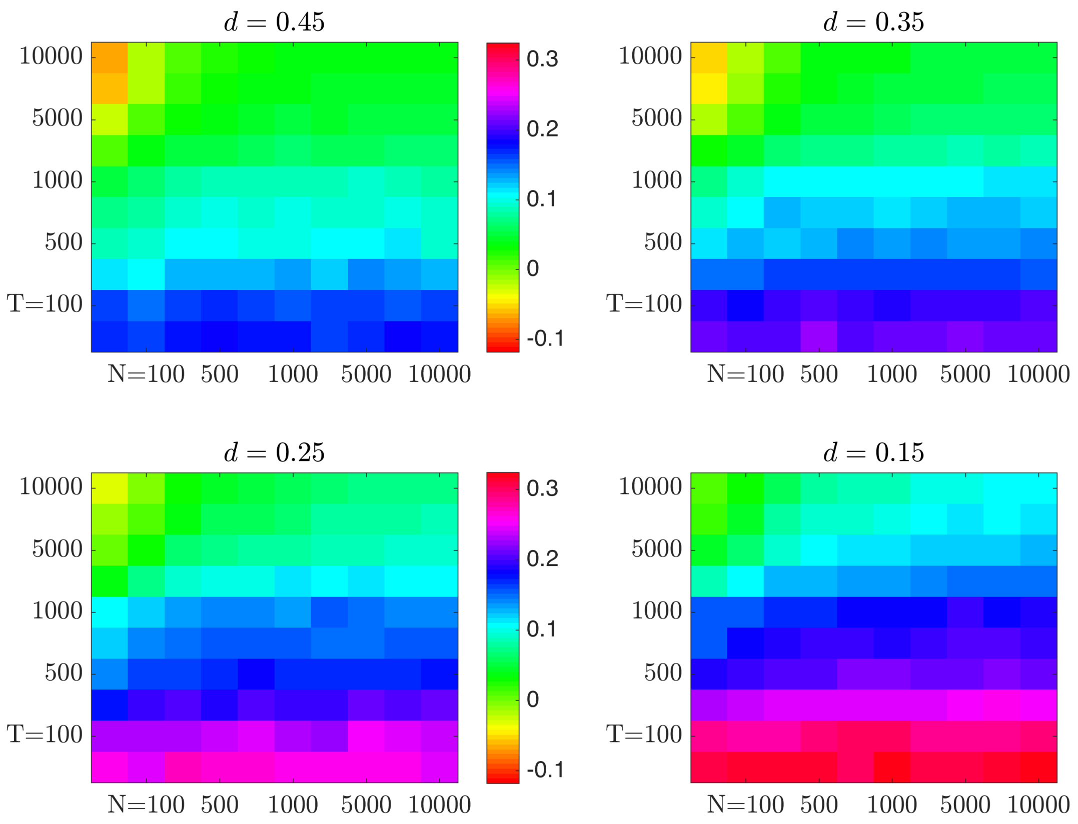


Moreover, the box-plots show that the cut-off between small and large scale aggregation varies with the density of the autoregressive coefficients. In general, with a sample size of 10,000, for larger degrees of memory, we need at least 250 $AR(1)$ series so that the median of the simulations is close to the theoretical values, while for smaller degrees of memory, as Table 1 showed, we are still far away even with $N = 10,000$ micro units. Moreover, estimation uncertainty is still significant in all cases.

Finally, to study the interaction between the sample size and the cross-section dimension,

Figure 3 presents the heat-maps of the mean of the GPH estimates in deviations from their theoretical values $(\hat{d} - d)$, while varying T and N. As noted previously, we need $T \to \infty$ for the estimator to be asymptotically valid and we need $N \to \infty$ to ensure a degree of memory in accordance with the theory. $T, N \to \infty$ thus plays separate roles. To construct the figures, $R = 1,000$ replications were considered and again we consider four theoretical values of $d \in \{0.45, 0.35, 0.25, 0.15\}$ corresponding to $q \in \{1.1, 1.3, 1.5, 1.7\}$.

Figure 3: Heat-map of the mean of $(\hat{d} - d)$ for the GPH estimator with $R = 1000$ replications, $T, N \in \{50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000\}$.



The figure shows that for smaller sample sizes we are always overshooting the true long memory parameter. This suggests that when working with a small sample size, the estimators do not have enough information to discern the true nature of the process in terms of memory. On the other hand, as the sample size T increases, more cross-sectional units are needed to approximate the asymptotic result: the estimator is becoming more precise, but more cross-sectional units are needed to eliminate the estimation bias centered around the true asymptotic memory. This quantifies the cut-off between small and large scale aggregation. The simulations indicate that if we were to use aggregation as a way to simulate long memory, we need to increase the cross-sectional dimension significantly, and the time dimension needs to be large as well for the estimator to be sufficiently precise.

4 Cross-Sectional Aggregation and $ARFIMA$ processes

Theorems 1 and 2 together with Lemma 1 show that cross-sectional aggregated processes share key properties with $ARFIMA$ processes. Both processes satisfy all of the definitions of long memory considered in this paper and both have $MA(\infty)$ representations with hyperbolic decaying coefficients.

These shared properties may explain why several authors have assumed that cross-sectional aggregated processes are of the $ARFIMA$ type. For instance, Balcilar (2004) and Gadea and Mayoral (2006) refer to cross-sectional aggregation as a possible explanation behind long memory found in inflation and fit $ARFIMA$ models using parametric methods.

Granger (1980), in his original article, also noted that although aggregated series were not $ARFIMA$, the $ARFIMA$ specification could provide a good approximation.

Others have suggested that the long memory of the cross-sectional aggregated series can be eliminated by fractional differencing. Diebold and Rudebusch (1989) allude to aggregation as the origin of long memory in output. They estimate the long memory parameter by the GPH method, fractionally difference the series, and subsequently estimate an $ARMA$ model. Kumar and Okimoto (2007) refer to aggregation as the motive behind long memory and use the Shimotsu and Phillips (2005) estimator for the long memory parameter. This method relies on fractional differencing.

Recall from (1) that an $ARFIMA$ process is a fractionally differenced $ARMA$ process. Thus, if we were to take a d-th difference, $(1-L)^d$, of an $ARFIMA(a, d, b)$ process we would recover the underlying $ARMA(a, b)$ process. However, as Lemma 1 shows, the cross-sectional aggregated process is a generalized fractional process. Thus, it may not appear from fractional differencing. As a way to give an answer to this question, Theorem 3 presents the autocovariance function of a fractionally differenced cross-sectionally aggregated process.

Theorem 3. *Let $y_t = (1-L)^d x_t$ where x_t is defined as in (4) with $N \to \infty$ and $\gamma_y(k) = E[y_t y_{t-k}]$ $\forall k \in \mathbb{N}$. Then,*

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

where

$$\gamma^*(k) = \sigma_\varepsilon^2 \frac{\Gamma(1+2d)}{\Gamma(-d)\Gamma(1+d)} \frac{\Gamma(-d-k)}{\Gamma(1+d-k)},$$

is the autocovariance function of an $I(-d)$ process with innovations with variance σ_ε^2 and

$$\begin{aligned}
F_1(k) \;&:=\; F\left[\left\{1, p, \frac{1-d+k}{2}, \frac{-d+k}{2}\right\}, \left\{p+q-1, \frac{2+d+k}{2}, \frac{1+d+k}{2}\right\}, 1\right] + \\
&\; F\left[\left\{1, p, \frac{1-d-k}{2}, \frac{-d-k}{2}\right\}, \left\{p+q-1, \frac{2+d-k}{2}, \frac{1+d-k}{2}\right\}, 1\right], \\
F_2(k) \;&:=\; \frac{-d+k}{1+d+k} * \\
&\; F\left[\left\{1, p+\frac{1}{2}, \frac{1-d+k}{2}, \frac{2-d+k}{2}\right\}, \left\{p+q-\frac{1}{2}, \frac{2+d+k}{2}, \frac{3+d+k}{2}\right\}, 1\right] \\
&\; + \frac{-d-k}{1+d-k} * \\
&\; F\left[\left\{1, p+\frac{1}{2}, \frac{1-d-k}{2}, \frac{2-d-k}{2}\right\}, \left\{p+q-\frac{1}{2}, \frac{2+d-k}{2}, \frac{3+d-k}{2}\right\}, 1\right],
\end{aligned}$$

where $F[\cdot]$ is the generalized hypergeometric function.

Proof: See appendix.

Two main points can be drawn from Theorem 3.

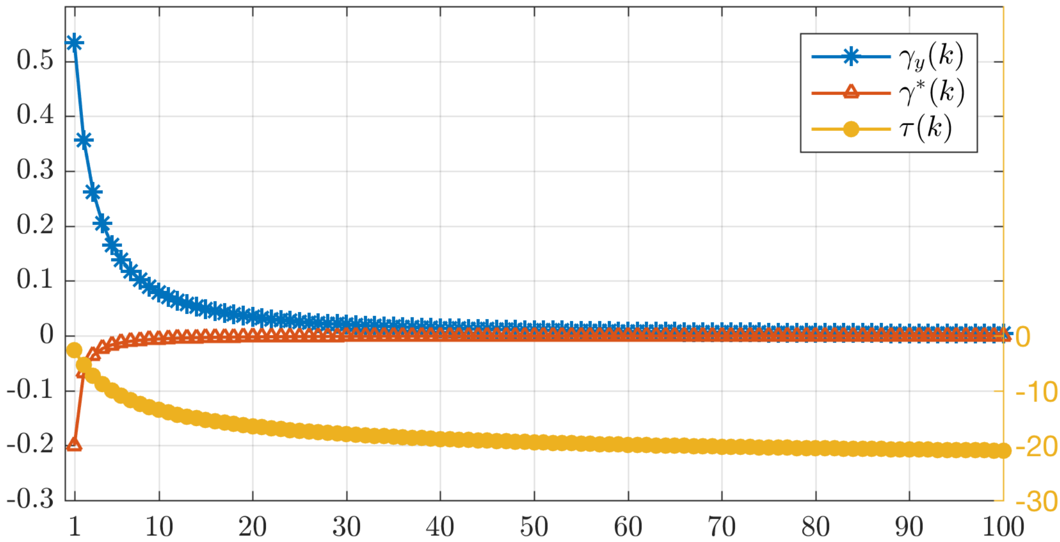
First, looking at the resulting autocovariance function, after fractional differencing, we find that it retains some memory even for large lags. In particular, it does not belong to the class of autocovariance functions for linear $ARMA$ processes. This has implications for modeling and estimation. In particular, Maximum Likelihood estimators rely on the fact that the resulting series after differencing is of the $ARMA$ type.

Second, as the proof of Theorem 3 shows, we are calculating the autocovariance function of cross-sectionally aggregated $ARFIMA(1, -d, 0)$ series. Hence, the individual series are antipersistent with parameter $-d$ and the cross-sectionally aggregated AR processes are overdifferenced.

The autocovariance function of the overdifferencing filter $(1-L)^d$ is given by $\gamma^*(k)$ in Theorem 3 which is a negative function in k.

Figure 4 displays the shape of the autocorrelation function for the fractionally differenced cross-sectionally aggregated process $\gamma_y(k)$, the autocorrelation of the antipersistent component $\gamma^*(k)$, and its ratio $\tau(k) := \gamma_y(k)/\gamma^*(k)$.

Figure 4: Autocorrelation function for the fractionally differenced cross-sectionally aggregated series $\gamma_y(k)$, the $I(-d)$ process $\gamma^*(k)$ (left scale), and its ratio $\tau(k)$ (right scale). $p = 1.4$, $q = 1.5$ so that $d = 0.25$.



The following Corollary shows that the function $\tau(k)$ is a negative slowly varying function in k and thus the autocovariance of the fractionally differenced cross-sectionally aggregated process shows hyperbolic decay.

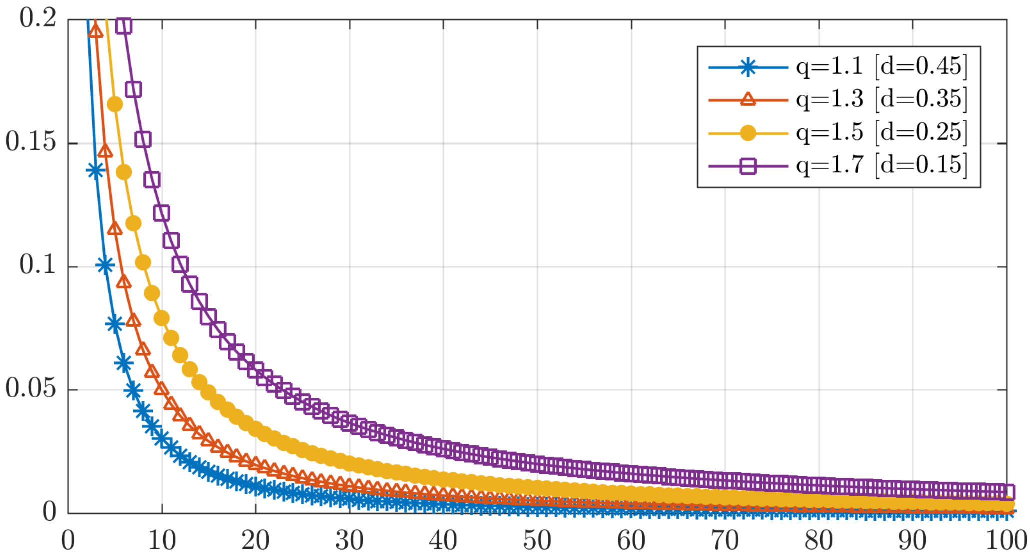
Corollary 1. *As $k \to \infty$, $\gamma_y(k) \approx \tau(k)k^{-1-2d}$, where $\tau(k)$ is a slowly-varying function in the sense that, for $c > 0$, $\lim_{k\to\infty} \tau(ck)/\tau(k) = 1$. Moreover, the autocorrelations are absolutely summable, that is, $\sum_{i=0}^{\infty} |\rho_y(k)| = \sum_{i=0}^{\infty}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

Proof: See appendix.

As demonstrated in Figure 4 and proved in Corollary 1, the autocovariance function $\gamma_y(k)$ decays at a hyperbolic rate similar to the rate for antipersistent processes. However, the sign of the function is positive as opposed to antipersistent processes, which is a feature induced by the cross-sectional aggregation. Despite the hyperbolic rate, the decay is still fast in the sense that the autocorrelations are summable and hence satisfy the condition for $I(0)$ described by Davidson (2009).

Note from the expression of $\gamma_y(k)$ given in Theorem 3 that autocovariances for finite k depend on both of the parameters p and q associated with the Beta distribution. The high frequency dynamics are thus much more dependent upon the particular aggregation scheme under consideration whereas the low frequency dynamics and long memory of the aggregated process allows a much broader class of aggregation schemes. Figure 5 displays the autocorrelation functions for $p = 1.4$ and $q \in \{1.2, 1.4, 1.6, 1.8\}$. Small values of q (and hence large memory) result in relatively small autocorrelations for finite k. As q increases, and hence memory declines, the fractionally differenced series tend to have rather significant autocorrelations for small as well as for moderately large lags.[4] This will clearly have a major impact on the properties of estimated parametric models of the $ARFIMA$ type which in general will be misspecified.

Figure 5: Autocorrelation functions for the fractionally differenced cross-sectionally aggregated series, $\gamma_y(k)$, for $p = 1.4$ and $q \in \{1.1, 1.3, 1.5, 1.7\}$.

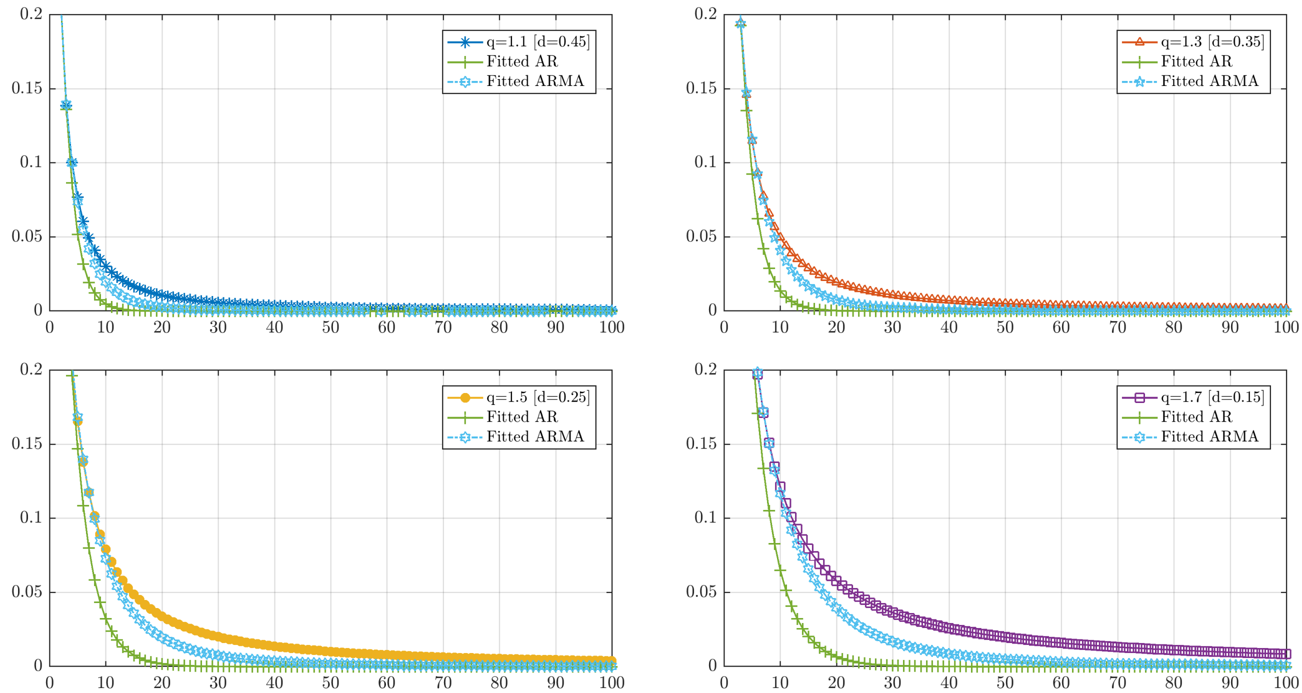


For illustration and to get an idea about this misspecification problem, Figure 6 displays the autocorrelation functions for the fractionally differenced cross-sectionally aggregated processes (i.e., those in Figure 5) together with fitted short memory models. By (Quasi)Maximum Like-

[4] We also constructed graphs similar to Figure 5 while varying p. They show that the autocorrelations increase in size as p increases.

lihood we estimate both pure AR and $ARMA$ processes for each degree of memory. We select the number of lags using the Bayesian Information Criterion (BIC), see Beran et al. (1998), and allow initially a maximum of 100 lags for the pure AR specification and a maximum of 4 lags for either the AR and MA components of the $ARFIMA$ specification.

Figure 6: Autocorrelation functions for the fractionally differenced cross-sectionally aggregated series, $\gamma_y(k)$, for $p = 1.4$ and $q \in \{1.1, 1.3, 1.5, 1.7\}$, and the average of fitted AR and $ARMA$ models estimated by QML.



For a sample of 1,000 fractionally differenced cross-sectionally aggregated series, the BIC consistently selected either 3 or 4 lags for the pure AR, and typically an $ARMA(2, 1)$ for the $ARMA$ alternative.

As the figure shows, the short memory models capture the first autocorrelations well, particularly the $ARMA$ specification. Nevertheless, for longer autocorrelations the fitted short memory models start to diverge from the theoretical ones. Moreover, the discrepancy increases as the theoretical long memory becomes smaller.

In Table 2 we compute the number of lags necessary for the autocorrelations to be below certain values approaching zero. The table shows that the $ARMA$ specification takes almost the same number of lags to fall below 10^{-1}, with the pure AR a close second. However, the table

Table 2: Number of lags needed for the autocorrelation function to fall below the given threshold value.

Memory d		ACF threshold value			
		10^{-1}	10^{-2}	10^{-3}	10^{-4}
0.45	Fitted AR	4	9	14	20
	Fitted $ARMA$	4	14	27	48
	Theoretical	5	22	87	331
0.35	Fitted AR	5	11	18	25
	Fitted $ARMA$	6	19	39	60
	Theoretical	6	32	143	> 500
0.25	Fitted AR	7	15	23	32
	Fitted $ARMA$	8	27	62	102
	Theoretical	9	51	258	> 500
0.15	Fitted AR	9	19	29	41
	Fitted $ARMA$	12	38	91	108
	Theoretical	13	88	> 500	> 500

shows that both short memory alternatives fall below 10^{-2} almost twice as fast as the theoretical ones, and the divergence gets exacerbated from there. In the best setting, $d = 0.45$, it takes more than 300 lags for the theoretical autocorrelations to fall below 10^{-4}, while the short memory alternatives fall below that value before 50 lags. This of course is not surprising given the geometric rate of decline of the autocorrelation function for the AR and $ARMA$ specifications.

5 Conclusions

In many empirical studies, long memory is modeled as $ARFIMA$ processes and often the motivation used in this research relies on the Granger (1980) argument that cross-sectional aggregation can lead to long memory. In this paper, we argue that both $ARFIMA$ processes and long memory processes generated according to Granger's aggregation scheme satisfy a range of long memory definitions. Despite these similarities, the two classes of processes have features that are somewhat different. First of all, one should be aware that cross-sectional aggregation leading to long memory is an asymptotic feature that applies for the cross-sectional dimension tending to infinity. In finite samples and for moderate cross-sectional dimensions the observed memory of the series can be rather different from the theoretical memory. Moreover, the aggregation result seems to be most apparent when the memory tends to be relatively high, and hence the distribution of the individual $AR(1)$ micro units has concentrated mass near but strictly less than one. Secondly, we have shown that when taking a fractional difference of a cross-sectionally aggregated long memory process, then the resulting process is not an $ARMA$ process. The fractionally differenced process has autocorrelations that are summable and the process is $I(0)$ according to Davidson's (2009) definition, but the autocorrelations still decay at a hyperbolic rate rather than a geometric rate. Especially when the memory is moderate the autocorrelations are more persistent than observed for $ARMA$ processes. Granger (1980) noted that cross-sectional aggregated long memory processes are likely to be well approximated as $ARFIMA$ processes in most cases. Our study shows that care should be taken regarding this common belief. In many cases, $ARFIMA$ specifications will not provide a satisfactory description of the short run dynamics even though the long memory can be effectively removed by fractional differencing.

To derive closed-form expressions of our results, we assumed that the $AR(1)$ coefficients are sampled from a Beta distribution. Nonetheless, Zaffaroni (2004) showed that the qualitative results apply to a broader class of distributions. Moreover, even though we considered a simplified noise structure, it follows from Granger 's (1980) analysis that our general results apply when weak assumptions about the dependence across time and cross-sectional dimensions are allowed for.

Acknowledgments

The authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

We appreciate helpful comments from James Davidson, Javier Hualde, and Peter Robinson, as well as a Referee and an Associate Editor of the Journal.

A Appendix

Proof of Theorem 1

Let x_t be defined as in (4).

To prove (i), note that x_t has zero mean and thus its variance is given by

$$\begin{aligned}\gamma_x(0) = E[x_t^2] &= E\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_{i,t}\right)^2\right] = \frac{1}{N}E\left[\left(\sum_{i=1}^{N} x_{i,t}\right)^2\right] \\ &= \frac{\sigma_\varepsilon^2}{N}\sum_{i=1}^{N} E\left[\frac{1}{1-\alpha_i^2}\right],\end{aligned}$$

where the third equality follows from the independence assumption.

Note that $\forall i \in \{1, 2, \cdots, N\}$, unconditionally,

$$E\left[\frac{1}{1-\alpha_i^2}\right] = \int_0^1 \frac{1}{1-\alpha^2}\mathcal{B}(\alpha)d\alpha = \int_0^1 \frac{x^{p-1}(1-x)^{q-2}}{B(p,q)}dx = \frac{B(p,q-1)}{B(p,q)},$$

which shows that each series has long memory in the covariance sense.

As previously discussed, (2) is not ergodic in the sense that realizations depend on the draw of α_i. To solve the ergodicity violation we consider the cross-sectional aggregated series noting that,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{1-\alpha_i^2} = \int_0^1 \frac{1}{1-\alpha^2}\mathcal{B}(\alpha)d\alpha$$

regardless of the conditioning on the autoregressive coefficients.

As for the autocovariances, similar calculations show that

$$\gamma_x(k) = E[x_t x_{t-k}] = \frac{\sigma_\varepsilon^2}{N}\sum_{i=1}^{N} E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right] = \sigma_\varepsilon^2\frac{B(p+k/2,q-1)}{B(p,q)},$$

for $k \in \mathbb{N}$, which, by Stirling's approximation,

$$\gamma_x(k) = \sigma_\varepsilon^2\frac{B(p+k/2,q-1)}{B(p,q)} = \sigma_\varepsilon^2\frac{\Gamma(q-1)}{B(p,q)}\frac{\Gamma(p+k/2)}{\Gamma(p+q+k/2-1)} \approx \sigma_\varepsilon^2\frac{\Gamma(q-1)}{B(p,q)}k^{1-q}.$$

So the aggregated series shows hyperbolic decaying autocovariances[5] $\gamma_x(k) \approx C_x k^{1-q}$ and is long memory in the covariance sense with parameter $d = 1 - q/2$.

To prove (ii), note that given the autocorrelation function $\rho_x(k) = \gamma_x(k)/\gamma_x(0)$ with $\gamma_x(k), \gamma_x(0)$ computed above, Theorem 1.3 in Beran et al. (2013) shows that the spectral density has a pole at the origin.

To prove (iii),

$$
\begin{aligned}
Var\left(\sum_{t=1}^{T} x_t\right) &= E[(x_1 + x_2 + \cdots + x_T)^2] \\
&= E\left[x_1^2 + \cdots + x_T^2 + 2(x_1 x_2 + \cdots + x_{T-1} x_T)\right] \\
&= TE[x_1^2] + 2E\left[\left(\sum_{t=2}^{T} x_1 x_t + \cdots + \sum_{t=T-1}^{T} x_1 x_t\right)\right] \\
&= TE[x_1^2] + 2\left((T-1)E[x_1 x_2] + \cdots + E[x_1 x_T]\right) \\
&= T\gamma_x(0) + 2\left((T-1)\gamma_x(1) + \cdots + \gamma_x(T-1)\right) \\
&= O_p(T^{3-q}) = O_p(T^{1+2d}),
\end{aligned}
$$

where in the last line we have used the asymptotic behavior for $\gamma_x(\cdot)$ calculated in (i).

Finally, to prove (iv), we need to analyze the series while considering temporal aggregation. Let $m \in \mathbb{N}$ and define

$$x_i^{(m)} = \frac{1}{m}(x_{im-m+1} + \cdots + x_{im}),$$

for $i = \{1, 2, \cdots\}$. That is, let $x_i^{(m)}$ be a temporal aggregation of x_t at level m. Then, note that $\forall t \in \mathbb{N}$ and for large $k \in \mathbb{N}$

$$
\begin{aligned}
E[x_t^{(m)} x_{t+k}^{(m)}] &= \frac{1}{m^2} E[(x_{tm-m+1} + \cdots + x_{tm})(x_{(t+k)m-m+1} + \cdots + x_{(t+k)m})] \\
&= \frac{1}{m^2} E[\underbrace{x_{tm-m+1} x_{(t+k)m-m+1} + \cdots + x_{tm-m+1} x_{(t+k)m} + \cdots + x_{tm} x_{(t+k)m}}_{m^2 \text{ terms}}] \\
&= \frac{1}{m^2}\left(\underbrace{\gamma_x(km) + \cdots + \gamma_x(km+m-1) + \cdots \gamma_x(km)}_{m^2 \text{ terms}}\right).
\end{aligned}
$$

[5]Note that the result relies upon calculating $E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right]$. The parametric assumption we make regarding the Beta distribution allows us to obtain closed-form expression for these terms. If we relax the parametric assumption and assume a broader class of distributions as in Zaffaroni (2004), we would obtain the same hyperbolic rate of decay. The same argument applies for definitions (ii) through (iv).

Factorizing terms and replacing $\gamma_x(|j-i|)$ for its asymptotic behavior calculated in (i),

$$\begin{aligned} E[x_t^{(m)} x_{t+k}^{(m)}] &= \frac{1}{m^2}\left(\gamma_x(km-m+1)+\cdots+m\gamma_x(km)+\cdots+\gamma_x(km+m-1)\right) \\ &\approx \frac{C_x}{m^2}\left((km-m+1)^{1-q}+m(km)^{1-q}+\cdots+(km+m-1)^{1-q}\right), \end{aligned}$$

dividing both sides by k^{1-q},

$$\begin{aligned} \frac{1}{k^{1-q}} E[x_t^{(m)} x_{t+k}^{(m)}] &\approx \frac{C_x}{m^2}\left(m^{1-q}+\cdots+mm^{1-q}+\cdots+m^{1-q}\right) \\ &= \frac{C_x}{m^2}\left(1+\cdots+m+\cdots+1\right) m^{1-q} \\ &= \frac{C_x}{m^2} m^2 m^{1-q} = C_x m^{1-q}, \end{aligned}$$

where in the first line we used that $m/k \to 0$ as $k \to \infty$.

Thus, with $d = 1 - q/2$, $m^{1-2d}\mathrm{Cov}(x_t^{(m)}, x_{t+k}^{(m)}) \approx C k^{2d-1}$ as $k, m \to \infty$, $m/k \to 0$.

Proofs of Lemma 1 and Theorem 2

Let x_t be defined as in (4). Using the infinite series representation of each $AR(1)$ process defined as in (2) note that x_t can be written as

$$x_t = \sum_{j=0}^{\infty}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i^j \varepsilon_{i,t-j}\right).$$

Given the additional assumption on $\varepsilon_{i,t-j}$, the classical Central Limit Theorem holds sideways and thus, $\forall j \in \mathbb{N}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i^j \varepsilon_{i,t-j} \sim \mathbb{N}(0, \sigma_\varepsilon^2 B(p+j,q)/B(p,q)),$$

We have used analogous derivations as in the proof above to obtain the variance terms. Note in particular that, in contrast to the proofs of Zaffaroni (2004), the parametric assumption on the distribution of the autoregressive coefficient allows us to obtain closed-form expressions for these terms.

The above suggests an infinite series representation for the aggregated process of the form

$$x_t = \sum_{j=0}^{\infty} \phi_j \nu_{t-j},$$

where $\nu_j \sim N(0, \sigma_\varepsilon^2)$ and $\phi_j = (B(p+j,q)/B(p,q))^{1/2}$, $\forall j \in \mathbb{N}$. Note that ν_j inherits the white noise properties of $\varepsilon_{i,t-j}$. Moreover, given Stirling's approximation, the coefficients show a hyperbolic rate of decay with parameter $d = 1 - q/2$, that is, $\phi_j \approx j^{-q/2} = j^{d-1}$ as $j \to \infty$.

Once we have proved that the cross-sectional aggregated series can be expressed as a generalized fractional process, Theorem 2 is a direct consequence of Theorem 4.6 in Beran et al. (2013).

Proof of Theorem 3 and Corollary 1

Let $y_t = (1-L)^d x_t$ where x_t is defined as before, then

$$\begin{aligned} E[y_t^2] &= E\left[\left((1-L)^d x_t\right)^2\right] = E\left[\left((1-L)^d \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}\right)^2\right] \\ &= E\left[\frac{1}{N}\left(\sum_{i=1}^{N} (1-L)^d x_{i,t}\right)^2\right] = \frac{1}{N} E\left[\sum_{i=1}^{N} \left((1-L)^d x_{i,t}\right)^2\right], \end{aligned}$$

where the last equality is due to independence across units. Note that the term $(1-L)^d x_{i,t}$ is an ARFIMA(1,$-d$,0); thus the variance of y_t depends on the expected value of the $AR(1)$ coefficient of an $ARFIMA(1, -d, 0)$ process.

Let $\gamma_i(k) = E\left[(1-L)^d x_{i,t} (1-L)^d x_{i,t-k}\right]$ be the autocovariance function of $(1-L)^d x_{i,t}$. From Sowell (1992) it follows that for $k \in \mathbb{N}$

$$\gamma_i(k|\alpha_i) = \gamma^*(k) \frac{1}{1-\alpha_i^2} \left(F[\{-d+k, 1\}, 1+d+k; \alpha_i] + F[\{-d-k, 1\}, 1+d-k; \alpha_i] - 1\right),$$

where

$$\gamma^*(k) = \sigma_\varepsilon^2 \frac{\Gamma(1+2d)}{\Gamma(-d)\Gamma(1+d)} \frac{\Gamma(-d-k)}{\Gamma(1+d-k)},$$

is the autocovariance function of an $I(-d)$ process with innovations with variance σ_ε^2 and $F[\cdot]$

is the hypergeometric function.

Thus,

$$\begin{aligned}
\gamma_y(k) &= E\left[\gamma_i(k|\alpha_i)\right] \\
&= E\left[\frac{\gamma^*(k)}{1-\alpha_i^2}\left(F[\{-d+k,1\},1+d+k;\alpha_i]+F[\{-d-k,1\},1+d-k;\alpha_i]-1\right)\right] \\
&= \frac{\gamma^*(k)}{B(p,q)}\left[\int_0^1 (1-x)^{q-2}x^{p-1}F[\{-d+k,1\},1+d+k;x^{\frac{1}{2}}]dx+\right. \\
&\left.\int_0^1 (1-x)^{q-2}x^{p-1}F[\{-d-k,1\},1+d-k;x^{\frac{1}{2}}]dx-\int_0^1 (1-x)^{q-2}x^{p-1}dx\right] \\
&= \frac{\gamma^*(k)}{B(p,q)}\left[B(p,q-1)\left(F_1(k)-1\right)+B(p+\frac{1}{2},q-1)F_2(k)\right],
\end{aligned}$$

where

$$\begin{aligned}
F_1(k) &:= F\left[\left\{1,p,\frac{1-d+k}{2},\frac{-d+k}{2}\right\},\left\{p+q-1,\frac{2+d+k}{2},\frac{1+d+k}{2}\right\},1\right]+ \\
&\quad F\left[\left\{1,p,\frac{1-d-k}{2},\frac{-d-k}{2}\right\},\left\{p+q-1,\frac{2+d-k}{2},\frac{1+d-k}{2}\right\},1\right] \\
F_2(k) &:= \frac{-d+k}{1+d+k}* \\
&\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d+k}{2},\frac{2-d+k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d+k}{2},\frac{3+d+k}{2}\right\},1\right] \\
&\quad +\frac{-d-k}{1+d-k}* \\
&\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d-k}{2},\frac{2-d-k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d-k}{2},\frac{3+d-k}{2}\right\},1\right].
\end{aligned}$$

Note that in the calculations above we have used

$$\begin{aligned}
\int_0^1 F[\{a,1\},b;x^{\frac{1}{2}}]x^{p-1}(1-x)^{q-2}dx &= \int_0^1\left[\sum_{i=0}^{\infty}\frac{(a)_i}{(b)_i}x^{\frac{i}{2}}\right]x^{p-1}(1-x)^{q-2}dx \\
=\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\int_0^1 x^{p-1+\frac{i}{2}}(1-x)^{q-2}dx\right] &= \sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}B\left(p+\frac{i}{2},q-1\right)\right].
\end{aligned}$$

Now,

$$\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}B\left(p+\frac{i}{2},q-1\right)\right] = \sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\frac{\Gamma(p+\frac{i}{2})\Gamma(q-1)}{\Gamma(p+q-1+\frac{i}{2})}\right]$$

$$
\begin{aligned}
&= \Gamma(q-1)\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\frac{\Gamma(p+\frac{i}{2})}{\Gamma(p+q-1+\frac{i}{2})}\right] \\
&= \Gamma(q-1)\left(\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}}{(b)_{2i}}\frac{\Gamma(p+i)}{\Gamma(p+q-1+i)}\right]+\right. \\
&\qquad \left.\sum_{i=0}^{\infty}\left[\frac{(a)_{2i+1}}{(b)_{2i+1}}\frac{\Gamma(p+\frac{1}{2}+i)}{\Gamma(p+q-\frac{1}{2}+i)}\right]\right) \\
&= \Gamma(q-1)\left(\frac{\Gamma(p)}{\Gamma(p+q-1)}\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}}{(b)_{2i}}\frac{(p)_i}{(p+q-1)_i}\right]+\right. \\
&\qquad \left.\frac{\Gamma(p+\frac{1}{2})}{\Gamma(p+q-\frac{1}{2})}\sum_{i=0}^{\infty}\left[\frac{(a)_{2i+1}}{(b)_{2i+1}}\frac{(p+\frac{1}{2})_i}{(p+q-\frac{1}{2})_i}\right]\right) \\
&= B\left(p,q-1\right)\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}(p)_i}{(b)_{2i}(p+q-1)_i}\right]+ \\
&\qquad B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}\sum_{i=0}^{\infty}\left[\frac{(a+1)_{2i}(p+\frac{1}{2})_i}{(b+1)_{2i}(p+q-\frac{1}{2})_i}\right] \\
&= B\left(p,q-1\right)\sum_{i=0}^{\infty}\left[\frac{(\frac{a}{2})_i(\frac{a+1}{2})_i(p)_i}{(\frac{b}{2})_i(\frac{b+1}{2})_i(p+q-1)_i}\right]+ \\
&\qquad B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}\sum_{i=0}^{\infty}\left[\frac{(\frac{a+1}{2})_i(\frac{a+2}{2})_i(p+\frac{1}{2})_i}{(\frac{b+1}{2})_i(\frac{b+2}{2})_i(p+q-\frac{1}{2})_i}\right] \\
&= B\left(p,q-1\right)f_1+B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}f_2,
\end{aligned}
$$

where

$$
\begin{aligned}
f_1 &= F\left[\left\{1,p,\frac{a}{2},\frac{a+1}{2}\right\},\left\{p+q-1,\frac{b}{2},\frac{b+1}{2}\right\},1\right], \\
f_2 &= F\left[\left\{1,p+\frac{1}{2},\frac{a+1}{2},\frac{a+2}{2}\right\},\left\{p+q-1,\frac{b+1}{2},\frac{b+2}{2}\right\},1\right],
\end{aligned}
$$

$(z)_i := \frac{\Gamma(z+i)}{\Gamma(z)}$ is the Pochhammer symbol, and noting that $(a)_{2i} = (\frac{1}{2})^{-2i}(\frac{a}{2})_i(\frac{a+1}{2})_i$, $i \in \mathbb{N}$.

For the corollary note that $\gamma_y(k)$ can be written as

$$
\begin{aligned}
\gamma_y(k) &= \frac{\gamma^*(k)}{B(p,q)}\left[-B(p,q-1)+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d+k+i)\Gamma(1+d+k)}{\Gamma(-d+k)\Gamma(1+d+k+i)}\right)B(p+i/2,q-1)\right. \\
&\qquad \left.+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d-k+i)\Gamma(1+d-k)}{\Gamma(-d-k)\Gamma(1+d-k+i)}\right)B(p+i/2,q-1)\right].
\end{aligned}
$$

Let

$$\tau(k) \;:=\; \frac{1}{B(p,q)}\left[-B(p,q-1)+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d+k+i)\Gamma(1+d+k)}{\Gamma(-d+k)\Gamma(1+d+k+i)}\right)B(p+i/2,q-1)\right.$$
$$\left.+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d-k+i)\Gamma(1+d-k)}{\Gamma(-d-k)\Gamma(1+d-k+i)}\right)B(p+i/2,q-1)\right],$$

and note that, by Stirling's approximation, for large k and $c>0$, $\Gamma(1+d+ck)/\Gamma(-d+ck)\approx (ck)^{1+2d}$, $\Gamma(-d+ck+i)\Gamma(1+d+ck+i)\approx (ck)^{-1-2d}$ and analogous approximations for the terms in the second series show that

$$\tau(ck) \;\approx\; \frac{1}{B(p,q)}\left[-B(p,q-1)+2\sum_{i=0}^{\infty}B(p+i/2,q-1)\right].$$

This, in turn, shows that $\lim_{k\to\infty}\tau(ck)/\tau(k)=1$.

Hence, for large k,

$$\gamma_y(k)=\tau(k)\gamma^*(k)\approx\tau(k)k^{-1-2d},$$

where $\lim_{k\to\infty}\tau(ck)/\tau(k)=1$.

Finally, note that $\sum_{i=0}^{\infty}|\rho_y(k)|=\sum_{i=0}^{\infty}|\gamma_y(k)/\gamma_y(0)|\approx\sum_{i=0}^{\infty}k^{-1-2d}=\zeta(-1-2d)$ where $\zeta(z)$ is the Euler-Riemann zeta function which converges for $z<1$.

References

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics*, 2(6):1095–1107.

Balcilar, M. (2004). Persistence in inflation: does aggregation cause long memory? *Emerging Markets Finance and Trade*, 40(5):25–56.

Beran, J., Bhansali, R. J., and Ocker, D. (1998). On unified model selection for stationary and non stationary short and long memory autoregressive processes. *Biometrika*, 85(4):921–934.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long Memory Processes*. Springer Verlag.

Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, 39(4):1053–1072.

Davidson, J. (2009). When is a time-series I(0)? In Castle, J. L. and Shephard, N., editors, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pages 322–342. Oxford University Press Oxford.

Davidson, J. and de Jong, R. M. (2000). The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes. *Econometric Theory*, 16(5):643–666.

Davidson, J. and Sibbertsen, P. (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of Econometrics*, 128(2):253–282.

Davydov, Y. (1970). The Invariance Principle for Stationary Processes. *Theory of Probability and its applications*, 15(3):487–499.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Diebold, F. X. and Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24(2):189–209.

Gadea, M. D. and Mayoral, L. (2006). The Persistence of Inflation in OECD Countries: A Fractionally Integrated Approach. *International Journal of Central Banking*, 2:1–55.

Geweke, J. and Porter Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Granger, C. W. (1966). The Typical Spectral Shape of an Economic Variable. *Econometrica*, 34(1):150–161.

Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227–238.

Granger, C. W. (1990). Aggregation of time series variables: a survey. In Barker, T. and Pesaran, M., editors, *Disaggregation in econometric modelling*, pages 17–34. Routledge, London.

Granger, C. W. and Joyeux, R. (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.

Jensen, A. N. and Nielsen, M. Ø. (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis*, 35(5):428–436.

Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Doklady Akademii Nauk Sssr*, 30(4):299–303.

Kumar, M. S. and Okimoto, T. (2007). Dynamics of persistence in international inflation rates. *Journal of Money, Credit, and Banking*, 39(6):1457–1479.

Künsch, H. (1986). Discrimination between Monotonic Trends and Long-Range Dependence. *Journal of Applied Probability*, 23(4):1025–1030.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437.

Parke, W. (1999). What is fractional integration? *The Review of Economics and Statistics*, 81(4):632–638.

Pearson, K. (1902). On the mathematical theory of errors of judgment, with special reference to the personal equation. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character*, 198:235–299.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5(3):163–168.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, 23(5):1630–1661.

Shimotsu, K. and Phillips, P. C. (2005). Exact local Whittle estimation of fractional integration. *The Annals of statistics*, 33(4):1890–1933.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165–188.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120(1):75–102.