

Generating Univariate Fractional Integration within a Large VAR(1)

Guillaume Chevillon, Alain Hecq, Sébastien Laurent

▶ **To cite this version:**

Guillaume Chevillon, Alain Hecq, Sébastien Laurent. Generating Univariate Fractional Integration within a Large VAR(1). 2018. halshs-01944588

HAL Id: halshs-01944588
https://shs.hal.science/halshs-01944588

Preprint submitted on 4 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Generating Univariate Fractional Integration within a Large VAR(1)

Guillaume Chevillon
Alain Hecq
Sébastien Laurent

WP 2018 - Nr 44

Generating Univariate Fractional Integration within a Large VAR(1)*

Guillaume Chevillon† Alain Hecq‡ Sébastien Laurent§
ESSEC Business School Maastricht University Aix-Marseille University

This version: December 4, 2018

Abstract

This paper shows that a large dimensional vector autoregressive model (VAR) of finite order can generate fractional integration in the marginalized univariate series. We derive high-level assumptions under which the final equation representation of a VAR(1) leads to univariate fractional white noises and verify the validity of these assumptions for two specific models.

Keywords: Long memory, Vector Autoregressive Model, Marginalization, Final Equation Representation.

JEL: C10, C32, C55.

*We would like to thank the participants at the 6th French Econometrics Conference, the 2014 Econometrics Conference at the Oxford Martin School, CFE 2013, the 15th OxMetrics User Conference at CASS Business School, the 2015 Symposium of the SNDE, the 4th Long-Memory Symposium at CREATES, the 2016 Summer Forum of Barcelona GSE, the 2016 symposium of the IAAE, the 2016 AMES in Kyoto, Aix-Marseille University, CREATES and CREST seminars, and in particular Karim Abadir, Richard Baillie, Niels Haldrup, David Hendry, Jurgen Doornik, Sophocles Mavroeidis, Bent Nielsen, Anders Rahbek, Peter Robinson and Paolo Zaffaroni. We also thank the Editor, Oliver Linton, and two anonymous referees for their helpful comments and suggestions. Financial support from EIF and Labex Louis Bachelier is gratefully acknowledged by the three authors. Chevillon also benefited from support by CREST and the ESSEC Foundation.

†Corresponding Author: Department of Information Systems, Decision Sciences and Statistics, ESSEC Business School, Avenue Bernard Hirsch, 95021 Cergy-Pontoise, France. Email: chevillon@essec.edu

‡Department of Quantitative Economics, School of Business and Economics, Maastricht University, The Netherlands. Email: a.hecq@maastrichtuniversity.nl

§Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE & Aix-Marseille Graduate School of Management-IAE, France. Email: sebastien.laurent@univ-amu.fr

1 Introduction and motivations

Long memory is commonly observed in many fields, dating back at least to Smith (1938), Cox and Townsend (1947) and Hurst (1951). However, as argued by Cox (2014), the origin of long memory is unclear. Müller and Watson (2008) show that this may be due to the fact that very large samples are needed to discriminate between the various models generating strong dependence at low frequencies. Hence several competing models of long range dependence have been proposed in the literature, see the recent overview by Haldrup and Vera-Valdés (2017) and references therein. For a covariance stationary process z_t, long memory of degree d is often defined, as in Beran (1994) or Baillie (1996), through the behavior of its spectral density $f_z(\omega)$ about the origin: $f_z(\omega) \sim c_f \omega^{-2d}$, as $\omega \to 0^+$, for some positive c_f. Since Granger and Joyeux (1980), fractional integration of order d, denoted $\mathsf{I}(d)$, has proved the most pervasive example of long memory processes in econometrics and statistics. The prototypical example of an $\mathsf{I}(d)$ process is the fractional white noise $z_t = (1-L)^{-d} \epsilon_t$, where L denotes the lag operator and ϵ_t is a white noise sequence. The class of fractionally integrated processes extends to cases where ϵ_t admits a covariance stationary ARMA representation.

To the best of our knowledge five reasons have been put forward in the literature so far to explain the presence of long range dependence: (i) aggregation across heterogeneous series, frequencies or economic agents (inter alia Granger, 1980, Gonçalves and Gouriéroux, 1988, Chambers, 1998, Abadir and Talmain, 2002, Zaffaroni, 2004); (ii) linear modeling of a nonlinear underlying process (e.g., Davidson and Sibbertsen, 2005, Miller and Park, 2010); (iii) structural change (Parke, 1999, Diebold and Inoue, 2001, Perron and Qu, 2007); (iv) learning by economic agents in forward looking models of expectations (Chevillon and Mavroeidis, 2017) and (v) network effects (Schennach, 2013).

The contribution of this paper is to provide conditions under which fractional integration arises as the result of the marginalization of a large dimensional linear multivariate system, hence being caused by hidden dependence across variables within a system. Our result is about the *existence* of such conditions and we do not aim to give a general framework for the conditions under which marginalization generates fractional integration. More specifically, we provide an asymptotic para-

metric framework under which the variables entering an n-dimensional vector autoregressive process of finite order (here a VAR(1)) can be individually modelled as fractional white noises as $n \to \infty$. Long memory may therefore be a feature of univariate or low dimensional models that vanishes when considering larger systems. The source of long memory identified here differs distinctly from the five sources listed above, and in particular, from the aggregation mechanism à la Granger (1980). Indeed, the mechanism that leads here to long memory does not rely on heterogeneity assumptions, nor does it involve aggregation.

The intuition behind our theoretical result is the following. We consider a simple VAR(1) model $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, for $t = 1, ..., T$, where $(\mathbf{A}_n)$ denotes a sequence of n-dimensional square Toeplitz matrices.[1] We use the final equation representation of this model (as proposed by Zellner and Palm, 1974) to prove that some of the marginalized processes x_{jt} belonging to $\mathbf{x}_t$ (for $j = 1, ..., n$) may converge weakly, as $n \to \infty$ but holding j fixed, to a long memory process of order $\delta \in (0, 1)$. Our result is not about empirical inference about the process but relates to the population properties of the marginalized series. We introduce a high-level assumption concerning $(\mathbf{A}_n)$. Under this assumption, the moving average lag polynomial associated with x_{jt} is asymptotically (throughout the paper, this means as $n \to \infty$) proportional to the ratio of $\det(\mathbf{I}_{n-1} - \mathbf{A}_{n-1}L)$ over $\det(\mathbf{I}_n - \mathbf{A}_n L)$. We parameterize $\mathbf{A}_n$ by defining a scalar sequence (δ_n) with $\delta_n \in (0, 1)$ such that $\lim_{n\to\infty} \delta_n = \delta \in (0, 1)$, and a circulant matrix $\mathbf{C}_n$ such that the polynomials $\det(\mathbf{I}_n - \mathbf{A}_n z) \sim \det(\mathbf{I}_n - \mathbf{C}_n z)$ as $n \to \infty$. $\mathbf{C}_n$ is assumed to possess close to a fraction $n\delta$ of unit eigenvalues. We then use the *first theorem* of Szegö (1915) to prove that under our high-level assumptions, as $n \to \infty$, $\det(\mathbf{I}_{n-1} - \mathbf{A}_{n-1} z) / \det(\mathbf{I}_n - \mathbf{A}_n z) \to (1 - z)^{-\delta}$. From this, we prove weak convergence of the process x_{jt} to the fractional white noise $(1 - L)^{-\delta} \epsilon_t$.

We then show that the high level assumption is satisfied for at least two specific examples of VAR(1) models. In the first parameterization, $\mathbf{A}_n$ denotes a Toeplitz matrix with diagonal elements converging to $\delta = 1/2$ as $n \to \infty$, and with vanishing off-diagonal elements. Importantly, the off-

[1] The class of Toeplitz matrices is chosen for the technical reason that we use in our proofs the well-established First Theorem of Szegö (1915). The theory of Toeplitz matrices also has the advantage that it provides a simple way to reduce the dimensionality of the VAR system, from n^2 parameters in a generic matrix $\mathbf{A}_n$ to n in the Toeplitz case.

diagonal elements decrease at an $O\left(n^{-1}\right)$ rate and the sum of each row equals 1 at all n. We show that as $n \to \infty$, each series x_{jt} of this system behaves as an ARFIMA$(0, 1/2, 0)$. In the second example, we consider a similar setting but with limiting value $\delta \in (0, 1)$ on the main diagonal of $\mathbf{A}_n$ and with the additional assumption that one innovation (say ϵ_{jt}) dominates the others in terms of magnitude. In this case, we prove that the dominant series j tends to an ARFIMA$(0, \delta, 0)$ for $\delta \in (0, 1)$. Our results exemplify that vanishing interaction coefficients in a multivariate system can give rise to long memory in individual series.

This phenomenon can be referred to as "hidden cross-section dependence" for two reasons. First, long memory appears through the marginalization mechanism and therefore in the univariate series or by extension, when estimating the model on a small subpart of a large system. The cross-section dependence appearing in the large system is therefore hidden in the univariate models. Second, because the off-diagonal elements of the VAR(1) may be so small that, in finite samples, the process is likely to be indistinguishable from a diagonal VAR(1) on the sole basis of the parameter estimates. The hidden dependencies induce modeling issues that were pointed out, inter alia, in Hendry (2009).

The rest of this paper is organized as follows. Section 2 provides our main theoretical results and we give two examples where they hold. Section 3 presents some Monte Carlo simulations. Section 4 concludes. The appendix contains the proofs of the main result, and an online supplement provides the proofs corresponding to the examples as well as an additional technical proof.

In the paper, we use the following notation. $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of, respectively, natural integers, real and complex scalars. For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the floor and ceiling of x. For any sequences a_n, b_n and c_n of real-valued scalars $a_n = O(b_n)$, $b_n = o(c_n)$, and $a_n \sim b_n$ imply, respectively, that as $n \to \infty$, $|a_n/b_n|$ is bounded, $b_n/c_n \to 0$, and $a_n/b_n \to 1$. For stochastic processes, $\Rightarrow$ denotes weak convergence of the associated probability measures. For any complex-valued square matrix $\mathbf{A}$, $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ its trace, $\widetilde{\mathbf{A}}$ its adjugate matrix, $\overline{\mathbf{A}}'$ its conjugate transpose and $|\mathbf{A}| = \sqrt{\mathrm{tr}\left(\overline{\mathbf{A}}'\mathbf{A}\right)}$ its weak (Frobenius) norm. For two sequences $(\mathbf{A}_n)$ and $(\mathbf{B}_n)$ of square matrices with bounded maximal eigenvalues, $\mathbf{A}_n \sim \mathbf{B}_n$ means that $|\mathbf{A}_n - \mathbf{B}_n| \to 0$ as $n \to \infty$.

$1_{\{\cdot\}}$ is the indicator function which takes value one if $\{\cdot\}$ is true and zero otherwise.

2 Theory

2.1 Set-up

We consider the n-vector $\mathbf{x}_t = (x_{1t}, ..., x_{nt})'$ which admits a vector autoregressive, VAR(1), representation:

$$\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t, \ t = 1, \ldots, T \tag{1a}$$

$$\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{nt})' \sim \mathsf{NID}\left(\mathbf{0}, \Sigma_n\right), \tag{1b}$$

where $\mathsf{IID}\left(\mathbf{0}, \Sigma_n\right)$ denotes a process that is identically and independently distributed with zero expectation and variance-covariance matrix Σ_n. Σ_n is not necessarily diagonal, although it is in the examples considered below, but we denote its diagonal by $\sigma_n^2 = \left(\sigma_{n,1}^2, ..., \sigma_{n,n}^2\right)$ such that $\sigma_{n,k} > 0$ for $k = 1, ..., n$. We also assume throughout that $\mathbf{A}_n$ does not admit eigenvalues with modulus strictly greater than unity.

We consider sequences of square Toeplitz matrices $(\mathbf{T}_n)_{n \in \mathbb{N}}$ that are n-dimensional with parameters given by the sequence $t_k^{(n)}$, for $-(n-1) \leq k \leq n-1$, such that coefficients $t_k^{(n)}$ and $t_{-k}^{(n)}$ (with $k > 0$) denote, respectively, the values of the kth upper and lower sub-diagonals, i.e.

$$\mathbf{T}_n = \begin{bmatrix} t_0^{(n)} & t_1^{(n)} & \cdots & t_{n-1}^{(n)} \\ t_{-1}^{(n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1^{(n)} \\ t_{-(n-1)}^{(n)} & \cdots & t_{-1}^{(n)} & t_0^{(n)} \end{bmatrix}. \tag{2}$$

Let δ denote a parameter and $(\mathbf{T}_{\delta,n})$ a sequence of n-dimensional Toeplitz matrices whose entries, $t_{\delta,k}$, do not depend on n (so only the dimension of $\mathbf{T}_{\delta,n}$ does). To $(\mathbf{T}_{\delta,n})$ we can associate a *symbol*, defined as the function $g_\delta(\cdot) : z \in \mathbb{C} \to g(\delta, z) = \sum_{k=-\infty}^{\infty} t_{\delta,k} z^k$. The function $\omega \in \mathbb{R} : \omega \to g_\delta\left(e^{i\omega}\right)$ is often referred to as "spectral density" of $\mathbf{T}_{\delta,n}$ (whose entries are obtained by the Fourier transform

of the spectral density) but to avoid confusion with the spectral density of the processes x_{jt}, we do not use this terminology. Yet with a slight abuse of notation, we refer to $g_\delta\left(e^{i\omega}\right)$ as the function $\omega \to g_\delta\left(e^{i\omega}\right)$.

We consider in this paper two examples which constitute applications of our main theorem and provide primitive conditions for its high-level assumptions.

Example (1)**.** *The vector* $\mathbf{x}_t$ *is generated by Model (1) with* $\mathbf{A}_n = \mathbf{T}_n^* + \eta_n \mathbf{D}_n$*, where* $\mathbf{T}_n^*$ *is the Toeplitz matrix with entries*

$$t_k^{*(n)} = \mathrm{Re}\left[\frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta_n, e^{i\frac{2\pi j}{n}}\right) e^{-\frac{2i\pi jk}{n}}\right], \tag{3}$$

where $g(\cdot,\cdot)$ *is defined for* $\delta \in (0,1)$ *and* $\omega \geq 0$ *as*

$$g\left(\delta, e^{i\omega}\right) = 1_{\{0\leq u<\pi\delta\}} + 1_{\left\{\pi\left(\frac{3}{2}-\delta\right)<u\leq\frac{3\pi}{2}\right\}}, \quad \omega = u \ mod\ 2\pi, \tag{4}$$

and where $\omega \to g\left(\delta, e^{i\omega}\right)$ *is even; the sequence* δ_n *satisfies* $\delta_n = \frac{1}{2} + o\left(n^{-2}\right)$*, with* $\delta_n < \frac{1}{2}$*;* η_n *is a real scalar sequence that satisfies* $\eta_n = o\left(n^{-2}\right)$*,* $\mathbf{D}_n$ *is a real antisymmetric Toeplitz matrix with absolutely summable entries. All innovations* ϵ_{jt} *have finite variance:* $\Sigma_n = \sigma^2 \mathbf{I}_n$ *with* $\sigma^2 > 0$.[2]

Example (2)**.** *The vector* $\mathbf{x}_t$ *is generated by Model (1) with* $\mathbf{A}_n = \mathbf{T}_n^*$ *where* $\mathbf{T}_n^*$ *and* g *are defined as in (3) and (4) where* $\delta_n \equiv \delta \in (0,1)$*.* Σ_n *is diagonal and there exists* $j \in \mathbb{N}$ *such that the variances of the innovations satisfy*

$$\begin{cases} \sigma_{n,j}^2 = \sigma_j^2 > 0, \\ \sigma_{n,k}^2 = o\left(n^{-1}\right), \quad k \neq j, \end{cases}$$

In both examples $\mathbf{A}_n$ is Toeplitz. In Example (1), the system is such that all processes entering the VAR present similar dynamics (i.e., the results are invariant by rotations of $\mathbf{A}_n$). This example stresses the fact that asymmetry, or heterogeneity, is not necessary for long memory to arise. We

[2]We assume that all the diagonal elements of Σ_n are equal for notational ease but this has no incidence on our results. We can relax the assumption to heterogenous and positive diagonal elements.

show in the online appendix that the entries $t_k^{*(n)}$ of $\mathbf{T}_n^*$ satisfy, as $n \to \infty$, and for $(n-1)/4 \in \mathbb{N}$,

$$t_0^{*(n)} = 1/2 + o\left(n^{-1}\right), \tag{5a}$$

$$t_k^{*(n)} = O\left(n^{-1}\right), \quad k \neq 0. \tag{5b}$$

Hence, as $n \to \infty$, the off-diagonal entries of $\mathbf{A}_n$ individually tend to zero. Yet the convergence is slow enough to ensure that for all n, $\mathbf{A}_n$ induces persistence that differs from that induced by a diagonal matrix. Indeed the off-diagonal elements of each row admit a nonzero sum: $\sum_{k=1}^{n-1} t_k^*(n) = 1/2 + O\left(n^{-1}\right)$. Hence, although the dependence across individual units x_{jt} and x_{kt} vanishes, the system retains sufficient dependence to differ from the one where $\mathbf{A}_n$ is diagonal for finite n. In Example (2), by contrast, the variance of one innovation ϵ_{jt} dominates the others so the system is asymmetric. The off-diagonal entries of $\mathbf{A}_n$ do not necessarily tend to zero. This constitutes an heterogenous case where the results are not symmetric for all x_{jt} and the distinction with the case of long memory arising from aggregation is clear.

We provide next a preliminary result relating to Toeplitz matrix theory that we use in our subsequent analysis.

2.2 A preliminary lemma

Consider sequences of square Toeplitz matrices $(\mathbf{T}_n)_{n\in\mathbb{N}}$ that are n-dimensional with parameters given as in (2) by the sequence $t_k^{(n)}$, for $-(n-1) \leq k \leq n-1$. We make a set of assumptions on the elements $t_k^{(n)}$ of the Toeplitz sequence that we collect in Assumption T (T stands for Toeplitz).

Assumption T.

(*i*) *There exists a real-valued function* $g(\cdot,\cdot)$*, defined on* $(0,1) \times \{\zeta \in \mathbb{C}, |\zeta| \leq 1\}$ *which is continuous with respect to its first argument and such that*

(*i.a*) $g(\cdot,\cdot) \leq 1$;

(*i.b*) $\int_0^{2\pi} \left|g\left(\cdot, e^{i\omega}\right)\right| d\omega < \infty$;

(*i.c*) $\forall (\delta, z) \in (0,1) \times (-\infty, 1)$, $\frac{1}{2\pi}\int_0^{2\pi} \log\left(1 - g\left(\delta, e^{i\omega}\right) z\right) d\omega = \delta \log(1-z)$.

(*ii*) $\forall \delta \in (0,1)$, *defining* $t_{\delta,k}^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta, e^{i\frac{2\pi j}{n}}\right) e^{-2i\pi jk/n}$, *the sequence* $(t_{\delta,k})$, *with* $t_{\delta,k} = \lim_{n\to\infty} t_{\delta,k}^{(n)}$, *satisfies* $\sum_{k=-\infty}^{\infty} |t_{\delta,k}| < \infty$;

(*iii*) *There exists a nondecreasing sequence* $\delta_n \in (0,1)^{\mathbb{N}} \to \delta \in (0,1)$ *as* $n \to \infty$ *such that*

(*iii.a*) *the parameters of the matrix sequence* $(\mathbf{T}_n)$ *satisfy*

$t_k^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta_n, e^{i\frac{2\pi j}{n}}\right) e^{-2i\pi jk/n}$;

(*iii.b*) $n^2\left(t_{\delta,k}^{(n)} - t_k^{(n)}\right) \to 0$ as $n \to \infty$.

Assumption T(*i*) characterizes the symbol of the Toeplitz matrix sequence $\mathbf{T}_{\delta,n}$ with entries $t_{\delta,k}$. Since the identity matrix $\mathbf{I}_n$ is Toeplitz with symbol $g_{\mathbf{I}}(\cdot) = 1$, hence for all $z < 1$ and $\delta \in (0,1)$, the matrix $\mathbf{I}_n - \mathbf{T}_{\delta,n} z$ is Toeplitz with symbol $1 - g_\delta(\cdot) z$. Under T(*ii*), $\mathbf{T}_{\delta,n}$ belongs to the so-called Wiener class of Toeplitz matrices with absolutely summable entries, so we can use associated results (Gray, 2006, in particular, Section 4.4 and Theorem 4.2) regarding the convergence of functions of $t_{\delta,k}$. As shown in the appendix, Assumptions T(*i.a-b*) and T(*ii*) ensure that the First Theorem of Szegö (1915) holds for $\mathbf{I}_n - \mathbf{T}_{\delta,n} z$, with $z \neq 1$, i.e.

$$\frac{\det\left(\mathbf{I}_n - \mathbf{T}_{\delta,n+1} z\right)}{\det\left(\mathbf{I}_n - \mathbf{T}_{\delta,n} z\right)} \underset{n\to\infty}{\to} \exp\left\{\frac{1}{2\pi}\int_0^{2\pi} \log\left(1 - g_\delta\left(e^{i\omega}\right) z\right) d\omega\right\}.$$

The latter limit is, under Assumption T(*i.c*) equal to $(1-z)^{\delta}$. Assumption T(*iii.a*) defines the sequence $t_k^{(n)}$, and hence the sequence of Toeplitz matrices $\mathbf{T}_n$. Assumption T(*iii.b*) ensures that the proximity of $t_k^{(n)}$ to $t_{\delta,k}$ is sufficient to ensure that $\frac{\det(\mathbf{I}_n - \mathbf{T}_{\delta,n+1} z)}{\det(\mathbf{I}_n - \mathbf{T}_{\delta,n} z)}$ and $\frac{\det(\mathbf{I}_n - \mathbf{T}_{n+1} z)}{\det(\mathbf{I}_n - \mathbf{T}_n z)}$ admit the same limit as shown in the following lemma, whose proof is in the appendix, Section 5.1.

Lemma 1. *Under Assumption T and for all* $z < 1$, *the sequence of Toeplitz matrices* $(\mathbf{T}_n)_{n\in\mathbb{N}}$ *with elements* $t_k^{(n)}$ *satisfies, as* $n \to \infty$,

$$\frac{\det\left(\mathbf{I}_{n-1} - \mathbf{T}_{n-1} z\right)}{\det\left(\mathbf{I}_n - \mathbf{T}_n z\right)} \to (1-z)^{-\delta}.$$

The lemma will be used to study the properties of marginalized series within vector autoregressive processes. In particular, Lemma 1 gives rise to the polynomial $(1-z)^{\delta}$, $\delta \in (0,1)$, that is pervasive when considering the dynamics of fractionally integrated processes. We will use this result in relation to lag polynomials in moving average representations.

2.3 Main result

In order to link Model (1) to its marginalized univariate components, we use the final equation representation (FER) as studied by Zellner and Palm (1974). They show how the elements of vector processes can be marginalized to yield univariate ARMA representations; see also Cubadda, Hecq and Palm (2009) in the context of factor models and Hecq, Laurent and Palm (2016) for an application to multivariate volatility processes. The FER of Model (1) is

$$\det\left(\mathbf{B}_n(L)\right)\mathbf{x}_t = \widetilde{\mathbf{B}_n(L)}\ \epsilon_t, \tag{6}$$

where $\mathbf{B}_n\left(L\right) = \mathbf{I}_n - \mathbf{A}_n L$, with L the lag operator and, in our notation, $\widetilde{\mathbf{B}_n(L)}$ denotes the adjugate of $\mathbf{B}_n(L)$. If $\mathbf{A}_n$ admits unitary eigenvalues, we implicitly assume that $\epsilon_t = 0$ for $t \leq 0$ and $\mathbf{x}_0 = 0$. In the spirit of Johansen and Nielsen (2016), we therefore consider truncations of lag polynomials and convergence to fractional processes with a fixed start (hence of type II). For $t > 0$, and for any polynomial $P\left(z\right)$, we let $\left[P\left(z\right)\right]^{+}$ denote the truncation of $P\left(z\right)$ for degrees of z^k strictly less than t so that $P\left(L\right)\mathbf{x}_t$ only involves $\mathbf{x}_k$ at dates $k > 0$.[3] Expression (6) shows that elements x_{jt} (for $j = 1\ldots, n$) obtained by marginalizing the n-dimensional VAR(1), admit a finite ARMA$(n, n-1)$ representation with a common AR lag polynomial. Hence, as $n \to \infty$, the univariate process x_{jt} without roots cancellation follows an ARMA(∞, ∞). Root cancellation may yield an ARMA representation that remains of finite order as $n \to \infty$ (Zellner and Palm, 204) but our paper explicitly considers the case where this does not arise, see below. For clarity of the exposition, consider the following trivariate example:

Example. $\mathbf{x}_t$ *is a trivariate VAR(1) specified as follows:*

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ b & a & b \\ 0 & b & a \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{2t} \end{bmatrix},$$

[3] $\left[P\left(z\right)\right]^{+}$ in effect depends on t, but we fix the latter throughout so the relation is implicit to avoid being cumbersome. For z_0 such that $P\left(z_0\right) = 0$, we define $\left[P\left(z\right)^{-1}\right]^{+}_{z=z_0} = \lim_{z\to z_0}\left[P\left(z\right)^{-1}\right]^{+}$, such that $\left|\left[P\left(z\right)^{-1}\right]^{+}_{z=z_0}\right| < \infty$.

where $\mathsf{E}[\epsilon_{jt}\epsilon_{kt}] = 0$ *for* $j \neq k$. *The FER of* $\mathbf{x}_t$ *is* $\det(\mathbf{B}_n(L))\,\mathbf{x}_t = \widetilde{\mathbf{B}_n(L)}\,\epsilon_t$, *where*

$$\det(\mathbf{B}_n(L)) = \frac{1}{(a^2-2b^2)}(1-aL)\left(1-\left(a-\sqrt{2}b\right)L\right)\left(1-\left(a+\sqrt{2}b\right)L\right),$$

$$\widetilde{\mathbf{B}_n(L)} = \begin{bmatrix} (1-aL)^2 - b^2L^2 & bL(1-aL) & b^2L^2 \\ bL(1-aL) & (1-aL)^2 & bL(1-aL) \\ b^2L^2 & bL(1-aL) & (1-aL)^2 - b^2L^2 \end{bmatrix}.$$

Hence $x_{jt} \sim ARMA(3,2)$ *for* $j = 1,2,3$ *when* $a \neq 0$ *and* $b \neq 0$, *while* $x_{jt} \sim ARMA(2,1)$ *when* $a = 0$ *and* $b \neq 0$ *and* $x_{jt} \sim AR(1)$ *when* $a \neq 0$ *and* $b = 0$.

Denote the jth, $1 \leq j \leq n$, row of $\widetilde{\mathbf{B}_n(L)}$ by $\widetilde{\mathbf{B}_n(L)}_{j.}$ such that

$$\widetilde{\mathbf{B}_n(L)}_{j.} = \left[\widetilde{\mathbf{B}_n(L)}_{j1} \quad \widetilde{\mathbf{B}_n(L)}_{j2} \quad ... \quad \widetilde{\mathbf{B}_n(L)}_{jn}\right],$$

where, for any matrix $\mathbf{D}$, we denote by $\mathbf{D}_{jk}$ its entry at the intersection of the jth row and kth column. Hence x_{jt} admits the ARMA representation

$$\det(\mathbf{B}_n(L))\,x_{jt} = \widetilde{\mathbf{B}_n(L)}_{j.}\epsilon_t = \sum_{k=1}^{n}\widetilde{\mathbf{B}_n(L)}_{jk}\,\epsilon_{kt}.$$

The process x_{jt} therefore admits the representation

$$x_{jt} = \sum_{k=1}^{n}\det(\mathbf{B}_n(L))^{-1}\,\widetilde{\mathbf{B}_n(L)}_{jk}\epsilon_{kt}. \tag{7}$$

Expression (7) constitutes the basis of our theoretical argument. To keep the generality of the exposition, we implicitly consider that (7) is replaced by $x_{jt} = \sum_{k=1}^{n}\left[\det(\mathbf{B}_n(L))^{-1}\,\widetilde{\mathbf{B}_n(L)}_{jk}\right]^{+}\epsilon_{kt}$ when $\det(\mathbf{B}_n(1)) = 0$.[4]

We make Assumption P for the parameters of the model (P stands for parameters).

Assumption P. *There exists* $j \in \mathbb{N}$ *for which the parameters of Model (1) satisfy for all finite* t, *and as* $n \to \infty$,

(i) $\quad \mathsf{E}\left[\left(\sum_{k=1,\ k\neq j}^{n}\det(\mathbf{B}_n(L))^{-1}\,\widetilde{\mathbf{B}_n(L)}_{jk}\epsilon_{kt}\right)^2\right] \to 0;$

[4]The degree of the polynomial $\left[\det(\mathbf{B}_n(z))^{-1}\,\widetilde{\mathbf{B}_n(z)}_{jk}\right]^{+}$ is finite (at most $t-1$). Therefore $\left|\left[\det(\mathbf{B}_n(z))^{-1}\,\widetilde{\mathbf{B}_n(z)}_{jk}\right]^{+}\right| < \infty$ for all z.

(*ii*) $\widetilde{\mathbf{B}_n(z)}_{jj} \sim \det(\mathbf{B}_{n-1}(z))$;

(*iii*) $\sigma^2_{n,j} \to \sigma^2_j > 0$;

(*iv*) *there exists a sequence of Toeplitz matrices* $(\mathbf{T}_n)$ *that satisfies Assumption T and such that for any real* $z < 1$,

$$\frac{\det(\mathbf{B}_{n-1}(z))}{\det(\mathbf{B}_n(z))} \sim \frac{\det(\mathbf{I}_{n-1} - \mathbf{T}_{n-1}z)}{\det(\mathbf{I}_n - \mathbf{T}_n z)}.$$

Assumption P(i) ensures that in the moving average representation for x_{jt}, the only innovation that retains a contribution as $n \to \infty$ is ϵ_{jt}, the others play no role asymptotically. In this sense, P(i) is the key assumption that ensures that the mechanism generating long memory here differs completely from that of aggregation, see the discussion that follows Theorem 1. P(i) may for instance hold when the off-diagonal elements of $\mathbf{B}_n(z)$ are negligible compared to those on the main diagonal (Example (1)), or when the the variance of one error ϵ_{jt} dominates all others (Example (2)). Assumption P(ii) provides a recursive formula on the dimension of the system in the FER for x_{jt}. It relates $\widetilde{\mathbf{B}_n(z)}_{jj}$ – which expresses the contribution relative to ϵ_{jt} in the MA lag polynomial of the n-dimensional system, see Expression (6) – to the AR lag polynomial of the $n-1$ dimensional system. Assumption P(iii) ensures that the distribution of ϵ_{jt} is nondegenerate as $n \to \infty$, so that of x_{jt} is nondegenerate too. Finally, Assumption P(iv) ensures that we can make use of Lemma 1: as $n \to \infty$, $[\det(\mathbf{B}_n(L))]^{-1} \det(\mathbf{B}_{n-1}(L)) \to (1-L)^{-\delta}$ where $\delta \in (0,1)$ is defined in Assumption T(iii).

The main result of the paper resides in the following theorem, whose proof is in the appendix, Section 5.2. In order to emphasize that our results involve Type II fractional integration, we introduce the truncated fractional difference operator $\Delta_+^c = [\Delta^c]^+$ for any $c \in \mathbb{R}$ (where, as usual, $\Delta = 1 - L$).

Theorem 1. *Let the real vector* $\mathbf{x}_t$ *of dimension* n *be defined by Model (1). Then under Assumption P, and as* $n \to \infty$, *the process* $(x_{jt})_{t\in\mathbb{N}}$ *satisfies*

$$(x_{jt})_{t\in\mathbb{N}} \Rightarrow \left(\Delta_+^{-\delta}\epsilon_{jt}\right)_{t\in\mathbb{N}}.$$

In Theorem 1, the marginalized univariate process x_{jt} tends weakly to an I(δ) type II fractional white noise as the dimension n of the system diverges to infinity.

Remark 1. *In Theorem 1, the limit of x_{jt} is only driven by one innovation, the others' impacts appear to vanish. Hence, this is in particular distinctly different from the heuristic example of Granger (1980, Section 4) where he generalizes his argument about long memory via aggregation of heterogenous micro-units to a large scale dynamic model similar to our VAR(1) model $\mathbf{B}_n(L)\, x_t = \epsilon_t$. He notices that $\mathbf{B}_n^{-1}(L)$ is given by Expression (7) above, i.e., the sum of n moving averages of the innovations ϵ_{kt}, $k = 1, ..., n$, and he argues that there is a resemblance with his model of aggregation. It may indeed be possible that some specific assumptions about the parameters may lead to long memory by aggregating heterogenous processes in such a setting. Our framework differs here and, in fact, precludes it: we specify in Assumption P(i) that the contribution of the moving averages of ϵ_{kt}, for $k \neq j$, vanishes as $n \to \infty$, so they do not play a role in the long memory of x_{jt}. Theorem 1 shows clearly this distinction since the limiting fractional white noise is $\Delta_+^{-\delta}\epsilon_{jt}$, so it is entirely driven by the innovation sequence ϵ_{jt} without any contribution from ϵ_{kt}, for $k \neq j$.*

Remark 2. *Theorem 1 can be used to generate richer short-run dynamics. Consider the process $\mathbf{y}_t$ defined as $\mathbf{F}_n(L)\, \mathbf{y}_t = \mathbf{x}_t$ where $\mathbf{F}_n(L)$ is a matrix lag polynomial of dimension n and degree k, and $\mathbf{x}_t$ satisfies the assumptions of Theorem 1. Then $\mathbf{F}_n(L)\,(\mathbf{I}_n - \mathbf{A}_n L)\, \mathbf{y}_t = \epsilon_t$, so $\mathbf{y}_t$ follows a VAR(k + 1). There exist conditions on $\mathbf{F}_n(L)$ such that $\alpha(L)\, y_{jt} = \beta(L)\, x_{jt}$, for some polynomials α, β of fixed degrees, say p and q, i.e., y_{jt} follows a classical autoregressive distributed lag (ADL) model in x_{jt}, see, e.g., Hendry, Pagan and Sargan (1984). Then, under the assumptions of Theorem 1 and as $n \to \infty$, y_{jt} tends to a type II ARFIMA(p, δ, q) that satisfies $\alpha(L)\,(1 - L)^{\delta}\, y_{jt} = \beta(L)\, \epsilon_{jt}$. For instance, if the elements of $\mathbf{F}_n(L)$ are diagonal then y_{jt} is asymptotically ARFIMA$(p, \delta, 0)$; and if the elements of $\mathbf{F}_n(L)$ are of reduced rank, and with zero coefficients for the impact of $y_{kt-\ell}$ to y_{jt}, where $k \neq j$ and $\ell \geq 0$, then an ARFIMA(p, δ, q) structure may asymptotically arise, see Cubadda* et al. *(2009) and the examples provided in the online supplement.*

Theorem 1 implies that fractional integration arises in Examples (1) and (2) as follows. Proofs are provided in the Supplementary Appendix where we show that both Assumptions T and P hold for n such that $(n-1)/4 \in \mathbb{N}$.

Example (1 continued)**.** *Theorem 1 implies that, as* $n \to \infty$, $(n-1)/4 \in \mathbb{N}$, *for all* $k \in \mathbb{N}$, *processes* x_{kt} *defined for* $t \geq 0$ *tend weakly to uncorrelated type II fractional white noises of order* $1/2$: *for all* $k \in \mathbb{N}$

$$(x_{kt}) \Rightarrow \left(\Delta_+^{-1/2}\epsilon_{kt}\right). \tag{8}$$

In this example, all marginalized series within an n-dimensional system present the exact same fractional degree of integration as $n \to \infty$. Since the entries of $\mathbf{A}_n - \frac{1}{2}\mathbf{I}_n$ tend to zero as $n \to \infty$, the cross section dependence between series x_{jt} vanishes asymptotically. Yet, as $\sum_{k=1}^{n-1} t_k^*(n)$ remains nonzero, the dependence across individual series is sufficient to generate long memory in each of the marginal processes.

Example (2 continued)**.** *Theorem 1 implies that, as* $n \to \infty$, $(n-1)/4 \in \mathbb{N}$, *for all* $\delta \in (0,1)$, *the process* x_{jt} *tends to a type II fractional white noise:*

$$(x_{jt}) \Rightarrow \left(\Delta_+^{-\delta}\epsilon_{jt}\right). \tag{9}$$

In Example (2), only one innovation ϵ_{jt} with nonzero variance remains in the system as $n \to \infty$. Following Diebold and Yılmaz (2009), if we interpret our VAR setting as a network, then there are n nodes, $k = 1, ..., n$, in the system which are in state x_{kt-1} at the end of any period $t-1$. At time t, each node k combines $\mathbf{x}_{t-1}$ with an additional idiosyncratic signal ϵ_{kt} to produce $\mathbf{x}_t$. Since, for $k \neq j$, the ratio $\sigma^2_{\epsilon_j}/\sigma^2_{\epsilon_k} \to \infty$ as $n \to \infty$, we can interpret our framework as providing an example of impact of a relatively "large" innovation sequence within a network. Albeit different, such a setting bears some resemblance with the famous "Noah effect" of Mandelbrot and Wallis (1968) and Mandelbrot (1997) where they show that the existence of outliers arising from heavy tailed distributions can be a source of self-similarity (usually in the marginal distribution). Because here self-similarity takes the form of long memory (the "Joseph effect"), this interpretation of dynamic networks constitutes a bridge between the Noah and Joseph effects which are considered different forms of non-similarity in the literature.[5]

[5]The mechanism also bears some similarity with that considered by Schennach (2013), in the case of the impact of a

3 Simulation evidence

In this section, we evaluate our key theoretical results via a Monte Carlo simulation. We provide simulations that examine the validity of our theoretical asymptotic results when the dimensions of the system and of the sample become large.

An n-dimensional VAR(1), as defined in Equations (1a)-(1b), is used to generate data for different choices of T, n and model parameters. We first compare the simulated data and estimated degree of fractional integration. To save space, we only report in Figures 1 to 5 the results for $n = 201$ series and $T = 4,000$ observations.

As a benchmark, we consider in our first experiment the case of a diagonal matrix, $\mathbf{A}_n = d\mathbf{I}_n$, where the parameter d is set to 0.499. The first panel of Figure 1 shows the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$), i.e., when $\mathbf{A}_n = d\mathbf{I}_n$, $a_k^{(n)} = 0.499$ for $k = 0$ and 0 otherwise. In this simple setting, the derived univariate processes have short memory and follow a stationary AR(1) model with an autoregressive parameter of 0.499 for each series.

Panel 2 of Figure 1 plots the empirical distribution (over 1,000 replications) of the long memory parameter of series x_{1t} estimated using three popular estimation methods, i.e., the log periodogram regression (GPH) of Geweke and Porter-Hudak (1983), the Local Whittle Likelihood Estimator (LWLE) of Robinson (1995), both with bandwidth $T/2$ and the MLE of an ARFIMA$(1, d, 0)$. We deliberately choose a large bandwidth, as implemented by default in Doornik and Ooms (2001) to reduce the variability of the estimators. As expected the estimated long memory parameters are concentrated around 0 suggesting that there is no evidence of long memory in the individual series. This is confirmed by the third panel of Figure 1 which reports the ACF of x_{1t} for the first replication.

In the next two experiments, we consider a symmetric Toeplitz matrix $\mathbf{A}_n = \mathbf{T}_n^*$, under the assumptions of Example 1 with $\eta_n = 0$. We denote by d the value taken by δ_n : we choose two values

single input that transits through a network (Schennach's process is scalar). Contrary to Schennach, all the coefficients of $\mathbf{A}_n$ are strictly less than unity in absolute value, so the signal is transmitted from x_{kt-1} to x_{jt} with attenuation (she requires that some nodes transmit without attenuation, see her Section 3).

of d close to 1/2, i.e., respectively $d = 0.499$ in Figure 2, and $d = 0.45$ in Figure 3. The structure of these figures is similar to that of Figure 1 except that now, since d is close to 1/2, i.e., to the nonstationary region of an $\mathsf{I}(d)$ process, we follow the approach of Abadir, Distaso and Giraitis (2007) and apply the three long memory estimators to $(1-L)^d x_{1t}$ (for the values we report, we have added d ex-post to the estimate).[6] The first panel of these figures emphasizes that the diagonal elements are near d while the off-diagonal elements of $\mathbf{A}_n$ are small for $d = 0.45$ and very small for $d = 0.499$. The sum of each row of $\mathbf{T}_n^*$ is unity by construction and therefore although the off-diagonal elements of $\mathbf{A}_n$ can be very small when d is close to 1/2, they are nonzero. Unlike in Figure 1, long memory is detected in x_{1t}, with a Monte Carlo mean (over the 1,000 replications) of 0.444, 0.484 and 0.488 respectively for the GPH, LWLE and ARFIMA$(0, d, 0)$ methods for $d = 0.499$ and $0.417, 0.451$ and 0.465 for $d = 0.45$. The ACF of x_{1t} in the first replication also suggests the presence of long memory. These figures show that although $\mathbf{A}_n$ is near diagonal, the very small off-diagonal elements play a crucial role in the apparition of long memory.

Next, we evaluate the robustness of the previous result using the asymmetric Toeplitz matrix $\mathbf{A}_n = \mathbf{T}_n^* + \eta_n \mathbf{D}_n$, with $d = 0.499$, $\eta_n = \frac{1}{n^2 \log(n)}$, and where the elements of $\mathbf{D}_n$ in the upper triangle are drawn independently from a standard normal distribution. Figure 4 suggests that results are qualitatively the same as in the case of the symmetric Toeplitz matrix in the sense that long memory is detected in x_{1t} with a parameter estimate close to d.

Theorem 1 states that not only x_{1t} but all variables belonging to $\mathbf{x}_t$ should tend to fractional white noises when $n \to \infty$. Our next experiment illustrates this finding for the case of a symmetric Toeplitz matrix with $d = 0.499$, as investigated in Figure 2. Figure 5 plots the empirical distribution of the long memory parameter estimated on all series, i.e., on $x_{1t}, \ldots, x_{201t}$, for the three estimation methods. We only report the results for four replications, each row in the figure corresponding to a replication. Results suggest that the estimated long memory parameters do not vary much across series and are all concentrated in a region close to 1/2, especially for the LWLE and MLE of the

[6] Throughout the simulations, we use the fast fractional difference algorithm of Jensen and Nielsen (2014).

ARFIMA$(0, d, 0)$.

Finally, following the suggestion of an anonymous referee, we apply two tests allowing to discriminate between exact fractional integration and spurious long memory since Theorem 1 gives rise to the former as $n \to \infty$. The test of Shimotsu (2006) builds upon the idea that if an observed time series is generated by an I(d) process, then it is also true on every subsample with the same value of d and its dth differences is I(0). Haldrup and Kruse (2014) propose an alternative test whose principle is that nonlinear transformations of stationary fractionally integrated Gaussian processes decrease the order of memory in a specific way which is determined by the Hermite rank of the transformation. They rely on the Wald test of Shimotsu (2007) to test the null hypothesis that a vector time series of properly transformed variables is I(0). Their test has reasonable power against various models generating spurious long memory.

Therefore, we simulate again $T = 4,000$ observations of a VAR(1) with $\mathbf{A}_n = \mathbf{T}_n^*$, under the assumptions of Example 1 (with $\eta_n = 0$), $d = 0.499$ and values of n ranging between $1,001$ and $15,001$. The first panel of Figure 6 reports the average (across 1,000 replications) of the estimated long memory parameter of x_{1t} obtained by LWLE estimated on four non-overlapping samples of 1,000 observations. Results suggest that the estimated d parameters are remarkably stable across the samples. The dotted and dashed lines in the second panel of Figure 6 correspond to the rejection frequencies of the null hypothesis of fractional integration against the alternative of spurious long memory of x_{1t} for the test of Shimotsu (2006) for respectively 2 and 4 sub-samples while the solid line corresponds to the rejection frequencies of the test of Haldrup and Kruse (2014). The nominal size of both tests is 5%. As expected, the rejection frequencies of the tests converge to the nominal size of 5% as $n \to \infty$, suggesting that x_{1t} is fractionally integrated.

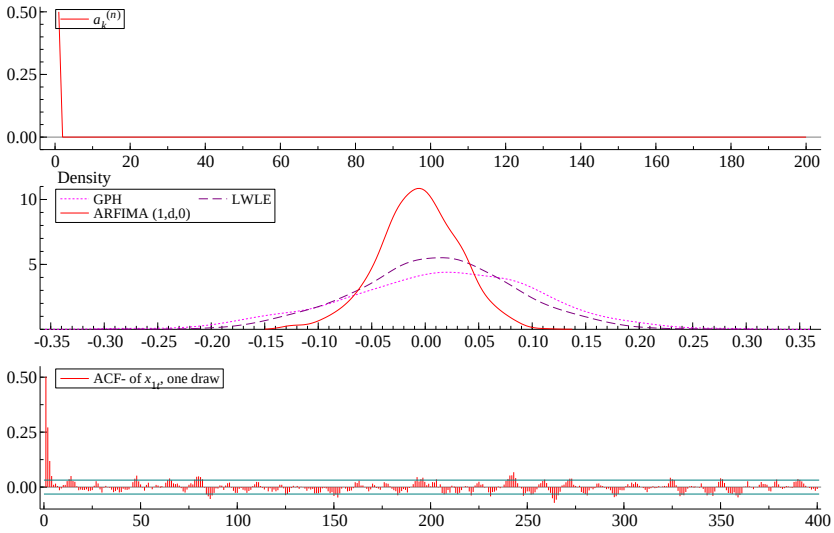


Figure 1: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = d\mathbf{I}_n$, where $d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(1, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

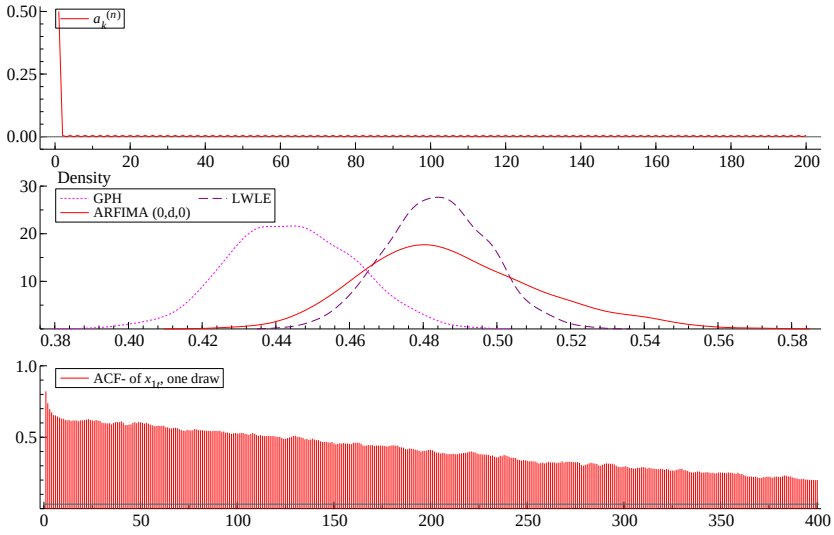


Figure 2: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, where $\mathbf{A}_n = \mathbf{T}_n^*$ is given in Example 1, with $\delta_n = d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

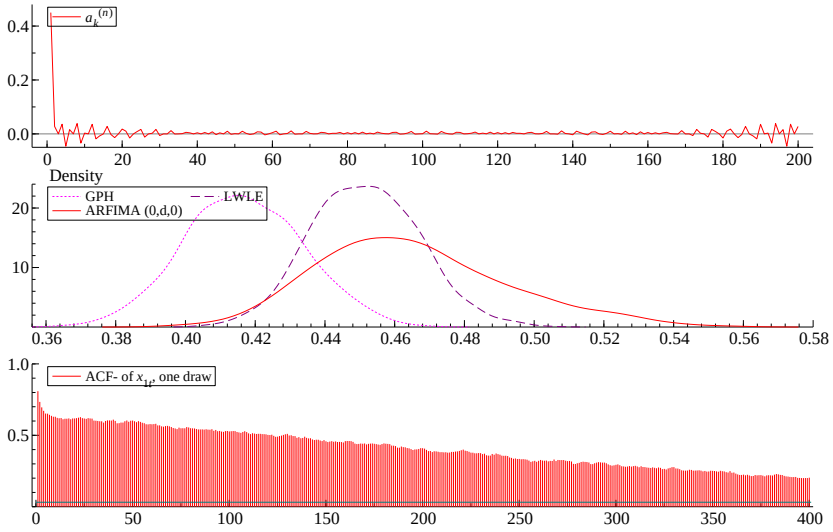


Figure 3: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n\mathbf{x}_{t-1} + \epsilon_t$, where $\mathbf{A}_n = \mathbf{T}_n^*$ is given in Example 1, with $\delta_n = d = 0.45$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

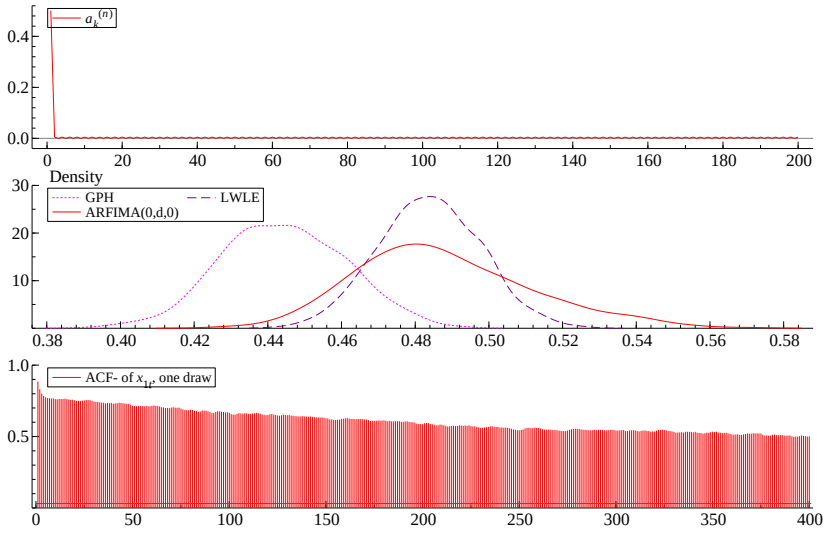


Figure 4: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n\mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^* + \eta_n\mathbf{D}_n$, where $\mathbf{T}_n^*$ is given in Example 1, with $\delta_n = d = 0.499$, $\eta_n = 1/(n^2\log(n))$, $\mathbf{D}_n$ is an antisymmetric Toeplitz matrix with above diagonal elements drawn from a standard normal distribution, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

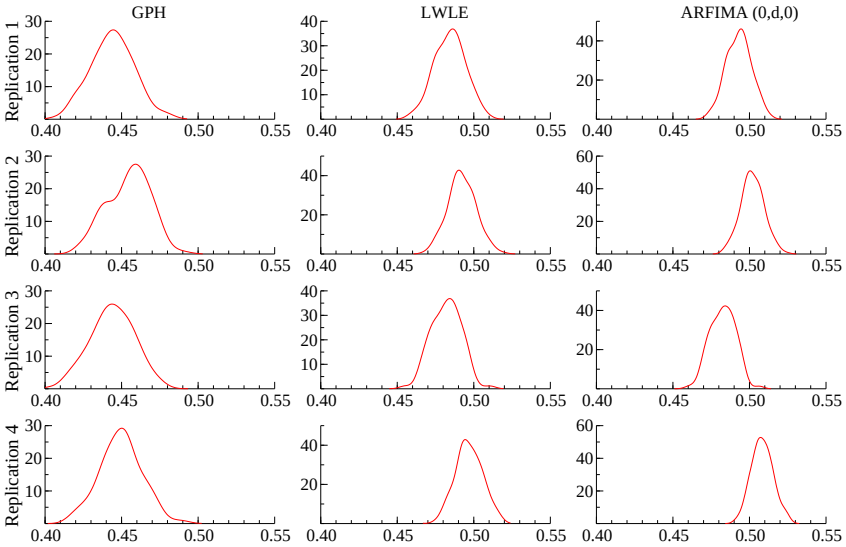


Figure 5: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, where $\mathbf{A}_n = \mathbf{T}_n^*$ as given in Example 1, with $\delta_n = d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The figure plots the empirical distribution of the long memory parameter estimated on all series, i.e., on $x_{1t}, \ldots, x_{201t}$, using GPH, LWLE and ARFIMA$(0, d, 0)$. Each row corresponds to a separate replication.

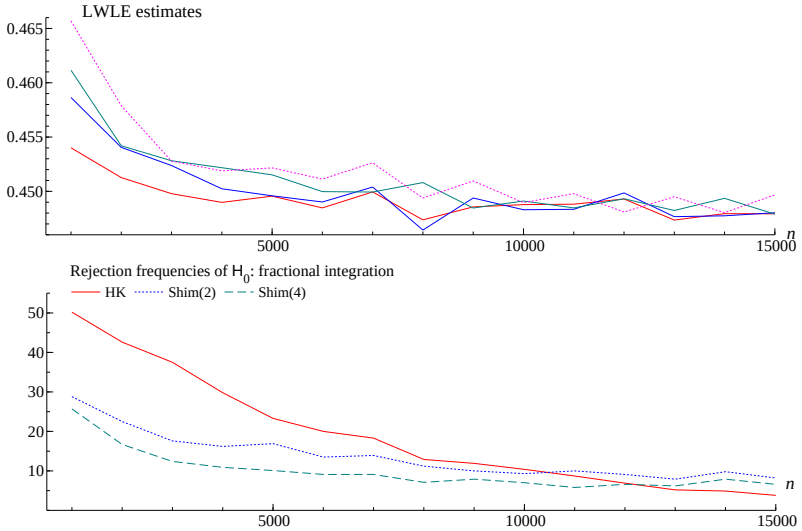


Figure 6: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^*$, where $\mathbf{T}_n^*$ is given in Example 1, with $\delta_n = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 1001, 2001, \ldots, 15001$ and $t = 1, \ldots, 4000$. The top panel reports the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the LWLE method estimated on four non-overlapping samples of 1000 observations. The bottom panel corresponds to the rejection frequencies of the null hypothesis of fractional integration (vs. spurious long memory) of x_{1t} at the nominal level of 5% for the test of Haldrup and Kruse (2014) and the test of Shimotsu (2006) for 2 and 4 sub-samples.

4 Conclusion

Our paper contributes to the time series literature investigating the mechanisms generating slowly decaying autocorrelations and low frequency variability, in particular those leading to long memory processes. We show that an n-dimensional vector autoregressive model of order 1 can generate long memory in the marginalized univariate series. To achieve this goal, we consider the Final Equation Representation of this model and obtain the n univariate implied ARMA$(n,\, n-1)$ models whose MA lag polynomial is expressed as a sum that is derived from the determinant and the adjugate of the matrix lag polynomial of the VAR. We then develop high-level assumptions ensuring that at least one of the elements of the vector process converges in probability to a fractional white noise of degree $\delta \in (0,1)$. We show that this assumption is satisfied for two specific examples of an n-dimensional VAR(1) model where either (i) all univariate processes tend in probability to an $\mathsf{I}\left(\frac{1}{2}\right)$ fractional white noise as $n \to \infty$, or (ii) one univariate process tends to an $\mathsf{I}(\delta)$ fractional white noise.

Our approach is general enough to allow extending it to groups of time series sharing within each group the properties we study in this paper and where each group is orthogonal to others. This would be the case for instance in a large dimensional block-diagonal VAR or in a GVAR, see Pesaran, Schuermann and Weiner (2004). Another route for extending our results would consist in replacing $\mathbf{A}_n$, the matrix parameter of the VAR(1) , with $\mathbf{V}_n \mathbf{A}_n \mathbf{V}_n^{-1}$ where $\mathbf{V}_n$ denotes a sequence of matrices. This would modify the adjugate matrices but not the determinant of $\mathbf{B}_n(L)$.

5 Appendix

Throughout the proofs, κ denotes a generic positive real number and N a generic integer. Entries of a Toeplitz matrix Ψ_n of dimension n, at the intersection of row $k = 1, ..., n$ and column $\ell = 1, ..., n$,

are denoted by $\psi_{\ell-k}$ such that

$$\mathbf{\Psi}_n = \begin{bmatrix} \psi_0^{(n)} & \psi_1^{(n)} & \cdots & \psi_{n-1}^{(n)} \\ \psi_{-1}^{(n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \psi_1^{(n)} \\ \psi_{-(n-1)}^{(n)} & \cdots & \psi_{-1}^{(n)} & \psi_0^{(n)} \end{bmatrix}.$$

We present below the proofs of Lemma 1 and Theorem 1. Proofs related to the validity of Assumptions T and P for Examples 1 and 2, together with additional technical proofs, are placed in the online supplementary material.

5.1 Proof of Lemma 1

To prove the lemma, we first need to show that, for $z \neq 1$ and as $n \to \infty$,

$$\frac{\det\left(\mathbf{I}_{n-1} - \mathbf{T}_{\delta,n-1} z\right)}{\det\left(\mathbf{I}_n - \mathbf{T}_{\delta,n} z\right)} \to (1-z)^{-\delta}. \tag{10}$$

For this, we start by recalling the First Theorem of Szegö, as stated in Grenander and Szegö (1958, reprinted in 2001, Section 5.2 p. 64). Grenander and Szegö's class L (2001, Section 1.2, p. 4) denotes the set of all complex-valued functions F which are measurable in the Lebesgue sense and for which[7] $\int_0^{2\pi} |F(u)|\, du$ exists. We denote by $\lambda_k^{(n+1)}$, $1 \leq k \leq n+1$, the eigenvalues associated with the $(n+1)$-dimensional Toeplitz Ψ_{n+1} matrix with entries ψ_k, where $\psi_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx} f\left(e^{iu}\right) du$ and $f(\cdot)$ denotes the symbol of (Ψ_n).[8] We are ready to state the required theorem.

Theorem 2 (Grenander and Szegö, 2001, Sec. 5.2)**.** *Let $x \to f\left(e^{ix}\right)$ be a real-valued function of the class L. We denote by m and M the 'essential' lower and upper bounds of $x \to f\left(e^{ix}\right)$, respectively, and assume that m and M are finite. If $F(\lambda)$ is any continuous function defined in the finite interval $m \leq \lambda \leq M$, we have*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n+1} F\left(\lambda_k^{(n)}\right)}{n+1} = \frac{1}{2\pi}\int_0^{2\pi} F\left[f\left(e^{ix}\right)\right] dx.$$

[7] The original definition is for the integral over $[-\pi, \pi]$. To match the setting of this paper, we shift it to $[0, 2\pi]$.

[8] Grenander and Szegö let f denote the spectral density, here it denotes the symbol of (Ψ_n) for notational consistency. The spectral density evaluated at ω coincides with the symbol evaluated at $e^{i\omega}$. In Grenander and Szegö, the eigenvalues of Ψ_{n+1} are denoted $\lambda_k^{(n)}$. For notational consistency, we denote them by $\lambda_k^{(n+1)}$ here.

As a corollary, using $F(\cdot) = \log(\cdot)$, the determinant of Ψ_{n+1} is $\det(\Psi_{n+1}) = \prod_{j=1}^{n+1} \lambda_j^{(n+1)}$ so, provided all eigenvalues are strictly positive, $\lim_{n\to\infty} \log\det(\Psi_{n+1})^{\frac{1}{n+1}} = \frac{1}{2\pi}\int_0^{2\pi} \log f\left(e^{iu}\right) du$ and $\lim_{n\to\infty} \det(\Psi_{n+1})^{\frac{1}{n+1}} = \exp\left\{\frac{1}{2\pi}\int_0^{2\pi} \log f\left(e^{iu}\right) du\right\}$. Hence

$$\frac{\det(\Psi_{n+1})}{\det(\Psi_n)} \sim \det(\Psi_n)^{\frac{1}{n}} \tag{11}$$

see e.g., Gray (2006, p. 8). Hence $\lim_{n\to\infty} \det(\Psi_{n+1}) / \det(\Psi_n) = \exp\left\{\frac{1}{2\pi}\int_0^{2\pi} \log f\left(e^{iu}\right) du\right\}$. Now we need to show that the symbol associated with $\Psi_n = \mathbf{I}_n - \mathbf{T}_{\delta,n} z$ is of class L with $m > 0$ for all $z \in (-\infty, 1)$. For this, we notice that the required symbol is $1 - g_\delta(\cdot) z$ which is of class L if g_δ is of class L too and its minimum is $1 - z > 0$. Hence the eigenvalues of Ψ_n are strictly positive. Now $\exp\left\{\frac{1}{2\pi}\int_0^{2\pi} \log\left(1 - g_\delta\left(e^{i\omega}\right) z\right) d\omega\right\} = (1-z)^\delta$ hence, as $n \to \infty$, Expression (10) follows.

We now show that the same results hold, replacing $\mathbf{T}_{\delta,n}$ with $\mathbf{T}_n$. For this, we show that

$$\frac{\det(\mathbf{I}_{n-1} - \mathbf{T}_{n-1} z)}{\det(\mathbf{I}_n - \mathbf{T}_n z)} - \frac{\det(\mathbf{I}_{n-1} - \mathbf{T}_{\delta,n-1} z)}{\det(\mathbf{I}_n - \mathbf{T}_{\delta,n} z)} \to 0. \tag{12}$$

Assumption T(i) states that $g(\cdot,\cdot)$ is real-valued, which implies that for $d \in (0,1)$, $t_{d,k} = \int_0^{2\pi} g\left(d, e^{i\omega}\right) e^{-ik\omega} d\omega$ $= \overline{t_{d,-k}}$, i.e., $\mathbf{T}_{d,n}$ is Hermitian. This entails in particular that $t_k^{(n)} + t_{-k}^{(n)} \in \mathbb{R}$ with the notations of Assumption T. Also $g_d(\cdot) = g(d,\cdot)$ being bounded ensures $(\mathbf{T}_{d,n})$ and the associated matrices below are uniformly bounded in strong norm. Assumption T(ii) ensures that $\sum_{k=-(n-1)}^{n-1} t_{d,k} e^{ik\omega}$ converges uniformly to $g_d\left(e^{i\omega}\right)$ which characterizes $\mathbf{T}_{d,n}$ as belonging to the Wiener class (Gray, 2006, p. 40).

Let $\mathbf{T}_{\delta_n,n}$ denote the matrix with entries $t_{\delta_n,k}$ defined as $t_{d,k}$ evaluated at $d = \delta_n$. Under Assumption T(iii), $\mathbf{T}_n$ and $\mathbf{T}_{\delta_n,n}$ are asymptotically equivalent under the weak norm, which is denoted $\mathbf{T}_n \sim \mathbf{T}_{\delta_n,n}$ (see Gray, 2006, Section 2.3 for the definition of equivalent matrices).

To any Toeplitz matrix $\mathbf{T}_{\delta,n}$ within the Wiener class, we can associate a Circulant matrix $\mathbf{C}_{\delta,n}$ such that $\mathbf{C}_{\delta,n} \sim \mathbf{T}_{\delta,n}$ defined as

$$\mathbf{C}_{\delta,n} = \begin{bmatrix} c_{\delta,0}^{(n)} & c_{\delta,1}^{(n)} & \cdots & c_{\delta,n-1}^{(n)} \\ c_{\delta,n-1}^{(n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_{\delta,1}^{(n)} \\ c_{\delta,1}^{(n)} & \cdots & c_{\delta,n-1}^{(n)} & c_{\delta,0}^{(n)} \end{bmatrix}. \tag{13}$$

The sequence $(\mathbf{C}_{\delta,n})$ is not uniquely defined, see Grenander and Szegö (1958, Section 7.6). Gray (2006, Lemma 4.6, p. 52) shows for instance that choosing $t_{\delta,-k} + t_{\delta,n-k}$ for the entries $c_{\delta,k}^{(n)}$ yields a matrix which is asymptotically equivalent to $\mathbf{T}_{\delta,n}$ but this is not how we define $c_{\delta,k}^{(n)}$ here. Instead, we define two Hermitian circulant matrices $\mathbf{C}_{\delta,n}$ and $\mathbf{C}_n$ with entries $c_{\delta,k}^{(n)}$ and $c_k^{(n)}$ such that, respectively,

$$c_{\delta,k}^{(n)} = t_{\delta,-k}^{(n)} + t_{\delta,n-k}^{(n)} \tag{14a}$$

$$c_k^{(n)} = t_{-k}^{(n)} + t_{n-k}^{(n)} \tag{14b}$$

with $t_{\delta,k}^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta, e^{i\frac{2\pi j}{n}}\right) e^{-2i\pi jk/n}$. Since $t_k^{(n)}$, defined in Assumption T($iii.a$), converges to $t_{\delta,k}$, the matrix $\mathbf{C}_n$ is by construction asymptotically equivalent to that with entries $t_{\delta_n,-k} + t_{\delta_n,n-k}$. Asymptotic equivalence is transitive (see Gray, 2006, Theorem 2.1), hence $\mathbf{C}_n \sim \mathbf{T}_{\delta_n,n}$ and $\mathbf{C}_n \sim \mathbf{T}_n$. It also holds that $\mathbf{I}_n - \mathbf{C}_n z \sim \mathbf{I}_n - \mathbf{T}_n z$ and, since they are Hermitian, and with positive eigenvalues for $z \in (-\infty, 1)$ (by Assumption T($i.a$)), Corollary 2.4 of Gray (2006, p. 23) implies $\left[\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)\right]^{1/n}$ $\sim \left[\det\left(\mathbf{I}_{n-1} - z\mathbf{T}_{n-1}\right)\right]^{1/n}$. Now $\left[\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)\right]^{1/n} \sim \left[\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1}\right)\right]^{1/(n-1)}$. As in Expression (11), it follows that

$$\frac{\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)}{\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1}\right)} \sim \left[\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)\right]^{\frac{1}{n}}. \tag{15}$$

Similarly, replacing $\mathbf{C}_n$ with $\mathbf{T}_n$, we obtain, by transitivity of the equivalence, that

$$\frac{\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)}{\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1}\right)} \sim \frac{\det\left(\mathbf{I}_n - z\mathbf{T}_n\right)}{\det\left(\mathbf{I}_{n-1} - z\mathbf{T}_{n-1}\right)}. \tag{16}$$

Notice that $\mathbf{C}_{\delta,n}$ is constructed so that Gray (2006, Lemma 4.6) implies that $\mathbf{C}_{\delta,n} \sim \mathbf{T}_{\delta,n}$ (the notation for $\mathbf{C}_{\delta,n}$ in Gray's lemma is $C\left(\hat{f}_n\right)$). Hence, using the same arguments as in Expression (15), together with Expression (10),

$$\frac{\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1}\right)}{\det\left(\mathbf{I}_n - z\mathbf{C}_{\delta,n}\right)} \sim \frac{\det\left(\mathbf{I}_n - z\mathbf{T}_{\delta,n}\right)}{\det\left(\mathbf{I}_{n-1} - z\mathbf{T}_{\delta,n-1}\right)}. \tag{17}$$

We prove next that as $n \to \infty$,

$$\frac{\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1}\right)}{\det\left(\mathbf{I}_n - z\mathbf{C}_{\delta,n}\right)} - \frac{\det\left(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1}\right)}{\det\left(\mathbf{I}_n - z\mathbf{C}_n\right)} \to 0. \tag{18}$$

Expressions (16), (18) and (17) yield (12) which, together, prove the lemma since $\frac{\det(\mathbf{I}_n - z\mathbf{T}_{\delta,n})}{\det(\mathbf{I}_{n-1} - z\mathbf{T}_{\delta,n-1})}$ $\underset{n\to\infty}{\to} (1-z)^{-\delta}$.

There only remains to provide the proof of (18), which is as follows. First notice that $\det(\mathbf{I}_n - z\mathbf{C}_n)$ $= \det(\mathbf{I}_n - z\mathbf{C}_{\delta,n} + z(\mathbf{C}_{\delta,n} - \mathbf{C}_n))$, so

$$\det(\mathbf{I}_n - z\mathbf{C}_n) = \det(\mathbf{I}_n - z\mathbf{C}_{\delta,n}) \det\left(\mathbf{I}_n + z(\mathbf{I}_n - z\mathbf{C}_{\delta,n})^{-1}(\mathbf{C}_{\delta,n} - \mathbf{C}_n)\right)$$

and

$$\begin{aligned}\frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_n)} &= \frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_{\delta,n})} \\ &\times \frac{\det\left(\mathbf{I}_{n-1} + z(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})^{-1}(\mathbf{C}_{\delta,n-1} - \mathbf{C}_{n-1})\right)}{\det\left(\mathbf{I}_n + z(\mathbf{I}_n - z\mathbf{C}_{\delta,n})^{-1}(\mathbf{C}_{\delta,n} - \mathbf{C}_n)\right)}.\end{aligned} \tag{19}$$

We now consider the limit of the expression on the second line of (19).

Let $\mathbf{H}_n(z) = z(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})^{-1}(\mathbf{C}_{\delta,n-1} - \mathbf{C}_{n-1})$. For all $z \in (-\infty, 1)$, by Assumption T$(iii.b)$, $t_{\delta,k}^{(n)} - t_k^{(n)} = o\left(n^{-2}\right)$ so $c_{\delta,k}^{(n)} - c_k^{(n)} = o\left(n^{-2}\right)$ and the diagonal elements of $\mathbf{H}_n(z)$ are $o\left(n^{-1}\right)$ so $\operatorname{tr} \mathbf{H}_n(z) = o(1)$. Let $\mu_k^{(n)}$, $k = 1, ..., n$, denote the eigenvalues of $\mathbf{H}_n(z)$. Then, as $n \to \infty$, $\mu_k^{(n)} \to 0$ since all the elements of $\mathbf{H}_n(z)$ tend to zero, hence for n large enough, $\log\det(\mathbf{I}_n + \mathbf{H}_n(z))$ exists and we may use Jacobi's formula (Abadir and Magnus, 2005, results 12.30 p. 339 and 13.36 p. 372) that implies $\det(\mathbf{I}_n + \mathbf{H}_n(z)) \leq \exp\operatorname{tr}\mathbf{H}_n(z) \to 1$. Since $\det(\mathbf{I}_n + \mathbf{H}_n(z)) = \prod_{k=1}^n \left(1 + \mu_k^{(n)}\right)$, with $\mu_k^{(n)} \to 0$, there exists N such that for $n > N$, $\det(\mathbf{I}_n + \mathbf{H}_n(z)) > 0$. Therefore, there exists $\kappa \in (0, \infty)$ such that $\det(\mathbf{I}_n + \mathbf{H}_n(z)) \to \kappa$ and, for all $\eta > 0$, there exists a value N such that for $n > N$,

$$\left|\frac{\det\left(\mathbf{I}_{n-1} + z(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})^{-1}(\mathbf{C}_{\delta,n-1} - \mathbf{C}_{n-1})\right)}{\det\left(\mathbf{I}_n + z(\mathbf{I}_n - z\mathbf{C}_{\delta,n})^{-1}(\mathbf{C}_{\delta,n} - \mathbf{C}_n)\right)} - 1\right| < \eta$$

implying that

$$\left|\frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_n)} - \frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_{\delta,n})}\right| < \left|\frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_{\delta,n})}\right| \eta.$$

Since $\lim_{n\to\infty} \frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_{\delta,n})} = (1-z)^{-\delta}$, it follows that $\forall \zeta > 0$ and $z \in (-\infty, 1)$, there exist $\eta = \zeta(1-z)^{\delta}$ and N such that for $n > N$

$$\left|\frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_n)} - \frac{\det(\mathbf{I}_{n-1} - z\mathbf{C}_{\delta,n-1})}{\det(\mathbf{I}_n - z\mathbf{C}_{\delta,n})}\right| < \zeta.$$

This proves Expression (18) and concludes the proof of the lemma.

5.2 Proof of Theorem 1

We first prove that under Assumption P, for all $\eta > 0$, as $n \to \infty$,

$$\Pr\left(\left|x_{jt} - \frac{\det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right| > \eta\right) \underset{n\to\infty}{\to} 0. \tag{20}$$

Assume first that $\det\left(\mathbf{B}_n\left(1\right)\right) \neq 0$ and $\det\left(\mathbf{B}_n\left(1\right)\right) \not\to 0$ as $n \to \infty$. Then for n large enough, Chebyshev's inequality implies that, for $\eta > 0$,

$$\begin{aligned}\Pr\left(\left|x_{jt} - \frac{\det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right| > \eta\right) &\leq \frac{1}{\eta^2}\mathsf{E}\left[\left(x_{jt} - \frac{\det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right)^2\right] \\ &= \frac{1}{\eta^2}\mathsf{E}\left[\left(\frac{\widetilde{\mathbf{B}_n\left(L\right)}_{jj} - \det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt} + \sum_{k\neq j}\frac{\widetilde{\mathbf{B}_n\left(L\right)}_{jk}}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{kt}\right)^2\right],\end{aligned}$$

where the latter follows from Expression (7). Minkowski's inequality then implies that

$$\begin{aligned}\Pr\left(\left|x_{jt} - \frac{\det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right| > \eta\right) \leq \frac{1}{\eta^2}\Bigg(&\mathsf{E}\left[\left(\sum_{k\neq j}\frac{\widetilde{\mathbf{B}_n}\left(L\right)_{jk}}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right)^2\right]^{1/2} \\ &+ \mathsf{E}\left[\left(\frac{\widetilde{\mathbf{B}_n\left(L\right)}_{jj} - \det\left(\mathbf{B}_{n-1}\left(L\right)\right)}{\det\left(\mathbf{B}_{n}\left(L\right)\right)}\epsilon_{jt}\right)^2\right]^{1/2}\Bigg)^2.\end{aligned}$$

Assumption P(i) states that $\mathsf{E}\left[\left(\sum_{k\neq j}\frac{\widetilde{\mathbf{B}_n}(L)_{jk}}{\det(\mathbf{B}_n(L))}\epsilon_{jt}\right)^2\right] \to 0$. Assumption P($ii$) implies that, as $n \to \infty$, $\frac{\widetilde{\mathbf{B}_n(z)}_{jj} - \det(\mathbf{B}_{n-1}(z))}{\det(\mathbf{B}_n(z))} \to 0$ for all z such that the denominator remains bounded away from zero. Hence, since $\det\left(\mathbf{B}_n\left(1\right)\right) \not\to 0$, $\mathsf{E}\left[\left(\frac{\widetilde{\mathbf{B}_n(L)}_{jj} - \det(\mathbf{B}_{n-1}(L))}{\det(\mathbf{B}_n(L))}\epsilon_{jt}\right)^2\right] \to 0$ in particular. Expression (20) follows.

Now, if the polynomial $\det\left(\mathbf{B}_n\left(1\right)\right) = 0$ or $\det\left(\mathbf{B}_n\left(1\right)\right) \to 0$, then when t is held fixed, $\left[\det\left(\mathbf{B}_n\left(1\right)\right)^{-1}\right]^+ = O\left(1\right)$ and Assumption P(ii) implies that $\mathsf{E}\left[\left(\frac{\widetilde{\mathbf{B}_n(L)}_{jj} - \det(\mathbf{B}_{n-1}(L))}{\det(\mathbf{B}_n(L))}\epsilon_{jt}\right)^2\right]^+ \to 0$, so Expression (20) holds here too.

Now, under assumption P(iv), Expression (20) implies that, for all $\eta > 0$,

$$\Pr\left(\left|x_{jt} - \frac{\det\left(\mathbf{I}_{n-1} - \mathbf{T}_{n-1}L\right)}{\det\left(\mathbf{I}_{n} - \mathbf{T}_{n}L\right)}\epsilon_{jt}\right| > \eta\right) \underset{n\to\infty}{\to} 0. \tag{21}$$

Therefore, $\Pr\left(\left|x_{jt}-\Delta^{-\delta}\epsilon_{jt}\right|>\eta\right)\leq\Pr\left(\left|x_{jt}-\frac{\det(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}L)}{\det(\mathbf{I}_n-\mathbf{T}_nL)}\epsilon_{jt}\right|>\eta\right)+\Pr\left(\left|\left(\frac{\det(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}L)}{\det(\mathbf{I}_n-\mathbf{T}_nL)}-\Delta^{-\delta}\right)\epsilon_{jt}\right|>\eta\right)$.

Denoting, $\frac{\det(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}z)}{\det(\mathbf{I}_n-\mathbf{T}_nz)}-(1-z)^{-\delta}=\sum_{k=0}^{\infty}\theta_{n,k}z^k$, then

$$\left|\left(\frac{\det\left(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}L\right)}{\det\left(\mathbf{I}_n-\mathbf{T}_nL\right)}-\Delta^{-\delta}\right)\epsilon_{jt}\right|\leq t\max_{0\leq k\leq t}\left|\theta_{n,k}\right|\max_{0\leq k\leq t}\left|\epsilon_{jk}\right|.$$

Lemma 1 shows that as $n\to\infty$, $\theta_{n,k}\to 0$ for all k. Hence, by Chebyshev's inequality,

$$\begin{aligned}\Pr\left(\left|\left(\frac{\det\left(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}L\right)}{\det\left(\mathbf{I}_n-\mathbf{T}_nL\right)}-\Delta^{-\delta}\right)\epsilon_{jt}\right|>\eta\right)&\leq\Pr\left(\max_{0\leq k\leq t}\left|\epsilon_{jk}\right|\geq\frac{\eta}{t\max_{0\leq k\leq t}\left|\theta_{n,k}\right|}\right)\\&\leq\left(\max_{0\leq k\leq t}\left|\theta_{n,k}\right|\right)^2\left(\frac{t}{\eta}\right)^2\mathsf{E}\left[\left(\max_{0\leq k\leq t}\left|\epsilon_{jk}\right|\right)^2\right].\end{aligned}$$

There exists κ such that, for fixed t, $\left(\frac{t}{\eta}\right)^2\mathsf{E}\left[\left(\max_{0\leq k\leq t}\left|\epsilon_{jk}\right|\right)^2\right]\leq\kappa$ yet $\left(\max_{0\leq k\leq t}\left|\theta_{n,k}\right|\right)^2\to 0$ as $n\to\infty$. Hence $\Pr\left(\left|\left(\frac{\det(\mathbf{I}_{n-1}-\mathbf{T}_{n-1}L)}{\det(\mathbf{I}_n-\mathbf{T}_nL)}-\Delta^{-\delta}\right)\epsilon_{jt}\right|>\eta\right)\underset{n\to\infty}{\to}0$. Together with (21), the latter expression implies that, as $n\to\infty$, $\Pr\left(\left|x_{jt}-\Delta^{-\delta}\epsilon_{jt}\right|>\eta\right)\to 0$.

This establishes pointwise convergence in probability for every finite t. Hence joint convergence in probability holds for $(x_{jt})_{t\in\mathbb{T}}$ and any finite subset $\mathbb{T}\in\mathbb{N}$. Uniform convergence of $(x_{jt})_{t\in\mathbb{T}}$ to $\left(\Delta_+^{-\delta}\epsilon_{jt}\right)_{t\in\mathbb{T}}$ also holds since pointwise convergence over a finite set implies uniform convergence (in probability and hence weakly). Now let $\mathbb{T}_T=\mathbb{N}\cap[1,T]$ the subset of integers no greater than T. Pollard (1984, Definition 22 and Theorem 23, p. 108) shows that uniform convergence of the finite-dimensional distributions $(x_{jt})_{t\in\mathbb{T}_T}$, for all $T\in\mathbb{N}$, implies weak convergence of the process:[9] $(x_{jt})_{t\in\mathbb{N}}\Rightarrow\left(\Delta_+^{-\delta}\epsilon_{jt}\right)_{t\in\mathbb{N}}$.

References

Abadir, K., W. Distaso, and L. Giraitis (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics 141*(2), 1353–1384.

Abadir, K. and J. R. Magnus (2005). *Matrix Algebra*. Cambridge University Press.

Abadir, K. M. and G. Talmain (2002). Aggregation, persistence and volatility in a macro model. *Review of Economic Studies 69*(4), 749–79.

[9] Note that uniform convergence of the discrete time series implies the same for its cadlag continuous extension $X_j(t)=x_{j\lfloor t\rfloor}$, $t\in\mathbb{R}_+$, and that the space D_∞ of cadlag functions on $[0\,\infty)$ is complete and separable (Billingsley, 1999, Theorem 16.3, p.170). The discrete time fractional noise $\left(\Delta^{-\delta}\epsilon_{jt}\right)_{t\in\mathbb{N}}$ can be defined as in Mandelbrot and Wallis (1969) as a discretized version of the fractional Brownian motion or, directly, as in Hosking (1981).

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

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley.

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

Chevillon, G. and S. Mavroeidis (2017). Learning can generate long memory. *Journal of Econometrics 198*, 1–9.

Cox, D. R. (2014). Scaling. Symposium presentation and video, Nuffield College, Oxford.

Cox, D. R. and M. W. H. Townsend (1947). The use of the correlogram in measuring yarn irregularities. *Proceedings of the Royal Society of Edinburgh, Section A 63*, 290–311.

Cubadda, G., A. Hecq, and F. C. Palm (2009). Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics 148*(1), 25–35.

Davidson, J. and P. Sibbertsen (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of Econometrics 128*(2), 253–82.

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

Diebold, F. X. and K. Yılmaz (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal 119*, 158–171.

Doornik, J. A. and M. Ooms (2001). *A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima Package 1.01 for Ox*. Package manual.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*, 221–38.

Gonçalves, E. and C. Gourieroux (1988). Agrégation de processus autorégressifs d'ordre 1. *Annales d'Economie et Statistiques 12*, 127–149.

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

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*(1), 15–29.

Gray, R. M. (2006). *Toeplitz and circulant matrices: A review*. NOW Publishers Inc.

Grenander, U. and G. Szegö (1958). *Toeplitz forms and their applications*, Volume 321. University of California Press.

Grenander, U. and G. Szegö (2001). *Toeplitz forms and their applications*. Providence, RI: AMS Chelsea Publishing. 2nd ed.

Haldrup, N. and R. Kruse (2014). Discriminating between fractional integration and spurious long memory. Technical report, CREATES Research Paper 2014-19.

Haldrup, N. and J. E. Vera-Valdès (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics 199*, 1–11.

Hecq, A., S. Laurent, and F. C. Palm (2016). On the univariate representation of BEKK models with common factors. *Journal of Time Series Econometrics 8*(2), 91–113.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills and K. D. Patterson (Eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, Chapter 1, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F., A. R. Pagan, and J. D. Sargan (1984). Dynamic specification. *Handbook of econometrics 2*, 1023–1100.

Hosking, J. (1981). Fractional differencing. *Biometrika 68*, 165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng. 116*, 770–808.

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

Johansen, S. and M. N. Nielsen (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory 32*, 10951139.

Mandelbrot, B. (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer.

Mandelbrot, B. and J. Wallis (1968). Noah, Joseph and operational hydrology. *Water Resources Research 4*, 909–918.

Mandelbrot, B. and J. Wallis (1969). Some long-run properties of geophysical records. *Water Resources Research 5*, 321–340.

Miller, J. I. and J. Y. Park (2010). Nonlinearity, nonstationarity, and thick tails: How they interact to generate persistence in memory. *Journal of Econometrics 155*(1), 83 – 89.

Müller, U. K. and M. W. Watson (2008). Testing models of low-frequency variability. *Econometrica 76*(5), 979–1016.

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

Perron, P. and Z. Qu (2007). An analytical evaluation of the log-periodogram estimate in the presence of level shifts. Working paper, Boston University.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics 22*(2), 129–162.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer-Verlag. Springer series in statistics, New York - Berlin - Heidelberg - Tokyo.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics 23*, 1630–61.

Schennach, S. (2013). Long memory via networking. Working paper CWP13/13, CEMMAP.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Technical report, Working Paper 1101, Queens University.

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics 137*, 277–310.

Smith, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *The Journal of Agricultural Science 28*(01), 1–23.

Szegö, G. (1915). Ein genzwertsatz über die toeplitzschen determinanten einer reellen positiven funktion. *Mathematische Annalen 76*(4), 490–503.

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

Zellner, A. and F. Palm (1974). Time series analysis and simultaneous equation econometric models. *Journal of Econometrics 2*(1), 17–54.