Effect of Temporal Aggregation on Persistence and Integration

Uwe Hassler*
Goethe University Frankfurt†

First version, July 24, 2009

Abstract

The impulse response function and related persistence measures are discussed for fractionally integrated processes, where the order of integration also covers the nonstationary case. Then we obtain a general result that characterizes the effect of temporal aggregation in the frequency domain (aliasing) for arbitrary stationary processes. Temporal aggregation includes here cumulation of flow variables as well as systematic skip sampling of stock variables. Next, the general result is applied to fractionally integrated processes. In particular, it is investigated whether typical assumptions made when analyzing fractional integration statistically are closed with respect to aggregation. It turns out that they are closed with respect to cumulating time series, but not with respect to skip sampling. Finally, we discuss proposals repairing the shortcoming in case of lack of closedness.

Keywords: long memory, cumulating time series, skip sampling, closedness of assumptions

*A first version of this paper was written while visiting the University of California San Diego and was presented at Texas A&M University, the 3rd ETSERN Meeting, Nottingham, and the Conference on Recent Developments in Macroeconomics, Mannheim. I am grateful to Graham Elliot, Joon Park, Philippe Soulier and Yixiao Sun for support and helpful comments.

†hassler@wiwi.uni-frankfurt.de

JEL classification: C14 (Semiparametric and Nonparametric Methods), C22 (Time-Series Models), C82 (Methodology for Collecting, Estimating, and Organizing Macroeconomic Data)

1 Introduction

The effect of temporal aggregation of economic time series has troubled econometrician for decades. Early results for autoregressive moving-average [ARMA] models were obtained by Brewer (1973) and Weiss (1984), and by Geweke (1978) for stationary dynamic regression models. A treatment of integrated (of order one) ARIMA models was provided by Wei (1981) and Stram and Wei (1986), for skip sampling and cumulating, respectively. In this paper we understand by temporal aggregation both, the systematic skip sampling of stock variables as well as the cumulation of flow variables. In particular, skip sampling can be embedded in the more general problem of missing observations, see Palm and Nijman (1984) for an investigation of dynamic regression models. In the frequency domain, skip sampling is typically accompanied by the so-called aliasing effect, which is well known under discrete-time sampling from a continuous-time process, see e.g. Sims (1971) and Hansen and Sargent (1983). Moreover, the aspect of temporal aggregation and forecasting has been addressed by Lütkepohl (1987). The potential interaction of seasonal integration and integration at frequency zero due to temporal aggregation was studied by Granger and Siklos (1995), see also Pons (2006).

The present paper focusses on the effect of time aggregation on the persistence in economic time series and in particular on the properties of a fractionally integrated time series model. The effect of temporal cumulation (on inflation persistence) has recently been addressed by Paya, Duarte and Holden (2007), see also Christiano, Eichenbaum and Marshall (1991) for empirical evidence in the context of the permanent income hypothesis. Chambers (1998) and Hwang (2000) contribute to the theoretical understanding of fractional integration under temporal aggregation; notice, however, the correction in Souza (2005). We add three aspects to this literature: a discussion of impulse response related persistence measures under fractional integration, a general characterization of time aggregation in the frequency domain, and an investigation how assumptions of a fractionally integrated model are affected under time aggregation.

In greater detail our contributions are the following. First, we discuss how the order d of fractional integration can be used for measuring persistence in economic time series. We spell out conditions under which the impulse response function at lag j turns out to be approximately proportional to j^{d-1}, and this result continues to hold even in the nonstationary region as long as $d < 1$ (Proposition 2.1). We discuss simple impulse response related measures of persistence and show that they can be approximated in terms of d alone, irrespective of eventual short memory parameters. Second, we study the effect of temporal aggregation (cumulating flow variables or systematic skip sampling stock variables) of an arbitrary stationary process in terms of spectral densities. The effect of aggregation is investigated and discussed in the frequency domain (Theorem 3.1 and Corollary 3.2). Third, the theorem is applied to fractionally integrated processes. In particular, we investigate whether typical assumptions on fractionally integrated processes, which are made in the literature to obtain statistical properties, are closed with respect to aggregation. In other words: If $\{y_t\}$ satisfies a set of assumptions $\mathcal{A}$ (which are sufficient to prove properties of some estimator or test), does the temporal aggregate fulfill $\mathcal{A}$, too? If not, then we should be worried, because in most cases we do not know the frequency of the "true" data generating process (DGP), i.e. our observed data may well be aggregates. If they do not satisfy $\mathcal{A}$ although the disaggregated unobserved process does, then we lose grounds for reliable inference. It turns out that typical spectral assumptions made in the semiparametric long memory literature are closed with respect to cumulating the data. Fourth, it is established that certain spectral assumptions are not closed with respect to skip sampling fractional integration. Fifth, we have a simple repair proposal to this shortcoming by linking the topic of skip sampling of long memory to the field of estimation of long memory from perturbed processes ("long memory plus noise"). In fact, skip sampling has in the frequency domain the same effect on long memory as adding noise. Hence, conditions under which certain procedures remain valid irrespective of skip sampling or not are available.

The rest of the paper is organized as follows. Section 2 discusses the impulse response function and related persistence measures for fractional integration. The third section is dedicated to the general aggregation result in the frequency domain. In Section 4, this result is applied to fractional integration, and the effects of temporal aggregation on integration are discussed in some detail. The last section contains concluding remarks. All technical proofs are relegated to the Appendix.

Finally a word on notation. For sequences a_j and b_j, let $a_j \sim b_j$ denote $a_j/b_j \to 1$ as $j \to \infty$, while for functions, $a(x) \sim b(x)$ is short for $a(x)/b(x) \to 1$ as $x \to 0$. Further, $a(x) = O(x^c)$ means that $a(x)\, x^{-c}$ is bounded as $x \to 0$.

2 Persistence and fractional integration

2.1 Impulse response of fractionally integrated processes

An essential ingredient to measure persistence of economic time series is the so-called impulse response function. Assume that the process $\{y_t\}$, $t \in \mathbb{Z}$ with $\mathbb{Z}$ denoting the set of all integers, is covariance stationary and given by

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}\,, \quad c_0 = 1\,,\ \sum_{j=0}^{\infty} c_j^2 < \infty\,. \tag{1}$$

The series of past innovations $\{\varepsilon_t\}$ driving $\{y_t\}$ is assumed to be white noise. If a unit shock occurred j periods ago, then the impulse response (IR) function c_j measures its influence on the present value of y_t:

$$IR_j = \frac{\partial\, y_t}{\partial\, \varepsilon_{t-j}} = c_j\,.$$

If the impulse responses c_j are absolutely summable, then

$$CIR := \sum_{j=0}^{\infty} c_j$$

as the cumulated impulse response function is a classical measure of persistence, see Campbell and Mankiw (1987). It is directly related to the limiting version of the variance ratio by Cochrane (1988, eq. (10)), which is proportional to CIR^2.

In this section we consider $\{y_t\}$ given by filtering a stationary process $\{e_t\}$. More precisely, $\{e_t\}$ is integrated of order 0, $I(0)$, in that it is stationary with $CIR \neq 0$, and a purely stochastic process with absolutely summable Wold coefficients ρ_j. In particular, the summability condition is made more precise in the following assumption, that is satisfied e.g. by all stationary and invertible autoregressive moving-average (ARMA) processes.

Assumption 1 *Let $\{e_t\}$, $t \in \mathbb{Z}$, be a linear process,*

$$e_t = \sum_{j=0}^{\infty} \rho_j \, \varepsilon_{t-j} \, ,$$

where $\{\varepsilon_t\}$ is white noise with mean 0 and variance σ^2, $\rho_0 = 1$, and

$$\rho(1) = \sum_{j=0}^{\infty} \rho_j \neq 0.$$

The process is assumed to be s-summable in that

$$\sum_{j=0}^{\infty} j^s |\rho_j| < \infty \, ,$$

for some $s > 1$.

In particular, we study fractionally integrated processes $\{y_t\}$ constructed from the filter

$$(1 - L)^{-d} = \sum_{j=0}^{\infty} \psi_{j,d} L^j$$

where L is the usual lag operator, and

$$\psi_{j,d} = \frac{\Gamma(j+d)}{\Gamma(j+1)\,\Gamma(d)} = \frac{j-1+d}{j}\,\psi_{j-1,d}\,, \quad j \geq 1\,, \quad \psi_{0,d} = 1\,, \tag{2}$$

such that

$$\psi_{j,d} \sim \frac{j^{d-1}}{\Gamma(d)}\,, \quad j \to \infty\,, \quad d \neq 0, -1, -2, \ldots\,. \tag{3}$$

Here, $\Gamma(x)$ denotes the Gamma function with $\Gamma(1) = \Gamma(2) = 1$, $\Gamma(x+1) = x\Gamma(x)$, which is not defined for $x = 0, -1, \ldots$, but $\Gamma(0)/\Gamma(0) = 1$. For $d < 0.5$, $\psi_{j,d}^2$ is summable, and

$$\begin{aligned} y_t &= (1-L)^{-d}\, e_t = \sum_{j=0}^{\infty} \psi_{j,d}\, e_{t-j} \\ &= \sum_{j=0}^{\infty} c_{j,d}\varepsilon_{t-j}\,, \qquad \text{for } -1 < d < 0.5\,, \end{aligned} \tag{4}$$

defines a stationary and invertible[1] process, where the impulse response function is given by convolution,

$$c_{j,d} = \sum_{k=0}^{j} \rho_k\, \psi_{j-k,d}\,, \tag{5}$$

cf. see Granger and Joyeux (1980) or Hosking (1981). There seems to be a common understanding that $c_{j,d}$ is dominated by $\psi_{j,d}$ and inherits a behaviour like in (3), such that $c_{j,d} \sim \gamma j^{d-1}$ for some constant γ, see Baillie and Kapetanios (2008) for a recent example of such a view. Of course this holds only true if the sequence ρ_k is well-behaved. Proposition 2.1 below states a sufficient condition on the s-summability in Assumption 1. Poskitt (2007, eq. (5)) states $c_{j,d} \sim \gamma j^{d-1}$ for $d < 0.5$ without detailed proof. He assumes absolute summability for the short memory component, which corresponds

[1] Expanding the fractional differences $(1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j$ with $\pi_j = \frac{j-1-d}{j}\pi_{j-1}$, the process is invertible as long as it has an AR(∞) representation. For $d > -0.5$, π_j^2 is summable. Assuming ARFIMA processes (where $\{e_t\}$ is ARMA), invertibility is guaranteed, see Hosking (1981). The range of invertibility has recently been extended for the ARFIMA case to $d > -1$ by Bondon and Palma (2007), see also Odaki (1993).

to $s = 0$ in Assumption 1. Our proof will require a stronger assumption with $s = 2(1-d)$: The larger d, the more dominant is the decay rate from (3), and the less restrictive is the required summability condition on $\{e_t\}$ from Assumption 1. Moreover, we will extend our result to $d \geq 0.5$.

We want to measure persistence also for nonstationary processes. To that end we define the $I(\delta)$ proceess $\{z_t\}$ as

$$z_t = \begin{cases} \sum_{i=1}^{t} y_i \,, \ t = 1, \ldots, T & \text{if } 0.5 \leq \delta < 1.5 \\ y_t \,, \ t = 0, \pm 1, \pm 2, \ldots & \text{if } \delta = d < 0.5 \end{cases} \,, \tag{6}$$

with $y_t = (1-L)^{-d} e_t$ given in (4). The process $\{z_t\}$ is hence stationary for $\delta = d < 0.5$, and it consists of a cumulation of stationary $I(d)$ differences with $d = \delta - 1$ for $0.5 \leq \delta < 1.5$. Such a process is sometimes called fractionally integrated of "type I", see Marinucci and Robinson (1999) and Robinson (2005). In the nonstationary case, the impulse responses are given as

$$IR_j = \frac{\partial z_t}{\partial \varepsilon_{t-j}} = \sum_{k=0}^{j} c_{k,\delta-1} \,, \quad j = 0, 1, \ldots, t-1 \,, \tag{7}$$

with $c_{k,\delta-1}$ given in (5). The approximate behaviour of the impulse response function is given as γ_j in Proposition 2.1.

Proposition 2.1 *Let $\{z_t\}$ from (6) be integrated of order δ, $\delta \leq 1$, $\delta \neq 0, -1, \ldots$, and $\{e_t\}$ is from Assumption 1 with*

$$s = \begin{cases} 2 - 2\delta & \textit{if } \delta < 0.5 \\ 2 - 2(\delta - 1) & \textit{if } 0.5 \leq \delta \end{cases} \,.$$

It then holds (with $\rho(1)$ from Assumption 1)

$$IR_j \sim \gamma_j := \gamma \, j^{\delta-1} \,, \quad \gamma := \frac{\rho(1)}{\Gamma(\delta)} \,,$$

as $j \to \infty$, where IR_j is from (5) or (7), depending on whether $\{z_t\}$ is stationary or not.

PROOF See Appendix.

REMARK A The parameter values $\delta \in \{0, -1, \ldots\}$ deserve special care. They have been excluded since $\Gamma(0) = \Gamma(-1) = \ldots = \infty$, and γ is not well defined, see also (3). The case $\delta = 0$ means $c_{j,0} = \rho_j$ in (4) with ρ_j from Assumption 1 being absolutely summable. Therefore $\delta \to 0$ cannot mean that $IR_j = c_{j,0}$ is proportional to $1/j$ asymptotically. Nevertheless, those integer cases can be embedded as limiting cases with the convention $\Gamma(x) \sim x^{-1}$ as $x \to 0$. Hence, for $\delta \to 0$ we find $\gamma \to 0$, simply meaning that the hyperbolical decay law $j^{\delta-1}$ does not hold for $\delta = 0$, and similarly for all other negative integers.

The interpretation of the other parameter values is straightforward. For $\delta = 1$, past shocks have a permanent effect that does not die out, while for $0.5 \leq \delta < 1$ we observe nonstationarity with transitory shocks[2], $\gamma_j \to 0$ as $j \to \infty$. Finally for $0 < \delta < 0.5$ the impulse responses die out fast enough to be square-summable resulting in a stationary process, but still they vanish so slowly that γ_j is not summable. This case has been called long memory. Denote the autocovariance function as $\gamma(h) = \mathrm{E}(y_t\, y_{t+h})$, $h = 0, 1, \ldots$. From Palma (2007, Theorem 3.1) we learn that Proposition 2.1 implies a hyperbolical decay with some constant c,

$$\gamma(h) \sim c\, h^{2\delta-1} \quad \text{as} \quad h \to \infty\,. \tag{8}$$

Consequently, $\gamma(h)$ is not summable if $\delta > 0$, which defines long memory. For $\delta < 0$ the process is stationary with short memory.

Note that it holds for the fractionally integrated process

$$CIR \approx \sum_{j=0}^{J} c_{j,d} \to \begin{cases} 0\,, & d < 0 \\ \infty\,, & d > 0 \end{cases}\,, \qquad J \to \infty\,.$$

Consequently, Hauser, Pötscher and Reschenhofer (1999) criticized CIR or related measures in the presence of fractional integration as being meaningless. Instead, we suggest the parameter d itself to measures the degree of

[2]Such a feature is sometimes called "mean-reversion" although Phillips and Xiao (1999) argue that this is a misnomer given the nonstationarity of the process.

persistence. Proposition 2.1 reconfirms the interpretation of the order of integration as memory parameter that characterizes the degree of persistence in the stationary as well as in the nonstationary case up to $\delta \leq 1$. More colourful measures of persistence in terms of δ will be discussed next, building on the fact that the proxy γ_j to the impulse response function condenses all parameters of the short memory component $\{e_t\}$ in the multiplicative constant γ. Hence, we find approximate measures depending on the order of integration δ alone. This has the advantage that it is sufficient to employ some semiparametric estimation of the order of integration to perform a persistence analysis.

2.2 Measuring persistence

We continue to consider a potentially nonstationary $I(\delta)$ process $\{z_t\}$ like in Proposition 2.1. Since, the cumulated impulse response function is not meaningful for $\delta \neq 0$, we suggest as a related measure a half-life indicator. It simply counts how many periods it takes for a unit shock to be reduced by 50% in absolute value. We define more generally the half-life indicator HL_j counting the number of periods h during which the effect of a unit shock after j periods displays at least half of the absolute value $|IR_j|$ for another h periods. Given IR_j, we hence search h such that

$$|IR_{j+h}| \geq \frac{|IR_j|}{2} \quad \text{and} \quad |IR_{j+h+1}| < \frac{|IR_j|}{2}$$

or

$$HL_j := \max_{h \in \{0,1,\ldots\}} \left\{ h : |IR_{j+h}| \geq \frac{|IR_j|}{2} \right\} .$$

The approximate version thereof becomes in light of Proposition 2.1

$$hl_j := \max_{h \in \{0,1,\ldots\}} \left\{ h : |\gamma_{j+h}| \geq \frac{|\gamma_j|}{2} \right\} .$$

It is elementary to verify

$$hl_j = \left\lfloor j \left(2^{\frac{1}{1-\delta}} - 1 \right) \right\rfloor , \quad \delta < 1 , \tag{9}$$

where $\lfloor x \rfloor$ denotes the integer part of x. For $\delta > 0$ it holds

$$2^{\frac{1}{1-\delta}} - 1 > 1 \, .$$

Consequently, hl_j is at least j in case of long memory. This means: if a unit shock has reduced to e.g. 1/2 after j periods, $\gamma_j = 1/2$, then it takes at least another j periods to halve it again. Generally, the larger $\delta > 0$, the longer it takes to reduce from $|\gamma_j|$ to $|\gamma_j|/2$. Finally, we learn from (9) that hl_j does not converge with j. This becomes clear when rewriting the definition,

$$hl_j = \max_{h \in \{0,1,\dots\}} \left\{ h : \left(1 + \frac{h}{j}\right)^{1-\delta} \leq 2 \right\} ,$$

such that h may grow with j. We conclude that the half-life indicator from (9) is only meaningful for a sufficiently large but finite j.

We close this section with some numerical demonstration. In Figure 1 through 3 we compare the exact impulse response function IR_j with the approximation γ_j from Proposition 2.1 for weak long memory ($d = 0.2$), strong long memory ($d = 0.45$), and a nonstationary process ($d = 0.8$) . We consider ARFI(1,d) processes where $e_t = \rho e_{t-1} + \varepsilon_t$ and observe a very good approximation for $j \geq 10$ as long as the autoregressive parameter is moderate (not larger than 0.5 in absolute value). Hence we believe that hl_j is a useful approximation to the true half-life. For a real life numerical example we refer to the investigation of US inflation data by Kumar and Okimoto (2007). With monthly observations they estimate from 1960 until 1982 $\widehat{d} \approx 0.5$, while the sample after 1982 until 2003 yields only roughly $\widehat{d} \approx 0.25$. We computed corresponding half-life values after one year ($j = 12$), and also included the case $d = 0.75$:

d	0.25	0.5	0.75
hl_{12}	18	36	180

We observe that at the border of (non)stationarity ($d = 0.5$), it takes 3 years to halve the effect of a shock after one year. An increase from $d = 0.25$

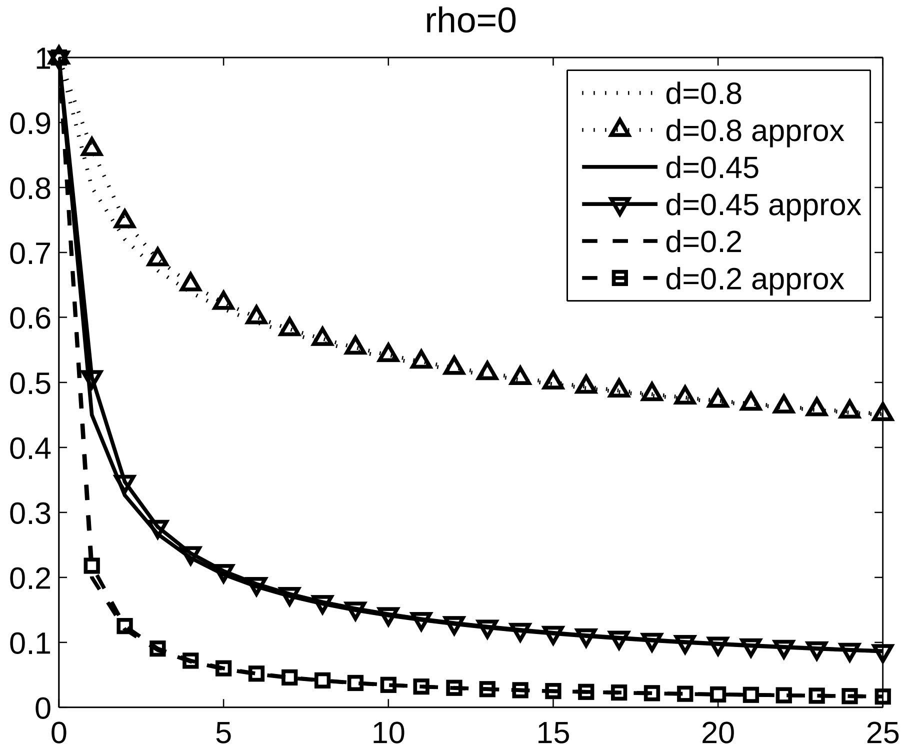

Figure 1: Impulse response function for $(1 - L)^d y_t = \varepsilon_t$

to $d = 0.5$ doubles the half-life (after one year) from 18 months to 3 years, which nicely illustrates the effect of the "great moderation". If inflation was nonstationary with $d = 0.75$, then it would take 15 years to reduce the effect of a shock after one year by 50%.

3 Aggregation in the frequency domain

3.1 Notation and assumptions

Let y_t, $t = 1, 2, \ldots, T$, denote some series that is aggregated over p periods. For simplicity we assume $T = p\,N$ for some integer N, and the aggregate is

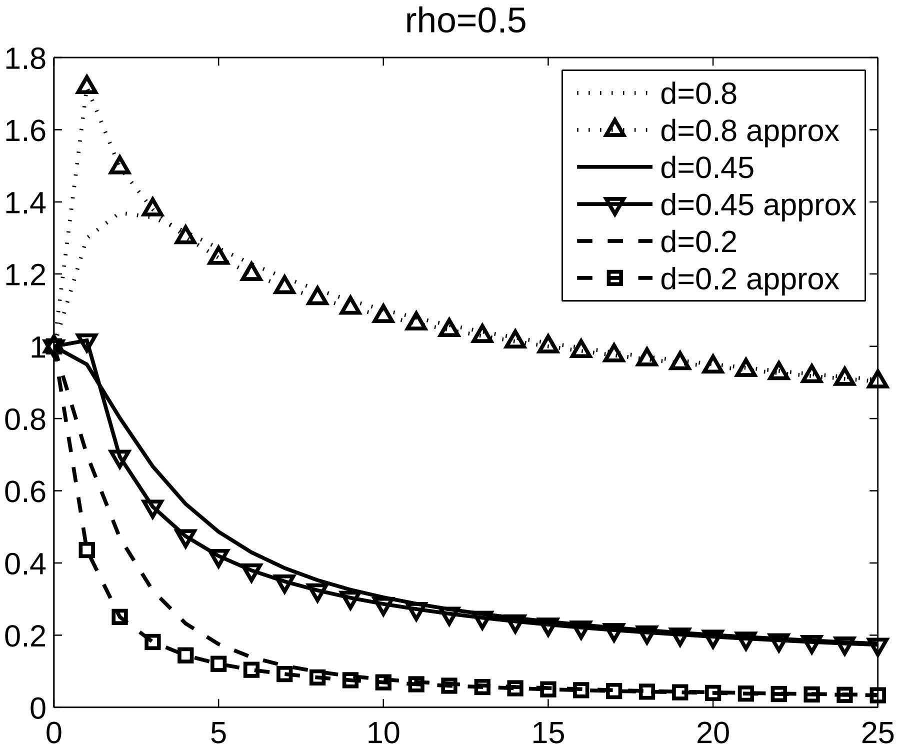

Figure 2: Impulse response function for $(1-\rho L)(1-L)^d y_t = \varepsilon_t$

constructed for the new time scale $\tau = 1, \ldots, N$. In case of stock variables, aggregation or systematic sampling means *skip sampling* where only every p'th data point is observed,

$$\dot{y}_\tau := y_{p\tau}, \quad \tau = 1, 2, \ldots, \tag{10}$$

where for the rest of the paper $p \geq 2$ is a finite integer. Flow variables are aggregated by *cumulating* p neighbouring observations that do not overlap to determine the total flow over p sub-periods,

$$\begin{aligned} \widetilde{y}_\tau &:= y_{p\tau} + y_{p\tau-1} + \ldots + y_{p(\tau-1)+1} \\ &= S_p(L)\, y_{p\tau}, \quad \tau = 1, 2, \ldots, \end{aligned} \tag{11}$$

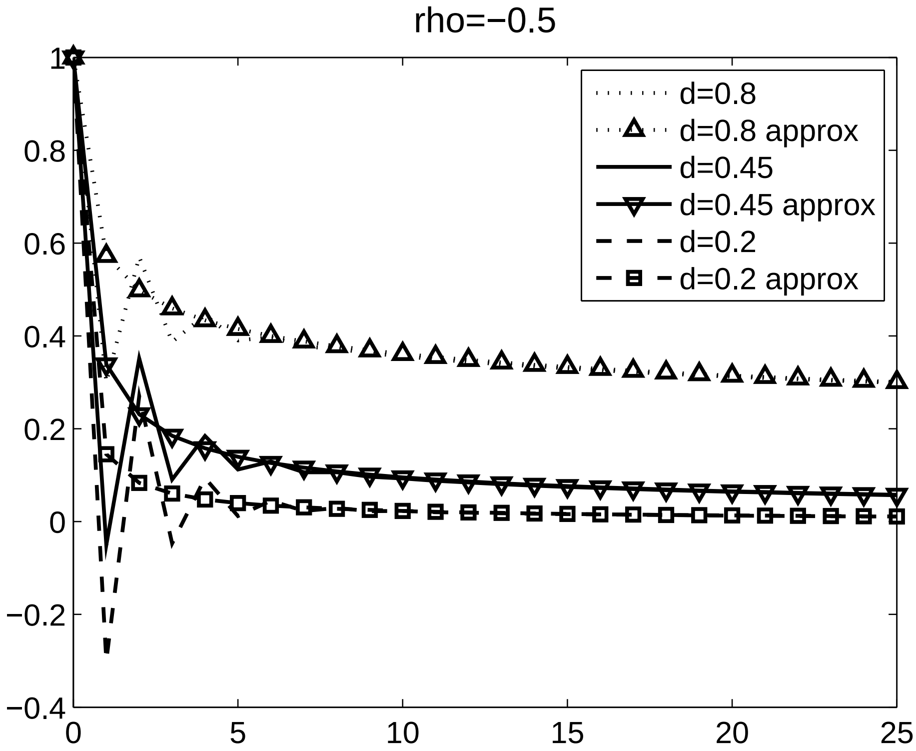

Figure 3: Impulse response function for $(1-\rho L)(1-L)^d y_t = \varepsilon_t$

where

$$S_p(L) := 1 + L + \cdots + L^{p-1}$$

is the moving average filter of order p. The link between the two aggregates is given by a moving average of order p where observations overlap,

$$y_t^{ma} \quad := \quad S_p(L)\, y_t\,, \tag{12}$$

because $\widetilde{y}_\tau$ is obtained from skip sampling the moving average, which amounts to

$$\dot{y}_\tau^{ma} = S_p(L)\, y_{p\tau} = \widetilde{y}_\tau\,, \quad \tau = 1, 2, \ldots\,.$$

Sometimes stock variables are aggregated by averaging over p non-overlapping observations, such that p sub-periods are replaced by the mean of the past p

values; obviously this is directly connected to cumulation from (11):

$$\overline{y}_\tau := \frac{\widetilde{y}_\tau}{p}, \quad \tau = 1, 2, \ldots . \tag{13}$$

The main result on the effect of temporal aggregation in the frequency domain holds for any stationary process $\{y_t\}$ with autocovariances $\gamma(h) = E(y_t y_{t+h})$ and spectral density f. For simplicity we assume $E(y_t) = 0$. The link between the time domain (autocovariances) and the spectral density $f(\lambda)$ in the frequency domain is given by Fourier transformation for $|\lambda| \leq \pi$:

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-i\,\lambda\, h)\,, \quad i^2 = -1\,, \\ \gamma(h) &= \int_{-\pi}^{\pi} f(\lambda) \exp(i\,\lambda\, h) d\lambda\,. \end{aligned}$$

Since f is an even and 2π-periodic function, the definition of the spectral density can be extended to the whole real range, and we can focus on the interval $[0, \pi]$ in the following assumption.

Assumption 2 *The process $\{y_t\}$ is covariance stationary with autocovariances $\gamma(h)$ and a spectral density $f(\lambda)$ on Π, where $\Pi = [0, \pi]$ if f is well defined on the whole interval, or $\Pi = [0, \pi] \setminus \{\lambda_0\}$ if f has a pole at some frequency $\lambda_0 \in [0, \pi]$.*

We only require that f is integrable over $[0, \pi]$, although it does not have to exist everywhere. In particular, a pole at $\lambda_0 = 0$ might come from fractional integration with long memory ($d > 0$). For the fractionally integrated process $\{y_t\}$ from (6) we know

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d} f_e(\lambda) = 4^{-d} \left(\sin \frac{\lambda}{2} \right)^{-2d} f_e(\lambda)$$

with f_e denoting the spectral density of $\{e_t\}$ from Assumption 1. Note that Assumption 1 implies $0 < f_e(0) < \infty$. Equivalently (because $|1 - e^{i\lambda}|^{-2d} = \lambda^{-2d}(1 + o(1))$ fractional integration is characterized through the assumption

$$f(\lambda) = \lambda^{-2d} f_e(\lambda)\,, \quad d < 0.5\,. \tag{14}$$

Similarly, we might allow for k poles (having e.g. so-called k-factor Gegenbauer processes in mind, see Woodward, Cheng and Gray, 1998).

Under the stronger assumption that the spectral density is "well behaved" on $(0, \pi]$, we will be able to establish how the properties of f are inherited by the spectral densities of the aggregates at frequency zero, see Corollary 3.2 below. This stronger assumption reads as follows.

Assumption 3 *The process $\{y_t\}$ from Assumption 2 has a spectral density $f(\lambda)$ on Π, which at frequencies $2\pi\, j/p$, $j = 1, \ldots, (p-1)$, is bounded, bounded away from zero and continuously differentiable with derivative f'.*

3.2 Results and discussion

Let the spectral densities of the aggregates $\{\dot{y}_\tau\}$ from (10) and $\{\widetilde{y}_\tau\}$ from (11) be denoted as $\dot{f}(\lambda)$ and $\widetilde{f}(\lambda)$, respectively.

Theorem 3.1 *Under Assumption 2 is holds for the spectral densities of the aggregates of $\{y_t\}$:*

a) *in case of skip sampling*

$$\dot{f}(\lambda) = \frac{1}{p} \sum_{j=0}^{p-1} f\left(\frac{\lambda + 2\,\pi\, j}{p}\right), \quad \lambda \in \Pi\,;$$

b) *in case of cumulating*

$$\widetilde{f}(\lambda) = \frac{1}{p} \sum_{j=0}^{p-1} f\left(\frac{\lambda + 2\,\pi\, j}{p}\right) \phi_j(\lambda)\,, \quad \lambda \in \Pi\,,$$

where

$$\phi_j(\lambda) = \frac{\sin^2\left(\frac{\lambda}{2} + \pi\, j\right)}{\sin^2\left(\frac{\frac{\lambda}{2}+\pi\, j}{p}\right)}\,, \quad \lambda > 0\,,$$

$$\begin{aligned} \phi_0\,(\lambda) &\sim p^2\,, \\ \phi_j\,(\lambda) &\sim \frac{\lambda^2}{4\,\sin^2\left(\frac{\pi\, j}{p}\right)}\,, \quad j = 1, \ldots, p-1\,, \end{aligned}$$

as $\lambda \to 0$. Further, ϕ_j is continuously differentiable with ($\lambda \to 0$)

$$\phi_j'(\lambda) = O(\lambda)\,, \quad j = 0, \ldots, p-1\,.$$

PROOF See Appendix.

REMARK B The summation over the frequencies $\frac{\lambda+2\pi j}{p}$, $j = 0, 1, \ldots, p-1$, in Theorem 3.1a) corresponds to the well known aliasing effect that occurs when observing a continuous time process at discrete points in time, see e.g. Hansen and Sargent (1983), or the discussions in Bloomfield (1976, p.205) and Priestley (1981, p.224, p.506). The intuition behind aliasing as found in Theorem 3.1a) runs as follows: Cycles with a length ℓ or frequency λ in the original data become cycles of length ℓ/p or frequency $p\,\lambda$ upon skip sampling. Hence, the frequencies $\frac{\lambda+2\pi j}{p}$ in time scale t become $\lambda + 2\pi j$ in the aggregate time τ. Due to periodicity, the frequencies $\frac{\lambda+2\pi j}{p}$ hence all show up at frequency λ after skip sampling. The cumulated aggregate is subject to aliasing, too (Theorem 3.1b)), simply because $\{\widetilde{y}_\tau\}$ is constructed from a moving average through skip sampling. In this case, however, aliasing is superimposed by the factors ϕ_j due to the moving average filter.

3.3 Implications at frequency zero

If the impulse responses c_j of $\{y_t\}$ from (1) are absolutely summable, then

$$f(0) = \frac{1}{2\pi}\left(\sum_{j=0}^{\infty} c_j\right)^2 = \frac{1}{2\pi}\, CIR^2\,.$$

Therefore, the spectral density at frequency zero is directly linked to classical persistence measures as the cumulated impulse response function. Moreover, the spectral density of a fractionally integrated processes of order d is dominated by d in a neighbourhood of zero, see (14). Therefore, it is of particular interest how spectral properties of $\{y_t\}$ at frequency zero carry over to the aggregates.

Corollary 3.2 *If Assumption 2 is replaced by Assumption 3 in Theorem 3.1, then*

$$\dot{f}(\lambda) \sim \frac{1}{p} f\left(\frac{\lambda}{p}\right) + \frac{1}{p}\sum_{j=1}^{p-1} f\left(\frac{2\pi j}{p}\right),$$

and

$$\begin{aligned} \widetilde{f}(\lambda) &= p\, f\left(\frac{\lambda}{p}\right) + O(\lambda^2), \\ \widetilde{f}'(\lambda) &= f'\left(\frac{\lambda}{p}\right) + O(\lambda), \end{aligned}$$

as $\lambda \to 0$.

Proof Obvious from Theorem 3.1 and therefore omitted.

Remark C If $f(0) = 0$, then $\dot{f}(0) > 0$ under the assumptions of Corollary 3.2. Hence, fractional integration of order $d < 0$ is not closed in the frequency domain with respect to skip sampling, see (14), which corrects differing claims made in Chamber (1998) and Hwang (2000). This has been observed already by Souza (2005): Only if $d \geq 0$, the skip sampled aggregate inherits the spectral properties of the original series at frequency zero. It is a remarkable result, since fractional processes are known to be self-similar in that stretching the time scale leaves distributional properties unchanged upon rescaling the process, see e.g. Mandelbrot and van Ness (1968). In fact, it holds in the time domain with (8)

$$E(\dot{y}_\tau\, \dot{y}_{\tau+h}) = \gamma(p\, h) \sim c(ph)^{2d-1}, \quad h \to \infty.$$

Hence, the hyperbolic decay of the autocovariance is inherited by the skip sampled process irrespective of the sign of d, while the power law in (14) is lost for $d < 0$.

Cumulating has a different effect according to Corollary 3.2. We learn that persistence measures related to the spectral density at frequency zero are inflated by p in case of temporal aggregation of flow variables, while the slope properties are preserved at the origin.

It is interesting to study a related result for the overlapping moving average from (12). Its spectral density f_{ma} is given in (21) in the Appendix, from where we learn

$$f_{ma}(\lambda) \sim f(\lambda)\, p^2 \quad \text{as } \lambda \to 0\,.$$

Consequently, persistence measured at frequency zero is increased by p^2 through averaging. This effect may easily occur in practice when usual differences $1-L$ of seasonal data are replaced by annual differences $1-L^p$. Assume as true DGP the integrated process $z_t = z_0 + \sum_{i=1}^t y_i$ with p observations per year, where $\{y_t\}$ is stationary. Annual differencing results in a moving average of the stationary series,

$$(1-L^p)z_t = S_p(L)(1-L)z_t = S_p(L)y_t = y_t^{ma}\,.$$

For a time domain discussion of spurious persistence in such surroundings with an application to seasonal inflation data, see Hassler and Demetrescu (2005).

Finally, let $\overline{f}$ stand for the spectral density of $\{\overline{y}_\tau\}$ from (13), where obviously $\overline{f} = \widetilde{f}/p^2$. With Corollary 3.2 one obtains immediately

$$\overline{f}(\lambda) \;=\; \frac{1}{p}\, f\left(\frac{\lambda}{p}\right) + O(\lambda^2)\,,$$

as $\lambda \to 0$. Hence, persistence measures related to frequency zero are reduced by computing non-overlapping averages.

3.4 Nonstationary integration

Similarly as in (6) we want again to allow for nonstationarity where the high frequency variable $\{z_t\}$ is $I(d+1)$:

$$z_t = z_0 + \sum_{i=1}^{t} y_i\,, \quad t = 1,2,\ldots,T\,, \tag{15}$$

with $\{y_t\}$ satisfying the previous assumptions, or $(1-L)\, z_t = y_t$. Following Velasco (1999a), one may define a pseudo spectral density of the $I(d+1)$ process as

$$f_z(\lambda) = |1-e^{i\lambda}|^{-2} f_y(\lambda)\,, \quad \lambda > 0\,,$$

where $f_y(\lambda)$ belongs to $\{y_t\}$. This is just another way of stating that the spectral density of Δz_t (with $\Delta = 1-L$) is f_y:

$$f_{\Delta z}(\lambda) = f_y(\lambda)\,, \quad \lambda \in \Pi\,.$$

Now, what is the effect of aggregating $\{z_t\}$, where $\{\dot{z}_\tau\}$ and $\{\widetilde{z}_\tau\}$ are constructed as in (10) and (11), respectively? This is most easily answered in terms of differences. To this end, we define the lag operator $\mathcal{L}$ that operates on the aggregate time scale τ, such that $\mathcal{L} = L^p$ with L operating on t (see e.g. Wei (1990, Ch.16)). Results for $(1-\mathcal{L})\,\dot{z}_\tau$ and $(1-\mathcal{L})\,\widetilde{z}_\tau$ are readily available. For the skip sampled aggregate ($\tau = 1, 2, \ldots$)

$$(1-\mathcal{L})\dot{z}_\tau = S_p(L)\,\dot{y}_\tau = \dot{y}_\tau^{ma} = \widetilde{y}_\tau\,, \tag{16}$$

and similarly for flow variables:

$$(1-\mathcal{L})\widetilde{z}_\tau = S_p(L)\,\widetilde{y}_\tau = S_p(L)\,y_{p\tau}^{ma} = \widetilde{y_\tau^{ma}}\,.$$

Hence, differencing and temporal aggregation are not interchangeable. Instead, differencing the aggregates results in moving averaging differences, which holds true because

$$(1-\mathcal{L}) = (1-L^p) = S_p(L)\,(1-L)\,.$$

4 Aggregation of fractional integration

In this section we apply the previous results to fractionally integrated models characterized by (14). The case of discrete-time sampling from a continuous-time long memory process has been covered by Chambers (1996). The effect of temporal cumulation with p getting large was treated in Man and Tiao (2006).

4.1 Assumptions

Chambers (1998) and Souza (2005) assume (4) as data generating process. If we assume an ARFIMA model ($\{e_t\}$ in (4) is ARMA), and if we know p, then it is straightforward from Theorem 3.1 to write down and maximize the approximate likelihood function of the aggregate in the frequency domain. This would amount to a so-called Whittle estimation. We claim that this approach is not practical in many situations, since we do not know the "true" frequency, at which the data are generated, hence we do not know p. E.g. with monthly observations, p could be 4 (four weeks) or p could be 21 (like 21 week days). Not knowing the "true" frequency of the DGP and hence not knowing p, we are particularly interested in semiparametric methods, relying on spectral assumptions at frequency zero. At the same time knowledge about d alone suffices to compute rc_j and hl_j discussed in Section 2.

Papers on semiparametric inference of long memory typically assume that the observed process has a spectral density like in (14) where the short memory component f_e is characterized by assumptions $\mathcal{A}$ as weak as possible. We consider typical spectral assumptions next.

Assumption 4 *Let $\mathcal{A}$ be a set of assumptions for $f(\lambda) = \lambda^{-2d} f_e(\lambda)$, $d < 0.5$, including*

(A0) *f_e is bounded and bounded away from zero at frequency zero;*

(A1) *f_e has a finite first derivative f_e' in a neighboorhood $(0, \delta)$ of zero and*

$$f_e'(\lambda) = O(\lambda^{-1}), \quad \lambda \to 0;$$

(A2) *f_e has a finite first derivative f_e' at zero.*

The first assumption **(A0)** that $f_e(0)$ is bounded and positive is minimal in order to identify d from (14). Next, some papers work under the assumption that f_e' exists in a neighbourhood of zero but may diverge at appropriate rate as getting close to zero, see Assumption **(A1)**. Although put slightly differently such an assumption is found in Robinson (1995, Assumption A2) and Shimotsu and Phillips (2005, Assumption 2) when establishing consistency of the local Whittle (LW) estimator and the so-called exact LW estimator, respectively[3]. Other papers assume a stronger degree of smoothness of f_e at frequency zero in that they demand the first derivative $f_e'(0)$ to be zero or at least to be finite[4], which is our assumption **(A2)**. Hurvich, Deo and Brodsky (1998) for instance assume $f_e'(0) = 0$ when deriving the asymptotic mean squared error and limiting distribution of the log-periodogram regression (LPR) by Geweke and Porter-Hudak (1983), while Andrews and Guggenberger (2003) discuss properties of a bias-reduced version under a smoothness assumption requiring $f_e'(0)$ to exist, see also Guggenberger and Sun (2006).

We wish to investigate which set of assumptions is closed in the following sense.

Definition 1 *A set of assumptions on* $\{y_t\}$ *is called closed with respect to temporal aggregation (skip sampling, averaging or cumulating), if* $\{\dot{y}_\tau\}$, $\{\overline{y}_\tau\}$ *or* $\{\widetilde{y}_\tau\}$, *respectively, satisfy the same set of assumptions for any finite positive integer* $p \geq 2$, *too.*

For practical purposes procedures with properties established under assumptions that are closed with respect to aggregation are desirable, because in many practical situations the frequency of the "true" DGP is not known. Let us assume the high frequency process satisfies $\mathcal{A}$ justifying a semiparametric procedure, then the procedure cannot be safely applied to an aggregate, unless $\mathcal{A}$ is closed with respect to temporal aggregation.

4.2 Results

In what follows we work under Assumption 3 that f is positive, finite and differentiable at $2\pi/p, \ldots, 2\pi(p-1)/p$ because aggregation amounts to sum-

[3]See also the assumption $|f_e'(\lambda)| \leq c\,\lambda^{-1}$ for $\lambda > 0$ in Moulines and Soulier (1999, Assumption 2), and similar although slightly weaker in Soulier (2001, Assumption 1).

[4]Note that e_t from Assumption 1 satisfies **(A2)**, since in $e_t = \sum_{j=0}^{\infty} \rho_j \varepsilon_{t-j}$ with absolutely summable MA coefficients implies $f_e'(0) = 0$.

ming over those frequencies, see Theorem 3.1. The usual long memory literature not addressing the aggregation issue does not need Assumption 3. In case of flow variables with (14), Corollary 3.2 motivates the following closedness properties.

Proposition 4.1 *Let $\{y_t\}$ be a stationary process with spectral density as in (14) with Assumption 3. It then holds with respect to cumulated sampling of flow variables:*

a) ***(A0)*** *is closed if and only if $d \geq -1$;*

b) ***(A1)*** *is closed if and only if $d \geq -1$;*

c) ***(A2)*** *is closed if and only if $d \geq -0.5$.*

Proof See Appendix.

The proof relies on a decomposition of the spectral density into a part due to integration and a short memory component $\widetilde{\varphi}$,

$$\widetilde{f}(\lambda) = \lambda^{-2d}\widetilde{\varphi}(\lambda)\,.$$

The latter turns out to be so smooth at the origin that **(A0)** and **(A1)** or even **(A2)** hold for $\widetilde{\varphi}$ as long as they hold for f_e; hence these conditions are closed in the sense of Definition 1 with respect to aggregation of flow variables wit $d \geq -0.5$. In particular, Proposition 4.1a) confirms the finding by Souza (2005) that the order of fractional integration is maintained under cumulated aggregation of flow variables (as long as $d \geq -1$). Similarly, assumptions about the spectral slope behaviour are inherited by the aggregate from the original series, so that usual semiparametric estimators like LPR or LW may be safely applied. In fact, Ohanissian, Russell and Tsay (2008) proposed a test on whether there is true long memory (i.e. fractional integration) or not, that builds on the LPR. It compares differences of estimators obtained from M different cumulated aggregates with $p_1 < \cdots < p_M$ of a series. Under the null hypothesis the difference between the estimators from the aggregates will vanish. The proof in Ohanissian, Russell and Tsay (2008) heavily relies on Soulier (2001), whose Assumption 1 (similar to our **(A1)**) is closed with respect to cumulated sampling.

In case of averaging stock variables as defined in (13), the spectral density becomes $\overline{f} = \widetilde{f}/p^2$. Hence, the properties of $\widetilde{f}$ that establish Proposition 4.1 carry over to $\overline{f}$. The next corollary follows immediately without further proof.

Corollary 4.2 *Let $\{y_t\}$ be a stationary process with spectral density as in (14) with Assumption 3. It then holds with respect to averaging of stock variables as in (13):*

a) ***(A0)*** *is closed if and only if $d \geq -1$;*

b) ***(A1)*** *is closed if and only if $d \geq -1$;*

c) ***(A2)*** *is closed if and only if $d \geq -0.5$.*

The results for skip sampling very much differ from the previous ones. The reason can be seen from Theorem 3.1a): The effect of the frequencies $\frac{\lambda+2\pi j}{p}$ on $\dot{f}$ is not negligible in a vicinity of zero. The consequences on fractional integration are given in the next proposition.

Proposition 4.3 *Let $\{y_t\}$ be a stationary process with spectral density as in (14) with Assumption 3 and **(A0)**. In case of skip sampling it holds that*

a) *the spectral density is given as*

$$\dot{f}(\lambda) = \lambda^{-2d}\dot{\varphi}(\lambda)$$

$$\dot{\varphi}(\lambda) \sim \gamma_0 + \gamma_1 \lambda^{2d}, \tag{17}$$

as $\lambda \to 0$ with $\gamma_0 = p^{2d-1} f_e(0)$, $0 < \gamma_1 < \infty$;

b) ***(A0)*** *is closed if and only if $d \geq 0$;*

c) ***(A1)*** *is closed if and only if* $d \geq 0$*;*

d) ***(A2)*** *is not closed for all values of* d*.*

Proof See Appendix.

Remark D Skip sampling preserves Assumption **(A1)** for $d \geq 0$. However, the short memory component $\dot{\varphi}$ displays an unbounded derivative at the origin for all $d < 0.5$ even if $f_e'(0)$ is finite, and consequently **(A2)** is never closed. Sufficient conditions for consistency or limiting normality of some estimators mentioned before are hence violated under skip sampling. This is a serious drawback at first glance. However, the behaviour of $\dot{\varphi}$ at the origin given in (17) is such that a repair proposal suggests itself, see the next subsection.

Remark E Proposition 4.3b) deserves separate consideration. With $\dot{\varphi}$ being of order λ^{2d}, **(A0)** is not closed for negative d. For $d < 0$, we get $\dot{f}(0) = \gamma_1 > 0$, and the aggregate $\dot{y}_\tau$ loses the spectral properties of $y_t \sim I(d)$ with $d < 0$, which confirms the point made by Souza (2005) and seen from Corollary 3.2 already, see Remark C. How relevant is this lack of closedness in the frequency domain in practice? Assume that some variable z_t is integrated of order δ between 0.75 and 1, such that limiting normality of estimators from the LPR or LW is no longer guaranteed, see Velasco (1999 a,b). Consequently, people have worked with differences $y_t = z_t - z_{t-1}$ in applied papers, where y_t is $I(d)$ with $d = \delta - 1 < 0$. What happens in case of skip sampling? In the unlucky case where you skip sample the differences $y_t \sim I(d)$ and try to estimate d from $\dot{y}_\tau$, any frequency-domain based estimate $\widehat{d}$ will tend towards zero even asymptotically because $\dot{y}_\tau$ is $I(0)$. While this seems worrisome at first glance, we argue that this is an unlikely case in practice. More typically, one aggregates the level, $\dot{z}_\tau$. As we know from (16), the differences thereof behave like $\widetilde{y}_\tau$. Consequently, the order $d+1$ of z_t can be discovered consistently from $(1 - L^p)\dot{z}_\tau = \widetilde{y}_\tau$, see eq. (16).

4.3 Perturbed fractional integration

Motivated by Remark D after Proposition 4.3, we now briefly review long memory ($0 < d$) estimation under perturbed integration.

Let $\{x_t\}$ be fractionally integrated perturbed by some $I(0)$ process $\{u_t\}$,

$$x_t = y_t + u_t \,, \tag{18}$$

where we assume that $\{u_t\}$ is independent of the unobservable process $\{y_t\}$. Given $\{y_t\}$ is fractionally integrated with (14) it holds in the frequency domain

$$f_x(\lambda) = \lambda^{-2d} f_e(\lambda) + f_u(\lambda) = \lambda^{-2d} \varphi(\lambda)$$

where the short memory component of the observable $\{x_t\}$ becomes

$$\begin{aligned} \varphi(\lambda) &= f_e(\lambda) + f_u(\lambda)\, \lambda^{2d} \\ &\sim c_0 + c_1\, \lambda^{2d} \,, \quad \lambda \to 0 \,, \end{aligned} \tag{19}$$

with $c_0 = f_e(0)$ and $c_1 = f_u(0)$. For $0 < d$, the perturbed process $\{x_t\}$ is fractionally integrated of order d where the short memory component behaves like in case of skip sampling, see (17). Therefore, methods taylored to the estimation of d from $\{x_t\}$ in (18) are candidates for the estimation of d from skip sampled series. For that reason, a short and informal review of some related papers is provided.

Most papers dealing with perturbed fractional integration (also called "long memory plus noise") are related to the so-called long memory stochastic volatility model (LMSV) introduced by Breidt, Crato and de Lima (1998) or the FIEGARCH model by Bollerslev and Mikkelsen (1996), see also the introductory section by Hurvich, Moulines and Soulier (2005). Such volatility models assume for return processes $\{r_t\}$ that

$$\log r_t^2 = \mu + y_t + \varepsilon_t \,, \tag{20}$$

where the perturbation term $\{\varepsilon_t\}$ is white noise. Deo and Hurvich (2001) establish consistency and limiting normality of the LPR estimator within an

LMSV framework. To hold true the number of included harmonic frequencies (or bandwidth m) has to be bounded by $T^{4d/(4d+1)}$, which is all the more problematic the closer d is to zero. The same kind of bound is found by Sun and Phillips (2003) for the more general model (18) under Gaussianity. Further, Sun and Phillips (2003) propose a improved nonlinear version of the LPR estimator that accounts explicitly for the effect of perturbation. Arteche (2004) studied the model (18) without the assumption of Gaussianity and found the LW estimator to be consistent and asymptotically normal with the bandwidth obeying the same restriction as in Deo and Hurvich (2001). Hurvich and Ray (2003) proposed a modification of the LS estimator adjusted explicitly for the noise effect of model (20); further refinements are provided by Hurvich, Moulines and Soulier (2005) in that correlation between y_t and ε_t is allowed for. Finally, it should be noted that the so-called broadband log-periodogram regression by Moulines and Soulier (1999) remains valid for a Gaussian LMSV model, see Iouditsky, Moulines and Soulier (2001).

5 Concluding remarks

Appendix

Proof of Proposition 2.1

The proof considers three cases separately.
1) The stationary case ($\delta = d < 0.5$): The impulse response IR_j from (5) is split into two sums,

$$\begin{aligned} IR_j &= \Sigma_{1,j} + \Sigma_{2,j} \\ &= \sum_{k \le \sqrt{j}} \rho_k \, \psi_{j-k,d} + \sum_{k > \sqrt{j}} \rho_k \psi_{j-k,d} \,. \end{aligned}$$

For the second sum we obtain due to s-summability with $s = 2 - 2\,d$

$$\begin{aligned} j^{1-d}\,|\Sigma_{2,j}| &\leq \sum_{k>\sqrt{j}} k^{2-2\,d}|\rho_k\,\psi_{j-k,d}| \\ &\leq \sup\{|\psi_{j-k,d}|\} \sum_{k>\sqrt{j}} k^{2-2\,d}|\rho_k| \to 0\,, \end{aligned}$$

because $\sum_{k\geq 0} k^{2-2\,d}|\rho_k| < \infty$. For the first sum one gets with (3) because of $\psi_{j,d} \sim \psi_{j-k,d}$ as $k \leq \sqrt{j}$:

$$j^{1-d}\Sigma_{1,j} = \sum_{k\leq\sqrt{j}} \rho_k\, j^{1-d}\,\psi_{j-k,d} \to \sum_{k=0}^{\infty} \rho_k\,\frac{1}{\Gamma(d)}\,.$$

Hence

$$j^{1-d}c_{j,d} \to \frac{\rho(1)}{\Gamma(d)} \quad \text{as } j \to \infty\,,$$

as required for the stationary case.

2) The nonstationary transitory case ($0.5 \leq \delta < 1$): Define $d = \delta - 1$, and note that

$$\sum_{k=0}^{\infty} c_{k,d} = \sum_{k=0}^{\infty} \rho_k z^k\,(1-z)^{|d|}\Bigg|_{z=1} = 0\,.$$

Therefore with IR_j from (7),

$$IR_j = -\sum_{k=j+1}^{\infty} c_{k,d} \sim -\frac{\rho(1)}{\Gamma(d)} \sum_{k=j+1}^{\infty} k^{d-1}\,,$$

where the validity of the approximation has just been established under $s = 2 - 2\,d$. Since k^{d-1} is monotonically decreasing, it is straightforward to see that

$$j^{-d} \sum_{k=j+1}^{\infty} k^{d-1} \to -\frac{1}{d}\,, \quad j \to \infty\,.$$

This in turn establishes

$$j^{1-\delta} IR_j \to \frac{\rho(1)}{\Gamma(d)d} = \frac{\rho(1)}{\Gamma(\delta)}\,,$$

as required.
3) The $I(1)$ case ($\delta = 1$): With $\delta - 1 = 0$ one obtains $\psi_{k,0} = 0$ for $k > 0$, and hence $c_{k,0} = \rho_k$, such that from (7)

$$IR_j = \sum_{k=0}^{j} \rho_k \to \frac{\rho(1)}{\Gamma(1)},$$

which completes the proof.

Proof of Theorem 3.1

a) The proof of the first result is similar to the one in Bloomfield (1976, p.205) for the time continuous case. By symmetry and periodicity we have

$$\gamma(h) = 2\int_0^{\pi} f(\lambda)\cos(\lambda\, h)d\lambda = \int_0^{2\pi} f(\lambda)\cos(\lambda\, h)d\lambda\,.$$

Let $\dot\gamma(h)$ denote the autocovariances of $\dot y_\tau$. Hence, $\dot f(\lambda)$ can be discovered from $\dot\gamma(h) = \int_0^{2\pi} \dot f(\lambda)\cos(\lambda\, h)d\lambda$. Elementary considerations yield

$$\begin{aligned}
\dot\gamma(h) &= \gamma(p\, h) = \int_0^{2\pi} f(\lambda)\cos(\lambda\, p\, h)d\lambda \\
&= \sum_{j=0}^{p-1} \int_{\frac{2\pi j}{p}}^{\frac{2\pi(j+1)}{p}} f(\lambda)\cos(\lambda\, p\, h)d\lambda \\
&= \sum_{j=0}^{p-1} \int_0^{2\pi} \frac{1}{p} f\left(\frac{\omega + 2\,\pi\, j}{p}\right)\cos(\omega\, h + 2\,\pi\, j\, h)d\omega \\
&= \int_0^{2\pi} \left\{\sum_{j=0}^{p-1} \frac{1}{p} f\left(\frac{\omega + 2\,\pi\, j}{p}\right)\right\}\cos(\omega\, h)d\omega\,.
\end{aligned}$$

Hence, $\dot f$ in braces has the required shape.
b) Since $\widetilde{y}_\tau$ is obtained from skip sampling an overlapping moving average, we study the effect of a moving average filter first. We express the moving average in terms of filter polynomials in the lag operator L:

$$y_t^{ma} = S_p(L)\, y_t = \frac{1 - L^p}{1 - L}\, y_t\,.$$

The corresponding spectral density is given as (see e.g. Priestley, 1981, p.268)

$$\begin{aligned} f_{ma}(\lambda) &= \left|S_p\left(e^{i\lambda}\right)\right|^2 f(\lambda) = \frac{|1-e^{ip\lambda}|^2}{|1-e^{i\lambda}|^2} f(\lambda) \\ &= \begin{cases} f(\lambda)\,\frac{\sin^2(\frac{p\,\lambda}{2})}{\sin^2(\frac{\lambda}{2})}\,, & \lambda \in \Pi \setminus \{0\} \\ f(\lambda)\,p^2\,, & \lambda \to 0 \end{cases} . \end{aligned} \tag{21}$$

The spectral density $\widetilde{f}$ is now obtained by applying **a)** to the moving average spectral density:

$$\widetilde{f}(\lambda) = \frac{1}{p}\sum_{j=0}^{p-1} f_{ma}\left(\frac{\lambda + 2\,\pi\,j}{p}\right) .$$

One may now verify that the ϕ_j coincide with the so-called Fejer kernel F_p of order p,

$$\phi_j(\lambda) = p\,F_p\left(\frac{\lambda + 2\pi j}{p}\right) , \tag{22}$$

where F_p is constructed from the Dirichlet kernel D_n of order n as follows (see e.g. Priestley, 1981, p.400)

$$\begin{aligned} D_n(x) &= \sum_{k=-n}^{n} \cos(k\,x) \\ &= \frac{\sin\left((n+0.5)\,x\right)}{\sin(x/2)}\,, \quad x \neq 2m\pi\,, \end{aligned}$$

$$F_p(x) = \frac{1}{p}\sum_{n=0}^{p-1} D_n(x) \tag{23}$$

$$= \frac{1}{p}\left(\frac{\sin\left(\frac{p\,x}{2}\right)}{\sin\left(\frac{x}{2}\right)}\right)^2 , \quad x \neq 2m\pi\,. \tag{24}$$

Properties of F_p will establish the remaining results. Using

$$\sin\left(\frac{\lambda}{2} + \pi\,j\right) \sim \frac{\lambda}{2}\cos(\pi\,j)\,, \quad \lambda \to 0\,, \tag{25}$$

for $j = 0, 1, \ldots, p-1$ it follows

$$\begin{aligned} F_p\left(\frac{\lambda}{p}\right) &\sim p\,, \\ F_p\left(\frac{\lambda + 2\pi j}{p}\right) &\sim \frac{\lambda^2}{4\,p\,\sin^2\left(\frac{\pi j}{p}\right)}\,, \quad j = 1, \ldots, p-1\,. \end{aligned} \tag{26}$$

From (23) it is obvious that F_p is everywhere continuously differentiable, and $F_p'(\lambda/p) = O(\lambda)$. Elementary manipulations of (24) establish

$$F_p'\left(\frac{\lambda + 2\pi j}{p}\right) = O(\lambda)\,, \quad j = 1, \ldots, p-1\,. \tag{27}$$

Hence the proof is complete.

Proof of Proposition 4.1

Theorem 3.1b) provides under (14)

$$\widetilde{f}(\lambda) = \lambda^{-2d} p^{2d-1} f_e\left(\frac{\lambda}{p}\right)\,\phi_0(\lambda) + R(\lambda)$$

where with (22)

$$R(\lambda) = \sum_{j=1}^{p-1} f\left(\frac{\lambda + 2\pi j}{p}\right)\,F_p\left(\frac{\lambda + 2\pi j}{p}\right)\,.$$

Under Assumption 3 we obtain for $R(\lambda)$ and its derivative from (26) and (27)

$$R(\lambda) = O(\lambda^2)\,,\ R'(\lambda) = O(\lambda)\,, \quad \lambda \to 0\,.$$

Consequently, the spectral density can be decomposed into a part due to integration and a short memory component $\widetilde{\varphi}$, $\widetilde{f}(\lambda) = \lambda^{-2d}\widetilde{\varphi}(\lambda)$. The latter is given by

$$\widetilde{\varphi}(\lambda) \ = \ p^{2d-1} f_e\left(\frac{\lambda}{p}\right)\,\phi_0(\lambda) + \lambda^{2d} R(\lambda)\,. \tag{28}$$

With the derivative R' it further holds under Assumption 3

$$\widetilde{\varphi}'(\lambda) \;=\; p^{2d-1} f_e' \left(\frac{\lambda}{p}\right) \frac{\phi_0(\lambda)}{p} + O(\lambda) + O(\lambda^{2d+1})\,.$$

If **(A0)** holds for f_e, than $\widetilde{\varphi}$ from (28) is finite if and only if $d \geq -1$, which hence proves **a)**. The results **b)** and **c)** are verified analogously by studying the behaviour of $\widetilde{\varphi}'$ given f_e' satisfies **(A1)** or **(A2)**, respectively.

Proof of Proposition 4.3

a) Theorem 3.1 yields

$$\dot{f}(\lambda) \;=\; \frac{1}{p}\left[\left[\frac{\lambda}{p}\right]^{-2d} f_e\left(\frac{\lambda}{p}\right) + \sum_{j=1}^{p-1} f\left(\frac{\lambda + 2\,\pi\, j}{p}\right)\right], \tag{29}$$

such that

$$\dot{\varphi}(\lambda) = p^{2d-1} f_e\left(\frac{\lambda}{p}\right) + \frac{\lambda^{2d}}{p}\sum_{j=1}^{p-1} f\left(\frac{\lambda + 2\,\pi\, j}{p}\right). \tag{30}$$

Hence, γ_1 becomes

$$0 < \gamma_1 = p^{-1}\sum_{j=1}^{p-1} f\left(\frac{2\,\pi\, j}{p}\right) < \infty\,,$$

and the definition of γ_0 is obvious. The statement follows from (30) under Assumption 3.

b) Under Assumption **(A0)** it follows with Assumption 3 from (17) that $\dot{\varphi}(0)$ is finite if and only if $d \geq 0$; which proves the result.

c) The derivative of $\dot{\varphi}$ is

$$\dot{\varphi}'(\lambda) = p^{2d-2} f_e'\left(\frac{\lambda}{p}\right) + 2d\,\gamma_1 \lambda^{2d-1} + O(\lambda^{2d})\,. \tag{31}$$

Given the derivative of f_e' satisfies Assumption **(A1)** one obtains:

$$\dot{\varphi}'(\lambda) = O(\lambda^{-1}) + O(\lambda^{2d-1}) + O(\lambda^{2d}) = O(\lambda^{\min(-1,2d-1)})\,;$$

which proves the result.

d) For f_e' satisfying Assumption **(A2)** we obtain directly from (31) that $\dot{\varphi}'(\lambda)$ diverges since $d < 0.5$. This completes the proof.

References

[1] Andrews, D.W.K., and P. Guggenberger (2003), A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter; *Econometrica* 71, 675-712.

[2] Arteche, J. (2004), Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models; *Journal of Econometrics* 119, 131-154.

[3] Baillie, R.T., and G. Kapetanios (2008), Nonlinear Models for Strongly Dependent Processes with Financial Applications; *Journal of Econometrics* 147, 60-71.

[4] Bloomfield, P. (1976), *Fourier Analysis of Time Series: An Introduction*; Wiley.

[5] Bollerslev, T., and H.O. Mikkelsen (1996), Modeling and Pricing Long Memory in Stock Market Volatility; *Journal of Econometrics* 73, 151-184.

[6] Bondon, P., and W. Palma (2007), A Class of Antipersistent Processes; *Time Series Analysis* 28, 261-273.

[7] Breidt, F.J., N. Crato and P. de Lima (1998), The Detection and Estimation of Long Memory in Stochastic Volatility; *Journal of Econometrics* 83, 325-348.

[8] Brewer, K.R.W. (1973), Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models; *Journal of Econometrics* 1, 133-154.

[9] Campbell, J.Y., and N.G. Mankiw (1987), Are Output Fluctuations Transitory? *Quarterly Journal of Economics* 102, 857-880.

[10] Chambers, M.J. (1996), The Estimation of Continuous Parameter Long-Memory Time Series Models; *Econometric Theory* 12, 374-390.

[11] Chambers, M.J. (1998), Long Memory and Aggregation in Macroeconomic Time Series; *International Economic Review* 39, 1053-1072.

[12] Chambers, M.J. (2004), Testing for Unit Roots with Flow Data and Varying Sampling Frequency; *Journal of Econometrics* 119, 1-18.

[13] Choi, I. (1992), Effects of data aggregation on the Power of Tests for a Unit Root; *Economics Letters* 40, 397-401.

[14] Christiano, L.J., Eichenbaum, M., and D. Marshall (1991), The Permanent Income Hypothesis Revisited; *Econometrica* 59, 397-423.

[15] Cochrane, J.H. (1988), How Big is the Random Walk in GNP?; *Journal of Political Economy* 96, 893-920.

[16] Deo, R.S., and C.M. Hurvich (2001), On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models; *Econometric Theory* 17, 686-710.

[17] Geweke, J. (1978), Temporal Aggregation in the Multiple Regression Model; *Econometrica* 46, 643-661.

[18] Geweke, J., and S. Porter-Hudak (1983), The Estimation and Application of Long Memory Time Series Models; *Journal of Tima Series Analysis* 4, 221-238.

[19] Granger, C.W.J., and R. Joyeux (1980), An Introduction to Long-Memory Time Series Models and Fractional Differencing; *Journal of Time Series Analysis* 1, 15-29.

[20] Granger, C.W.J., and L. Siklos (1995), Systematic Sampling, Temporal Aggregation, Seasonal Adjustment, and Cointegration: Theory and Evidence; *Journal of Econometrics* 66, 357-370.

[21] Guggenberger, P., and Y. Sun (2006), Bias-Reduced Log-Periodogram and Whittle Estimation of the Long-Memory Parameter Without Variance Inflation; *Econometric Theory* 22, 863-912.

[22] Hansen, L.P., and Th.J. Sargent (1983), The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities; *Econometrica* 51, 377-387.

[23] Hassler, U., and M. Demetrescu (2005), Spurious Persistence and Unit Roots due to Seasonal Differencing: The Case of Inflation Rates; *Journal of Economics and Statistics* 225, 413-426.

[24] Hauser, M.A., B.M. Pötscher, and E. Reschenhofer (1999), Measuring Persistence in Aggragate Output: ARMA Models, Fractionally Integrated ARMA Models and Nonparametric Procedures; *Empirical Economics* 24, 243-269.

[25] Hosking, J.R.M. (1981), Fractional Differencing; *Biometrika* 68, 165-176.

[26] Hurvich, C.M., and B.K. Ray (2003), The Local Whittle Estimator of Long-Memory Stochastic Volatility; *Journal of Financial Econometrics* 1, 445-470.

[27] Hurvich, C.M., and P. Soulier (2002), Testing for Long Memory in Volatility; *Econometric Theory* 18, 1291-1308.

[28] Hurvich, C.M., R. Deo, and J. Brodsky (1998), The mean squared error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series; *Journal of Time Series Analysis* 19, 19-46.

[29] Hurvich, C.M., E. Moulines, and Ph. Soulier (2005), Estimating Long Memory in Volatility; *Econometrica* 73, 1283-1328.

[30] Hwang, S. (2000), The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and Their Finite Sample Properties; *Econometric Theory* 16, 347-372.

[31] Iouditsky, A., E. Moulines and Ph. Soulier (2001), Adaptive Estimation of the Fracional Differencing Coefficient; *Bernoulli* 7, 699-731.

[32] Kumar, M.S., and T. Okimoto (2007), Dynamics of Persistence in International Inflation Rates; *Journal of Money, Credit and Banking* 39, 1457- 1479.

[33] Lükepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*; Springer.

[34] Man, K.S., and G.C. Tiao (2006), Aggregation Effect and Forecasting Temporal Aggregates of Long Memory Process; *International Journal of Forecasting* 22, 267-281.

[35] Mandelbrot, B.B., and J.W. van Ness (1968), Fractional Brownian Motion, Fractional Noises and Applications; *SIAM Review* 10, 422-437.

[36] Marinucci, D., and P.M. Robinson (1999), Alternative Forms of Fractional Brownian Motion; *Journal of Statistical Planning and Inference* 80, 111-122.

[37] Moulines, E., and Ph. Soulier (1999), Broadband Log-Periodogram Regression of Time Series with Long-Range Dependence; *The Annals of Statistics* 27, 1415-1439.

[38] Ng, S. (1995), Testing for Unit Roots in Flow Data Sampled at Different Frequencies; *Economics Letters* 47, 237-242.

[39] Odaki, M. (1993), On the Invertibility of Fractionally Differenced ARIMA Processes; *Biometrika* 80, 703-709.

[40] Ohanissian, A., J.R. Russell and R.S. Tsay (2008), True or Spurious Long Memory? A New Test; *Journal of Business & Economic Statistics* 26, 161-175.

[41] Palm, F.C., and Th.E. Nijman (1984), Missing Observations in the Dynamic Regression Model; *Econometrica* 52, 1415-1435.

[42] Palma, W. (2007), *Long-Memory Time Series: Theory and Methods*; Wiley.

[43] Paya, I., Duarte, A., and K. Holden (2007), On the Relationship between Inflation Persistence and Temporal Aggregation; *Journal of Money, Credit and Banking* 39, 1521-1531.

[44] Phillips, P.C.B., and Z. Xiao (1999), A Primer on Unit Root Testing, *Journal of Economic Surveys* 12, 423-470.

[45] Priestley, M.B. (1981), *Spectral Analysis and Time Series, Vol. 1*; Academic Press.

[46] Perron, P. (1991), Test Consistency with Varying Sampling Frequency; *Econometric Theory* 7, 341-368.

[47] Pons, G. (2006), Testing Monthly Seasonal Unit Roots with Monthly and Quarterly Information; *Journal of Time Series Analysis* 27, 191-210.

[48] Poskitt, D.S. (2007), Autoregressive Approximation in Nonstandard Situations: The Fractionally Integrated and Non-invertible Cases; *Annals of the Institute of Statistical Mathematics* 59, 697-725.

[49] Robinson, P.M. (1995), Gaussian Semiparametric Estimation of Long Range Dependence; *Annals of Statistics* 23, 1630-1661.

[50] Robinson, P.M. (2005), The Distance between Rival Nonstationary Fractional Processes; *Journal of Econometrics* 128, 283-300.

[51] Shiller, R.J., and P. Perron (1985), Testing the Random Walk Hypothesis; *Economics Letters* 18, 381-386.

[52] Shimotsu, K., and P.C.B. Phillips (2005), Exact Local Whittle Estimation of Fractional Integration; *The Annals of Statistics* 33, 1890-1933.

[53] Sims, C.A. (1971), Discrete Approximations to Continuous Time Distributed Lags in Econometrics; *Econometrica* 39, 545-563.

[54] Soulier, Ph. (2001), Moment Bounds and Central Limit Theoren for Functions of Gaussian Vectors; *Statistics & Probability Letters* 54, 193-203.

[55] Souza, L.R. (2005), A Note on Chambers's "Long Memory and Aggregation in Macroeconomic Time Series"; *International Economic Review* 46, 1059-1062.

[56] Souza, L.R. (2007), Temporal Aggregation and Bandwidth Selection in Estimating Long Memory; *Journal of Time Series Analysis* 28, 701-722.

[57] Souza, L.R. (2008), Why Aggregate Long Memory Time Series?; *Econometric Reviews* 27, 298-316.

[58] Souza, L.R., and J. Smith (2002), Bias in the Memory Parameter for Different Sampling Rates; *International Journal of Forecasting* 18, 299-313.

[59] Stram, D.O., and W.W.S. Wei (1986), Temporal aggregation in the ARIMA Process; *Journal of Time Series Analysis* 7, 279-292.

[60] Sun, Y. and P.C.B. Phillips (2003), Nonlinear Log-Periodogram Regression for Perturbed Fractional Processes; *Journal of Econometrics* 115, 355-389.

[61] Velasco, C. (1999a), Gaussian Semiparametric Estimation of Non-stationary Time Series; *Journal of Time Series Analysis* 20, 87-127.

[62] Velasco, C. (1999b), Non-Stationary Log-Periodogram Regression; *Journal of Econometrics* 91, 325-372.

[63] Wei, W.W.S. (1981), Effect of Systematic Sampling on ARIMA Models; *Communications in Statistics - Theory & Methods* A10 (23), 2389-2398.

[64] Wei, W.W.S. (1990), *Time Series Analysis: Univariate and Multivariate Methods*; Addison-Wesley.

[65] Weiss, A.A. (1984), Systematic sampling and temporal aggragation in time series models; *Journal of Econometrics* 26, 271-282.

[66] Woodward, W.A., Q.C. Cheng, and H.L. Gray (1998), A k-Factor GARMA Long-Memory Model; *Journal of Time Series Analysis* 19, 485–504.