cemmap

centre for microdata methods and practice

Long memory via networking

Susanne M. Schennach

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP49/18

Long memory via networking

Susanne M. Schennach*
Economics Dept., Brown University, Providence, RI 02912.
smschenn@brown.edu

This version: June 12, 2018 — First draft: June 18, 2012.

Abstract

Many time-series exhibit "long memory": Their autocorrelation function decays slowly with lag. This behavior has traditionally been modeled via unit roots or fractional Brownian motion and explained via aggregation of heterogenous processes, nonlinearity, learning dynamics, regime switching or structural breaks. This paper identifies a different and complementary mechanism for long memory generation by showing that it can naturally arise when a large number of simple linear homogenous economic subsystems with short memory are interconnected to form a network such that the outputs of the subsystems are fed into the inputs of others. This networking picture yields a type of aggregation that is not merely additive, resulting in a collective behavior that is richer than that of individual subsystems. Interestingly, the long memory behavior is found to be almost entirely determined by the geometry of the network, while being relatively insensitive to the specific behavior of individual agents.

Keywords: Long memory, fractionally integrated processes, spectral dimension, networks, fractals.

1 Introduction

It is widely recognized that many economic and financial time-series data exhibit "long memory" (e.g., Mandelbrot and Ness (1968), Granger and Ding (1996), Baillie (1996), Comte and Renault (1996)), so that shocks have a persistent effect. Long memory can equivalently be characterized via a slow rate of decay of the autocorrelation function with lag or by a divergence of the power spectrum near the origin (Baillie (1996)). Explaining and modeling these features has led to an active literature on fractional Brownian motion (e.g., Mandelbrot and Ness (1968), Granger and Ding (1996), Comte and Renault (1996), Baillie (1996)), aggregation (e.g., Granger (1980), Zafaroni (2004), Abadir and Talmain (2002), Chambers (1998)), structural

*The author would like to thank Lars Hansen, Eric Renault, Adam McCloskey, Florian Gunsilius, seminar participants at Cemmap/UCL, at the University of Cambridge and at the Harvard/MIT Econometrics seminar and anonymous referees for useful comments and acknowledges support from NSF grants SES-1061263/1156347, SES-1357401 and SES-1659334.

breaks and/or regime switching (e.g., Diebold and Inoue (2001), Perron (1989), Perron and Qu (2007), Davidson and Sibbertsen (2005), Granger and Ding (1996)) unit roots (e.g., Hall (1978), Nelson and Plosser (1982), Perron (1988), Phillips (1987)), learning dynamics (e.g., Alfarano and Lux (2005), Chevillon and Mavroeidis (2011)), nonlinearity (e.g., Chen, Hansen, and Carrasco (2010), Miller and Park (2010)), as well as other mechanisms (e.g., Parke (1999), Calvet and Fisher (2002)). While these approaches all identify plausible mechanisms generating a long memory behavior, the search for a simple structural explanation for long memory is still actively ongoing (especially for the popular "fractionally integrated" processes). The goal of this paper is to identify a new, different and arguably more universal mechanism.

We demonstrate that long memory can naturally arise when a large number of simple economic subsystems (or agents) are interconnected to form a network such that the outputs of each of the subsystems are fed into the inputs of others. The agents are "simple" in the sense that they are linear, have a short memory and are homogenous (although our results are also robust to the presence of heterogeneity). Networking yields a type of aggregation that is not merely additive, resulting in a collective behavior that is richer than that of individual subsystems. The long memory behavior is found to be mainly determined by the network geometry, while being relatively insensitive to the specific behavior of individual agents.

We show that the key geometric factor, called the *spectral dimension*, can be calculated for general classes of networks. These classes include not only simple periodic networks, but also more general *fractal* networks, which provide a useful description of social and economic networks (Song, Havlin, and Makse (2005), Inaoka, Ninomiya, Taniguchi, and Takayasu (2004)). Fractals (Mandelbrot and Ness (1968)) are mathematical objects that exhibit some form of self-similarity across scales, thus mimicking people's natural tendency of aggregating into hierarchical structures (e.g., work groups, departments, firms, conglomerate, sectors, etc.). Drawing from the literature on diffusion on fractals (Havlin and Ben-Avraham (1987)), we show that a variety of plausible network structures exhibit a wide range of spectral dimensions and thus generate long memory processes with a wide range of power spectra characteristics.

Our results are distinct from the known fact that long-memory fractionally integrated processes can arise from the additive aggregation of an infinite number of heterogenous times series (Granger (1980)), when some individual series approach a unit-root behavior arbitrarily closely. In contrast, in our framework, all subsystems, on their own, have short memory, thus demonstrating that the aggregation via the network structure is the sole source of the long memory behavior.

Our framework also differs from recent efforts directed at connecting network structure and the propagation of adverse shocks, which focus on the "contagion" of catastrophic events in specific sectors, such as bank failure (e.g. Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), Elliott, Golub, and Jackson (2014), Gouriéroux, Héam, and Monfort (2012), among others) over just a few time periods. In contrast, our model generates long memory even from every day shocks, and not just through rare catastrophic events, and makes specific predictions regarding the network's spectral response within an infinite-horizon framework.

The implications of the economy's network structure on aggregate fluctuations is also receiving considerable attention (e.g., Long and Plosser (1983), Horvath (1998), Dupor (1999), Gabaix (2011), Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012)). This strand of literature does not seek to generate long memory behavior, however, and instead centers on explaining why micro-level noise does not simply average out in the aggregate or how business cycles can arise.

The present paper generates general classes of long memory behavior by considering a general dynamic model in the limit of large networks characterized by scaling laws (including, but not limited to, fractal networks). In this limit, the effect of network geometry on the small-frequency spectrum dominates the effect of individual subsystems, a feature that could not be captured by earlier finite network models. We thus make a direct link between so far distinct literatures, the study of long memory and of the structure of economic networks.

In the sections below, we first develop a general method to calculate a network's spectral response as a function of a simple parameter with a natural geometric interpretation. We then calculate this parameter for infinite periodic networks in any number of dimensions, before focusing on fractal networks, which enable a richer range of possible long memory behaviors. The Supplemental Material (Schennach (2018)) provides a simple empirical example to illustrate the theory.

2 Vector autoregressive formulation

We construct the generating process via a collection of elementary short-memory subsystems (the nodes) interconnected as a network (see Figure 1). Each subsystem takes a number of "input" variables as given (e.g. supply of various input goods) and decides the value of output variables (e.g. quantity produced). Without loss of generality, we consider agents that have only one output (since multiple outputs can merely be modeled as multiple agents taking the same inputs but yielding different outputs). The terms "input" or "output" do not necessarily refer to goods being purchased or sold. "Input" denotes information the system takes as given and cannot change while an "output" denotes variables the subsystem can decide and that provides information that can propagate to other subsystems. We place no fundamental restrictions in the direction of the flow of information (except when considering specific examples). If the "output" of subsystem A goes to subsystem B, the output could be sent to another subsystem C or back to the same subsystem A.

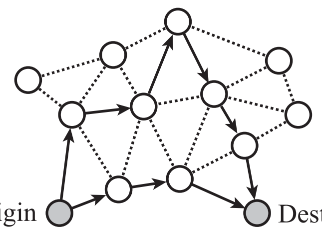

Figure 1: General ideas underying of the approach. Exogenous short-memory noise is fed in to a network of short-memory subsystems at the "origin". This noise is then propagated, through numerous paths of various geometries and lengths, to the "destination". It is the sum of all of these indirect effects that generates the long memory property of the noise monitored at the "destination".

We consider networks consisting of linear subsystems (that is, their output is linear in the input history). If we further assume that the dynamic response of each system to noise is invariant with respect to time shifts, we can then model the response of each subsystem via a convolution (i.e., a linear translation-invariant filter).[1] Working in the linear limit not only makes the problem analytically tractable, but also offers the advantage of illustrating that nonlinearity is not necessary to generate long memory within our framework. One can also interpret our linear approach as a linearization of the network's nonlinear subsystems that is justified in the limit of small noise.

In a discrete time framework, the behavior of such interconnected linear agents can be fully expressed as a vector autoregressive process:[2]

$$X_t = \sum_{s=0}^{\infty} W_s X_{t-s} + U_t \tag{1}$$

where X_t is a $N \times 1$ vector of each individual agent's output at time t, W_s is a $N \times N$ matrix of coupling coefficients for a given lag s and U_t is a $N \times 1$ vector of idiosyncratic zero mean shocks (whose covariance structure will be specified later). Vector autoregressive dynamics arise naturally as the solution to numerous utility maximization problems or as the linearization of such solutions around an equilibrium and are often used to describe model economies (e.g. Long and Plosser (1983), Foerster, Sarte, and

[1] Satisfying this translation invariance assumption may involve working with some deterministic transformation of the model, e.g. discounted present-values or logarithms of some variables.

[2] Our approach can easily be adapted to continuous processes, since our proofs rely on a spectral representation — see ealier version of the present paper Schennach (2013).

Watson (2011), Özgür and Bisin (2013)). For finite N, the dynamics of such system is well-characterized,[3] but considering the $N \to \infty$ limit opens the way to a broader range of interesting dynamics.

To present the main ideas more transparently, we assume that the effect of all inputs on the outputs have the same time-dependence, up to a multiplicative prefactor. This assumption is often satisfied when all the agents solve the same type of optimization problem.

Assumption 1 *The sequence of matrices W_s factor as $W_s = r_s W$, where r_s is a lag-dependent scalar and W is a constant $N \times N$ matrix satisfying*[4] $\sum_{j=1}^{N} W_{ij} = 1$ *for* $i = 1, \ldots, N$.

This factorization implies that the network structure is encoded in the W matrix, while the individual dynamic response of an agent is encoded in the r_s sequence. We relax this assumption, to allow for some form of heterogeneity, in Section C.4 of the Supplemental Material.

To fix the ideas, it is helpful to provide a specific idealized example of impulse response function r_s for a simple and stylized variant of the classic model economy of Long and Plosser (1983). In this model, each agent i of the network produces one good i using other goods $j \neq i$ as inputs, according to a Cobb-Douglas production function with constant returns to scale and a random multiplicative productivity shock (unknown to the agents at the time of making production decisions). The vector X_t in Equation (1) contains the log output of each good (up to an additive constant shift). Goods are perishable (i.e. last only one time period) and agents chose the allocation of goods to optimize expected production.

Unlike Long and Plosser's model, labor inputs are here also provided through a network and treated symmetrically with the other inputs. In this approach, the constraint of a constant total labor force is not imposed, which can be interpreted as labor being measured in productivity-weighted units that can evolve over time through networking interactions. Labor not entering into the production of goods can enter into the "production" of a more productive labor force (e.g. via training) or into the "production" of leisure, viewed as a consumption good. Our model can be solved in the same fashion as Long and Plosser's original model, by taking the limit of zero labor share and relabelling one good as labor.[5] The quantity dynamics of this

[3] In particular, finite networks can only generate long memory of a unit-root type: See Theorem 6 and 7 in the working paper version of the present paper (Schennach (2013)).

[4] As every agent is already assumed to have the same response function r_s, the condition $\sum_{j=1}^{N} W_{ij} = 1$ can be seen as a normalization to ensure a unique factorization.

[5] See Section D.1 of the Supplemental Material for details

economy follows Equation (1) and satisfies Assumption 1 with $r_s = \mathbf{1}\{s=1\}$, with a matrix W whose entries W_{ij} are equal the equilibrium cost shares of each commodity j in the production of another commodity i and with disturbances U_t related to the agents' random productivity shocks.

A simple "one-lag" autoregressive process (i.e. with $r_s = \mathbf{1}\{s=1\}$) is also sufficient to cover a broad range of model economies that include durable capital goods or labor (see Equation (10) and Section V in Foerster, Sarte, and Watson (2011)), after linearization of the model around the equilibrium. More fundamentally, our subsequent analysis actually holds for very general forms of the sequence r_s, which is helpful to consider more complex models of firm behavior. For instance, some models of learning often take the form of such general convolutions (Chevillon and Mavroeidis (2011)). Other examples would be when each agent can be described by a state space or dynamic latent variable model (Harvey, Koopman, and Shephard (2004)). Such a model would then admit a representation in the form of a general convolution (an autoregressive process that could have infinite order), when expressed solely in the terms of observable variables, even if the original formulation of the model had a single lag.

With some concrete examples of agent behavior in mind, we can proceed to study the network dynamics. Letting X denote the entire history of X_t (and similarly for U), we can introduce the convolution operator R (operating on a sequence of vectors), defined as $[RX]_t \equiv \sum_{s=0}^{\infty} r_s X_{t-s}$. (In the simple one-lag case, R is a standard lag operator.) With this notation, Equation 1 reduces to

$$X = RWX + U \tag{2}$$

where the convolution R and the multiplication by a matrix W actually commute ($RWX = WRX$), since they separately act in the time and spatial domains, respectively. By repeated substitution of X by its expression from (2), we directly obtain an infinite moving-average representation:

$$X = \sum_{n=0}^{\infty} W^n R^n U,$$

assuming this sum converges. In the absence of noise, this system adopts a nonrandom steady-state equilibrium $X_t = 0$. For simplicity of exposition, we consider how this equilibrium is perturbed by introducing a stationary short-memory common shock at one (or more) point(s) in the network (hereafter called the "origins" and labelled by o) and by measuring its impact at other arbitrary points in the network (hereafter called the "destinations" and labelled by d). To capture this noise source setup, we let $U = e^o Y$ where Y is a scalar sequence and e^o is a selection vector containing unit entries where perfectly correlated noise is to be introduced and zeros elsewhere.

(Multiple uncorrelated noise sources can be easily handled by calculating the resulting X for each source separately and adding the corresponding power spectra and/or autocorrelation functions. A more general noise covariance structure can be reduced to the uncorrelated case by appropriately redefining the network, as shown in Section C.2 of the Supplemental Material.)

The aggregate impact of the input noise(s) on many points of the network can be determined by introducing a selection vector e^d having unit entries for the destination point(s) of interest and zero elsewhere. We are thus interested in the quantity $Z \equiv \left(e^d\right)' X$, which can be written as:

$$Z = \left(e^d\right)' \sum_{n=0}^{\infty} W^n R^n e^o Y = \sum_{n=0}^{\infty} c_n R^n Y \tag{3}$$

where

$$c_n \equiv \left(e^d\right)' W^n e^o. \tag{4}$$

In this formalism, information regarding the geometry of the network (and the choice of destination vector e^d and origin vector e^o) is encoded in the scalar c_n coefficients. For the remainder of the paper, we will (i) see how the c_n determine whether the limiting process (3) has long memory and (ii) determine the behavior of the coefficients c_n for a range of economically motivated, yet idealized, examples of networks.

3 Long memory behavior

Since our building blocks are stationary processes and translation-invariant operators, it is natural to state our results in terms of spectral representations. Following standard practice (see, e.g., Lobato and Robinson (1996), Baillie (1996), Granger and Ding (1996)) we consider a divergence of the power spectrum at the origin as a signature of a process exhibiting long memory. In this section, we will see how the asymptotic rate of decay of the coefficients c_n very directly determines the rate of divergence of the power spectrum at the origin.

We first state our regularity conditions regarding the process Y and operator R.

Assumption 2 *The stochastic process Y_t admits the moving average representation $Y_t = \sum_{s=0}^{\infty} y_s G_{t-s}$ where G_s are independent $N(0,1)$ random variables (indexed by s) and where the real sequence y_t satisfies (i) $\sum_{t=0}^{\infty} |y_t| < \infty$ and (ii) $\sum_{t=0}^{\infty} y_t \neq 0$.*

Assumption 3 *The real sequence r_t defining the convolution R satisfies (i) $\sum_{t=0}^{\infty} |r_t| (1+t^2) < \infty$, (ii) $\sum_{t=0}^{\infty} r_t = 1$, (iii) $\sum_{t=0}^{\infty} r_t t \neq 0$ and (iv) $\sum_{t=0}^{\infty} r_t t^2 > \left(\sum_{t=0}^{\infty} r_t t\right)^2$.*

Although it is not necessary for the applicability of our approach, Assumption 2 singles out Gaussian processes for simplicity of exposition. This assumption also rules out the degenerate case making any divergence in the spectrum at the origin impossible because the input noise has no zero-frequency component.

Assumption 3(i) is a standard constraint on the tail behavior of r_t that implies that its spectrum $\tilde{r}(\lambda) \equiv \sum_{t=0}^{\infty} r_t e^{i\lambda t}$ is twice continuously differentiable. It also implies that r_t belongs to ℓ_1. Assumption 3(ii) imposes constant returns to scale (when combined with Assumption 1). Assumption 3(iii) rules out an exceptional case that would eliminate the leading term of one of our asymptotic expansions. Assumption 3(iv) is automatically satisfied if, in addition, $r_r \geq 0$, but holds more generally as well. Assumption 3(iv) implies that the spectrum $\tilde{r}(\lambda)$ does not exceed 1 in magnitude near the origin (and, in fact, can be replaced by that latter condition without affecting the results). It should be noted that, in our leading example of the stylized Long and Plosser-type model, $r_t = \mathbf{1}\{t = 1\}$ and r_t thus trivially satisfies Assumption 3.

The requirement that our sequences of coefficient y_t and r_t belong to ℓ_1 is a transparent way to ensure that all our building blocks have short-memory, so that any long-memory behavior must be due to the network structure. Note that ℓ_1 membership implies ℓ_2 membership, a property that is central to the theory of stochastic processes (Doob (1953)). A side-benefit is that ℓ_1 is closed under convolutions, so convolutions can be freely iterated without worries about domains of validity.

To circumvent well-known difficulties in defining the power spectrum of potentially nonstationary processes (Mandelbrot and Ness (1968), Flandrin (1989), Loyne (1968)), we view a long memory process as a limiting case of a sequence of stationary processes. Accordingly, we define a sequence $Z^{\bar{n}}$ of stationary processes.

Definition 1 *Let $Z^{\bar{n}} = \sum_{n=0}^{\bar{n}} c_n R^n Y$ where R satisfies Assumption 3 and Y satisfying Assumption 2 and define the corresponding spectrum $\tilde{z}^{\bar{n}}(\lambda) = \sum_{n=0}^{\bar{n}} c_n (\tilde{r}(\lambda))^n \tilde{y}(\lambda)$, where tilded symbols denote spectra associated with the corresponding process or convolution: $\tilde{y}(\lambda) = \sum_{t=0}^{\infty} y_t e^{i\lambda t}$ and $\tilde{r}(\lambda) = \sum_{t=0}^{\infty} r_t e^{i\lambda t}$.*

Each $Z^{\bar{n}}$ is associated with a corresponding well-defined power spectrum $|\tilde{z}^{\bar{n}}(\lambda)|^2$ and we study the behavior of $\lim_{\bar{n}\to\infty} |\tilde{z}^{\bar{n}}(\lambda)|^2 \equiv |\tilde{z}^{\infty}(\lambda)|^2$ as a function of the asymptotic behavior of the sequence of weights c_n. Here we consider the leading case of a power law behavior for c_n — more general behaviors are considered in Section C.1 of the Supplemental Material.

Theorem 1 *Let Assumptions 1-3 hold. If $|c_0| < \infty$ and $\sum_{n=1}^{\infty} |c_n - Cn^{-\gamma}| < \infty$ for some $\gamma \in \mathbb{R}^+$ and $C \in \mathbb{R}$, then there exists a neighborhood $\mathcal{N}$ of the origin such that, for all $\lambda \in \mathcal{N} \setminus \{0\}$, the limiting power spectrum of $Z^{\tilde{n}}$, defined as $|\tilde{z}^{\infty}(\lambda)|^2 \equiv \lim_{\bar{n}\to\infty} |\tilde{z}^{\bar{n}}(\lambda)|^2$, has the following properties:*
(i) If $\gamma \leq 1$ and $C \neq 0$, then

$$|\tilde{z}^{\infty}(\lambda)|^2 = A|\lambda|^{-2\alpha} + o\left(|\lambda|^{-2\alpha}\right) \quad (5)$$

for $\alpha = 1 - \gamma$ and some $A \in \mathbb{R}\setminus\{0\}$, (with the convention that $|\lambda|^{-2\alpha} \equiv |\ln|\lambda||^2$ for

$\alpha = 0$) and

(ii) if $\gamma > 1$ or if $C = 0$, then

$$|\tilde{z}^{\infty}(\lambda)|^2 = A + o(1) \tag{6}$$

for some $A \in \mathbb{R}$.

This result states conditions under which the resulting limiting power spectrum $|\tilde{z}_{\infty}(\lambda)|^2$ exhibits the same asymptotic behavior ($|\lambda|^{-2\alpha}$ as $\lambda \to 0$) as the a widely used fractionally integrated process of order α. Empirically, this behavior can be detected by observing a linear trend in a plot of (an estimated) log power spectrum $\ln|\tilde{z}_{\infty}(\lambda)|^2$ as a function of $\ln \lambda$ for small values λ.

The proof of this Theorem, given in Appendix A, can be informally outlined as follows: The spectral representation of the series $\sum_{n=0}^{\infty} c_n R^n$ is $\sum_{n=0}^{\infty} c_n (\tilde{r}(\lambda))^n$. For the sequence $c_n = n^{-(1-\alpha)}$ this series is very closely related to a Taylor series of the function $(1-x)^{-\alpha}$. Since $\tilde{r}(\lambda) = 1 - C\lambda + o(\lambda)$ for some $C \neq 0$ under our assumptions, combining these results yields a spectral representation of the form $(1 - 1 + C\lambda + o(\lambda))^{-\alpha} = C^{-\alpha}\lambda^{-\alpha} + o\left(\lambda^{-\alpha}\right)$, i.e., a power spectrum of the form $|\lambda|^{-2\alpha}$. This result can easily be shown to be unaffected by summable deviations of the power law $c_n = n^{-(1-\alpha)}$.

Intuitively, long memory arises because each additional convolution lengthens the tail of the impulse response, and because the additive contributions of infinitely many different paths yield a nonsummable aggregate impulse response, even though individual agents have a summable impulse response function. Note that the lengthening of the tail can occur even if the onset of the agents' response is instantaneous (i.e. $r(0) \neq 0$). Of course, long memory cannot arise if the agents only have an instantaneous response, but that situation is ruled out by Assumption 3(iii).

In the case where the limiting power spectrum $|\tilde{z}^{\infty}(\lambda)|^2$ is integrable ($\alpha < 1/2$), we can also establish a stronger form of convergence that implies the existence of a stationary long-memory limiting process $Z^{\infty}(t)$ with a power spectrum behaving as $|\lambda|^{-2\alpha}$ as $|\lambda| \to 0$.

Theorem 2 *Let the Assumptions of Theorem 1 hold. Assume that $|\tilde{r}(\lambda)| < 1$ for $\lambda \in]0, \pi]$ and that $|\tilde{y}(\lambda)|$ is uniformly bounded for $\lambda \in [0, \pi]$. If $\sum_{n=0}^{\infty} |c_n| < \infty$ or if $\alpha < 1/2$, there exists a stationary process $Z_{\infty}(t)$ with spectrum $\tilde{z}^{\infty}(\lambda) \equiv \lim_{n\to\infty} \tilde{z}^n(\lambda)$ and corresponding moving average representation $z^{\infty}(t)$ such that $\int_0^{\pi} |\tilde{z}^n(\lambda) - \tilde{z}^{\infty}(\lambda)|^2 d\lambda \to 0$, $\sum_{t=0}^{\infty} |z^n(t) - z^{\infty}(t)|^2 \to 0$ and $E\left[|Z^n(t) - Z^{\infty}(t)|^2\right] \to 0$ for almost any given $t \in \mathbb{R}$ and $\sum_{t=-\infty}^{\infty} E\left[|Z^n(t) - Z^{\infty}(t)|^2\right] w(t) \to 0$ for a given absolutely integrable, bounded and continuous weighting function $w(t)$.*

One can also establish a similar convergence result that covers both integrable

($\alpha < 1/2$) and non integrable ($\alpha \geq 1/2$) limiting power spectra $|\tilde{z}^{\infty}(\lambda)|^2$ by focusing on increments of the processes (see Section C.3 of the Supplemental Material).

Our results for infinite networks also have implications for the large but finite networks found in the real world. The following theorem establishes that behaviors of finite and infinite networks are similar in a way that makes them empirically difficult to distinguish.

Theorem 3 *Consider an infinite network* $\mathcal{N}_{\infty}$ *and let* $\tilde{z}^{\infty}(\lambda)$ *and* $c_n^{\infty}(\lambda)$ *respectively denote the spectrum and the coefficients obtained for a given set of origin nodes* $\mathcal{O}$ *and a set destination nodes* $\mathcal{D}$*. Consider a finite network* $\mathcal{N}_{*}$ *containing* $\mathcal{O} \cup \mathcal{D}$ *and all nodes of* $\mathcal{N}_{\infty}$ *that are within* n^* *hops of at least one node in* $\mathcal{O} \cup \mathcal{D}$*. Let* $\tilde{z}^{*}(\lambda)$ *and* c_n^{*} *respectively denote the spectrum and the coefficients associated with* $\mathcal{O}$ *and* $\mathcal{D}$ *in the finite network. Assume that* $|\tilde{r}(\lambda)| < 1$ *for* $\lambda \in]0, \pi]$*. Then, under Assumptions 1-3, for any given* $\lambda_{\min} > 0$*,*

$$\sup_{\lambda \in [\lambda_{\min}, \pi]} |\tilde{z}^{\infty}(\lambda) - \tilde{z}^{*}(\lambda)| \leq \frac{2\bar{C}}{1-\bar{r}} \bar{r}^{n^*+1} \tag{7}$$

where $\bar{r} = \sup_{\lambda \in [\lambda_{\min}, \pi]} |\tilde{r}(\lambda)| < 1$ *and* $\bar{C} = \sup_n \max\{|c_n^{\infty}|, |c_n^{*}|\} < \infty$*.*

This result follows from the fact that series of the form $\sum_{n=0}^{\infty} c_n (\tilde{r}(\lambda))^n$ (used in the proof of Theorem 1) converge exponentially fast for $|\lambda| \geq \varepsilon$ (since then $|\tilde{r}(\lambda)| < 1$). Hence, a truncated series (representing a set of finite pathways that can fit within a finite network), tends to be very close to its limiting value for an infinite network. The region $|\lambda| \geq \varepsilon$ where this fast convergence takes place is precisely the only portion of the spectrum that is empirically accessible, since the finite duration of recorded time series limits the smallest frequency for which the spectrum can be reliably determined.

4 Network models

Now that have characterized the connection between the spectrum of the network response and the asymptotic behavior of the coefficients c_n, thanks to Theorems 1 and 2, we turn to the question of determining these coefficients for natural and economically-motivated classes of network geometries and show that power law decays $c_n \propto n^{-\gamma}$ with any $\gamma \in]0, 1]$ can be realized.

To do this, we exploit the following simple geometric interpretation of the coefficients $c_n \equiv (e^d)' W^n e^o$. If the matrix W contains only nonnegative elements, it can be viewed as the transition matrix of a Markov chain, or random walk,[6] on a network. (It is only this geometric interpretation that relies on $W_{ij} \geq 0$ — the definition of c_n holds more generally.) Consider some vector e^d that has a single nonzero element

[6] Here, the term is used more broadly, since the random walk's node-to-node hops are not necessarily independent.

at i representing the starting point of the random walker. This walker then jumps to another node j with a probability W_{ij}. The probability distribution of the random walker will then be given by the row vector $\left(e^d\right)' W$. After n jumps, the distribution is $\left(e^d\right)' W^n$. The probability that a random walker lands on the source node[7] is then selected by multiplying by e^o, to yield $c_n = \left(e^d\right)' W^n e^o$. A similar interpretation holds when e^d or e^o have multiple nonzero elements: One then has multiple simultaneous random walks with different start and end points.[8]

As initial examples of network, we consider simple periodic networks in d dimensions in which the nodes are indexed by points $i \in \mathbb{Z}^d$, for a fixed positive integer d.

Theorem 4 *Consider a network with nodes on $\mathbb{Z}^d$, all of which are reachable. If e^d and e^o have a single nonzero element, the coupling coefficients matrix W_{ij} ($i, j \in \mathbb{Z}^d$) satisfies (i) $W_{ij} = W_{i+k,j+k}$ for all $i, j, k \in \mathbb{Z}^d$, (ii) $W_{ii} > 0$ for all $i \in \mathbb{Z}^d$, (iii) $W_{ij} = W_{ji}$ and $W_{ij} \geq 0$ for all $i, j \in \mathbb{Z}^d$ and (iv) for each $i \in \mathbb{Z}^d$, $W_{ij} \neq 0$ for a finite number of j, then $c_n = Cn^{-d/2} + O\left(n^{-1-d/2}\right)$ for some $C > 0$.*

Interestingly, $d = 1$ (a linear network) gives us c_n scaling as $n^{-1/2}$ and therefore a long memory process of order $1 - 1/2 = 1/2$ by Theorem 1. Similarly $d = 2$ gives an order of $1 - 2/2 = 0$ (i.e. a spectrum with a logarithmic divergence at the origin). For $d = 3, 4, \ldots$ the sequence $n^{-d/2}$ is absolutely summable, so that no long memory results. However, this does not imply that high-dimensional networks cannot generate any long memory behavior. The *aggregate output of a group of nodes* can exhibit long memory in networks of an arbitrarily high dimension. One can show that, if one considers the *sum of the nodes outputs* over a subspace of dimension h of the periodic lattice, then the power law from Theorem 4 becomes $c_n = Cn^{-(d-h)/2} + O\left(n^{-1-(d-h)/2}\right)$, so the dimension of the aggregate considered offsets the effect of the dimensionality of the network. The reason for this result is simply that the problem reduces to studying a random walk consisting of jumps across *different* hyperplanes, since jumps within one hyperplane are irrelevant. This effectively removes h dimensions from the random walk, which then behaves like a random walk on a $d - h$ dimensional lattice.

To fill in the gaps in the integral exponents generated by the periodic lattices, we

[7] That the random walk proceeds backward from the destination to the source is merely a consequence of the choice of normalization $\sum_{j=1}^{N} W_{ij} = 1$. One could alternatively consider models where columns of W add up to one and the natural random walk interpretation would then hold from the origin to the destination.

[8] It should be noted that, in the $N \to \infty$ limit, the most interesting cases arise when the fraction of nonzero elements in the origin vector e^o or in the destination vector e^d decays to zero as $N \to \infty$. Otherwise, the c_n may not decay to zero (since a constant vector is a eigenvector of W with unit eigenvalue, by Assumption 1).

would need networks that effectively have a "fractional dimension". Such mathematical objects, called *fractals* (Mandelbrot (1982)), have been constructed and derive their properties from the power law nature of their self-similarity across scales. Fractals have proven to be an effective tool to represent many natural and human-made phenomena (Mandelbrot and Ness (1968)) and actual social or economic networks have been observed to exhibit self-similarity across scales (Song, Havlin, and Makse (2005), Inaoka, Ninomiya, Taniguchi, and Takayasu (2004)).

Since there is a direct relationship between random walks on a network and the c_n coefficients, we can borrow a key result from the literature on random walks (or *diffusions*) in fractals (e.g., Havlin and Ben-Avraham (1987)): The probability c_n that a random walker visits a given point after n steps scales as $n^{-\tilde{d}/2}$ asymptotically, where $\tilde{d}$ is a positive real number known as the *spectral dimension* that is related to the geometry of the network (but not uniquely determined by other common descriptors, such as the degree distribution). There is therefore a rather direct correspondence with diffusion on periodic lattices in Euclidian space. This finding comes from a combination of formal analytical treatments of various self-similar fractals (such as the Sierpinski Gasket) as well as from thorough Monte Carlo simulations on random *statistically* self-similar fractals (such as those obtained via diffusion-limited aggregation (Witten and Sander (1981))) guided by renormalization arguments (ben Avraham and Havlin (2005), Given and Mandelbrot (1983), Havlin and Ben-Avraham (1987)).

Among the many examples of networks with a well-defined spectral dimension, we describe here in more detail examples of network classes that represent natural hierarchical extensions of network connectivities commonly used in theoretical economic models. For conciseness, we only report the relevant spectral dimensions $\tilde{d}$ (which yields the scaling $c_n \propto n^{-\tilde{d}/2}$), referring the reader to the original references for formal statements and proofs. Our examples cover the entire range of spectral dimensions $\tilde{d} \in]0, 2]$ that yield long memory processes.

The first class generalizes star networks that arise in certain network formation games (see, e.g., Proposition 3 in Jackson (2005)) or in studies of the effect of the simultaneous presence of highly and weakly connected agents (e.g., Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012)). Here we consider *hierarchical* star geometries (Figure 2): Not only can firms be connected via a star network, but so can sectors of the economy, at various levels of aggregation. Some models of network formation actually generate such networks: Optimal transportation networks often take the form of minimal spanning trees (Sharkey (1995)), which exhibit a statistically self-similar nature (Steele, Shepp, and Eddy (1987)). Simple hierarchical star networks can be

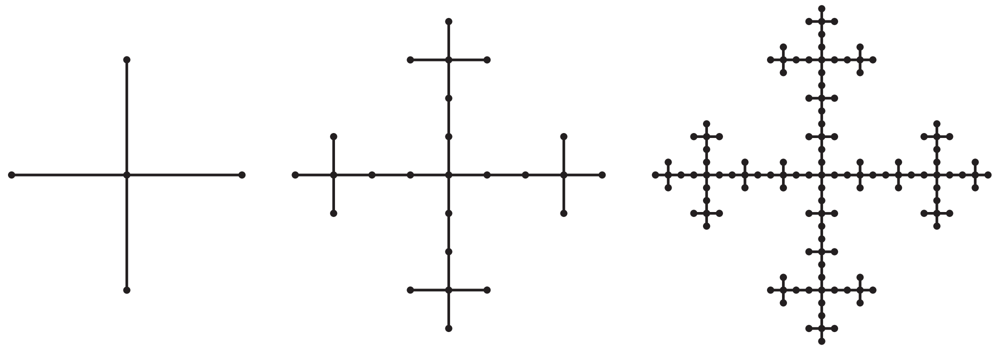

Figure 2: Example of a hierarchical star-like network ($m = 2$ case)).

constructed by starting with a node connected to $2m$ identical neighbors (located along each the m Cartesian axes). One then repeatedly applies the following *generating rule*: Replace each node by a star consisting of $2m + 1$ nodes, as illustrated in Figure 2. The spectral dimension of such a network has been calculated analytically (Christou and Stinchcombe (1986)):

$$\tilde{d} = \frac{2 \ln (2m + 1)}{\ln 3 + \ln (2m + 1)}. \tag{8}$$

Possible spectral dimensions thus range from $\tilde{d} = 1$ (for $m = 1$) to $\tilde{d} \to 2$ (in the limit as $m \to \infty$). While Equation (8) only yields discrete values of $\tilde{d}$ in $]1, 2[$, on can fill-in the whole continuum of values $\tilde{d} \in]1, 2[$ by simply alternating two different *generating rules* (corresponding to different m) at each step of the recursion to interpolate between the values of $\tilde{d}$ generated by Equation (8), as shown more formally in Theorem 5 of Appendix B. Other examples of star-like networks can be found in Given and Mandelbrot (1983).

Another common type of network is a ring, which is used to model weakly connected firms (e.g. Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015)). We consider here more general hierarchical ring networks, where, at each step of the generation process, the *generating rule* consists of replacing one link of the network by a ring of $u + v$ nodes such that the original nodes are u hops apart on the ring along one side of the ring and v hops apart along the other side (see Figure 3), with each link being of equal strength. The spectral dimension of such a network is (see Rozenfeld, Havlin, and ben Avraham (2007) and Appendix B):

$$\tilde{d} = 2 \frac{\ln (u + v)}{\ln (uv)} \tag{9}$$

with $u, v \in \{2, 3, 4, \ldots\}$. Possible spectral dimensions thus range from $\tilde{d} = 1$ (taking the limit as $u, v \to \infty$ with $u/v \to s \in \mathbb{R} \backslash \{0\}$) to $\tilde{d} \to 2$ (if $u/v \to s \in \{0, \infty\}$) and any values in between (again via Theorem 5 in Appendix B). This model can be generalized to K pathways of different lengths (Tejedor (2012)). Hierarchical ring

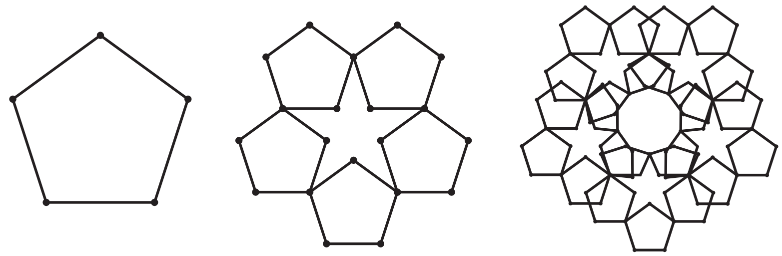

Figure 3: Example of a hierarchical ring-like network, a (2,3)-flower.

networks can model the fact that two sectors of the economy may appear connected by a single link when viewed at a coarse level of aggregation, while a finer level of disaggregation may actually reveal that the connection takes place via a number of intermediary links, possibly along multiple (competing) pathways. Another interesting connection is that hierarchical ring networks can generate a so-called *scale-free* degree distribution (Rozenfeld, Havlin, and ben Avraham (2007)) (i.e. the number of neighbors follows a power law, or Pareto, distribution) for which there is empirical evidence in economic networks (Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012)).

Another network topology with economic relevance is an idealized supply chain, which models the production of a good that requires consecutive steps performed by a sequence of firms on the "backbone" of the supply chain. Each of these firms also requires inputs from other firms located on side attachments (the "fibers"). The simplest example is the linear "comb" structure of Figure 4. One can consider more general structures where the backbone and the fibers are themselves arbitrary fractal networks with spectral dimensions $\tilde{d}_B$ and $\tilde{d}_F$, respectively, and the resulting spectral dimension, as shown by Cassi and Regina (1996) in the context of a diffusion problem unrelated to economic networks, is:

$$\tilde{d} = \tilde{d}_B + \tilde{d}_F - \frac{\tilde{d}_B \tilde{d}_F}{2}, \tag{10}$$

provided $\tilde{d}_B < 4$ and $\tilde{d}_F < 2$. Their results also enable the study the effect of aggregation: If the destination nodes e^d consist of an entire fiber, the appropriate exponent becomes:

$$\tilde{d} = \tilde{d}_B - \frac{\tilde{d}_B \tilde{d}_F}{2} = \tilde{d}_B \left(1 - \frac{\tilde{d}_F}{2}\right). \tag{11}$$

This setup illustrates a simple way to construct networks that produce aggregated signals with long memory having any order of power law decay that can approach the unit root case arbitrarily closely. Our hierarchical star and ring examples delivered fractal networks with any spectral dimensions in $]1,2[$ and we observe here that any value of $\tilde{d}$ in $]0,1[$ can be obtained via Equation (11) for some choice of $\tilde{d}_B, \tilde{d}_F \in]1,2[$.

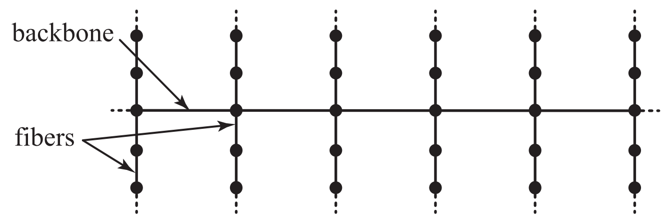

Figure 4: Simple example of a generalized supply chain: A "comb" structure.

In the examples so far, the nonzero elements of the coupling coefficients matrix W_{ij} were either identical or bounded away from zero. Nontrivial spectral dimensions can also be obtained by relaxing that constraint in an otherwise nonfractal network. For instance, consider a linear chain of nodes, each linked to its 2 nearest neighbors and where the jump probabilities $W_{i,i\pm1}$ are each drawn at random (but kept fixed over time) from the density $f(w) \propto w^{-\eta}\mathbf{1}\{w \in [0, 1/2]\}$ for some $\eta \in [0, 1[$. The associated spectral dimension is $\tilde{d} = (1-\eta)/(2-\eta)$ (ben Avraham and Havlin (2005)), thus showing that a range of long memory behaviors can also be obtained in simple networks with strong heterogeneity in the coupling coefficients.

We now have demonstrated simple plausible networks that can exhibit any spectral dimensions $\tilde{d} \in]0, 2]$. Hierarchical star and ring networks cover the $]1, 2[$ range, which is extended to $]0, 1[$ via a simple supply chain construction. The special cases $\tilde{d} = \{1, 2\}$ are covered by simple periodic lattices. Theorem 1, then leads to the conclusion that the divergent spectrum characteristic of fractionally integrated long memory processes of any order can be naturally obtained from the collective behavior of a population of linear homogenous agents interconnected through a (possibly) fractal network with idealized, yet economically motivated, geometries.

In empirical settings, if one has access to a specific observed network structure, it is unnecessary to attempt to recreate this network via generating rules. Instead, a suitable power law behavior can be directly detected as a linear trend (with a slope in the range $[-1, 0]$) in a plot of $(\ln(c_n), \ln n)$, with c_n computed from Equation (4). This method works best when one has access to a very "disaggregated" version of the network geometry data, since this enables a plot of $(\ln(c_n), \ln(n))$ over the widest possible range of values of $\ln n$ before finite size artifacts set in, which facilitates the identification of a linear trend. This type of evidence alone would suggest the applicability of our mechanism, independently of whether or not the network can be constructed via iteration of a simple generating rule.[9] Section D.2 of the Supplemental Material provides an empirical example of such an analysis, based on the "input-output accounts"

[9] One can also create examples of networks with a well-defined spectral dimension but that exhibit no self-similarity — see Appendix B.

database compiled by the Bureau of Economic Analysis and describing interactions between sectors of the US economy.

5 Conclusion

We show that long memory can naturally arise when a large number of simple linear homogenous economic subsystems with a short memory are interconnected to form a network. The long memory behavior then is largely determined by the geometry of the network while being relatively insensitive to the specific behavior of individual subsystems. Under weak regularity conditions, the power spectrum of the network's response to exogenous short-memory noise exhibits the same power spectrum signature as a fractionally integrated processes $I(\alpha)$, with α related to the scaling properties of the network (its *spectral dimension*). This work not only provides a plausible structural model for the generation of fractionally integrated long memory processes, but also demonstrates that long memory is possible without nonlinearity, heterogeneity, unit roots or near unit roots, learning or structural breaks (although these mechanisms can obviously play a role as well). The proposed approach also makes a direct connection between the literatures focusing on long memory processes, economic networks and diffusion on fractals. It also suggests that the spectral dimension would be a very useful descriptor to add to the list of commonly used summary statistics (e.g., de Paula (2016)) to characterize networks (degree distribution, centrality, betweenness, etc).

A Proofs

The following Lemmas summarize well-known results from the theory of stochastic processes (e.g., Doob (1953), Chap. XI, Section 9):

Lemma 1 *If $y \in \ell_1$ (and thus $y \in \ell_2$) then it also admits a spectral representation $\tilde{y}(\lambda) \equiv \sum_{t=0}^{\infty} y_t e^{i\lambda t}$ and an associated power spectrum $|\tilde{y}(\lambda)|^2$. Moreover, $\tilde{y}(\lambda)$ is a bounded and square-integrable function defined for any $\lambda \in \mathbb{R}$. A corresponding result holds with $Y, y, \tilde{y}$ replaced by $R, r, \tilde{r}$, respectively, with $\tilde{r}(\lambda) = \sum_{t=0}^{\infty} r(t) e^{i\lambda t}$.*

For conciseness, we often call the "spectral representation" simply the "spectrum", reserving the term "power spectrum" (or "spectral density") for its modulus square. The $\tilde{r}(\lambda)$ is traditionally called the gain function while the r is the usual impulse response function. The following lemma summarizes a simple form of convolution theorem.

Lemma 2 *Let $y^0 \in \ell_1$ and let $y^n = r^n \otimes \cdots \otimes r^1 \otimes y^0$ with $r^1, \ldots, r^n \in \ell_1$ for some $n \in \mathbb{N}$ and with $\otimes$ denoting convolutions. Then $y^n \in \ell_1$ and the spectral representation of these quantities are related through $\tilde{y}^n(\lambda) = \tilde{r}^n(\lambda) \cdots \tilde{r}^1(\lambda)\, \tilde{y}^0(\lambda)$.*

Note that Lemma 2 does not let us conclude that $\lim_{n\to\infty} y^n \in \ell_1$. In fact, it is precisely the fact that $\lim_{n\to\infty} y^n \notin \ell_1$ in general that allows us to consider long memory processes via a limiting process (since processes with summable moving average representation necessarily have short memory).

Definition 2 *To avoid ambiguities due to the multivalued nature of the fractional power function, we define:*

$$(i\lambda)^{\alpha} \equiv \begin{cases} |\lambda|^{\alpha} e^{\alpha i\pi/2} \text{ if } \lambda > 0 \\ |\lambda|^{\alpha} e^{-\alpha i\pi/2} \text{ if } \lambda < 0 \end{cases}$$

Moreover, the following convention for powers of $i\lambda$ is useful to avoid special cases: If $\alpha = 0$, then

$$(i\lambda)^{\alpha} \equiv -\ln(i\lambda) \equiv \begin{cases} \ln|\lambda| + i\pi/2 \text{ if } \lambda > 0 \\ \ln|\lambda| - i\pi/2 \text{ if } \lambda < 0 \end{cases}$$

Lemma 3 *Assumption 3 implies that (i) for some finite $A, B \in \mathbb{R}^{+}\backslash\{0\}$, $\tilde{r}(\lambda) = 1 + Ai\lambda + o(\lambda)$ and $|\tilde{r}(\lambda)|^2 = 1 - B\lambda^2 + o(\lambda^2)$ as $\lambda \to 0$ and (ii) there exists a neighborhood $\mathcal{N}$ of the origin such that $|\tilde{r}(\lambda)| < 1$ for all $\lambda \in \mathcal{N}\backslash\{0\}$.*

Proof. Assumption 3(i) implies that $\tilde{r}(\lambda)$ is everywhere twice continuously differentiable. Thus, in particular, near the origin, we have the expansion $\tilde{r}(\lambda) = A_0 - A_1 i\lambda - \frac{1}{2}A_2\lambda^2 + o(\lambda^2)$ with A_0, A_1, A_2 finite. Assumption 3(i) also implies that the moment theorem applies up to order 2, so that $A_j = \sum_{t=0}^{\infty} t^j r_t$. By Assumption 3(ii) $A_0 = 1$. Since r_t is real, the real part of $\tilde{r}(\lambda)$ is symmetric while its imaginary part is anti-symmetric. Therefore, A_1 and A_2 must be real. Assumption 3(iii) implies that $A_1 \in \mathbb{R}\backslash\{0\}$ and the first conclusion of the lemma follows. Next, we note that $|\tilde{r}(\lambda)|^2 = \left(1 - \frac{1}{2}A_2\lambda^2\right)^2 + A_1^2\lambda^2 + o(\lambda^2) = 1 - A_2\lambda^2 + \frac{1}{4}A_2^2\lambda^4 + A_1^2\lambda^2 + o(\lambda^2) = 1 - (A_2 - A_1^2)\lambda^2 + o(\lambda^2)$, where $A_2 - A_1^2 > 0$ by assumption 3(iv). It follows that $|\tilde{r}(\lambda)| < 1$ in some neighborhood of the origin. ■

Lemma 4 *Assumption 2 implies that $\tilde{y}(\lambda) = B + o(1)$ for some $B \in \mathbb{R}\backslash\{0\}$.*

Proof. Assumption 2 implies that the Fourier transform $\tilde{y}(\lambda)$ is continuous, thus implying an expansion of the form $B + o(1)$. Moreover y_t is real, so $B = \tilde{y}(0)$ is real as well and nonzero by Assumption 2. ■

Lemma 5 *Let Assumptions 2-3 hold. Let c_n and c_n' be two sequences such that $\sum_{n=0}^{\infty}|c_n - c_n'| < \infty$. Then, the corresponding $\tilde{z}^{\infty}(\lambda)$ and $\tilde{z}^{\infty\prime}(\lambda)$ are such that (i) $|\tilde{z}^{\infty}(\lambda) - \tilde{z}^{\infty\prime}(\lambda)|$ is continuous and uniformly bounded in a neighborhood $\mathcal{N}$ of the origin and (ii) whenever $|\tilde{z}^{\infty}(\lambda)|^2 = A|\lambda|^{-2\alpha} + o\left(|\lambda|^{-2\alpha}\right)$ (for $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}^{+}$) we also have $|\tilde{z}^{\infty\prime}(\lambda)|^2 = \tilde{A}|\lambda|^{-2\alpha} + o\left(|\lambda|^{-2\alpha}\right)$ for some $\tilde{A} \in \mathbb{R}$ (with $\tilde{A} = A$ if $\alpha > 0$).*

Proof of Lemma 5. Let $\Delta c_n \equiv c_n - c_n'$ and let $\Delta\tilde{z}^n(\lambda) = \tilde{z}^n(\lambda) - \tilde{z}^{n\prime}(\lambda)$ denote the corresponding spectrum. To prove the result, we exploit the fact that a uniformly convergent sequence of continuous functions converges to a continuous function. Since, by Assumption 3 and Lemma 3, $|\tilde{r}(\lambda)| \leq 1$ for $\lambda \in \mathcal{N}$, some neighborhood of the origin and since $\sum_{n=0}^{\infty}|\Delta c_n| < \infty$ by assumption, we can write, for $\lambda \in \mathcal{N}$,

$$|\Delta\tilde{z}^{\bar{n}}(\lambda) - \Delta\tilde{z}^{\infty}(\lambda)| = \left|\sum_{n=\bar{n}+1}^{\infty} \Delta c_n (\tilde{r}(\lambda))^n\right| \leq \sum_{n=\bar{n}+1}^{\infty} |\Delta c_n| |\tilde{r}(\lambda)|^n \leq \sum_{n=\bar{n}+1}^{\infty} |\Delta c_n| \to 0$$

as $\bar{n} \to \infty$. Therefore, $\Delta\tilde{z}^{\bar{n}}(\lambda)$ converges uniformly to $\Delta\tilde{z}^{\infty}(\lambda)$ as $\bar{n} \to \infty$ over all $\lambda \in \mathcal{N}$. This, combined with the fact that $\Delta\tilde{z}^{\bar{n}}(\lambda)$ is continuous in λ for any finite $\bar{n}$ and $\lambda \in \mathcal{N}$ (since it is a finite sum of continuous functions) implies that $\Delta\tilde{z}^{\infty}(\lambda)$ is continuous in $\mathcal{N}$ and we also have $\Delta\tilde{z}^{\infty}(\lambda) = B + o(1)$ as $\lambda \to 0$. It follows that, for

$\alpha \geq 0$, and as $\lambda \to 0$,
$\tilde{z}^{\infty\prime}(\lambda) = \tilde{z}^{\infty}(\lambda) + \Delta\tilde{z}^{\infty}(\lambda) = A|\lambda|^{-\alpha} + o\left(|\lambda|^{-\alpha}\right) + B + o(1) = \tilde{A}|\lambda|^{-\alpha} + o\left(|\lambda|^{-\alpha}\right)$
for some finite nonzero $\tilde{A}$ (that equals A if $\alpha > 0$). ■

Proof of Theorem 1. By Assumption 3 and Lemma 3, (i) for some finite $A \neq 0$, $\tilde{r}(\lambda) = 1 - Ai\lambda + o(\lambda)$ as $\lambda \to 0$ and (ii) there exists a neighborhood $\mathcal{N}$ of the origin such that $|\tilde{r}(\lambda)| < 1$ for all $\lambda \in \mathcal{N}\setminus\{0\}$. Also, by Lemma 4, $\tilde{y}(\lambda) = B + o(1)$ as $\lambda \to 0$ for some $B \in \mathbb{R}\setminus\{0\}$.

By Lemma 5, we can focus on the case where $c_n = Cn^{-\gamma}$ since absolutely summable deviations from such power law will only contribute to a constant in the spectrum near the origin and hence will not affect the type of divergence that occurs in the spectrum at the origin. Furthermore, we consider the case $C = 1$ without loss of generality to simplify the notation.

Consider first the special case[10] $\gamma = 0$ and hence $\alpha = 1$, so that $c_n = 1$. By Lemma 2, the spectrum of $R^n Y$ is given by $(\tilde{r}(\lambda))^n \tilde{y}(\lambda)$ and thus the spectrum of Z_n is $\sum_{m=0}^{n} (\tilde{r}(\lambda))^m \tilde{y}(\lambda)$ (and the corresponding power spectrum is $|\sum_{m=0}^{n} (\tilde{r}(\lambda))^m \tilde{y}(\lambda)|^2$). For all $\lambda \in \mathcal{N}\setminus\{0\}$, the series $\sum_{m=0}^{\infty} (\tilde{r}(\lambda))^m \equiv \lim_{n\to\infty} \sum_{m=0}^{n} (\tilde{r}(\lambda))^m$ is convergent because $|\tilde{r}(\lambda)| < 1$ and we can directly evaluate this geometric series:

$$\begin{aligned}
\tilde{z}^{\infty}(\lambda) &= \tilde{y}(\lambda)\sum_{n=0}^{\infty}(\tilde{r}(\lambda))^n = \tilde{y}(\lambda)\frac{1}{1-\tilde{r}(\lambda)} \\
&= (B + o(1))\frac{1}{1 - 1 + Ai\lambda + o(\lambda)} = (B + o(1))\frac{1}{Ai\lambda + o(\lambda)} \\
&= B(1 + o(1))\frac{A^{-1}}{i\lambda}\frac{1}{1 + o(\lambda)/\lambda} = B\frac{A^{-1}}{i\lambda}\frac{1 + o(1)}{1 + o(1)} \\
&= \frac{BA^{-1}}{i\lambda}(1 + o(1)) = \frac{BA^{-1}}{i\lambda} + o\left(\lambda^{-1}\right)
\end{aligned}$$

Next, we consider the more general cases where $\alpha \in \,]0,1[$. Consider the Taylor series $C(1-x)^{-\alpha} = \sum_{n=0}^{\infty} c'_n x^n$ for $|x| < 1$ for any nonzero constant C, where

$$c'_n = C\frac{1}{n!}\prod_{j=1}^{n}(\alpha + j - 1)$$

(with $c'_0 \equiv C$ by convention) and note that, for $\lambda \in \mathcal{N}\setminus\{0\}$,

$$\begin{aligned}
\tilde{z}^{\infty\prime}(\lambda) &\equiv \tilde{y}(\lambda)\sum_{n=0}^{\infty} c'_n (\tilde{r}(\lambda))^n = C\tilde{y}(\lambda)(1 - \tilde{r}(\lambda))^{-\alpha} = C(B + o(1))(1 - 1 + Ai\lambda + o(\lambda))^{-\alpha} \\
&= C(B + o(1))A^{-\alpha}(i\lambda + o(\lambda))^{-\alpha} = CBA^{-\alpha}(i\lambda)^{-\alpha}(1 + o(1))(1 + o(\lambda)/\lambda)^{-\alpha} \\
&= CBA^{-\alpha}(i\lambda)^{-\alpha}(1 + o(1)) = \frac{CBA^{-\alpha}}{(i\lambda)^{\alpha}} + o\left(|\lambda|^{-\alpha}\right)
\end{aligned}$$

There remains to show that c_n is sufficiently close to c'_n so that $\tilde{z}^{\infty}(\lambda)$ has the same asymptotic behavior as $\tilde{z}^{\infty\prime}(\lambda)$. By Lemma 5, it is sufficient to show that

[10] This case could be combined with the more general case $\alpha \in \,]0,1[$ below, but this simple case illustrates the idea of the proof with the least technical complications.

$\sum_{n=0}^{\infty}|c_n - c'_n| < \infty$. To this effect, note that

$$c'_n = C\prod_{j=1}^{n}\frac{(\alpha+j-1)}{j} = C\prod_{j=1}^{n}\left(1-\frac{\gamma}{j}\right)$$

where $\gamma \equiv 1-\alpha$. Let $a_n = \ln(n^\gamma c'_n/C)$ and observe that

$$\begin{aligned}
a_n &= \gamma\ln n + \sum_{k=1}^{n}\ln\left(1-\frac{\gamma}{k}\right) = \ln(1-\gamma) + \gamma\ln n + \sum_{k=2}^{n}\ln\left(1-\frac{\gamma}{k}\right)\\
&= \ln(1-\gamma) + \gamma\sum_{k=2}^{n}(\ln k - \ln(k-1)) + \sum_{k=2}^{n}\ln\left(1-\frac{\gamma}{k}\right)\\
&= \ln(1-\gamma) - \gamma\sum_{k=2}^{n}\ln\frac{k-1}{k} + \sum_{k=2}^{n}\ln\left(1-\frac{\gamma}{k}\right)\\
&= \ln(1-\gamma) - \sum_{k=2}^{n}\gamma\ln\left(1-\frac{1}{k}\right) + \sum_{k=2}^{n}\ln\left(1-\frac{\gamma}{k}\right)\\
&= \ln(1-\gamma) + \sum_{k=2}^{n}\left(\ln\left(1-\frac{\gamma}{k}\right) - \gamma\ln\left(1-\frac{1}{k}\right)\right) \qquad (12)
\end{aligned}$$

Note that since $\ln(1-x) = -x - \frac{1}{2}x^2 + O(x^3)$ as $x \to 0$, the summand in (12) is such that

$$\begin{aligned}
\ln\left(1-\frac{\gamma}{k}\right) - \gamma\ln\left(1-\frac{1}{k}\right) &= -\frac{\gamma}{k} - \frac{1}{2}\left(\frac{\gamma}{k}\right)^2 - \gamma\left(-\frac{1}{k} - \frac{1}{2}\frac{1}{k^2}\right) + O\left(k^{-3}\right)\\
= -\frac{\gamma}{k} - \frac{\gamma^2}{2}k^{-2} + \frac{\gamma}{k} + \frac{\gamma}{2}k^{-2} + O\left(k^{-3}\right) &= \frac{\gamma(1-\gamma)}{2}k^{-2} + O\left(k^{-3}\right) \qquad (13)
\end{aligned}$$

Since k^{-2} is a summable sequence, it follows that the series (12) converges, i.e. $a_\infty \equiv \lim_{n\to\infty} a_n$ is well-defined and finite. We can also conclude that

$$\begin{aligned}
a_n - a_\infty &= \sum_{k=n+1}^{\infty}\left(\ln\left(1-\frac{\gamma}{k}\right) - \gamma\ln\left(1-\frac{1}{k}\right)\right) = \sum_{k=n+1}^{\infty}\left(\frac{\gamma(1-\gamma)}{2}k^{-2} + O\left(k^{-3}\right)\right)\\
&\leq \int_n^\infty\left(\frac{\gamma(1-\gamma)}{2}k^{-2} + O\left(k^{-3}\right)\right)dk = O\left(n^{-1}\right).
\end{aligned}$$

Now, set the constant $C = \exp(-a_\infty)$ and consider $c_n = n^{-\gamma}$. We have

$$\begin{aligned}
c'_n - c_n &= c'_n - n^{-\gamma} = n^{-\gamma}\left(Cc'_n n^\gamma/C - 1\right)\\
&= n^{-\gamma}\left(C\exp\ln\left(c'_n n^\gamma/C\right) - 1\right) = n^{-\gamma}\left(C\exp(a_n) - 1\right)\\
&= n^{-\gamma}\left(\exp(a_n - a_\infty) - 1\right) = n^{-\gamma}\left(\exp\left(O\left(n^{-1}\right)\right) - 1\right) \qquad (14)\\
&= n^{-\gamma}\left(1 + O\left(n^{-1}\right) - 1\right) = O\left(n^{-\gamma-1}\right).
\end{aligned}$$

Since $\sum_{n=1}^{\infty} n^{-\gamma-1} < \infty$, we have $\sum_{n=1}^{\infty}|c'_n - c_n| < \infty$ and the result follows.[11]

[11] To cover the $\alpha > 1$ case (i.e. $\bar{\alpha} < 0$), one would need to consider the expansions (13) and (14) to higher order to obtain an expression for $c'_n - c_n$ of the form $K_1 n^{-\bar{\alpha}-1} + K_2 n^{-\bar{\alpha}-2} + \cdots + K_k n^{-\bar{\alpha}-k}$ with k sufficiently large so that $\bar{\alpha} + k > 1$. The corresponding spectrum $\tilde{z}_\infty(\lambda)$ would then admit the expansion $C_0\lambda^{-\alpha} + C_1\lambda^{-(\alpha-1)} + \cdots + C_k\lambda^{-(\alpha-k)}+$ finite terms and Equation (5) would still hold for $\alpha > 1$.

For $\alpha = 0$, consider $-\ln(1-x) = \sum_{n=1}^{\infty} c'_n x^n$ with $c'_n = \frac{1}{n}$ for $n \geq 1$ and $c'_0 = 0$. Note that, for $\lambda \in \mathcal{N} \backslash \{0\}$,

$$\begin{aligned}
\tilde{z}^{\infty\prime}(\lambda) &\equiv \sum_{n=0}^{\infty} c'_n (\tilde{r}(\lambda))^n = -\ln(1-\tilde{r}(\lambda)) = -\ln(1-1+Ai\lambda+o(\lambda)) \\
&= -\ln(Ai\lambda+o(\lambda)) = -\ln(Ai\lambda(1+o(1))) \\
&= -\ln(Ai\lambda)+\ln(1+o(1)) = -\ln(Ai\lambda)+o(1) \\
&= -\ln(i\lambda)-\ln(A)+o(1) = -\ln(i\lambda)+O(1) = -\ln(i\lambda)+o(|\ln|\lambda||)
\end{aligned}$$

The same conclusion holds for $\tilde{z}^{\infty}(\lambda)$ since c'_n and c_n differ only for $n = 0$, implying that $\sum_{n=0}^{\infty} |c_n - c'_n| < \infty$ and enabling the use of Lemma 5.

We now consider the final case where either $\gamma > 1$ or $\sum_{n=0}^{\infty} |c_n| < \infty$ (i.e. $C = 0$). In this case, conclusion (i) of Lemma 5 with $c'_n = 0$ delivers the desired result: $\tilde{z}^{\infty}(\lambda) = B + o(1)$. ■

Lemma 6 *Let $r_m \to 0$ be a real, positive and decreasing sequence and $\theta \in [-\pi, \pi]$, then for any $n \in \mathbb{N}$,*

$$\left|\sum_{m=1}^{n} r_m e^{i\theta m}\right| \leq \frac{\pi}{2} \sum_{m=1}^{\bar{m}} r_m,$$

where $\bar{m} \equiv \lceil 2\pi/|\theta| \rceil$ (where $\lceil\cdot\rceil$ denotes the "round up" operation).

Proof. Let $s(t) = \frac{i\theta}{1-e^{-i\theta}} \int_0^t r_{\lceil\tau\rceil} e^{i\theta\tau} d\tau$ for $t \in \mathbb{R}^+$ and note that $s(n)$ for $n \in \mathbb{N}^*$ matches the partial sum $\sum_{m=1}^{n} r_m e^{i\theta m}$:

$$\begin{aligned}
s(n) &= \frac{i\theta}{1-e^{-i\theta}} \int_0^n r_{\lceil\tau\rceil} e^{i\theta\tau} d\tau = \frac{i\theta}{1-e^{-i\theta}} \sum_{m=1}^{n} \int_{m-1}^{m} r_{\lceil\tau\rceil} e^{i\theta\tau} d\tau \\
&= \frac{i\theta}{1-e^{-i\theta}} \sum_{m=1}^{n} r_m \int_{m-1}^{m} e^{i\theta\tau} d\tau = \frac{i\theta}{1-e^{-i\theta}} \sum_{m=1}^{n} r_m \frac{e^{i\theta m} - e^{i\theta(m-1)}}{i\theta} = \sum_{m=1}^{n} r_m e^{i\theta m}.
\end{aligned}$$

Now, observe that $s(t)$ traces out a spiral in the complex plane as t increases and let $\mathcal{D}$ be the closed and finite region bounded by the curve $s(t)$ for $t \in [0, \bar{m}]$ and the segment joining $s(\bar{m})$ with the origin. That is, $\mathcal{D}$ contains the first complete "turn" of the spiral (which corresponds to terms 1 to $\bar{m}$ of the series). Since r_m is decreasing, the region $\mathcal{D}$ will also enclose all subsequent "turns" of the spiral and we can write

$$\begin{aligned}
\left|\sum_{m=1}^{n} r_m e^{i\theta m}\right| &\leq \max_{z\in\mathcal{D}} \|z\| = \sup_{t\in[0,\bar{m}]} |s(t)| \leq \sup_{t\in[0,\bar{m}]} \left|\frac{i\theta}{1-e^{-i\theta}}\right| \int_0^t \left|r_{\lceil\tau\rceil}\right| \left|e^{i\theta\tau}\right| d\tau \\
&= \left(\sup_{\theta\in[-\pi,\pi]} \left|\frac{i\theta}{1-e^{-i\theta}}\right|\right) \sup_{t\in[0,\bar{m}]} \int_0^t r_{\lceil t\rceil} dt \leq \frac{\pi}{2} \int_0^{\bar{m}} r_{\lceil t\rceil} dt = \frac{\pi}{2} \sum_{m=1}^{\bar{m}} r_m.
\end{aligned}$$

for any $n \in \mathbb{N}$. ■

Proof of Theorem 2. First note that $|\tilde{r}(\lambda)| < 1$ for $\lambda \in]0,\pi]$ implies that $\tilde{z}^n(\lambda) \to \tilde{z}^\infty(\lambda)$ pointwise for any $\lambda \in]0,\pi]$, since

$$|\tilde{z}^n(\lambda) - \tilde{z}^\infty(\lambda)| = \left|\tilde{y}(\lambda)\sum_{m=n+1}^{\infty} c_m \tilde{r}(\lambda)^m\right| \leq |\tilde{y}(\lambda)| \sum_{m=n+1}^{\infty} |c_m| |\tilde{r}(\lambda)|^m$$
$$\leq |\tilde{y}(\lambda)| \left(\sup_m |c_m|\right) \sum_{m=n+1}^{\infty} |\tilde{r}(\lambda)|^m = \left(\sup_m |c_m|\right) |\tilde{y}(\lambda)| (1 - |\tilde{r}(\lambda)|)^{-1} |\tilde{r}(\lambda)|^{n+1},$$

where $|\tilde{r}(\lambda)|^{n+1} \to 0$ as $|\tilde{r}(\lambda)| < 1$ for $\lambda \in]0,\pi]$ and where all the prefactors are finite by assumption.[12]

The proof then proceeds by first showing that $\int_0^\pi |\tilde{z}^\infty(\lambda)|^2 d\lambda < \infty$, thus implying that $\sum_{t=0}^\infty |z^\infty(t)|^2 < \infty$, which in turn implies, that there exists some stationary process $Z_\infty(t)$ with moving average representation $z^\infty(t)$ and with spectrum $\tilde{z}^\infty(\lambda)$. Then, we show that there exists some $\bar{z}(\lambda)$ also satisfying $\int_0^\pi (\bar{z}(\lambda))^2 d\lambda < \infty$ such that

$$|\tilde{z}^n(\lambda) - \tilde{z}^\infty(\lambda)|^2 \leq (\bar{z}(\lambda))^2$$

for all n, so that, by Lebesgue dominated convergence theorem, $\lim_{n\to\infty} \int_0^\pi |\tilde{z}^n(\lambda) - \tilde{z}^\infty(\lambda)|^2 d\lambda = \int_0^\pi \lim_{n\to\infty} |\tilde{z}^n(\lambda) - \tilde{z}^\infty(\lambda)|^2 d\lambda = 0$. This implies that $\sum_{t=0}^\infty |z^n(t) - z^\infty(t)|^2 \to 0$, from which the mean square convergence of $Z_n(t)$ to $Z_\infty(t)$ follows by standard arguments (e.g., Doob (1953), Chap. XI, Section 9).

The $\sum_{n=0}^\infty |c_n| \equiv C_1 < \infty$ case (including the $\alpha < 0$ case) is simple:

$$\begin{aligned}
|\tilde{z}^\infty(\lambda)| &\leq |\tilde{y}(\lambda)| \sum_{n=0}^{\infty} |c_n| |\tilde{r}(\lambda)|^n \leq |\tilde{y}(\lambda)| \sum_{n=0}^{\infty} |c_n| 1^n = |\tilde{y}(\lambda)| C_1 \equiv \bar{z}(\lambda) \\
|\tilde{z}^n(\lambda) - \tilde{z}^\infty(\lambda)| &= \left|\tilde{y}(\lambda)\sum_{m=n+1}^{\infty} c_m \tilde{r}(\lambda)^m\right| \leq |\tilde{y}(\lambda)| \sum_{m=n+1}^{\infty} |c_m| |\tilde{r}(\lambda)|^m \\
&\leq |\tilde{y}(\lambda)| \sum_{m=0}^{\infty} |c_m| |\tilde{r}(\lambda)|^m \leq |\tilde{y}(\lambda)| C_1 \equiv \bar{z}(\lambda)
\end{aligned}$$

where $\int_0^\pi |\tilde{y}(\lambda)|^2 d\lambda < \infty$.

For the $\alpha \in]0,1/2[$ case, we consider some small cutoff $\bar{\lambda} > 0$ and compute a separate bound for large ($|\lambda| \geq \bar{\lambda}$) and small ($|\lambda| \leq \bar{\lambda}$) frequencies.

To find a bound on $|\tilde{r}(\lambda)|$ for $|\lambda| \geq \bar{\lambda}$, we note that, by Assumption 3, and Lemma 3, $|\tilde{r}(\lambda)|^2 = 1 - C_2\lambda^2 + o(\lambda^2)$ for some $C_2 > 0$ as $\lambda \to 0$ and thus

$$|\tilde{r}(\lambda)| \leq 1 - C_3\lambda^2 \tag{15}$$

for some $C_3 \in]0, C_2/2[$ for all $|\lambda| \leq \bar{\lambda}$ sufficiently small. We can then show that for $\bar{\lambda}$ sufficiently small, the maximum of $|\tilde{r}(\lambda)|$ over the set $[\bar{\lambda}, \pi]$ is reached at $\lambda = \bar{\lambda}$. The maximum of $|\tilde{r}(\lambda)|$ in any set of the form $[\bar{\lambda}, \pi]$ for $\bar{\lambda} > 0$ is reached at some λ^*, by compactness of the set and continuity of $\tilde{r}(\lambda)$ (by Assumption 3(i)) and by Assumption (iv), $\tilde{r}(\lambda^*) < 1$. Such a $\tilde{r}(\lambda^*)$ would eventually be exceed by $|\tilde{r}(\bar{\lambda})|$ for

[12] Note that if there existed sequence m_n such that $|c_{m_n}| \to \infty$, then we would have $\sum_{n=0}^\infty |c_n| \geq \sum_{n=0}^\infty |c_{m_n}| \to \infty$. Having $c_n = n^{-(1-\alpha)}$ with $a < 1/2$ also rules out $|c_{m_n}| \to \infty$.

$\bar{\lambda}$ sufficiently small since $\tilde{r}\left(\bar{\lambda}\right) \to 1$ as $\bar{\lambda} \to 0$. This contradiction is avoided only if $\lambda^* = \bar{\lambda}$ for all $\bar{\lambda}$ sufficiently small. Hence $|\tilde{r}(\lambda)| \le 1 - C_3\bar{\lambda}^2$ for $|\lambda| \ge \bar{\lambda}$ for sufficiently small $\bar{\lambda}$.

Letting $\gamma = 1 - \alpha$, we can then write, for $|\lambda| \ge \bar{\lambda}$,

$$\begin{aligned}
|\tilde{z}^{\infty}(\lambda)| &= |\tilde{y}(\lambda)| \left| 1 + \sum_{m=1}^{\infty} m^{-\gamma} (\tilde{r}(\lambda))^m \right| \le |\tilde{y}(\lambda)| \left(1 + \sum_{m=1}^{\infty} m^{-\gamma} |\tilde{r}(\lambda)|^m \right) \\
&\le |\tilde{y}(\lambda)| \left(\sum_{m=0}^{\infty} \left(1 - C_3\bar{\lambda}^2\right)^m \right) = \frac{|\tilde{y}(\lambda)|}{1 - \left(1 - C_3\bar{\lambda}^2\right)} \\
&= \frac{|\tilde{y}(\lambda)|}{C_3\bar{\lambda}^2} \le C_4 |\tilde{y}(\lambda)| \equiv \bar{z}(\lambda)
\end{aligned}$$

and

$$\begin{aligned}
|z^n(\lambda) - z^{\infty}(\lambda)| &= |\tilde{y}(\lambda)| \left| \sum_{m=n+1}^{\infty} m^{-\gamma} (\tilde{r}(\lambda))^m \right| \le |\tilde{y}(\lambda)| \sum_{m=n+1}^{\infty} m^{-\gamma} |\tilde{r}(\lambda)|^m \\
&\le |\tilde{y}(\lambda)| \sum_{m=1}^{\infty} |\tilde{r}(\lambda)|^m \le |\tilde{y}(\lambda)| \sum_{m=0}^{\infty} \left(1 - C_3\bar{\lambda}^2\right)^m \\
&= |\tilde{y}(\lambda)| \frac{1}{1 - \left(1 - C_3\bar{\lambda}^2\right)} = \frac{|\tilde{y}(\lambda)|}{C_3\bar{\lambda}^2} \le C_4 |\tilde{y}(\lambda)| \equiv \bar{z}(\lambda)
\end{aligned}$$

for some $C_3, C_4 > 0$ and where $\tilde{y}$ is such that $\int_{|\lambda| \ge \bar{\lambda}} |\tilde{y}(\lambda)|^2 d\lambda \le \int |\tilde{y}(\lambda)|^2 d\lambda < \infty$ since $\tilde{y} \in \mathcal{L}_2(\mathbb{R})$ because $y \in \mathcal{L}_2(\mathbb{R}^+)$.

For $|\lambda| \le \bar{\lambda}$, since $\tilde{z}^{\infty}(\lambda) = C|\lambda|^{-\alpha} + o\left(|\lambda|^{-\alpha}\right)$, we have

$$|\tilde{z}^{\infty}(\lambda)| \le C_4 |\lambda|^{-\alpha}$$

which satisfies $\int_{|\lambda| \le \bar{\lambda}} |\lambda|^{-2\alpha} d\lambda < \infty$ for $\alpha \in [0, 1/2[$. Also, since $\tilde{r}(\lambda) = 1 + Ai\lambda + o(\lambda)$ (from Lemma 3), we have, by Lemma 6,

$$\begin{aligned}
&|\tilde{z}^n(\lambda) - \tilde{z}^{\infty}(\lambda)| \\
&= |\tilde{y}(\lambda)| \left| \sum_{m=n+1}^{\infty} m^{-\gamma} (\tilde{r}(\lambda))^m \right| = |\tilde{y}(\lambda)| |\tilde{r}(\lambda)|^n \left| \sum_{m=1}^{\infty} (m+n)^{-\gamma} (\tilde{r}(\lambda))^m \right| \\
&\le |\tilde{y}(\lambda)| |\tilde{r}(\lambda)|^n \sum_{m=1}^{\lceil C_5/|\lambda| \rceil} (m+n)^{-\gamma} |\tilde{r}(\lambda)|^m \le |\tilde{y}(\lambda)| |\tilde{r}(\lambda)|^n \sum_{m=1}^{\lceil C_5/|\lambda| \rceil} (m+n)^{-\gamma} |\tilde{r}(\lambda)|^m \\
&\le \frac{\pi}{2} \sum_{m=1}^{\lceil C_5/|\lambda| \rceil} m^{-\gamma} \le 2 \left(1 + \int_1^{C_5/|\lambda|} m^{-\gamma} d\lambda \right) = \frac{\pi}{2} \left(1 + \left[m^{1-\gamma} \right]_1^{C_5/|\lambda|} \right) \\
&= \frac{\pi}{2} \left(1 + \left(\frac{C_5}{|\lambda|} \right)^{1-\gamma} - 1 \right) = \frac{\pi}{2} \left(\frac{C_5}{|\lambda|} \right)^{1-\gamma} = C_6 |\lambda|^{-\alpha}
\end{aligned}$$

for some finite $C_5, C_6 > 0$ and where $\lceil x \rceil$ denotes the smallest integer no smaller than x. Hence, we can set $\bar{z}(\lambda) = C_6 |\lambda|^{-\alpha}$ for $|\lambda| \le \bar{\lambda}$, which is square integrable over $|\lambda| \le \bar{\lambda}$ for $\alpha \in [0, 1/2[$. ■

Proof of Theorem 3. We first observe that the coefficient c_n^* for the finite network satisfy $c_n^* = c_n^\infty$ for $n = 0, \ldots, n^*$ since paths shorter than $n^* + 1$ must be the same in the finite and in the infinite networks by construction For $\lambda \in [\lambda_{\min}, \pi]$, we then have

$$\begin{aligned}
\left|\tilde{z}^\infty(\lambda) - \tilde{z}^*(\lambda)\right| &= \left|\sum_{n=0}^{\infty} c_n^\infty (\tilde{r}(\lambda))^n - \sum_{n=0}^{\infty} c_n^* (\tilde{r}(\lambda))^n\right| \\
&= \left|\sum_{n=0}^{\infty} c_n^\infty (\tilde{r}(\lambda))^n - \sum_{n=0}^{n^*} c_n^* (\tilde{r}(\lambda))^n - \sum_{n=n^*+1}^{\infty} c_n^* (\tilde{r}(\lambda))^n\right| \\
&\leq \left|\sum_{n=0}^{\infty} c_n^\infty (\tilde{r}(\lambda))^n - \sum_{n=0}^{n^*} c_n^* (\tilde{r}(\lambda))^n\right| + \left|\sum_{n=n^*+1}^{\infty} c_n^* (\tilde{r}(\lambda))^n\right| \\
&= \left|\sum_{n=n^*+1}^{\infty} c_n^\infty (\tilde{r}(\lambda))^n\right| + \left|\sum_{n=n^*+1}^{\infty} c_n^* (\tilde{r}(\lambda))^n\right| \\
&\leq \sum_{n=n^*+1}^{\infty} |c_n^\infty| \, |\tilde{r}(\lambda)|^n + \sum_{n=n^*+1}^{\infty} |c_n^*| \, |\tilde{r}(\lambda)|^n \\
&\leq \sum_{n=n^*+1}^{\infty} \bar{C}\bar{r}^n + \sum_{n=n^*+1}^{\infty} \bar{C}\bar{r}^n = 2\bar{C}r^{n^*+1} \sum_{n=0}^{\infty} \bar{r}^n = \frac{2\bar{C}r^{n^*+1}}{1-\bar{r}}
\end{aligned}$$

where the infinite series converges, since $\tilde{r} < 1$ by assumption. Also note that $\bar{C} < \infty$ under Assumption 1. ■

Proof of Theorem 4. The fact that the network is a translation-invariant periodic network with nodes $i \in \mathbb{Z}^d$ and that $W_{ij} = W_{i+k,j+k}$ and $W_{ij} \geq 0$ for all $i, j, k \in \mathbb{Z}^d$ implies that the problem of determining the value of $\left(e^d\right)' W^n$ is equivalent to determining the distribution of a random variable X_n taking value in $\mathbb{Z}^d$ and generated according to $X_{m+1} = X_m + U_{m+1}$ for $m = 0, \ldots, n-1$ with increments U_{m+1} taking value in $\mathbb{Z}^d$, independent from $U_{m'}$ and $X_{m'}$ for $m' \leq m$ and identically distributed. The assumption that $W_{ij} = W_{ji}$ implies that the distribution of U_m is symmetric about the origin. The assumption that $W_{ii} > 0$ implies that $P[U_m = 0] > 0$ while the fact that $W_{ij} \neq 0$ for a finite number of j implies that U_m is supported on a finite number of points. The assumption that all nodes are reachable implies that $\mathrm{Var}[U_m]$ is nonsingular. The fact that e^d has a single nonzero element indicates that the initial condition is $X_0 = 0$ (without loss of generality, due to translation-invariance) while the fact that e^o has a single element implies that we need to calculate $P[X_n = x_0]$ for some fixed $x_0 \in \mathbb{Z}^d$.

Let F denote the distribution of U_n (the same for any n). Note that the distribution of X_n (denoted $F^{\otimes n}$, the n-fold convolution of F with itself) is supported on $\mathbb{Z}^d$, so that $P[X_n = x_0]$ can be written in the form

$$P[X_n = x_0] = \int_{\mathbb{R}^d} g(x - x_0) \, dF^{\otimes n}(x) \tag{16}$$

where $g : \mathbb{R}^d \mapsto \mathbb{R}$ is a continuous function such that $g(0) = 1$ and $g(x) = 0$ for $x \in \mathbb{Z}^d \backslash \{0\}$ (its value for $x \in \mathbb{R}^d \backslash \mathbb{Z}^d$ is not restricted, other than to satisfy continuity).

A convenient choice of $g(x)$ is

$$g(x) = \prod_{j=1}^{d} \frac{\sin(\pi x_j)}{\pi x_j}.$$

Note that $g(x)$ is continuous (even at $x = 0$), $\sin(\pi x_j) = 0$ for any integer x_j and $g(0) = 1$ (as defined via a limit). The function $g(x)$ is the inverse Fourier transform of a rectangular function on $[-\pi, \pi]^d$:

$$g(x) = (2\pi)^{-d} \int_{\xi \in [-\pi,\pi]^d} e^{-i\xi x} d\xi.$$

Using Parseval's identity, we can write (16) in terms of Fourier transforms:

$$P[X_n = x_0] = (2\pi)^{-d} \int_{\xi \in [-\pi,\pi]^d} e^{i\xi \cdot x_0} \left(\tilde{f}(\xi)\right)^n d\xi,$$

where $\tilde{f}(\xi)$ is the characteristic function of the probability measure F and, by the Convolution Theorem, $\left(\tilde{f}(\xi)\right)^n$ is the characteristic function of the probability measure $F^{\otimes n}$.

We can further decompose $P[X_n = x_0]$ as

$$P[X_n = x_0] = (2\pi)^{-d} \int_{\xi \in \mathcal{B}\left(n^{-1/2+\varepsilon}\right)} e^{i\xi \cdot x_0} \left(\tilde{f}(\xi)\right)^n d\xi + R_1 \tag{17}$$

where $\mathcal{B}(r)$ denotes an open ball of radius r centered at the origin, $\varepsilon \in]0, 1/8[$ and where R_1 is a remainder:

$$R_1 = (2\pi)^{-d} \int_{\xi \in [-\pi,\pi]^d \setminus \mathcal{B}\left(n^{-1/2+\varepsilon}\right)} e^{i\xi \cdot x_0} \left(\tilde{f}(\xi)\right)^n d\xi. \tag{18}$$

To bound R_1, we observe that, since U_n is supported on a finite subset of $\mathbb{Z}^d$, the characteristic function $\tilde{f}(\xi)$ is a sum of a finite number of terms of the form $e^{i\xi \cdot x}$, with $x \in \mathbb{Z}^d$. The assumption that $P[U_n = 0] > 0$ implies that the term $e^{i\xi \cdot 0} = 1$ is present in this sum. As a result, $\left|\tilde{f}(\xi)\right|$ can only reach the value 1 when all term $e^{i\xi \cdot x}$ have the same phase, i.e., if $\xi/(2\pi) \in \mathbb{Z}^d$. Hence, in the set $[-\pi, \pi]^d$, $\left|\tilde{f}(\xi)\right|$ can only reach 1 at $\xi = 0$. Since U_n is supported on a bounded set, any of its moments are finite and thus $\tilde{f}(\xi)$ is differentiable (any number of times) and, in particular, it admits a Taylor expansion about $\xi = 0$:

$$\tilde{f}(\xi) = 1 + \frac{1}{2}\xi' \tilde{f}^{(2)}(0)\, \xi + O\left(\|\xi\|^4\right) \tag{19}$$

where we exploit the facts that $\tilde{f}(0) = 1$ and that the distribution of U_n is symmetric about 0, so all odd terms vanish. Also the second derivative $\tilde{f}^{(2)}(0)$ is a negative-definite $d \times d$ matrix by the moment theorem, since $\mathrm{Var}[U_n]$ is positive-definite by assumption. The expansion (19) implies that there exists $\eta_1 > 0$ such that $\left|\tilde{f}(\xi)\right| \leq 1 - C_1 \|\xi\|^2$ for any $\xi \in \mathcal{B}(\eta_1)$ for some $C_1 > 0$. Let $\xi_0 \equiv \arg\max_{\xi \in [-\pi,\pi]^d \setminus \mathcal{B}(\eta_1)} \left|\tilde{f}(\xi)\right|$, which exists since $\tilde{f}(\xi)$ is continuous and $[-\pi, \pi]^d \setminus \mathcal{B}(\eta)$ is compact. Since $\left|\tilde{f}(\xi)\right|$ only reaches 1 at $\xi = 0$, we must have $\left|\tilde{f}(\xi_0)\right| < 1$. Let $f_1 = \left(1 + \left|\tilde{f}(\xi_0)\right|\right)/2$

and pick $\eta_2 \in \left]0, \eta_1\right]$ such that for any $\xi \in \mathcal{B}(\eta_2)$ we have $\left|\tilde{f}(\xi)\right| > f_1$. Such an η_2 always exists since $\left|\tilde{f}(\xi)\right| \leq 1 - C_1 \|\xi\|^2$ for $\xi \in \mathcal{B}(\eta_1)$. It follows that for any n such that $n^{-1/2+\varepsilon} < \eta_2$, we have $\left|\tilde{f}(\xi)\right| \leq 1 - C_1 \left(n^{-1/2+\varepsilon}\right)^2 = 1 - C_1 n^{-1+2\varepsilon}$ for any $\xi \in [-\pi, \pi]^d \backslash \mathcal{B}\left(n^{-1/2+\varepsilon}\right)$. We can now bound the $\left(\tilde{f}(\xi)\right)^n$ term in (18) as:

$$
\begin{aligned}
\sup_{\xi \in [-\pi,\pi]^d \backslash \mathcal{B}\left(n^{-1/2+\varepsilon}\right)} \left|\left(\tilde{f}(\xi)\right)^n\right| &= \sup_{\xi \in [-\pi,\pi]^d \backslash \mathcal{B}\left(n^{-1/2+\varepsilon}\right)} \left|\exp\left(n \ln \tilde{f}(\xi)\right)\right| \\
&\leq \exp\left(n \ln\left(1 - C_1 n^{-1+2\varepsilon}\right)\right) = \exp\left(n\left(-C_1 n^{-1+2\varepsilon} + O\left(n^{-2+4\varepsilon}\right)\right)\right) \\
&= \exp\left(-C_1 n^{2\varepsilon} + O\left(n^{-1+4\varepsilon}\right)\right) \leq \exp\left(-C_2 n^{2\varepsilon}\right)
\end{aligned}
$$

for some $C_2 \in \left]0, C_1\right[$ for all n sufficiently large. We then have

$$
\begin{aligned}
|R_1| &\leq (2\pi)^{-d} \int_{\xi \in [-\pi,\pi]^d \backslash \mathcal{B}\left(n^{-1/2+\varepsilon}\right)^d} \left|e^{i\xi \cdot x_0}\right| \exp\left(-C_2 n^{2\varepsilon}\right) d\xi \\
&\leq (2\pi)^{-d} \exp\left(-C_2 n^{2\varepsilon}\right) \int_{\xi \in [-\pi,\pi]^d} d\xi = \exp\left(-C_1 n^{2\varepsilon}\right),
\end{aligned}
$$

which goes to 0 faster than any negative power of n.

We now come back to $P[X_n = x_0]$ given by Equation (17), in which we now write $\left(\tilde{f}(\xi)\right)^n$ as $\exp\left(n\tilde{F}(\xi)\right)$ with $\tilde{F}(\xi) = \ln \tilde{f}(\xi)$. Note that since $\tilde{f}(\xi)$ is differentiable (any number of times) and since $\tilde{f}(\xi)$ is nonvanishing in a neighborhood of $\xi = 0$ (because we established above that $\tilde{f}(\xi) = 1 + O\left(\xi^2\right)$), $\tilde{F}(\xi)$ admits a Taylor expansion about $\xi = 0$:

$$
\begin{aligned}
\tilde{F}(\xi) &= \tilde{F}(0) + \tilde{F}^{(1)}(0)\,\xi + \frac{1}{2}\tilde{F}^{(2)}(0)\,\xi^2 + \frac{1}{6}\tilde{F}^{(3)}(0)\,\xi^3 + \frac{1}{24}\tilde{F}^{(4)}\left(\bar{\xi}\right)\xi^4 \\
&= \tilde{F}^{(2)}(0)\,\xi^2 + \tilde{F}^{(4)}\left(\bar{\xi}\right)\xi^4
\end{aligned}
$$

where $\bar{\xi} \in [0, \xi]$ is a mean value and, for simplicity, we let an expression such as $\tilde{F}^{(k)}\left(\bar{\xi}\right)\xi^k$ stand for $\sum_{j_1,\ldots,j_k} \tilde{F}^{(k)}_{j_1,\ldots,j_k}\left(\bar{\xi}\right)\xi_{j_1}\cdots\xi_{j_k}$. We used symmetry of the distribution of U_n to obtain the second expression. Note that $\tilde{F}^{(2)}(0)$ is negative-definite by the moment theorem and the nonsingularity of the variance of U_n. We then have

$$
P[X_n = x_0] = (2\pi)^{-d} \int_{\xi \in \mathcal{B}\left(n^{-1/2+\varepsilon}\right)} e^{i\xi \cdot x_0} \exp\left(n\tilde{F}^{(2)}(0)\,\xi^2 + n\tilde{F}^{(4)}\left(\bar{\xi}\right)\xi^4\right) d\xi + R_1
$$

Next, we make the change of variable $\xi = n^{-1/2}\tilde{\xi}$

$$
\begin{aligned}
P[X_n = x_0] = (2\pi)^{-d} \int_{\tilde{\xi} \in \mathcal{B}(n^{\varepsilon})} & e^{in^{-1/2}\tilde{\xi} \cdot x_0} \exp\left(n\tilde{F}^{(2)}(0)\, n^{-1}\tilde{\xi}^2 + n\tilde{F}^{(4)}\left(n^{-1/2}\bar{\xi}\right) n^{-2}\tilde{\xi}^4\right) \times \\
& \times n^{-d/2} d\tilde{\xi} + R_1 = (2\pi)^{-d} n^{-d/2} R_0 + R_1
\end{aligned}
$$

where

$$
R_0 = \int_{\tilde{\xi} \in \mathcal{B}(n^{\varepsilon})} e^{in^{-1/2}\tilde{\xi} \cdot x_0} \exp\left(\tilde{F}^{(2)}(0)\,\tilde{\xi}^2 + n\tilde{F}^{(4)}\left(n^{-1/2}\bar{\xi}\right) n^{-2}\tilde{\xi}^4\right) d\tilde{\xi}
$$

in which the mean value $\bar{\xi}$ lies in $\left[0, \tilde{\xi}\right]$. We then have

$$
\begin{aligned}
R_0 &= \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} e^{in^{-1/2}\tilde{\xi}\cdot x_0} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} + R_2 \\
&= \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \left(1 + in^{-1/2}\tilde{\xi}\cdot x_0 - \frac{n^{-1}}{2} e^{i\tilde{\xi}\cdot x_0}\left(\tilde{\xi}\cdot x_0\right)^2\right) \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} + R_2 \\
&= \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} + R_2 + R_3 + R_4 \\
&= C_0 + R_2 + R_3 + R_4 + R_5
\end{aligned}
$$

where we have introduced the remainder terms:

$$
\begin{aligned}
R_2 &= \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} e^{in^{-1/2}\tilde{\xi}\cdot x_0} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right)\left(\exp\left(\tilde{F}^{(4)}\left(n^{-1/2}\bar{\xi}\right)n^{-1}\tilde{\xi}^4\right) - 1\right) d\tilde{\xi} \\
R_3 &= -\frac{n^{-1}}{2}\int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} e^{i\tilde{\xi}\cdot x_0}\left(\tilde{\xi}\cdot x_0\right)^2 \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} \\
R_4 &= in^{-1/2}\int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \tilde{\xi}\cdot x_0 \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} \\
R_5 &= \int_{\tilde{\xi}\in\mathbb{R}^d\setminus\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi}
\end{aligned}
$$

and the constant $C_0 = \int_{\tilde{\xi}\in\mathbb{R}^d} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} > 0$. Considering each term in turn, we have

$$
\begin{aligned}
|R_2| &\leq \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \left|e^{in^{-1/2}\tilde{\xi}\cdot x_0}\right| \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right)\left(\exp\left(\tilde{F}^{(4)}\left(n^{-1/2}\bar{\xi}\right)n^{-1}\tilde{\xi}^4\right) - 1\right) d\tilde{\xi} \\
&= \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right)\left(\exp\left(\tilde{F}^{(4)}\left(n^{-1/2}\bar{\xi}\right)n^{-1}\tilde{\xi}^4\right) - 1\right) d\tilde{\xi}.
\end{aligned}
$$

Let $\bar{F}^{(4)} \equiv \sup_{\xi\in\mathcal{B}(\eta_3)}\left|\tilde{F}^{(4)}\left(\xi\right)\right|$ for some $\eta_3 > 0$. For n sufficiently large, we eventually have $n^{-1/2}\bar{\xi} \leq \eta_3$ and we can write

$$
\begin{aligned}
|R_2| &\leq \int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right)\left(\exp\left(\bar{F}^{(4)}n^{-1}n^{4\varepsilon}\right) - 1\right) d\tilde{\xi} \\
&= \left(\exp\left(\bar{F}^{(4)}n^{-1+4\varepsilon}\right) - 1\right)\int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} \\
&= \left(1 + \bar{F}^{(4)}n^{-1+4\varepsilon} + o\left(n^{-1+4\varepsilon}\right) - 1\right)\int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} \\
&= O\left(n^{-1+4\varepsilon}\right)\int_{\mathbb{R}^d} \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi}
\end{aligned}
$$

where the last integral is finite since $\tilde{F}^{(2)}\left(0\right)$ is negative-definite. Next,

$$
\begin{aligned}
|R_3| &\leq \frac{n^{-1}}{2}\int_{\tilde{\xi}\in\mathcal{B}(n^\varepsilon)} \left|e^{i\tilde{\xi}\cdot x_0}\right|\left(\tilde{\xi}\cdot x_0\right)^2 \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} \\
&\leq \frac{n^{-1}}{2}\int_{\tilde{\xi}\in\mathbb{R}^d} \left(\tilde{\xi}\cdot x_0\right)^2 \exp\left(\tilde{F}^{(2)}\left(0\right)\tilde{\xi}^2\right) d\tilde{\xi} = O\left(n^{-1}\right)
\end{aligned}
$$

where the last integral is finite since $\tilde{F}^{(2)}(0)$ is negative-definite.

Next, R_4 vanishes by the symmetry of $\exp\left(\tilde{F}^{(2)}(0)\,\tilde{\xi}^2\right)$ (in $\tilde{\xi}$). Finally,

$$\begin{aligned}|R_5| &\leq \int_{\tilde{\xi}\in\mathbb{R}^d\setminus\mathcal{B}(n^\varepsilon)} \exp\left(\lambda\left\|\tilde{\xi}\right\|^2\right) d\tilde{\xi} = S_d \int_{n^\varepsilon}^{\infty} \rho^d \exp\left(-\lambda\rho^2\right) d\rho \\ &\leq S_d \int_{n^\varepsilon}^{\infty} \exp\left(-C_2\rho\right) d\rho = \frac{S_d}{C_2}\exp\left(-C_2 n^\varepsilon\right)\end{aligned}$$

where λ is the smallest eigenvalue of $-\tilde{F}^{(2)}(0)$. In the second line, we have expressed the integral in polar coordinates with ρ being the radius and S_d is the $(d-1)$-dimensional "surface" of a hypersphere of radius 1. The second inequality holds for some $C_2 > 0$ for n sufficiently large and yields an expression that decays faster than any power of n.

Collecting the order of the remainders, we have, with $C = (2\pi)^{-d} C_0 > 0$,

$$\begin{aligned}P[X_n = x_0] &= (2\pi)^{-d} n^{-d/2}\left(C_0 + O\left(n^{-1+4\varepsilon}\right) + O\left(n^{-1}\right) + 0 + O\left(\exp\left(-C_2 n^{2\varepsilon}\right)\right)\right) + \\ &\quad + O\left(\exp\left(-C_1 n^\varepsilon\right)\right) = C n^{-d/2} + O\left(n^{-1-d/2}\right)\end{aligned}$$

■

B Miscellaneous results regarding spectral dimension

Equation (9) can be obtained from Equation 3.30 in Havlin and Ben-Avraham (1987): $\tilde{d} = 2d_f/d_w$, where d_f is the fractal dimension and d_w is the so-called *walk dimension*. For a (u,v)-flower (with $u \leq v$ by convention), we have $d_w = \ln(uv)/(\ln u)$ (see Section 4.3 in Rozenfeld, Havlin, and ben Avraham (2007)) and $d_f = \ln(u+v)/(\ln u)$ (see Equation (9) in Rozenfeld, Havlin, and ben Avraham (2007)).

Next, we state a general result regarding networks generated by combining different generating rules.

Theorem 5 *Let G_a and G_b be two mutually compatible*[13] *generating rules for two self-similar fractal networks with spectral dimension $\tilde{d}_a$ and $\tilde{d}_b$, respectively. The generating rule G_a increases the number of nodes by a factor β_a at each application (and similarly for G_b and β_b). Let a_m and b_m be two sequences of positive integers such that $a_m/(a_m+b_m) \to \rho \in \mathbb{R}^+$ and apply the following sequence of generating rules*

$$\cdots\left(G_b^{(b_m)} G_a^{(a_n)}\right)^{(m)}\left(G_b^{(b_{m+1})} G_a^{(a_{n+1})}\right)^{(m+1)}\cdots$$

where $G^{(m)}$ denotes m repetitions of rule G. Then, the resulting network has spectral dimension $\tilde{d} = \mu\tilde{d}_a + (1-\mu)\tilde{d}_b$, where $\mu = \left(1 + \frac{(1-\rho)}{\rho}\frac{\ln\beta_b}{\ln\beta_a}\right)^{-1}$. Note that the mapping from $\rho \in [0,1]$ to $\mu \in [0,1]$ is one-to-one and onto,[14] *so all values of $\tilde{d} \in \left[\tilde{d}_a, \tilde{d}_b\right]$ are reachable via suitable choices of the a_m and b_m sequences.*

---

[13] We say that two generating rules are *mutually compatible* if they can be applied sequentially in any order. Generating rules for the hierachical star networks for any m are mutually compatible and similarly for the generating rules of hierachical ring networks of any u, v. However, star and ring generating rules cannot be combined.

[14] if one includes, by convention, the limiting value $\mu = 0$ when $\rho = 0$

Proof. The self-similar network generated by G_a is characterized by a nested sequence of subsets $R_{a,n}$ of the network that are mutually identical up to a scaling factor. Let $T_{a,n}$ denote the expected residence time of a random walker in region $R_{a,n}$ and let $M_{a,n}$ denote the fraction of the network's nodes that lie in region $R_{a,n}$. By a standard renormalization group argument (see Section 2.1 in Havlin and Ben-Avraham (1987)), if a generating rule G_a for a self-similar fractal yields a spectral dimension of $\tilde{d}_a$, this indicates that these quantities satisfy $T_{a,n}/T_{a,n+1} = \alpha_a + o(1)$ and $M_{a,n}/M_{a,n+1} = \beta_a$ with

$$\tilde{d}_a = 2\frac{\ln \alpha_a}{\ln \beta_a}.$$

Similar definitions and results hold for the network generated by G_b.

We now define the generating rule G as the application of the following generating rule:

$$\underbrace{G_a, \ldots, G_a}_{a_m \text{ times}}, \underbrace{G_b, \ldots, G_b}_{b_m \text{ times}},$$

and consider the network obtained in the limit of the iterated application of G. We can then define a sequence of nested (not necessarily self-similar) subsets R_n of the network such that

$$\begin{aligned} \frac{T_n}{T_{n+1}} &= \alpha_a^{a_m} \alpha_b^{b_m} + o(1) \\ \frac{M_n}{M_{n+1}} &= \beta_a^{a_m} \beta_b^{b_m}. \end{aligned}$$

The spectral dimension associated with those sequences is then obtain via a limit (as $m \to \infty$):

$$\begin{aligned} 2\frac{\ln \alpha_a^{a_m} \alpha_b^{b_m}}{\ln \beta_a^{a_m} \beta_b^{b_m}} &= 2\frac{a_m \ln \alpha_a + b_m \ln \alpha_b}{a_m \ln \beta_a + b_m \ln \beta_b} = 2\frac{\frac{a_m}{a_m+b_m} \ln \alpha_a + \frac{b_m}{a_m+b_m} \ln \alpha_b}{\frac{a_m}{a_m+b_m} \ln \beta_a + \frac{b_m}{a_m+b_m} \ln \beta_b} \\ &\to 2\frac{\rho \ln \alpha_a + (1-\rho) \ln \alpha_b}{\rho \ln \beta_a + (1-\rho) \ln \beta_b} = \frac{2\left(\ln \alpha_a\right)/\left(\ln \beta_a\right)}{1 + \frac{(1-\rho)\ln \beta_b}{\rho \ln \beta_a}} + \frac{2\left(\ln \alpha_b\right)/\left(\ln \beta_b\right)}{1 + \frac{\rho \ln \beta_a}{(1-\rho)\ln \beta_b}} \\ &= \frac{\tilde{d}_a}{1+R} + \frac{\tilde{d}_b}{1+R^{-1}} = \frac{1}{1+R}\tilde{d}_a + \left(1 - \frac{1}{1+R}\right)\tilde{d}_b = \mu \tilde{d}_a + (1-\mu)\tilde{d}_b \end{aligned}$$

where $R = \frac{(1-\rho)\ln \beta_b}{\rho \ln \beta_a}$ and $\mu = (1+R)^{-1}$ and the result is shown. ■

This theorem actually constructs networks that do not exhibit self-similarity, since the number of times each rule is applied consecutively constantly changes across scales, and yet, they still have a well-defined spectral dimension. This shows that the classic case of a self-similar fractal network is not a necessary condition for our mechanism of long-memory generation to apply. One can even take two generating rules, each of which, in isolation, yields the same spectral dimension. Then, one could just apply one of these two rules at random or repeat each rule a random number of times. One can even randomize which rule is used in different portions of the network at each step of the generation algorithm. The resulting network still has a well-defined spectral dimension, but the randomness in the application of the generating rule make it impossible to even have statistical self-similarity.

References

Abadir, K., and G. Talmain (2002): "Aggregation, persistence and volatility in a macro model," *Review of Economic Studies*, 69, 749–779.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

Alfarano, S., and T. Lux (2005): "A Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory," Working Paper, University of Kiel.

Avarucci, M., and C. Velasco (2009): "A Wald test for the cointegration rank in nonstationary fractional systems," *Journal of Econometrics*, 151, 178–189.

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

ben Avraham, D., and S. Havlin (2005): *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, Cambridge.

Calvet, L., and A. Fisher (2002): "Multifractality in asset returns: Theory and evidence," *Review Of Economics and Statistics*, 84, 381–406.

Cassi, D., and S. Regina (1996): "Random Walks on Bundled Structures," *Physical Review Letters*, 76, 2914–2917.

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

Chen, X., L. P. Hansen, and M. Carrasco (2010): "Nonlinearity and Temporal Dependence," *Journal of Econometrics*, 155, 155–169.

Chevillon, G., and S. Mavroeidis (2011): "Learning generates Long Memory," Working Paper, University of Oxford.

Christou, A., and R. B. Stinchcombe (1986): "Anomalous diffusion on regular and random models for diffusion-limited aggregation," *J. Phys. A: Math. Gen.*, 19, 2625–2636.

Comte, F., and E. Renault (1996): "Long memory continuous time models," *Journal of Econometrics*, 73, 101–149.

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

de Paula, A. (2016): "Econometrics of Network Models," Working Paper CWP06/16, Cemmap.

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

Doob, J. L. (1953): *Stochastic Processes*. Wiley, New York.

Dupor, B. (1999): "Aggregation and irrelevance in multi-sector models," *Journal of Monetary Economics*, 43, 391–409.

Elliott, M., B. Golub, and M. O. Jackson (2014): "Financial Networks and Contagion," *American Economic Review*, 104, 3115–3153.

Flandrin, P. (1989): "On the Spectrum of Fractional Brownian Motions," *IEEE Transactions on information theory*, 35, 197–199.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119, 1–38.

Gabaix, X. (2011): "The Granular Origins Of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

Given, J. A., and B. B. Mandelbrot (1983): "Diffusion on fractal lattices and the fractal Einstein relation," *J. Phys. B: At. Mol. Phys.*, 16, L565–L569.

Gouriéroux, C., J.-C. Héam, and A. Monfort (2012): "Bilateral exposures and systemic solvency risk," *Canadian Journal of Economics*, 45, 12731309.

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

Granger, C. W. J., and Z. Ding (1996): "Varieties of long memory models," *Journal of Econometrics*, 73, 61–77.

Hall, R. E. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971–987.

Harvey, A., S. J. Koopman, and N. Shephard (2004): *State Space and Unobserved Component Models: Theory and Applications*. Cambridge University Press, UK.

Havlin, S., and D. Ben-Avraham (1987): "Diffusion in disordered media," *Advances in Physics*, 36, 695–798.

Horvath, M. (1998): "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*, 1, 781–808.

INAOKA, H., T. NINOMIYA, K. TANIGUCHI, AND T. S. H. TAKAYASU (2004): "Fractal Network derived from banking transaction — An analysis of network structures formed by financial institutions," Working Paper 04-E-04, Bank of Japan.

JACKSON, M. O. (2005): "A Survey of Models of Network Formation: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, ed. by G. Demange, and M. Wooders. Cambridge University Press, Cambridge.

LOBATO, I., AND P. M. ROBINSON (1996): "Averaged periodogram estimation of long memory," *Journal of Econometrics*, 73, 303–324.

LONG, J. B., AND C. I. PLOSSER (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

LOYNE, R. M. (1968): "On the Concept of the Spectrum for Non-stationary Processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, 30, 1–30.

MANDELBROT, B. B. (1982): *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York.

MANDELBROT, B. B., AND J. W. V. NESS (1968): "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, 10, 422–437.

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, 83–89.

NELSON, C., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

ÖZGÜR, O., AND A. BISIN (2013): "Dynamic Linear Economies with Social Interactions," Working Paper, Melbourne Business School.

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

PERRON, P. (1988): "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach," *Journal of Economic Dynamics and Control*, 12, 297–332.

——— (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

PERRON, P., AND Z. QU (2007): "An analytical evaluation of the log-periodogram estimate in the presence of level shifts," Working Paper, Boston University.

Phillips, P., and T. Magdalinos (2007): "Limit theory for moderate deviations from a unit root," *Journal of Econometrics*, 136, 115–130.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Rozenfeld, H. D., S. Havlin, and D. ben Avraham (2007): "Fractal and transfractal recursive scale-free nets," *New Journal of Physics*, 9, 1–16.

Schennach, S. M. (2013): "Long memory via networking," Working Paper CWP13/13, Cemmap, Second revise and resubmit at *Econometrica*.

——— (2018): "Supplement to 'Long memory via networking'," *Econometrica Supplemental Material.*

Sharkey, W. W. (1995): "Network Models in Economics," in *Handbooks in Operations Research and Management Science*, ed. by M. Ball, T. Magnanti, C. Monma, and G. Nemhauser, vol. 8, chap. 9, pp. 713–765. Elsevier.

Song, C., S. Havlin, and H. A. Makse (2005): "Self-similarity of complex networks," *Nature*, 433, 392–395.

Steele, J. M., L. A. Shepp, and W. F. Eddy (1987): "On the Number of Leaves of a Euclidean Minimal Spanning Tree," *Journal of Applied Probability*, 24, 809–826.

Tejedor, V. (2012): "Random walks and first-passage properties," Ph.D. thesis, Université Pierre et Marie Curie and of the Technische Universitä at München.

Witten, T. A., and L. M. Sander (1981): "Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon," *Physical Review Letters*, 47, 1400.

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

Supplement to "Long memory via networking"

Susanne M Schennach

Abstract

This Supplement Material includes various extension of the paper's main results, namely (i) deviations from power laws in the c_n coefficients (ii) the presence of multiple sources of noise in the network (iii) the possibility of non integrable limiting power spectra and (iv) heterogeneity in the agents' responses. It also includes the description of a simple and stylized variant of the Loss-Plosser model as well as a "toy" application based on the "input-output accounts" database compiled by the Bureau of Economic Analysis.

C Some Extensions

C.1 Deviations from power laws

The assumed power-law behavior for c_n in Theorem 1 may seem specific, but other natural possibilities either yield uninteresting or implausible results. One obvious generalization is $c_n = e^{\beta n} n^{-(1-\alpha)}$ for $\alpha, \beta \in \mathbb{R}$. However, the $\beta < 0$ case falls under case (ii) of Theorem 1 and yields a short memory process. The case $\beta > 0$ yields a spectrum that diverges at all λ such that $|\tilde{r}(\lambda)| > e^{-\beta}$ and not just at $\lambda = 0$. In that case, even a perturbation of a finite duration would be magnified by the network to such an extent that the overall economy would leave the local equilibrium considered in a finite time and visit another equilibrium. The process would then presumably repeat itself until a stable equilibrium (with non-explosive c_n) is found. In a sense, the economy should plausibly self-organize to rule out cases where $\tilde{z}^{\infty}(\lambda)$ diverges for $\lambda \neq 0$. In this sense, $\beta = 0$ is the only nontrivial and plausible case.[15]

While the results of Theorem 1 are already robust to deviations from exact power laws that are absolutely summable, we can also handle deviations of the c_n coefficients from a power law that are bigger than absolutely summable. For instance, consider the case where the c_n (for $n \geq 1$) admit an expansion of the form

$$c_n = \sum_{i=1}^{\bar{\imath}} A_i n^{-(1-\alpha_i)} + c'_n \tag{20}$$

where $\alpha_1 > \alpha_2 > \cdots > \alpha_{\bar{\imath}}$ and $\sum_{n=1}^{\infty} |c'_n| < \infty$. One can apply Theorem 1 to each individual term to yield the conclusion that the resulting power spectrum $|\tilde{z}^{\infty}(\lambda)|^2$ would then have the behavior

$$|\tilde{z}^{\infty}(\lambda)|^2 = \sum_{i=1}^{\bar{\imath}} O\left(|\lambda|^{-2\alpha_i}\right) = O\left(|\lambda|^{-2\alpha_1}\right) \text{ as } |\lambda| \to 0,$$

[15] It is straightforward to extend Theorem 1 to allow for $\alpha > 1$, thus covering cointegrated processes (e.g., Avarucci and Velasco (2009)) or "mildly explosive" processes (e.g., Phillips and Magdalinos (2007)). (The necessary adjustments are outlined in footnote 11 in the Appendix, to avoid cluttering the main proof with lengthy manipulations.)

since $\alpha_1 > \alpha_i$ for $i = 2, \ldots, \bar{\imath}$. Taking $\bar{\imath}$ finite is without much loss of generality, since eventually, for some α_i, the power law would become absolutely integrable (if consecutive exponents α_i are at least some finite distance from each other). Expansions of the form (20) can be obtained, for instance, if the c_n coefficients can be written as $c_n = g(n^{-1})$ where $g(\cdot)$ is a function such that $(g(u))^a$ admits a Taylor expansion around $u = 0$ for some real a, so this extension brings considerable generality.

C.2 Multiple sources of noise

In this section, we consider the effect of multiple sources of noise with an arbitrary covariance structure introduced at multiple points of the network. We maintain the Gaussian assumption. It turns out that the general covariance case can always be reduced to the uncorrelated noise case (across the spatial dimension) by a suitable redefinition of the network. Specifically, consider again our general vector autoregressive setup $X_t = \sum_{s=0}^{\infty} W_s X_{t-s} + V^{1/2} u_t$, but where the noise now has the general form $V^{1/2} u_t$ for some general correlation matrix V and with u_t being a $N(0, I)$ noise vector. This model can equivalently be written via an augmented state vector $(X_t', X_t^{*\prime})'$ as

$$\begin{bmatrix} X_t \\ X_t^* \end{bmatrix} = \sum_{s=0}^{\infty} \begin{bmatrix} W_s & V^{1/2}\mathbf{1}\{s=0\} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{t-s} \\ X_{t-s}^* \end{bmatrix} + \begin{bmatrix} 0 \\ u_t \end{bmatrix},$$

which has the same basic form as Equation (1) with a noise that is spatially uncorrelated. This construction amounts to building a network with twice the number of nodes containing the original network (as modeled via W_s) and an additional network (modeled via V) whose role is solely to propagate each component of the uncorrelated noise vector u_t to multiple nodes of the original network.

For uncorrelated noise sources, we can easily compute the c_n coefficients via Equation (4) associated with one source node i at the time (setting all but one element of e^o to zero) while considering a given fixed set of destination nodes (via e^d). Let $|\tilde{z}_i^{\infty}(\lambda)|^2$ denote the power spectrum obtained when only source node i is active. Since the noise sources are independent, the overall power spectrum is simply the sum of the individual power spectra $\sum_{i=1}^{n} |\tilde{z}_i^{\infty}(\lambda)|^2$.

C.3 Non integrable power spectra

One can also establish a convergence result similar to Theorem 2 that covers both integrable ($\alpha < 1/2$) and non integrable ($\alpha \geq 1/2$) limiting power spectra $|\tilde{z}^{\infty}(\lambda)|^2$ by focusing on increments of the processes. Working with increments is a standard technique (see Mandelbrot and Ness (1968) and Comte and Renault (1996), for instance) that offers the advantage of providing finite-variance quantities even in the presence of nonstationarity in the process.

Theorem 6 *Let the Assumptions of Theorem 1 hold. Assume that* $|\tilde{r}(\lambda)| < 1$ *for* $\lambda \in]0, \pi]$, *that* $|\tilde{y}(\lambda)|$ *is uniformly bounded for* $\lambda \in [0, \pi]$, *and consider the differenced process*

$$\Delta Z_n(t) \equiv Z_n(t) - Z_n(t - \Delta t)$$

for a given $\Delta t \in \mathbb{Z}$ *and any* $n \in \mathbb{N}$ *(with corresponding moving average representation* $\Delta z^n(t) \equiv z^n(t) - z^n(t - \Delta t)$ *and spectrum* $\Delta \tilde{z}^n(\lambda) \equiv \left(1 - e^{i\lambda\Delta t}\right) \tilde{z}^n(\lambda)$*). Let* $\tilde{z}^{\infty}(\lambda) \equiv \lim_{n\to\infty} \tilde{z}^n(\lambda)$ *with a corresponding moving average representation* $z^{\infty}(t)$*.*

Then, there exists a stationary process $\Delta Z_\infty(t)$ with moving average representation[16] $\Delta z^\infty(t) \equiv z^\infty(t) - z^\infty(t-\Delta t)$ *and spectrum* $\Delta\tilde{z}^\infty(\lambda) \equiv \left(1-e^{i\lambda\Delta t}\right)\tilde{z}^\infty(\lambda)$ *satisfying* $\int_0^\pi |\Delta\tilde{z}^n(\lambda) - \Delta\tilde{z}^\infty(\lambda)|^2 d\lambda \to 0$, $\sum_{t=0}^\infty |\Delta z^n(t) - \Delta z^\infty(t)|^2 \to 0$ *and* $E\left[|\Delta Z_n(t) - \Delta Z_\infty(t)|^2\right] \to 0$ *for almost any given* $t \in \mathbb{R}$ *and* $\sum_{t=-\infty}^\infty E\left[|\Delta Z_n(t) - \Delta Z_\infty(t)|^2\right] w(t) \to 0$ *for a given absolutely integrable, bounded and continuous weighting function* $w(t)$*.*

Proof. The proof is similar to the one of Theorem 2 and we focus here on the differences. It is clear that the differenced process $\Delta Z_n(t)$ admits the moving average representation:

$$\Delta Z_n(t) = \sum_{s=-\infty}^{t} \left(z^n(t-s) - z^n(t-\Delta t - s)\right) G(s)$$

where the kernel $z^n(t-s) - z^n(t-\Delta t-s)$ is absolutely integrable/summable since it is a difference of two absolutely integrable/summable terms. Its Fourier transform is thus well-defined and equal to:

$$\Delta\tilde{z}^n(\lambda) = \sum_{t=0}^{\infty} \left(z^n(t) - z^n(t-\Delta t)\right) e^{i\lambda t} = \tilde{z}^n(\lambda) - e^{i\lambda\Delta t}\tilde{z}^n(\lambda) = \left(1-e^{i\lambda\Delta t}\right)\tilde{z}^n(\lambda)$$

The pointwise limit of $\Delta\tilde{z}^n(\lambda)$ also poses no problem (as in Theorem 2):

$$\Delta\tilde{z}^\infty(\lambda) \equiv \lim_{n\to\infty} \left(1-e^{i\lambda\Delta t}\right)\tilde{z}^n(\lambda) = \left(1-e^{i\lambda\Delta t}\right)\tilde{z}^\infty(\lambda),$$

with the additional advantage that $\Delta\tilde{z}^n(0) = 0$ and therefore $\Delta\tilde{z}^\infty(0) = 0$ (so the $\lambda = 0$ point is no longer exceptional).

Now observe that, for some sufficiently small $\bar{\lambda} > 0$,

$$\begin{aligned}
\int_0^\pi |\Delta\tilde{z}^\infty(\lambda)|^2 d\lambda &= \int_{\lambda\le\bar{\lambda}} \left|\left(1-e^{i\lambda\Delta t}\right)\tilde{z}^\infty(\lambda)\right|^2 d\lambda + \int_{\bar{\lambda}}^\pi \left|\left(1-e^{i\lambda\Delta t}\right)\tilde{z}^\infty(\lambda)\right|^2 d\lambda \\
&\le \int_{\lambda\le\bar{\lambda}} C_1 |\Delta t\lambda|^2 |\lambda|^{-2\alpha} d\lambda + \int_{\bar{\lambda}}^\pi 2|\tilde{z}^\infty(\lambda)|^2 d\lambda \\
&\le \int_{\lambda\le\bar{\lambda}} C_1 |\lambda|^{2(1-\alpha)} d\lambda + \int_{\bar{\lambda}}^\pi 2|\tilde{y}(\lambda)|^2 d\lambda < \infty
\end{aligned}$$

for some finite constant $C_1 > 0$ and where $1-\alpha \ge 0$. Hence $\Delta\tilde{z}^\infty \in \mathcal{L}_2(\mathbb{R})$ and therefore the corresponding Δz^∞ is also in $\mathcal{L}_2(\mathbb{R}^+)$ and the corresponding process $\Delta Z_\infty(t)$ is stationary.

Next, we again make use of Lebesgue's dominated convergence theorem to show that $\int_0^\pi |\Delta\tilde{z}^n(\lambda) - \Delta\tilde{z}^\infty(\lambda)|^2 d\lambda \to 0$, which requires the existence of a square integrable $\bar{z}(\lambda)$ such that $|\Delta\tilde{z}^n(\lambda) - \Delta\tilde{z}^\infty(\lambda)| \le \bar{z}(\lambda)$. For $|\lambda| \ge \bar{\lambda}$, we proceed as in Theorem 2 after noting that the prefactor $\left(1-e^{i\lambda\Delta t}\right)$ is bounded in magnitude by 2. For $|\lambda| \le \bar{\lambda}$, we proceed as in Theorem 2, after noting that the prefactor $\left(1-e^{i\lambda\Delta t}\right)$ is bounded in magnitude by $C_2|\lambda|$ for some finite $C_2 > 0$. This leads to a $\bar{z}(\lambda)$ that has the form $|\lambda|^{1-\alpha}$ (instead of $|\lambda|^{-\alpha}$), which is clearly square integrable for $|\lambda| \le \bar{\lambda}$ for any $\alpha \in [0,1]$. ■

C.4 Heterogeneity

To allow for heterogeneity in the agents' responses, we relax Assumption 1 as follows.

[16] We take the convention that $z_\infty(t) = 0$ for $t < 0$.

Assumption 4 *The autoregressive coefficient matrix in Equation (1) factors as* $W_{s,ij} = r_{s,ij}W_{ij}$ *where the* W_{ij} *are fixed constants (satisfying* $\sum_{j=1}^{N} W_{ij} = 1$ *for* $i = 1, \ldots, N$.*) while the impulse response function* $r_{s,ij}$ *of each agent is chosen at random once at* $t = -\infty$ *and kept constant thereafter.*

The assumption allows for the effect of each input j on the output of each node i of the network to be characterized by a *different* convolution operation. We view the network structure as fixed (via the deterministic W_{ij}) and allow for heterogeneity in the agents (via the random impulse response functions $r_{s,ij}$). We place no specific assumption regarding the covariance structure of between the different elements of $r_{s,ij}$, although we will need to constrain the amount of possible dependence.

This section provides conditions under which the conclusion of Theorem 1 actually holds with probability 1 for such randomly constructed networks. A key feature of the result is the existence of an average spectral representation denoted $\bar{r}(\lambda)$. In essence, there are so many very long pathways that connect the origin and the destination, that the fluctuations in the $r_{s,ij}$ across the different i, j quickly average out to a single effective value representative of the whole network. To state our result, we introduce a few convenient definitions that are heterogenous analogues of previously defined quantities.

Definition 3 *Let* $\tilde{r}_{ij}(\lambda) = \sum_{s=0}^{\infty} r_{s,ij}e^{i\lambda s}$. *Let* $\mathcal{P}_n$ *denote the set of paths connecting the origin nodes to the destination nodes in* n *steps (each element* p *of* $\mathcal{P}_n$ *is a* $(n+1)$*-dimensional vector of integer specifying which sequence of nodes are visited by the path). For any maximum path length* $\bar{n} \in \mathbb{N}$, *the spectral representation of the aggregate output of the destination nodes is given by*

$$\tilde{z}^{\bar{n}}(\lambda) = \tilde{y}(\lambda) \sum_{n=0}^{\bar{n}} \sum_{p \in \mathcal{P}_n} \prod_{\ell=1}^{n} \left(\tilde{r}_{p_\ell p_{\ell+1}}(\lambda) W_{p_\ell p_{\ell+1}}\right). \tag{21}$$

and we let $c_n = \sum_{p \in \mathcal{P}_n} \prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}$ *(which coincides with the earlier definition via Equation (4) after expanding the matrix product).*

Equation (21) merely states that the output is the sum of the effect of the input noise (modeled via $\tilde{y}(\lambda)$) through the various possible pathways p, of lengths up to $\bar{n}$, joining the origin and the destination nodes. Along each path, the noise is filtered as it goes through the network. Going from node p_ℓ to node $p_{\ell+1}$, its spectral representation is multiplied by $\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)$ (the spectral response of node $p_{\ell+1}$) and weighted by the link strength $W_{p_\ell p_{\ell+1}}$.

Theorem 7 *Let* $\tilde{y}$ *satisfy Assumption 2 and let Assumption 4 hold. Let* $\bar{r}(\lambda) \equiv \lim_{n\to\infty} \left(\sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}\right) E\left[\prod_{\ell=1}^{n} \tilde{r}_{p_\ell p_{\ell+1}}(\lambda)\right]\right)^{1/n}$. *Assume that* $\bar{r}(\lambda)$ *exists, satisfies Assumption 3 and is such that*

$$E\left[\left(\sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}\right)\left(\prod_{\ell=1}^{n} \frac{\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)}{\bar{r}(\lambda)} - 1\right)\right)^2\right] \leq D n^{-3-\varepsilon} \tag{22}$$

for some $D, \varepsilon > 0$ *for all* λ *in some neighborhood of the origin. Then, the conclusion of Theorem 1 for* $\tilde{z}^{\bar{n}}(\lambda)$ *holds with probability* 1.

To prove this result, we first need a simple Lemma.

Lemma 7 *Let c_n be a deterministic sequence and let the corresponding $\tilde{z}^{\infty}(\lambda)$ satisfy $\tilde{z}^{\infty}(\lambda) = A(i\lambda)^{-\alpha} + o\left((i\lambda)^{-\alpha}\right)$ (for $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}^{+}$). Let c_n' be a random sequence such that $E\left[(c_n' - c_n)^2\right] \leq D(1+n)^{-3-\varepsilon}$ for some $\varepsilon, D > 0$, then the corresponding $\tilde{z}^{\infty\prime}(\lambda)$ satisfies $\tilde{z}^{\infty\prime}(\lambda) = A(i\lambda)^{-\alpha} + o\left((i\lambda)^{-\alpha}\right)$ with probability one.*

Proof. To simplify the notation let the sequence start at index $n = 1$ instead of 0. By Lemma 5, it suffices to show that $\sum_{n=1}^{\infty} |c_n' - c_n|$ is finite with probability one, i.e. $P\left[\sum_{n=1}^{\infty} |c_n' - c_n| \geq C\right] \to 0$ as $C \to \infty$. Let $\Delta c_n = c_n' - c_n$ and for a given C, let $c = C\left(\sum_{n=1}^{\infty} n^{-1-\varepsilon/3}\right)^{-1}$. Note that $\sum_{n=1}^{\infty} n^{-1-\varepsilon/3} < \infty$ and that $C \to \infty \implies c \to \infty$. Then note that $|\Delta c_n| \leq c n^{-1-\varepsilon/3}$ for all $n \in \mathbb{N}^*$ implies that $\sum_{n=1}^{\infty} |\Delta c_n| \leq C$. Taking the contrapositive of that statement yields that the event $\sum_{n=1}^{\infty} |\Delta c_n| \geq C$ implies the event $|\Delta c_n| \geq c n^{-1-\varepsilon/3}$ for some $n \in \mathbb{N}^*$. Then write

$$
\begin{aligned}
P\left[\sum_{n=1}^{\infty} |\Delta c_n| \geq C\right] &\leq P\left[|\Delta c_n| \geq c n^{-1-\varepsilon/3} \text{ for some } n \in \mathbb{N}^*\right] \\
&\leq \sum_{n=1}^{\infty} P\left[|\Delta c_n| \geq c n^{-1-\varepsilon/3}\right] = \sum_{n=1}^{\infty} P\left[|\Delta c_n|^2 \geq c^2 n^{-2-2\varepsilon/3}\right] \\
&\leq \sum_{n=1}^{\infty} \frac{E\left[|\Delta c_n|^2\right]}{c^2 n^{-2-(2/3)\varepsilon}} \leq \sum_{n=1}^{\infty} \frac{D n^{-3-\varepsilon}}{c^2 n^{-2-(2/3)\varepsilon}} = \frac{D}{c^2} \sum_{n=1}^{\infty} n^{-1-\varepsilon/3}
\end{aligned}
$$

where we have used, in turn, (i) the fact that if two events are such that $A \implies B$ then $P[B] \geq P[A]$, (ii) for any sequence of events A_i, we have $P[\cup_i A_i] \leq \sum_i P[A_i]$, (iii) monotonicity of the function x^2 for $x \geq 0$ (iv) Markov's inequality $P[X \geq x] \leq E[X]/x$ applied to the random variable $X = |\Delta c_n|^2$, (v) the assumption $E\left[|\Delta c_n|^2\right] \leq D n^{-3-\varepsilon}$. Since $\sum_{n=1}^{\infty} n^{-1-\varepsilon/3} < \infty$, it follows that, as $C \to \infty$, $c \to \infty$ and $P\left[\sum_{n=1}^{\infty} |X_n| \geq C\right] \to 0$, as desired. ■

Proof of Theorem 7. From Definition 3, we have $c_n = \sum_{p \in \mathcal{P}_n} \prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}$ and thus

$$
\begin{aligned}
\tilde{z}^{\bar{n}}(\lambda) &= \tilde{y}(\lambda) \sum_{n=0}^{\bar{n}} \sum_{p \in \mathcal{P}_n} \prod_{\ell=1}^{n} \left(\tilde{r}_{p_\ell p_{\ell+1}}(\lambda) W_{p_\ell p_{\ell+1}}\right) \\
&= \tilde{y}(\lambda) \sum_{n=0}^{\bar{n}} (\bar{r}(\lambda))^n \sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}\right) \left(\prod_{\ell=1}^{n} \frac{\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)}{\bar{r}(\lambda)}\right) \\
&= \tilde{y}(\lambda) \sum_{n=0}^{\bar{n}} (\bar{r}(\lambda))^n \sum_{p \in \mathcal{P}_n} \prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}} + \sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}\right) \left(\prod_{\ell=1}^{n} \frac{\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)}{\bar{r}(\lambda)} - 1\right) \\
&= \tilde{y}(\lambda) \sum_{n=0}^{\bar{n}} (\bar{r}(\lambda))^n \left(c_n + \sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}\right) \left(\prod_{\ell=1}^{n} \frac{\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)}{\bar{r}(\lambda)} - 1\right)\right) \\
&= \tilde{y}_0(\lambda) \sum_{n=0}^{\bar{n}} (c_n + \Delta c_n) (\bar{r}(\lambda))^n
\end{aligned}
$$

where

$$\Delta c_n = \sum_{p \in \mathcal{P}_n} \left(\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}} \right) \left(\prod_{\ell=1}^{n} \frac{\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)}{\bar{r}(\lambda)} - 1 \right)$$

Hence, Lemma 7 applies directly when Δc_n satisfies the variance bound assumed in the present Theorem. ■

Condition (22) is stated in somewhat high-level form for maximum generality, but it is relatively easy to realize that it is a weak restriction. This condition places a limit on the order of magnitude of the variance of a certain average. (The weighting factor $\prod_{\ell=1}^{n} W_{p_\ell p_{\ell+1}}$ sums up to one over all paths in $\mathcal{P}_n$, so the sum is a weighted average.) This average is taken over all possible pathways and effectively samples the spectral representation of the impulse response of large number of agents. Typically, the number of possible pathways of length n is an exponentially increasing function of n (because at each node there are certain number of possible ways to go and these alternative multiply to give the number of paths). Hence, unless the covariance of the summand across two pathways is extremely strong, the decrease of the variance of the average with n should often satisfies the bound (22).

Note that (22) bounds the heterogeneity in the response of paths, while placing only weak restrictions on the heterogeneity in the response of individual agents. Even if the economy is characterized by agents whose response $\tilde{r}_{ij}(\lambda)$ varies significantly with i and j, it is still plausible that the response $\prod_{\ell=1}^{n} \tilde{r}_{p_\ell p_{\ell+1}}(\lambda)$ of most paths $p \in \mathcal{P}_n$ could be very similar due to an averaging effect over the responses of many different agents sampled along the path. This assumption is plausible even in an economy with a mixture of very large firms (e.g. banks that are "too big to fail", such as some banks in the recent banking crisis.) and very small firms. In that case, as most paths will likely go through some of the same large firms, the responses $\prod_{\ell=1}^{n} \tilde{r}_{p_\ell p_{\ell+1}}(\lambda)$ of two paths would tend to be quite similar, since they would often include some identical $\tilde{r}_{p_\ell p_{\ell+1}}(\lambda)$ terms. The fact that only the average $\bar{r}(\lambda)$ needs to satisfy Assumption 3, and not the individual $\tilde{r}_{ij}(\lambda)$, brings considerable generality to the result. In particular, the constant results to scale assumption need not hold at the node level but only at a global level.

D A simplified Long and Plosser model

D.1 Model

In this section, we show how the Long and Plosser model (hereafter LP) and its solution can be specialized to our setup where there are no separate labor inputs. LP's production function has the form

$$q_{it} = \eta_{it} \ell_{it-1}^{b_i} \prod_{j=1}^{N} q_{ij,t-1}^{W_{ij}} \tag{23}$$

where ℓ_{it} is labor inputs for the production of good i and b_i is a parameter such that the constant returns to scale $b_i + \sum_{j=1}^{N} W_{ij} = 1$ constraint holds. All other variables are as in our model. LP's representative consumer maximizes his expected discounted utility:

$$u_t = E\left[\sum_{s=t}^{\infty} \beta^{t-s} Z_t^{\theta_0} \prod_{i=1}^{N} c_{is}^{\theta_i} \middle| q_{t-1}, \eta_{t-1}\right] \tag{24}$$

where Z_t is leisure, equal to $H - \sum_{j=1}^{N} \ell_{js}$ where H is the total labor available, and θ_0 is a parameter and all other variables are as in our model. Defining

$$\gamma_j \equiv \theta_j + \beta \sum_{i=1}^{N} \gamma_i W_{ij},$$

LP show that the solution to this model is

$$\begin{aligned}
c_{it} &= \left(\frac{\theta_i}{\gamma_i}\right) q_{it} \\
Z_t &= \theta_0 \left(\theta_0 + \beta \sum_{i=1}^{N} \gamma_i b_i\right)^{-1} H \\
q_{ij,t} &= \left(\frac{\beta \gamma_i W_{ij}}{\gamma_j}\right) q_{jt} \\
\ell_{it} &= \beta \gamma_i b_i \left(\theta_0 + \beta \sum_{j=1}^{N} \gamma_j b_j\right)^{-1} H \\
\ln q_t &= W \ln q_{t-1} + k + \ln \eta_t
\end{aligned}$$

where k is a vector of constants and the ln function is applied element-by-element.

Our production function is a special case of Equation (23) obtained in the limit as $b_i \to 0$ while adjusting W_{ij} to preserve the constant returns to scale constraint. As a result, the solution to our model reduces to:

$$\begin{aligned}
c_{it} &= \left(\frac{\theta_i}{\gamma_i}\right) q_{it} \\
Z_t &= H \\
q_{ij,t} &= \left(\frac{\beta \gamma_i W_{ij}}{\gamma_j}\right) q_{jt} \\
\ell_{it} &= 0 \\
\ln q_t &= W \ln q_{t-1} + k + \ln \eta_t
\end{aligned}$$

and substituting the solution $Z_t = H$ into the utility yields

$$u_t = E\left[\sum_{s=t}^{\infty} \beta^{t-s} H^{\theta_0} \prod_{i=1}^{N} c_{is}^{\theta_i} \middle| q_{t-1}, \eta_{t-1}\right],$$

which is equivalent to our utility (Equation (24)) up to an irrelevant multiplicative constant H^{θ_0}. Observe that the solution remains well-behaved in the limit of $b_i \to 0$. In particular, the form of the time-evolution of $\ln q_t$ is preserved; the only difference is that the coefficients W_{ij} must now satisfy $\sum_{j=1}^{N} W_{ij} = 1$ instead of $\sum_{j=1}^{N} W_{ij} = 1 - b_i < 1$. Within the original Long-Plosser model, when $b_i > 0$, labor's ability to adjust instantaneously effectively dampens the noise and always yields exponentially decaying c_n coefficients (since $\sum_{j=1}^{N} W_{ij} < 1$) and thus short-memory processes as solutions. The limit $b_i \to 0$ leads to more interesting long-memory dynamics in the large-network limit.

It should be noted that the absence of a separate labor input ($b_i \to 0$ limit) does not mean that the model does not allow for labor inputs. Labor can be supplied via the network and treated symmetrically as part of the remaining inputs q_t. The limit $b_i \to 0$ then implies that the fraction of labor input that can adjust instantaneously to shocks is infinitesimal, which is arguably no less plausible than assuming that the entire labor force can adjust instantaneously to shocks.

D.2 Empirical Example

One way to empirically assess if the proposed mechanism for long memory generation is plausible is to verify if the c_n coefficients in a toy model based on real economic network data indeed obey a power law with the appropriate exponent. For this purpose, we use the so-called "input-output accounts" database compiled by the Bureau of Economic Analysis describing interactions between sectors of the US economy. We use the most disaggregated version of this data since it already contains all the information about information propagation (or "diffusion") over all scales, small and large. This strategy enables a plot of $(\ln(c_n), \ln(n))$ over as many orders of magnitude as possible, thus facilitating the identification of a linear trend.

We construct the network following the same procedure as in Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), using a reconstructed Commodity-by-Commodity Direct Requirements table for year 2002, available in their supplementary material. These represent the equilibrium cost shares of each commodity j in the production of another commodity i. (Following Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), we use the terms industries and commodities interchangeably.) In the Long and Plosser-type model, these shares are equal to the Cobb-Douglass parameters W_{ij} of the production function (Equation (23) with $b_i = 0$). We include an additional node ℓ in the network to model labor supply. In the same spirit as in Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) (see p. 1998), and in accordance with our constant return to scale assumption, we set the labor share in the production of good i to $W_{i\ell} = 1 - \sum_{j \neq \ell} W_{ij}$.

To close the loop, the labor force must take input from the economy for their livelihood. We do not have quantitative data on this, hence we assume that the workers take inputs from all industries $j = 1, \ldots, (N-1)$ with equal equilibrium share $W_{\ell j} = \rho/(N-1)$ and from each other with share $W_{\ell\ell} = 1 - \rho$. We used $\rho = 0.75$, but the results are not very sensitive to this parameter.

In this empirical example, there is no reason to expect that the c_n coefficients should be the same for every choice of source and destination node. As an example, we pick the group of industries that are numbered, according to North American Industry Classification System (NAICS), with a leading "2". These correspond largely to primary sector industries (such as mining and utilities). We compute the c_n coefficients via Equation (4), setting both the destination vector e^d and origin vector e^o to be a vector selecting all industries in this group. This corresponds to computing the spectrum of the aggregate response of this group of industries to a common shock.

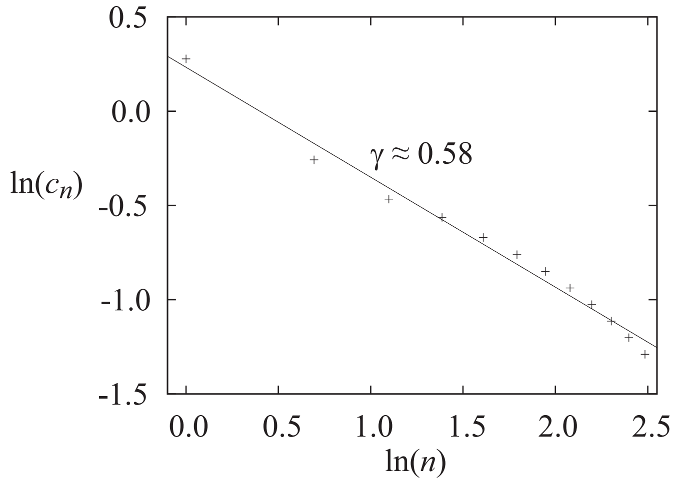

Figure 5: Evidence of power law scaling $n^{-\gamma}$ with $\gamma \approx 0.58$) in the c_n coefficients (i.e. the probability of reaching a given point of the network after n steps of a random walk) in a network representing the US economy as 418 "sectors".

The resulting c_n coefficients are shown in Figure 5 and reveal evidence of a power law $c_n = n^{-\gamma}$ in this industry group with an exponent of $\gamma \approx 0.58$, as obtained with a standard linear least squares regression of the data in logarithmic form. This corresponds to $\alpha = 1 - \gamma \approx 0.42$, i.e., a power spectrum behaving as $|\lambda|^{-2\alpha} = |\lambda|^{-0.84}$ near the origin, resulting in a long memory network behavior of a fractionally integrated nature of order $\alpha \approx 0.42$. Although this is, strictly speaking, a finite network, one can still observe a behavior that would be expected from an infinite network for "short" paths, because "short" paths do not "feel" the boundary of the network. Of course, if we increased the range of n, the graph would flatten out, as would be expected for a finite network (since the c_n would be asymptotically constant in that case).

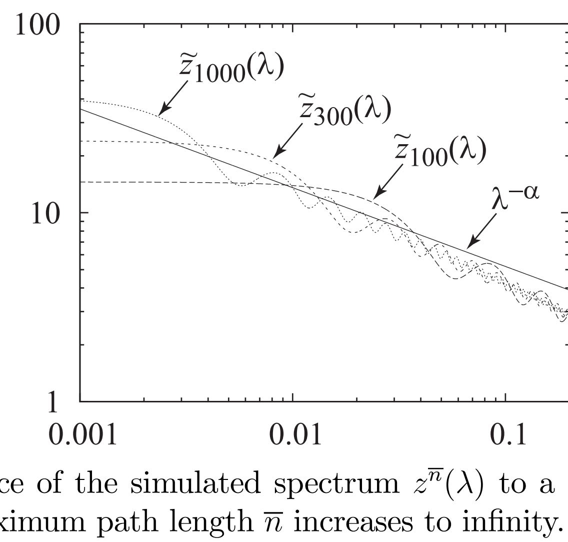

Figure 6: Convergence of the simulated spectrum $z^{\overline{n}}(\lambda)$ to a power law ($\lambda^{-\alpha}$, with $\alpha = 0.42$), as the maximum path length $\overline{n}$ increases to infinity.

We can pursue this example a bit further and explicitly calculate the spectrum associated with the power law $c_n \propto n^{-0.58}$ for our simplified Long-Plosser model. We employ the expression $\tilde{z}^{\bar{n}}(\lambda) = \sum_{n=0}^{\bar{n}} c_n (\tilde{r}(\lambda))^n \tilde{y}(\lambda)$, in which $\tilde{r}(\lambda) = e^{i\lambda}$ (since there is a single lag in the autoregressive representation in this model) and $\tilde{y}(\lambda) = 1$ (assuming a standard white noise as noise source). Figure 6 illustrates how $\tilde{z}^{\bar{n}}(\lambda)$ converges to a power law $\lambda^{-\alpha}$ as $\bar{n}$ increases. One can see that, as $\bar{n} \to \infty$, the oscillations around the limiting power law decrease in magnitude and the interval over which the spectrum is well described by a power law expands towards zero frequency.