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

Niels Haldrup† J. Eduardo Vera Valdés‡

Abstract

It is commonly argued that observed long memory in time series and financial 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. In a simulation study we show however that both the cross-section and time dimensions have to be rather large to reflect the true implied memory when using commonly used estimators, especially when the theoretical memory is not too high. Finally, we show that even though the aggregated process will converge to a generalized fractional Brownian motion in the limit, the fractionally differenced series will still have an autocorrelation function that exhibits hyperbolic decay, but at a rate that still ensures summability. The fractionally differenced series is thus $I(0)$ but standard $ARFIMA$ modelling may be invalid when the long memory is caused by aggregation.

Keywords: Long memory, Fractional Integration, Aggregation.

JEL classification: C2, C22.

*Corresponding address: Department of Economics and Business Economics, Aarhus University and CREATES. Fuglesangs Allé 4, 8210 Aarhus V, Denmark.

†nhaldrup@creates.au.dk

‡evera@creates.au.dk

1 Introduction

Without specifically talking about 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 correlations that decay hyperbolically instead of the standard geometric decay characterizing $ARMA$ processes.

This kind of behavior, along similar findings in other scientific areas, has given rise to several definitions of long memory. In this study, following Guégan (2005), 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 $\mathrm{Var}(\sum_t^T x_t) = O_p(T^{1+2d})$ as $T \to \infty$*

(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 $m \in \mathbb{N}$ and C_m a constant*

(v) in the **distribution sense** *if $X_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{[n\xi]} x_t \xrightarrow{d} B_H(\xi)$, 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 $\xrightarrow{d}$ denotes convergence in distribution.*

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 regression around 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 by focusing on the rate at which the variance of partial sums grows in time. 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.

Finally, definition (*i*), concerned with the behavior of the autocorrelation function for large lags, was one of the motivations behind the $ARFIMA$ model due to Adenstedt (1974), Granger and Joyeux (1980), and Hosking (1981). They extended the $ARMA$ model to account for fractional differencing. That is, for a stationary fractional process

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

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 in turn translates to slowly decaying autocorrelations.

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 this has become the canonical construction 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 the 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. Using 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 hyperbolic decay, rather than the standard geometric rate characterizing $ARMA$ processes. Thus, cross-sectional aggregation can produce long memory in the covariance sense, definition (i).

In this paper we focus on the aggregation argument leading to long memory. We address the particular specification considered by Granger because the Beta distribution is a rather flexible specification but the analysis could be extended to other aggregation schemes. 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 both the time series and the cross section dimensions have to be significant for the theoretical degree of memory to apply. Finite samples 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. It occurs that the fractionally differenced series, using the theoretical degree of differencing, 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, $d \geq 0.5$. Our findings may have implications for the argument that is often given for estimating $ARFIMA$ models, namely that the observed long memory of time series can occur due to cross-sectional aggregation.

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

Granger (1980) showed that aggregating $AR(1)$ processes with random coefficients can produce long memory according to definition (i). He considered N series generated as[1]

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

where $\varepsilon_{i,t}$ is a white noise process 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 $p, q > 1$ 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.

[1]Granger also considered the case with dependence across series and allowing for different variances across the cross-sectional units but for clarity we will focus on the scenario under independence and equal variance.

Furthermore, define the cross-sectional aggregated series as:

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

Granger showed that, as $N \to \infty$, the autocorrelations of x_t decay at a hyperbolic rate and hence generates long memory in the covariance sense according to definition (i) with parameter $d = 1 - q/2$. In Theorem, 1 we extend his 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).*

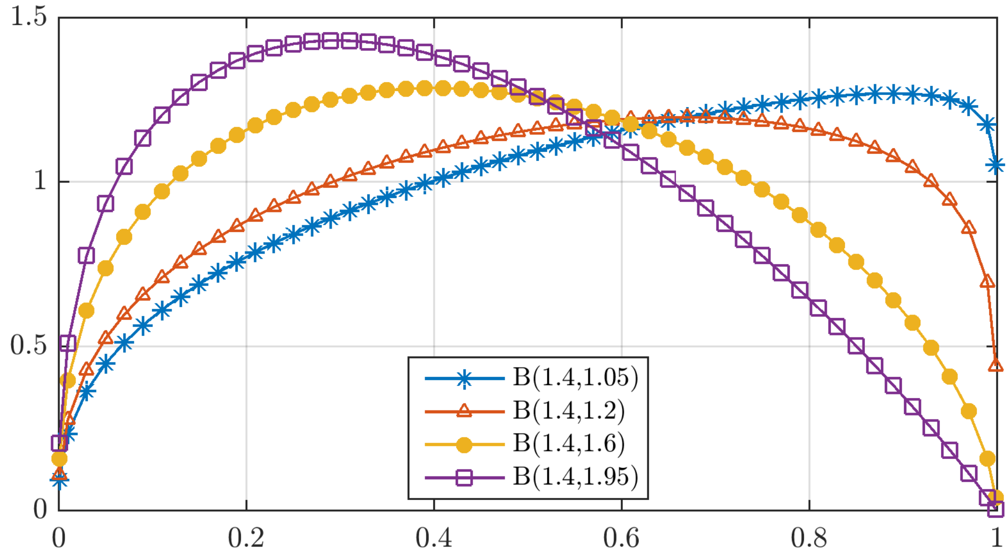
Proof: See appendix.

Theorem 1 shows that a cross-sectional aggregated series of infinitely many $AR(1)$ processes with squared autoregressive coefficients 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. Taking $q \in (1, 2)$ the long memory generated falls in the stationary range, $d \in (0, 1/2)$. We will focus on this range for the rest of the analysis. Furthermore, it appears that the value of p plays no role for this result as $N \to \infty$. As a consequence, Granger conjectured 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 greater degree of memory in the cross-sectionally aggregated series, x_t.

Granger's result has been extended by, among others Oppenheim and Viano (2004), allowing for $AR(s)$ processes (with $s \geq 1$) and Linden (1999) changing the Beta distribution to the Uniform; note that in Granger's setting the Uniform distribution was ruled out given that $p, q > 1$. Under the scenario of Oppenheim's et al. the aggregated series exhibits seasonal behavior along with long memory.

Granger's finding about the dependence of the result on the behavior of the distribution near one was further discussed by Zaffaroni (2004). He showed that if the distribution of the

Figure 1: Beta distribution.



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, it allows us to have 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.

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

$ARFIMA$ processes are fractional differenced $ARMA$ processes after adopting the $(1-L)^d$ filter. The MA series resulting from expansion 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. We can generalize this construction to series that still show hyperbolic decaying coefficients, yet, the coefficients do not come from 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 if $\varepsilon_{i,t}$ are *i.i.d.*, cross-sectional aggregated processes can be expressed as a generalized fractional process.

Lemma 1. *Let x_t be defined as in (4) with $N \to \infty$ and assume that $\varepsilon_{i,t}$ is an i.i.d. process. Then, 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 $\phi_j = (B(p+j,q)/B(p,q))^{1/2}$, $\forall j \in \mathbb{N}$.

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 without being 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 argues that the scaled partial sum of cross-sectional aggregated processes converges to fractional Brownian motion.

Theorem 2. *Let x_t be defined as in (4) with $N \to \infty$ and assume that $\varepsilon_{i,t}$ is an i.i.d. process.*

Consider the scaled partial sum of x_t defined as

$$X_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{[n\xi]} x_t,$$

with $\sigma_n^2 = \mathrm{E}[(\sum_{t=1}^{n} x_t)^2]$ and $\xi \in [0,1]$. Then, $X_n(\xi) \xrightarrow{d} B_H(\xi)$, where $B_H(\xi)$ is a fractional Brownian motion, $H = d + 1/2$, and $\xrightarrow{d}$ denotes convergence in distribution.

Proof: See appendix.

Theorem 2 is in line with the results from Zaffaroni (2004) when restricting the analysis to the Beta distribution. In this context, the parametric assumption 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, note that Theorem 2 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 can not be obtained. Nonetheless, the numerical finite sample implications of these conclusions should be quantified.

To shed some light on this question we generate x_t as in (4) under different parametric settings focusing on three main dimensions: the density of the autoregressive coefficient near one determined by the parameter q; the sample size T; and the cross-sectional dimension N, that is, the number of $AR(1)$ processes aggregated over.

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. The error terms, $\varepsilon_{i,t}$, are sampled from independent standard normals.
- Aggregate the individual series cross-sectionally according to equation (4).
- Estimate the long memory parameter by the GPH estimator, Geweke and Porter Hudak (1983). For robustness we also consider the local Whittle estimator of Robinson (1995) and Künsch (1986) $[LW]$ and the bias-reduction method for the GPH estimator suggested by Andrews and Guggenberger (2003) using second degree $[AND(2)]$ and fourth degree polynomials $[AND(4)]$.

We use these estimators of the long memory parameter since they do not depend on a full parametric assumption. The importance of this will be made clearer in Section 4 when discussing the relationship of cross-sectional aggregated series with $ARFIMA$ processes.

Throughout, we have used a bandwidth of $T^{0.5}$ as it is standard in the literature. As it is well known, the bandwidth affects the bias-precision tradeoff. Results with different bandwidths are available upon request showing this tradeoff; notwhistanding, the main conclusions maintain. 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.

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 $d = 1 - q/2$. We have conducted $R = 10,000$ replications with $T = N = 10,000$. Additionally, for comparison we also simulate 10,000 $FI(d)$ series using the exact algorithm of Jensen and Nielsen (2014).

Table 1: Mean and standard deviation in parentheses of the estimated long memory parameter. $T = N = R = 10,000$. The last three columns show comparable $FI(d)$ processes simulated according to Jensen and Nielsen (2014) algorithm.

Theoretical	Cross-sectional aggregated				$FI(d)$			
d	GPH	LW	$AND(2)$	$AND(4)$	GPH	LW	$AND(2)$	$AND(4)$
0.475	0.5117 (0.0711)	0.5067 (0.0563)	0.4967 (0.1126)	0.492 (0.1463)	0.4818 (0.0710)	0.4779 (0.0565)	0.4840 (0.1116)	0.4849 (0.1467)
0.45	0.4894 (0.0718)	0.4840 (0.0544)	0.4731 (0.1139)	0.4671 (0.1499)	0.4566 (0.0700)	0.4519 (0.0550)	0.4582 (0.1128)	0.4606 (0.1471)
0.4	0.4442 (0.0723)	0.4409 (0.0598)	0.4255 (0.1135)	0.4186 (0.1482)	0.4029 (0.0699)	0.4034 (0.0563)	0.4045 (0.1120)	0.4051 (0.1465)
0.35	0.4041 (0.0722)	0.4031 (0.0578)	0.3826 (0.1127)	0.3744 (0.1482)	0.3536 (0.0698)	0.3504 (0.0572)	0.3542 (0.1104)	0.3541 (0.1449)
0.3	0.3633 (0.0723)	0.3601 (0.0564)	0.3394 (0.1155)	0.3295 (0.1508)	0.3017 (0.0693)	0.3012 (0.0529)	0.3040 (0.1102)	0.3043 (0.1453)
0.25	0.3251 (0.0730)	0.3254 (0.0619)	0.2965 (0.1159)	0.2829 (0.1520)	0.2529 (0.0702)	0.2480 (0.0531)	0.2532 (0.1104)	0.2529 (0.1442)
0.2	0.2887 (0.0738)	0.2882 (0.0613)	0.2573 (0.1183)	0.2434 (0.1552)	0.2009 (0.0700)	0.1946 (0.0566)	0.2012 (0.1112)	0.2013 (0.1464)
0.15	0.2547 (0.0730)	0.2517 (0.0619)	0.2198 (0.1173)	0.2075 (0.1529)	0.1512 (0.0694)	0.1458 (0.0537)	0.1509 (0.1107)	0.1519 (0.1454)
0.10	0.2252 (0.0741)	0.2253 (0.0615)	0.1888 (0.1174)	0.1753 (0.1536)	0.1004 (0.0683)	0.0957 (0.0561)	0.1022 (0.1103)	0.1029 (0.1448)
0.05	0.1938 (0.0748)	0.1953 (0.0611)	0.1569 (0.1181)	0.1422 (0.1550)	0.0500 (0.0692)	0.0457 (0.0550)	0.0494 (0.1104)	0.0493 (0.1472)

Note. The estimators considered are GPH, Geweke and Porter Hudak (1983), LW, the local Whittle estimator of Robinson (1995) and Künsch (1986), $AND(2)$ and $AND(4)$ are the bias corrected GPH tests of Andrews and Guggenberger (2003) using second degree and fourth degree polynomials, respectively.

The table shows that for large degrees of memory we are close to the theoretical values but rather distant when the memory is low. Thus, it shows that the density of the autoregressive coefficient plays a key role in finite samples.[2] It suggests that using cross-sectional aggregation as a way to simulate long memory works poorly when dealing with a small memory index, d. In contrast, Table 1 shows that fractional differencing remains precise for all values of d. In particular, note that for a sample size of 10,000 and using 10,000 $AR(1)$ series, the cross-sectional aggregated series tends to show a larger degree of memory than the asymptotic result implies, and that of a comparable $FI(d)$ process.[3] This, coupled with the computational load required to generate the aggregated series, suggests that the aggregation scheme is clearly dominated by

[2]Note that the Andrews and Guggenberger (2003) estimates do reduce the bias, however, this is at the cost of more imprecise estimates.

[3]We need a sample size T and cross-sectional dimension N of more than $100,000$ to obtain results mimicking the $FI(d)$ simulations.

exact fractional differencing.

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 four theoretical degrees of long memory with the GPH estimation method.

Figure 2: Box-plot of the GPH long memory estimator for different levels of aggregation. $T = R = 10,000$. 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.

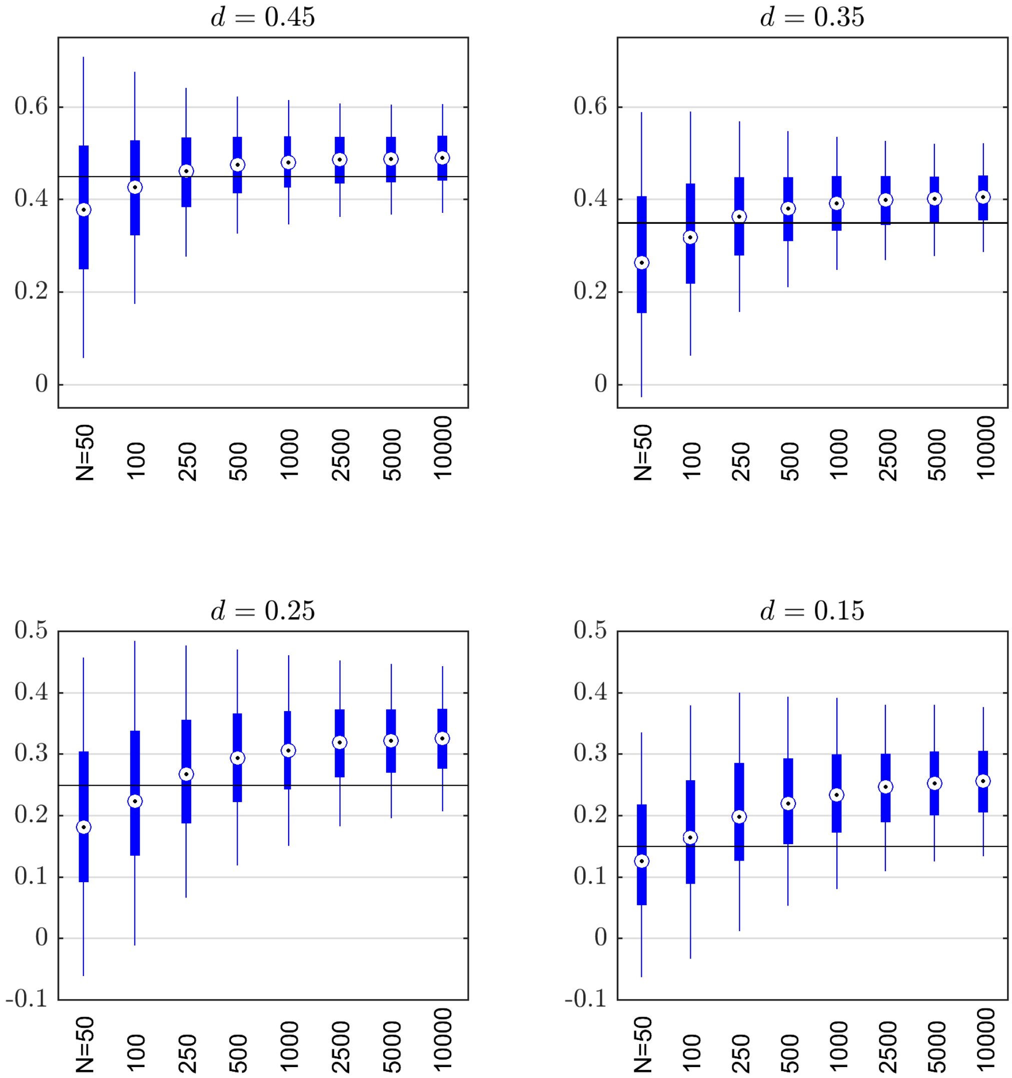


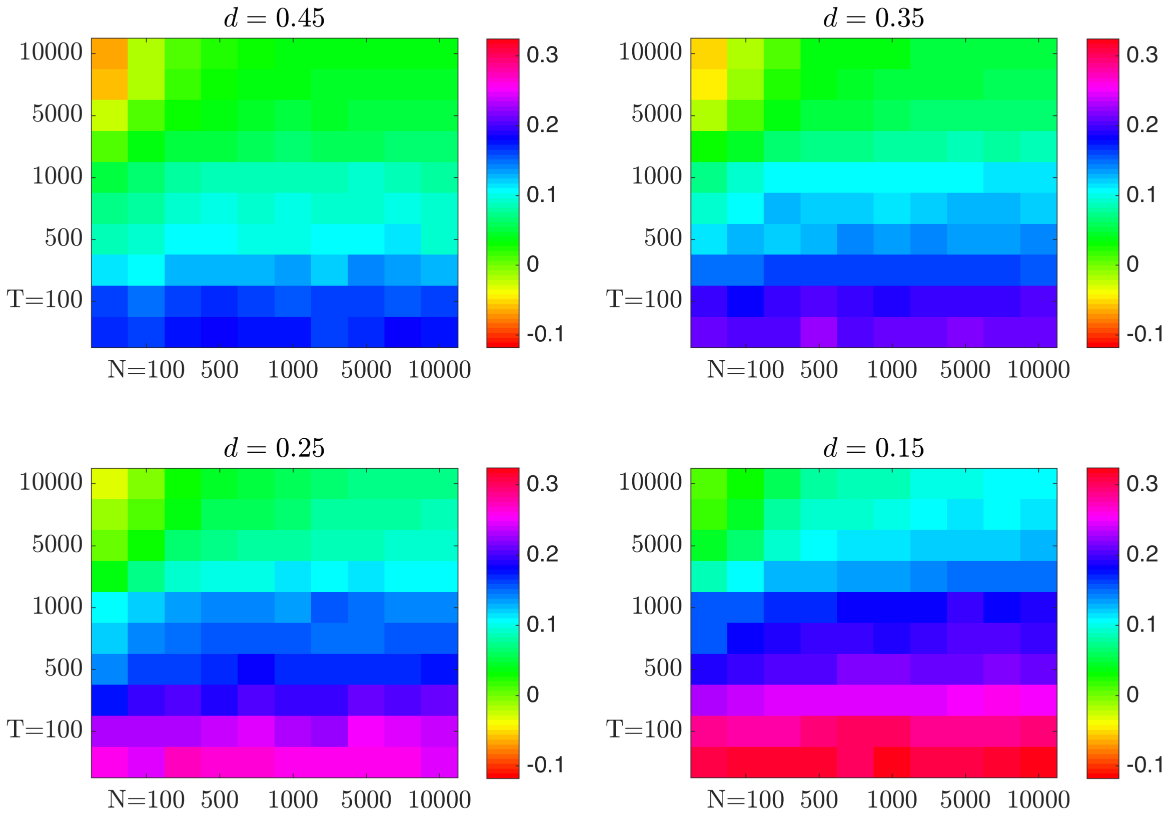
Figure 2 allows us to see how the long memory parameter evolves while increasing the cross-sectional dimension. It further shows the dependence of the result on the density of the autoregressive coefficient and the implied theoretical memory d. The larger the degree of memory (the denser the Beta distribution 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. Moreover, the box-plots show that the cutoff 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 10,000 $AR(1)$ series. Moreover, much estimation uncertainty is still present 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 estimated parameters for 1,000 replications minus their theoretical values while varying T and N. We consider four theoretical values of $d \in \{0.45, 0.35, 0.25, 0.15\}$.

The figure shows the interaction between the cross-sectional dimension and the sample size. 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. On the other hand, as the sample size T increases, more cross-sectional units are needed to approximate the asymptotic result. Thus, it quantifies the cutoff between small and large scale aggregation. This indicates that if we were to use aggregation as a way to simulate long memory we need to increase the cross-sectional dimension proportionally to the sample size, with the associated computational cost that it implies.

In summary, we find that the aggregation scheme to generate long memory can be rather imprecise and generally requires many time series observations and many cross-sectional units. In particular for small values of d.

Figure 3: Heat-map of the mean of the GPH estimator for $R = 1000$ replications minus the theoretical value; $T, N \in \{50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000\}$.



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*

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

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 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. The properties of the Quasi-Maximum Likelihood estimation of $ARFIMA$ models when the underlying process is a generalized fractional process remain an open question.

Second, note that as the proof of Theorem 3 shows, in reality 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.

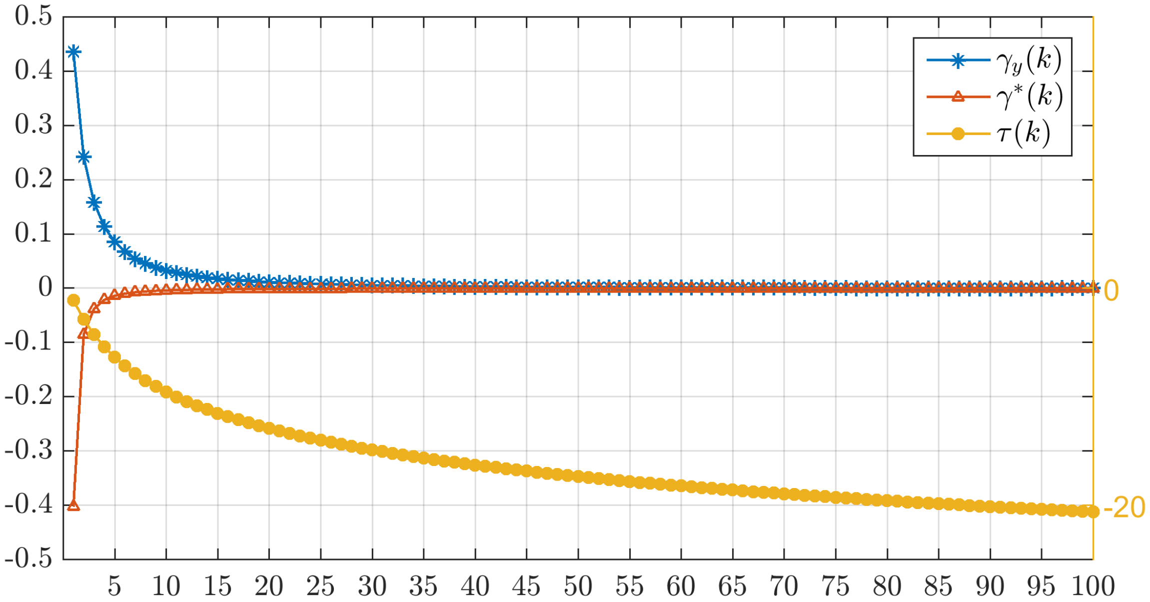
Figure 4 displays the shape of the autocovariance function for the fractionally differenced cross-sectionally aggregated process $\gamma_y(k)$, the autocovariance of the antipersistent component $\gamma^*(k)$, and its ratio $\tau(k) := \gamma_y(k)/\gamma^*(k)$.

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.

Figure 4: Autocovariance 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). $p = 1.4$, $q = 1.05$ so that $d = 0.475$.

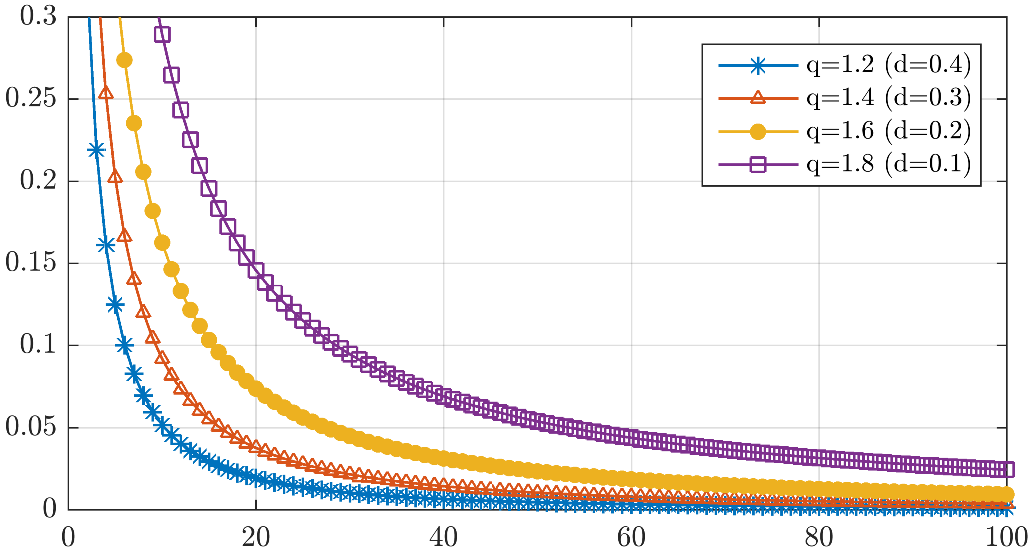


As seen in Figure 4 and proved in Corollary 1, the autocovariance function $\gamma_y(k)$ decays at a hyperbolic rate similar to the rate for antipersistent processes. However, the sign of the function is positive as opposed to antipersistent processes, which is a feature induced by the cross-sectional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

Note from the expression of $\gamma_y(k)$ given in Theorem 3 that autocovariances for finite k depend on the parameters p and q associated with the Beta distribution. Figure 5 displays the autocovariance 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 autocovariances for finite k. As q increases, and hence memory declines, the fractionally differenced series tend to have rather significant autocovariances 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.

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

Figure 5: Autocovariance functions for the fractionally differenced cross-sectionally aggregated series $\gamma_y(k)$ for $p = 1.1$ and $q \in \{1.2, 1.4, 1.6, 1.8\}$.



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 both the cross-sectional and the time dimensions 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 Beta distribution has concentrated mass around one. Secondly, we have shown that when taking a fractional difference of a cross-sectionally aggregated long memory process, 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 in $ARMA$ 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.

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 would like to thank the helpful comments from James Davidson and Peter Robinson.

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

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

where the third equality follows from the independence assumption and the last equality comes from the fact that as $N \to \infty$ the sum can be approximated by an integral and thus,

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

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 $k \in \mathbb{N}$. This in turn yields the autocorrelations

$$\rho_x(k) = \frac{\gamma_x(k)}{\gamma_x(0)} = \frac{B(p+k/2,q-1)}{B(p,q-1)} = \frac{\Gamma(q-1)}{B(p,q-1)}\frac{\Gamma(p+k/2)}{\Gamma(p+k/2+q-1)},$$

which, by Stirling's approximation shows that $\rho_x(k) \approx Ck^{1-q}$. So that the aggregated series shows hyperbolic decaying autocorrelations. That is, long memory in the covariance sense with parameter $d = 1 - q/2$.

To prove (ii), note that given the autocorrelation function, Theorem 1.3 in Beran et al. (2013) shows that the spectral density has a pole in the origin.

To prove (iii),

$$\begin{aligned}
Var\left(\sum_{t=1}^{T} x_t\right) &= \frac{1}{N}E[(x_1 + x_2 + \cdots + x_T)^2] \\
&= E[x_1^2 + \cdots + x_T^2 + 2(x_1x_2 + \cdots + x_{T-1}x_T)] \\
&= TE[x_1^2] + 2E\left[\left(\sum_{t=2}^{T} x_1 x_t + \cdots + \sum_{t=T-1}^{T} x_1 x_t\right)\right] \\
&= TE[x_1^2] + 2\left((T-1)E[x_1x_2] + \cdots + E[x_1x_T]\right) \\
&= 2\gamma_x(0)\left(\frac{T}{2} + 2\left((T-1)\rho(1) + \cdots + \rho(T-1)]\right)\right) \\
&\approx 2\gamma_x(0)\left((T-1) + (T-2)2^{1-q} + \cdots + (T-1)^{1-q}\right) \\
&= 2\gamma_x(0)\sum_{t=1}^{T}(T-t)t^{1-q} = O_p(T^{3-q}) = O_p(T^{1+2d}).
\end{aligned}$$

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 $k \in \mathbb{N}$

$$
\begin{aligned}
E[x_t^{(m)} x_{t+k}^{(m)}] &= \frac{1}{m^2} E[(x_{tm-m+1} + \cdots + x_{tm})(x_{(t+k)m-m+1} + \cdots + x_{(t+k)m})] \\
&= \frac{1}{m^2} E[x_{tm-m+1} x_{(t+k)m-m+1} + \cdots + x_{tm} x_{(t+k)m}] \\
&= \frac{\gamma_{x^m}(0)}{m^2} \left(\rho(km - 2m + 1) + \cdots + m\rho(km - m) + \cdots + \rho(km - 1)\right) \\
&\approx \frac{\gamma_{x^m}(0)}{m^2} \left((km - 2m + 1)^{1-q} + m(km - m)^{1-q} + \cdots + (km - 1)^{1-q}\right) \\
&\approx \frac{\gamma_{x^m}(0)}{m^2} \left((km)^{1-q} + \cdots + m(km)^{1-q} + \cdots + (km)^{1-q}\right) \\
&= \frac{\gamma_{x^m}(0)}{m^2} \left(1 + \cdots + m + \cdots + 1\right) (km)^{1-q} \\
&= \frac{\gamma_{x^m}(0)}{m^2} m^2 (km)^{1-q} = (\gamma_{x^m}(0) m^{1-q}) k^{1-q},
\end{aligned}
$$

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

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 and thus

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

$\forall j \in \mathbb{N}$. 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 $\phi_j = \left(B(p+j,q)/B(p,q)\right)^{1/2}$, $\forall j \in \mathbb{N}$. 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) we know that for $k \in \mathbb{N}$

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

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,

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

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

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

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

Let

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

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.

Andrews, D. W. K. and Guggenberger, P. (2003). A bias–reduced log–periodogram regression estimator for the long–memory parameter. *Econometrica*, 71(2):675–712.

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

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. (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.

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

Guégan, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews*, 24(2):113–149.

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.

Linden, M. (1999). Time series properties of aggregated AR (1) processes with uniformly distributed coefficients. *Economics Letters*, 64(1):31–36.

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

Oppenheim, G. and Viano, M. C. (2004). Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results. *Journal of Time Series Analysis*, 25(3):335–350.

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.