Inference for Noisy Long Run Component Process

Gourieroux, Christian and Jasiak, Joann

University of Toronto, York University

1 January 2010

Inference for Noisy Long Run Component Process

C., GOURIEROUX [(1)] and J., JASIAK [(2)]

March, 1, 2020

The first author gratefully acknowledges financial support from the ACPR chair "Regulation and Systemic Risk", and the ERC DYSMOIA, the second author thanks the Natural Sciences and Engineering Council of Canada (NSERC). We thank Maynard, A., Monfort, A., for helpful comments.

[1]Toronto University, Toulouse School of Economics, and CREST.
[1]York University.

Inference for Noisy Long Run Component Process

Abstract

This paper introduces a new approach to the modelling of a stationary long run component, which is an autoregressive process with near unit root and small sigma innovation. We show that a combination of a noise and a long run component can explain the long run predictability puzzle pointed out in Fama-French (1988). Moreover in the presence of a long run component, spurious regressions arise and misleading long run predictions are obtained when standard statistical approaches are applied. Cleaner asymptotic inference is provided for models with a noisy long run component .

Keywords : Long Run, Predictability Puzzle, Weak Identification, Deconvolution, Term Structure, Near Unit Root, Small Sigma.

1 Introduction

The idea of separating unobserved and independent transitory and permanent components, that represent distinct short and long run dynamics, of a time series appeared in the early economic literature [see e.g. Friedman (1957) for the notion of permanent income]. This distinction is especially relevant in impulse response analysis, which evaluates the consequences of shocks on the short run component (resp. the long run component) [see e.g. Bansal, Yaron (2004), where those components are present in the consumption and dividend growths [2]]. It is expected that such shocks can have a short term and also potentially a long term impact (resp. a long term impact only). If we are interested in the long term impacts, we need to "identify" the long run component, which is difficult if the variance of the short run component is large. This difficulty has been pointed out from a practical point of view in a series of papers by Fama, French (1988) a,b, (1989), and is known as the (long run) predictability puzzle. More precisely, they considered a series of stock returns and applied the standard approach of regressing the series y_t on its lagged value y_{t-1}. Then they observed an absence of a significant relationship supporting the efficient market hypothesis. However, when they regressed the averages $\frac{1}{h}(y_t + \ldots + y_{t+h-1}) \equiv y_t(h)$, say, over a different time unit, the regression coefficient becomes significant, for large horizon h. The (long run) predictability puzzle[3] is an empirical contradiction : "Efficient markets could be inefficient in the long run", or "profits cannot be sure in the short term, but become possible in the long term" [4].

To paraphrase Yule (1926) on nonsense correlations : "We sometimes obtain between quantities varying with the time quite (low) (serial) correlations [5] to which we cannot attach any (relevant) physical significance. It is important "clarify" how they arise and in what special cases".

The aim of this paper is to consider a model for noisy long run component

[2]Extensive literature on the Long Run Risk (L.R.R.) model has followed this work. See Hansen, Heaton, Li (2008), Malloy, Moskowitz, Vissing-Jorgensen (2009), Bansal, Kiku, Yaron (2012), Beeler, Campbell (2012), Croce, Lettau, and Ludvigson (2014), Gramming, Kuchlin (2016), Pohl et al. (2018).., where the L.R. component can also appear in the volatility or the model may include several L.R. components to account for business cycles.

[3]See e.g. Ferson et al. (2003) for a survey on stock return predictability puzzles.

[4]see also Bansal, Yaron (2004), Table 3, where variance ratios increase with horizon, Bandi, Perron (2008), Beeler, Campbell (2012), Table 4, or Bonomo et al. (2015).

[5]In the original Yule's text "low" is replaced by "high".

that can explain the observed predictability puzzle. The observed series is written as $y_t = y_{s,t} + y_{l,t}$, where the short run component $y_{s,t}$ is a pure noise and the long run component is an autoregressive process $y_{l,t} = \rho y_{l,t-1} + \sigma\varepsilon_{l,t}$, where ρ is close to one and σ close to zero, and the two components are independent. This double condition on ρ, σ allows to generate long run patterns in the trajectories, while conserving the stationarity of the long run (L.R.) component. If the short run component has a much larger "size" than the long run component, the standard statistical inference can hardly detect (identify) the L.R. component in such a noisy environment.

The paper is organized as follows. In Section 2, we discuss the properties of the standard autoregressive process of order 1 : $y_{l,t} = \rho y_{l,t-1} + \sqrt{1-\rho^2}\varepsilon_{l,t}$, where $\varepsilon_{l,t}$ is i.i.d, with zero-mean and unit variance. In particular we discuss the effect of a change of time unit, that is of a compression/dilatation of time, on the trajectories of such a process. This allows us to discuss the difference between the analysis in frequency domain and in term domain, and to define this latter notion. In Section 3, we explain how to identify semi-nonparametrically the parameters of the model, i.e. the scalar parameter ρ and the distributions of the two independent noises $y_{s,t}$ and $\varepsilon_{l,t}$. We compare in Section 4 theoretical and estimated regression models with long run components in both the dependent and regressor variables. This shows an underestimation bias, even if the standard patterns function of the term are still observed. Section 5 considers statistical inference. We first introduce closed form estimators of both the scalar (ρ) and functional parameters (the distributions of shocks). Next we explain why the standard estimators and tests provide spurious results for the noisy long run component. We develop a joint near unit root/small sigma asymptotics to get more reliable statistical analysis in this framework and give some insights on (large) finite sample properties. Section 5 concludes. Proofs are gathered in Appendices.

2 Dynamic properties

Let us first recall and interpret the dynamic properties of a noisy long run component. For expository purpose we consider a stationary series y_t that can be decomposed as :

$$y_t = y_{s,t} + y_{l,t}, \tag{2.1}$$

$$\text{where} \quad y_{l,t} = \sqrt{p_l} x_{l,t}, \; y_{s,t} = \sqrt{p_s} x_{s,t}, \tag{2.2}$$

$$\text{with} \quad p_l > 0, \; p_s > 0, \; p_l + p_s = 1, \tag{2.3}$$

where the components $x_{s,t}, x_{l,t}$ are a noise and an autoregressive process defined below :

$$x_{s,t} = \varepsilon_{s,t}, \; x_{l,t} = \sqrt{1-\rho^2} \frac{\varepsilon_{l,t}}{1-\rho L}, \; 0 \leq \rho < 1, \tag{2.4}$$

and $(\varepsilon_{s,t}), (\varepsilon_{l,t})$ are independent i.i.d. processes, with mean-zero and unit variance.

We are interested in the cases when the autoregressive coefficient ρ in the $x_{l,t}$ component is close to 1. As shown below, in this case $x_{l,t}$ can be interpreted as a long run component, and y_t as a noisy long run component process. The weight of $y_{l,t}$ depends on the value of p_l. The condition $p_l + p_s = 1$ ensures that the aggregate process has a unit variance, equal to the variance of each process $x_{s,t}$ and $x_{l,t}$. Therefore p_s (resp. p_l) is the fraction of the total variance due to the white noise component (resp. the L.R. component).

Our model differs from the nearly integrated, nearly white noise modelling introduced in Nabeya, Perron (1994) [see also Ng, Perron (1996), Deng (2004)], where $y_t = \rho y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \rho$ and θ being both close to 1. Their process has a single component that features persistence in finite sample and becomes white noise in large sample. As shown below, model (2.1)-(2.4) with ρ close to 1 behaves as a white noise in the short run, but reveals persistence in the long run.

2.1 The AR process

Let us focus on the process $x_{l,t} = \rho x_{l,t-1} + \sqrt{1-\rho^2} \varepsilon_{l,t}$. For any value of ρ, $0 < \rho < 1$, this process is strictly stationary with mean zero and unit variance.

Let us now consider the consequence of a variation in parameter ρ. For expository purpose, we assume henceforth that $\varepsilon_{l,t}$ is standard normal and the initial condition $x_{l,0}$, is drawn in the standard normal stationary distribution. In this case process $(x_{l,t})$ can be viewed as a time discretized Ornstein-Uhlenbeck process defined by :

$$dx_t = -kx_t dt + \eta dW_t, \tag{2.5}$$

where k, η denote the (infinitesimal) drift and volatility parameters, and (W_t) a Brownian motion. The parameters k and η are related to parameter ρ as follows [see e.g. Vasicek (1977), eq. (25)-(26)] :

$$\rho = \exp(-k) \Longleftrightarrow k = -\log \rho,\ \eta^2 = -2 \log \rho = 2k. \tag{2.6}$$

Let us now consider the consequences of a change of time unit of index t. If the initial time unit is h times the new one, we dilate (resp. compress) the time, if $h > 1$ (resp. $0 < h < 1$). By applying the change of time unit to stochastic diffusion equation (2.5)-(2.6), the process $\tilde{x}_t(h)$ expressed in the new time unit satisfies :

$$d\tilde{x}_t(h) = -\frac{k}{h}\tilde{x}_t(h)dt + \sqrt{\frac{2k}{h}}d\tilde{W}_t(h), \text{ say,} \tag{2.7}$$

with $\tilde{x}_t(1) = x_t$.

The infinitesimal drift and volatility parameters $k, \eta^2 = 2k$, and also the autoregressive parameter ρ depend on the time unit. More specifically the autoregressive parameter becomes $\rho^{1/h}$. Therefore a change of value of the autoregressive parameter from ρ to any other value ρ^* in (0,1) can be also interpreted as multiplying the time unit by $h = \log \rho / \log \rho^*$. The larger ρ^*, the smaller the new time unit. This time compression/dilatation effect largely explains the patterns of trajectories. By dilatation of time we transform a trajectory of a process $x_{l,t}$ associated with autoregressive coefficient ρ into a trajectory of a process associated with another autoregressive coefficient $\rho^* > \rho$. Then the new trajectory observed in the initial time unit appears smoother. There are two limiting cases :

i) For $\rho = 0$, we get the Gaussian white noise corresponding to an infinite compression of time, that is to the limiting case $h = 0$.

ii) For $\rho = 1$, we have an infinite dilatation, $k = \eta^2 = 0$, and the process has constant trajectories. More precisely if $x_{l,0}$ is drawn in the standard normal, we have $x_{l,t} = x_{l,0}, \forall t$. The trajectories are constant in time, but their level is stochastic. This limiting result can be surprising as a unit root process is usually considered as nonstationary. However, in our framework, the innovation variance of the autoregressive process $1 - \rho^2$ is adjusted with

ρ. When ρ increases, the shock diminishes in order to keep the stationary distribution unchanged.

The dilatation effect is illustrated in Figures 1a,b where we report six trajectories of length 5000 of AR(1) Gaussian processes $x_{l,t}$ with variance 1, equal initial value $x_{l,0} = 0$, and ρ set equal to $\rho = 1 - 1/K, K = 1, 10, 50, 100, 500, 1000$. They are based on a same underlying trajectory of $\varepsilon_{l,t}$. The first trajectory of $x_{l,t}$ corresponds to the standard Gaussian white noise. When ρ increases, the trajectories become smoother, due to the time dilatation.

As mentioned earlier, for ρ very close to 1, we expect the $x_{l,t}$ process to be close "in distribution" to a constant process with a stochastic level. This feature starts to appear in the last panel of Figure 1.b. Given a zero starting value, the trajectory remains rather close to zero as long as there is no $\varepsilon_{l,t}$ drawn in the "tail" of the Gaussian distribution. Then the trajectory remains close to this new value up to a next drawing in the "tail" and so on, while always preserving the same marginal distribution.

[Insert Figure 1 : Trajectories of Gaussian AR(1) Processes]

We report in Figure 2 the noisy long run component processes obtained by combining the first and last trajectories of Figures 1a,b. Two combinations are considered : one with equal weight : $p_s = p_l = 0.5$, and one with a smaller weight of the long run component : $p_s = 0.95, p_l = 0.05$. In process y_t, defined in (2.1)-(2.2) the relative weights $\sqrt{p_s/p_l}$ of the noise to the L.R. component are 1 and about 4 for the two combinations, respectively. These examples are in line with the calibration exercices in Bansal, Yaron (2004), Bansal, Kiku, Yaron (2007), or Beeler, Campbell (2012).

In the equal weights combination, the impact of the L.R. component becomes visible after $T = 2000$, i.e. when sufficiently many small shocks have been cumulated. In the combination with a large noise, this effect is less visible.

[Insert Figure 2 : Trajectories of Noisy Long Run Components]

The transformation $\rho \to \rho^{1/h}$ has an alternative interpretation. Indeed $\rho^{1/h}$ is the autoregressive coefficient for predicting the future value of process $x_{l,t}$ at horizon (term) $1/h$. Thus the inverse of a change of time unit can be interpreted as a term. Let us now discuss the difference between the term analysis and frequency analysis.

2.2 Term analysis versus frequency analysis

It is common to perform a second-order analysis of weakly stationary time-series by representing the time series as combination of sine and cosine functions with stochastic coefficients :

$$y_t \simeq \Sigma_\omega [A(\omega)\cos \omega t + B(\omega)\sin \omega t], \text{say}, \tag{2.8}$$

where argument $\omega, \omega \in (0, 2\pi)$, denotes the frequency. The coefficients $(A(\omega), B(\omega))$ are uncorrelated for different $\omega' s$, have mean zero and variances such that $V[A(\omega) + B(\omega)] \sim f(\omega)$ [see e.g. Gourieroux, Monfort (1977), Th. 8.21], where $f(\omega)$ is the spectral density :

$$f(\omega) = \sum_{h=-\infty}^{+\infty} \gamma(h) \exp(i\omega h), 0 \leq \omega \leq 2\pi, \tag{2.9}$$

and $\gamma(h)$ is the autocovariance function. The second-order analysis can be equivalently performed in the time domain by considering the pattern of the autocovariance function, or in the frequency domain by considering the pattern of the spectral density.

The analysis in frequency domain allows for decomposing a series into its frequency components : $y_t = \Sigma_\omega y_{\omega,t}$, say, by means of second-order properties. It has been proposed in the literature to interpret the low frequency component $\omega \sim 0$ as the long run component [see e.g. Beveridge, Nelson (1981), Calvet, Fisher (2007) and for more recent papers Ortu et al. (2013), Bandi et al. (2018) for a similar approach based on wavelets]. However, this approach provides frequency components that are deterministic functions of the same initial series and are mutually dependent. Therefore the components at different frequencies cannot be shocked separately in a causal impulse response analysis.

This is not the case with the decomposition introduced in model (2.1)-(2.2). In general y_t can have J autoregressive components $J \geq 2$:

$$y_t = \sum_{j=1}^{J} \left(\sqrt{p_j} \sqrt{1-\rho_j^2} \frac{\varepsilon_{j,t}}{1-\rho_j L} \right), \text{ say, with } p_j > 0, j = 1, \ldots, J, \sum_{j=1}^{J} p_j = 1, \tag{2.10}$$

where by construction the innovations $(\varepsilon_{j,t}), j = 1, \ldots, J$, and the components at different time units (different terms) are independent (not only uncorrelated) and can be shocked separately. Model (2.1)-(2.2) is the special case with $J = 2, \rho_1 = 0, \rho_2 = \rho$.

Let us now discuss the relationship between the term domain (in ρ) and frequency domain (in ω) analysis. The spectral density of process (y_t) in (2.10) is :

$$f(\omega) = \frac{1}{2\pi} \sum_{j=1}^{J} \left\{ p_j \frac{1-\rho_j^2}{|1-\rho_j \exp(i\omega)|^2} \right\} = \frac{1}{2\pi} \sum_{j=1}^{J} \left[p_j \frac{1-\rho_j^2}{1+\rho_j^2 - 2\rho_j cos\omega} \right]. \tag{2.11}$$

The spectral density contains the same information as the sequence of $\rho_j' s$ and weights $p_j' s$ (see Section 3.1).

Let us now examine the pattern of a baseline component of spectral density (2.11) :

$$f(\omega; \rho) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho cos\omega}. \tag{2.12}$$

This function of $\cos\omega$ has a maximum for $\omega = 0, 2\pi$, a minimum for $\omega = \pi$, and inflexion points at $\omega = \pi/2, 3\pi/2$ (see Figure 3). Its values at these points are : $\frac{1}{2\pi}\frac{1+\rho}{1-\rho}, \frac{1}{2\pi}\frac{1-\rho}{1+\rho}$ and $\frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2}$, respectively.

[Insert Figure 3 : Component Spectral Density, $\rho = 0.3, 0.8$].

We see the drawback of detecting the long run component in frequency domain. If the process satisfies (2.10), the spectral density (2.11) evaluated at $\omega \simeq 0$ is a mixture of AR(1) components with different $\rho' s$ (i.e. terms). Loosely speaking in a simple model such as a noisy long run component (2.1)-(2.4), we want to shock separately $\varepsilon_{s,t}$ and $\varepsilon_{l,t}$. In particular we want to distinguish the long run effect by focusing on $\varepsilon_{l,t}$, whereas a spectral density close to $w = 0$ is equal to $\frac{1}{2\pi} \Sigma_j \left(p_j \frac{1+\rho_j}{1-\rho_j} \right)$, that involves all the terms.

3 Identification

In model (2.10), with i.i.d. shocks $\varepsilon_{j,t}$, from a distribution g_j, which is not necessarily Gaussian, it is possible to identify semi-nonparametrically the parameters $p_j, \rho_j, j = 1, \ldots, J$, and the distributions $g_j, j = 1, \ldots, J$, whenever the $\rho_j' s$ are different with $|\rho_j| < 1, \forall j$ [see Gourieroux, Jasiak (2019)]. The proof of identification is constructive and used to derive consistent semi-parametric estimation methods (at least for ρ fixed). We apply below these results to the noisy long run component model (2.1)-(2.4).

3.1 Identification of parameters p_l, ρ

For model (2.10), the proof of identification is based on the uniqueness of the partial fraction decomposition of spectral density (2.11) [see e.g. Maravall (1979), Bradley, Cook (2012)]. In the special case of model (2.1)-(2.4), we have :

$$\begin{aligned} \gamma(h) &= p_s \gamma_s(h) + p_l \gamma_l(h) \\ &= p_l \rho^h, \text{ for any } h > 1. \end{aligned}$$

Therefore parameters p_l, ρ are directly identifiable from $\gamma(1), \gamma(2)$. We have : $\gamma(1)(= \rho(1)) = p_l \rho, \gamma(2)(= \rho(2)) = p_l \rho^2$. Therefore :

$$\rho = \gamma(2)/\gamma(1), \;\; p_l = \gamma(1)^2/\gamma(2). \tag{3.1}$$

This corresponds to the solutions of the two first Yule-Walker equations.

The expression of ρ corresponds to an instrumental variable approximation of the first-order autoregressive coefficient of y_t with instrument y_{t-2}.

3.2 Nonparametric identification of the distributions

Let us rewrite model (2.1)-(2.2) as :

$$\begin{aligned} y_t &= y_{s,t} + y_{lt} = \tilde{\varepsilon}_{s,t} + \frac{\tilde{\varepsilon}_{lt}}{1 - \rho L}, \\ \text{where } \tilde{\varepsilon}_{st} &= \sqrt{p_s} \varepsilon_{s,t}, \;\; \tilde{\varepsilon}_{l,t} = \sqrt{p_l} \varepsilon_{l,t}. \end{aligned} \tag{3.2}$$

The identification result is obtained from the second characteristic functions of $\tilde{\varepsilon}_{s,t}, \tilde{\varepsilon}_{l,t}, y_{l,t}$, defined by :

$$b_s(u) = \log E[\exp(iu\tilde{\varepsilon}_{s,t})],\ b_l(u) = \log E[\exp(iu\tilde{\varepsilon}_{l,t})], \tag{3.3}$$

and

$$c_l(u) = \log E[\exp(iuy_{l,t})]. \tag{3.4}$$

The lemma below shows that it is equivalent to identify the marginal distribution of $y_{l,t}$ (i.e. c_l), or the marginal distribution of $\tilde{\varepsilon}_{l,t}$ (*i.e.* b_l), when ρ is given.

Lemma 1 : We have $c_l(u) = c_l(\rho u) + b_l(u)$.

Proof : We have $y_{l,t} = \rho y_{l,t-1} + \tilde{\varepsilon}_{l,t}$, where $y_{l,t-1}$ and $\tilde{\varepsilon}_{l,t}$ are independent. Therefore, by independence :

$$E[\exp(iuy_{l,t})] = E[\exp(iu\rho y_{t-1})]E[\exp(iu\tilde{\varepsilon}_{l,t})]$$

$\Leftrightarrow c_l(u) = c_l(\rho u) + b_l(u)$, by stationarity of process (y_t).

$\square$

The identification is obtained by considering the nonlinear dependencies at order 1. Let us denote the pairwise second characteristic function at lag 1 :

$$\psi(u,v) = \log E[\exp(iuy_t + ivy_{t-1})], u, v \in I\!R. \tag{3.5}$$

Proposition 1 : We have :

$$i) \quad \frac{db_s(u)}{du} = \frac{\partial \psi}{\partial v}(-u/\rho, u).$$

$$ii) \quad \frac{dc_l(u)}{du} = \frac{\partial \psi(0,u)}{\partial v} - \frac{\partial \psi}{\partial v}(-u/\rho, u).$$

Proof : See Appendix 1.

Since function ψ is identifiable, the identification of b_s, c_l (and then of b_l by Lemma 1) follows by integration, noting that $b_s(0) = c_l(0) = 0$.

4 Theoretical Dynamic Properties of Noisy Long Run Component

Before discussing statistical inference, it is important to consider the properties of theoretical regression coefficients in regressions that involve L.R.R. components. This will provide a first explanation of the long run predictability puzzle. Then, in a second step, we compare the difference between the theoretical and estimated results to show that a detailed analysis of statistical inference in the case of L.R.R. components is needed.

4.1 Theoretical regression

Let us consider a noisy AR(1) process,

$$y_t = x_{1t} + \sigma\varepsilon_{2t}, \tag{4.1}$$

where $x_{1t} = \rho x_{1t-1} + \varepsilon_{1t}, E\varepsilon_{1t} = 0, V\varepsilon_{1t} = 1 - \rho^2$ ε_{2t} is independent of (ε_{1t}), such that $E\varepsilon_{2t} = 0, V\varepsilon_{2t} = 1$. This model includes a long-run component X_{1t}, if ρ is close to 1, and a short-term component $\sigma\epsilon_{2t}$, as in the analysis of permanent and temporary components of stock prices [Fama, French (1988), Table 1] or in the Long Run Risk consumption growth model [Bansal, Yaron (2004), Hansen et al. (2008), Gollier (2016), Section 6, Schorfheide, Song, Yaron (2018)]. Let us now run the theoretical autoregressions :

$$y_{t+1,t+h} = \beta_{hh} y_{t-h+1,t} + w_{ht},$$

where $y_{t+1,t+h} = y_{t+1} + \ldots + y_{t+h}$. We denote $\beta_{hh}(\rho, \sigma^2)$ the theoretical regression coefficient and $R_{hh}(\rho, \sigma^2)$ the theoretical correlation between the dependent and explanatory variables.

Then we have [see Appendix 2] :

$$\beta_{hh}(\rho, \sigma^2) = R_{hh}(\rho, \sigma^2) = \frac{\gamma(h, h, \rho)}{\gamma(h, \rho) + h\sigma^2}, \tag{4.2}$$

where : $\gamma(h, h, \rho) = \dfrac{\rho(1 - \rho^h)^2}{(1 - \rho)^2}, \gamma(h, \rho) = \dfrac{1 + \rho}{1 - \rho} h - \dfrac{2\rho}{(1 - \rho)^2}(1 - \rho^h)$. In particular :

$$R_{1,1}(\rho,\sigma^2)=\frac{1}{\rho+\sigma^2}, R_{2,2}(\rho,\sigma^2)=\frac{\rho(1+\rho)^2}{2(1+\rho)+2\sigma^2}.$$

Corollary 1: Let us assume $\rho>0$, then $R_{2,2}(\rho,\sigma^2)>R_{1,1}(\rho,\sigma^2)$, iff $(1+\rho)^2\geq 2$ and $\sigma^2\geq\dfrac{(1-\rho)^2}{(1+\rho)^2-2}$.

Proof: The condition on correlations is :

$$\frac{\rho(1+\rho)^2}{2(1+\rho)+2\sigma^2}>\frac{\rho}{1+\sigma^2}$$

$$\iff\quad (1+\rho)^2(1+\sigma^52)>2(1+\rho)+2\sigma^2$$

$$\iff\quad [(1+\rho)^2-2]\sigma^2>(1+\rho)(1-\rho)>0.$$

The result follows. QED

The sequence of correlations is first increasing in h, if ρ is sufficiently large $\rho>\sqrt{2}-1$, and σ^2 is large too. When ρ is close to 1, the constraint on σ^2 becomes : $\sigma^2>0$ and is always satisfied. Thus it seems important to analyze the behaviour of $R_{h,h}(\rho,\sigma^2)$ close to a unit root.

Corollary 2: Let us assume $\rho=1-\delta/h$, where $\delta>0$ is fixed. Then, $\lim_{h\to\infty}R_{h,h}(1-\delta/h;\sigma^2)=\dfrac{[1-\exp(-\delta)]^2}{2[\exp(-\delta)-1+\delta]}\equiv R_\infty(\delta)$.

The function $R_\infty(\delta)$ is a decreasing function of δ, from 1, for $\delta=0$, to 0 for $\delta=\infty$.

Proof: See Appendix 2 iii).

Corollary 2 implies that one can expect a hump-shaped pattern of the theoretical correlation. It suggests that the hump may arise for a rather large h and its size is determined by parameter δ. The model in Corollary 2 resembles the dynamic model of consumption growth in Bansal, Yaron (2004), where X_{1t} is "a small persistent predictable component which determines the conditional expectation of consumption growth". The hump-shaped patterns are illustrated in Figure 4 below for parameters set equal to $\rho=0.8,\sigma^2=10;\rho=0.9,\sigma^2=20;\rho=0.99,\sigma^2=100,\rho=0.995,\sigma^2=10$.

[Insert Figure 4 : $\beta_{h,h} = R_{h,h}$ function of h]

Such patterns are evidenced in the empirical literature [see e.g. Bandi et al. (2018) for a similar curve, and Bandi, Perron (2008), Fig 2, for a curve for small h]. Predictability is not observed in the short run (h small), as the small persistent component is concealed by the large noise. When h increases the effect of the noise diminishes. Hence, for very large h, function $R_{h,h}$ decreases slowly to zero due to the persistent component.

4.2 Estimated versus theoretical regressions

However in practice we do not know the truth and the L.R. predictatbility puzzle is observed on estimated regressions. Thus it is important to see if there is a systematic estimation bias.

To illustrate the size of this bias, we consider a series of simulated data with $T = 400$ observations and parameters $\rho = 0.99, \sigma^2 = 10$. The simulated series is displayed in Figure 5.

[Insert Figure 5 : Simulated Series]

Despite the high value of the autoregressive coefficient, we do not observe a trend and the trajectory is rather erratic. The estimates of ρ, σ^2 in step 1 are : $\hat{\rho}_T = 0.904, \hat{\sigma}_T^2 = 10.23$. In Figure 6 we plot the true values of R_{hh} (dashed line), the values of $R_{hh}(\hat{\rho}_T, \hat{\sigma}_T^2)$ (dotted line), as well as the values of the standard empirical estimator $\tilde{R}_{h,h}$ (solid line) as functions of h . As expected, the $\tilde{R}_{h,h}$ line lies considerably above the true value of R_{hh} and is close to 1 for large lags h because of the overlapping observations and the decreasing number of observations from which it is calculated when lag h increases. The estimator $R_{hh}(\hat{\rho}_T, \hat{\sigma}_T^2)$ provides the hump shaped curve, but underestimates the true value of the determination coefficient (see the discussion in the next Section).

[Insert Figure 6 : True and Estimated Coefficients]

5 Statistical Inference

In this section, we first introduce consistent estimators of parameters p_l, ρ and of the distributions of shocks in model (2.1)-(2.4), when p_l, ρ are fixed.

Next we show that in practice spurious results are obtained when the true value of ρ is close to 1, in the long run interpretation of component $y_{l,t}$. This is a consequence of interpreting ρ as an autoregressive coefficient rather than a change of time unit, that follows from the reparametrization of the standard AR(1) model. Indeed this reparametrization implies a lack of local identifiability for $\rho = 1$. We also consider the case of ρ approaching one when the number of observations tends to infinity, that is a joint near unit root and small sigma asymptotics [see Phillips (1987), Chan, Wei (1987) for the introduction of near unit root asymptotics, Kadane (1971) for the introduction of small sigma asymptotics in the econometric literature and Gospodinov (2009), Gospodinov, Maynard and Pesavento (2017), Assumption B, for both high persistence and low signal to noise ratio in cointegrated systems]. We provide the asymptotic and (large) finite sample behaviours of estimated parameters of interest. In the last part of this Section, we explain how the standard Bayesian approach [see e.g. Schorfheide et al. (2018)] has also to be modified. These new asymptotic properties call into question the results obtained in the literature on long run risk model, when standard asymptotic results are automatically applied [see e.g. Bansal, Yaron (2004), Bansal, Kiku, Yaron (2007), Schorfheide et al. (2018)].

5.1 Estimators

The estimators of scalar parameters p_l, ρ and of distributions of shocks are derived by using equation (3.1) and equations i), ii) in Proposition 1, that is by finding the solutions of these equations after replacing reduced form parameters $\gamma(1), \gamma(2)$ [resp. $\frac{\partial \psi(u,v)}{\partial v}$] by their sample counterparts :

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=3}^{T} y_t y_{t-h}, \; h = 1, 2 \;\; [\text{ resp. } \frac{\partial \hat{\psi}(u,v)}{\partial v}].$$

The solution of equation (3.1) is the estimator of ρ :
$\hat{\rho} = \hat{\gamma}(2)/\hat{\gamma}(1)$.

Next we estimate the partial derivative of the pairwise second-characteristic function :

$$\frac{\partial \psi(u,v)}{\partial v} = \frac{E[iy_{t-1} \exp(iuy_t + ivy_{t-1})]}{E[\exp(iuy_t + ivy_{t-1})]}, \tag{5.1}$$

by its sample counterpart :

$$\frac{\partial\hat{\psi}(u,v)}{\partial v} = i\frac{\Sigma_t[y_{t-1}\exp(iuy_t + ivy_{t-1})]}{\Sigma_t[\exp(iuy_t + ivy_{t-1})]}. \tag{5.2}$$

Then the sample counterparts of equations i), ii) in Proposition 1 are solved for $\frac{db_s}{du}, \frac{dc_l}{du}$ after separating their real (Re) and imaginary (Im) parts as :

$$\begin{aligned}
Re\frac{\widehat{db_s}(u)}{du} &= Re\frac{\widehat{\partial\psi}}{\partial v}(-u/\hat{\rho}, u), \\
Im\frac{\widehat{db_s}(u)}{du} &= Im\frac{\widehat{\partial\psi}}{\partial v}(-u/\hat{\rho}, u), \\
Re\frac{\widehat{dc_l}(u)}{du} &= Re\frac{\widehat{\partial\psi}}{\partial v}(0, u) - Re\frac{\widehat{db_s}(u)}{du}, \\
Im\frac{\widehat{dc_l}(u)}{du} &= Im\frac{\widehat{\partial\psi}}{\partial v}(0, u) - Im\frac{\widehat{db_s}(u)}{du},
\end{aligned}$$

where $\hat{\rho}$ is the estimator of ρ obtained in the first step.

The closed form expressions of the real and imaginary parts of $\frac{\widehat{\partial\psi}}{\partial v}(u,v)$ are given in Appendix 3.

The estimator $\hat{\rho}$ of the scalar parameter and the estimators of the functional parameters $\frac{\partial\hat{b}_1}{du}, \frac{\partial\hat{c}_l(u)}{du}$ are consistent, asymptotically normal, when T tends to infinity and all (scalar and functional) parameters are fixed. For instance, $\sqrt{T}(\hat{\rho} - p_l\rho)$ is asymptotically normal with mean zero and variance $1 - p_l^2\rho^2$.

These asymptotic properties of estimators are no longer valid when $\rho = \rho_T$ tends to 1, when T tends to infinity.

5.2 Parametrization and identifiability

Before discussing the inference, let us consider two alternative parametrizations of a Gaussian AR(1) model :

$$\begin{aligned} x_t &= r x_{t-1} + \eta \varepsilon_t, && (5.3) \\ \text{and } x_t &= \rho x_{t-1} + \sigma\sqrt{1-\rho^2}\varepsilon_t. && (5.4) \end{aligned}$$

It is very important to distinguish these parametric models in which the parameters have different economic interpretations, that are a serial correlation for r and a change of time unit for ρ. This distinction is hidden when the same notation is used for r and ρ (resp: η and σ). For instance, the model is written as $x_t = \rho x_{t-1} + (\varphi\sigma)\varepsilon_t$ in Bansal, Yaron (2004) and the major part of the L.R.R. literature, but it is written as : $x_t = \rho x_{t-1} + \sqrt{1-\rho^2}(\varphi\sigma)\varepsilon_t$ in Schorfheide, Song, Yaron (2018).

In model (4.3), the parametric efficiency bound for (r, η) is : $B(r, \eta) = \begin{pmatrix} 1-r^2 & 0 \\ 0 & 2\eta^4 \end{pmatrix}$. For $r = 1$, we get zero as the first diagonal element. This corresponds to the limiting case of a random walk in which the speed of convergence of the maximum likelihood estimator of r is $1/T$, not $1/\sqrt{T}$.

The results are different with the parametrization in (4.4). Indeed the Jacobian to transform (r, η) into (ρ, σ) is : $J = \begin{pmatrix} 1 & 0 \\ \dfrac{\rho\sigma}{1-\rho^2} & \dfrac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$. The last row of the Jacobian tends to infinity, when ρ tends to 1. This reveals a weak identification in a neighbourhood of $\rho = 1$. This local nonidentification explains the use of local alternatives to $\rho = 1$ in Section 4.4 to analyze the behaviour of standard estimators.

Similarly, biases in $\hat{\rho}$ have to be interpreted with caution. A downward bias in $\hat{\rho}$ has an impact on the time unit (time deformation), and therefore a double effect diminishing the persistence and increasing the size of the shocks.

5.3 Spurious inference

The framework of near unit root/small sigma can induce spurious results when standard methods of statistical analysis are used. Table 1 gives summary statistics, including the sample mean, sample variance, 5% and 95% sample quantiles and first-order correlations for the trajectories given in Figures 1-2, with $\rho = 0.999$, $p_s = 1$ (white noise), $p_s = 0.5$ and $p_s = 0.95$. They are computed for $T = 250, 500, 1000, 2000, 5000$, i.e for time span of 1 year, 2 years, 4 years, 8 years, and 20 years (of opening days).

[Table 1 : Summary Statistics]

The first row of Tables 1.a-1.e corresponds to the white noise and as expected provides the following results : the sample mean tends to 0, the sample variance to 1, the quantiles to ± 1.64 and the autocorrelation to zero. The next row reports results for the equiweighted combination. We observe that the sample mean, sample variance, quantiles, and the first-order correlation do not seem to converge to their theoretical counterparts. In fact, they converge theoretically, but this convergence becomes visible for much larger number of observations. The last row of each table reports the results on large noise process. These results are similar to those in the first row corresponding to the white noise, except for $T = 5000$, where the effect of the L.R. component starts to be revealed.

Let us now consider the ACF of y_t for the series with $p_s = 0.5$ and $p_s = 0.95$.

[Insert Figure 7: ACF of Noisy Long Run Components]

The ACF is inside the standard confidence band up to $T = 5000$ for the large noise combination and up to $T = 2000$ for the equiweighted combinations except at a lag of order 25, where it is slightly significant. This explains why in practice the (weak) white noise hypothesis is not rejected (i.e. the market efficiency hypothesis in case of stock returns), despite that the true (y_t) is not a (weak) white noise. This spurious result is due to :

i) **A misleading interpretation of $\gamma(1)$.**

From equation (3.1) we have : $\gamma(1) = p_l \rho \simeq p_l$, if ρ is close to 1. Thus we estimate the weight p_l, which is small in the simulation. In other words the fact that $\hat{\gamma}(1)$ is small simply means that the L.R. component of y_t is small.

ii) **An inadequate null hypothesis.**

The standard test of $\rho(1) = 0$ is performed under a fixed probability of type I error for the null hypothesis $H_0 = \{y_t = y_{s,t}\} = \{y_{l,t} = 0\}$, whereas to get relevant long run predictions, we would need to avoid overrejecting the alternative $\tilde{H}_0 = H_0^c = \{y_{l,t} \neq 0\}$. Thus, when the interest is in long run predictions, the hypothesis of predictability has to be considered as the null hypothesis, not as the alternative hypothesis [compare with Hjalmarsson (2008), Section 5, Hjalmarsson (2011)].

This remark has important practical implications. Usually, when estimating an ARMA model for prediction purpose, there is a tendency to keep the model parsimonious and retain low autoregressive and moving average orders to avoid (in sample) overfitting. This is equivalent to underparametrize the model [see Hirano, Wright (2017) for a recent example of this practice in selecting prediction models]. In our L.R. setting it may be preferable to overparametrize the model in order to get some information on the hidden L.R. component. We suggest a slight overparametrization that would include nonsignificant parameters, without an extensive (in sample) overfitting. The slight overparametrization will improve the predictive power in the long run.

Moreover, some methods introduced to reduce the variance of forecasts are not robust to the presence of a small L.R. component, and become irrelevant. As an example, let us consider the bootstrap aggregating (bagging) [Breiman (1996), Efron (2014)]. If the ACF misleadingly indicates white noise, the bagging is invalid. In this case the data misleadingly considered as a white noise are resampled; then the forecasting methods are reapplied to the resampled data and the results are averaged over all the bootstrap samples. Clearly the resampling neglects the small L.R. component and destroys the persistence in the L.R. component. Therefore a bagging approach will increase the difficulty in revealing the L.R. component.

Next the ACF of the equiweighted combination for $T = 5000$ [see the South-East panel in Figure 7.a] has a pattern suggesting an integrated process, whereas our process is stationary with just a L.R. component.

Let us now consider the implementation of estimation methods of subsection 4.1. We provide in Table 2 the Yule-Walker estimates of ρ computed for the two noisy L.R. component series and for different values of T. Table 1e shows that the estimated first-order correlations are close to zero for T up to 2000 (as well as the second-order correlation not reported here). This explains the unexpected values of the Yule-Walker estimates of ρ. It can even be difficult to calculate $\hat{\rho}$ for $T = 250$ and large noise [see the X in Table 2].

[Table 2 : Estimates of ρ]

We also provide in Figure 8, functional estimates of the distribution of shocks for $T = 1000, T = 5000$ and the two noisy L.R. components. Since the true distributions are zero-mean Gaussian, we expect second-characteristic functions c_l, b_s, b_l, b_e to be real quadratic functions of u, and their derivatives to be real and linear in u. Thus, graphically we expect the real part for these derivatives to be a straight line in u and their imaginary part to be zero.

[Insert Figure 8 : Functional Estimates of the Distributions]

Tables 1-2 have revealed biases in the estimates of the mean, variance,..., and parameter ρ. Other biases can also exist in the nonparametric functional estimates of the second-characteristic functions. For the functional estimators, two types of biases can arise, which are a direct bias existing even if ρ is evaluated at its true value (close to 1), and an indirect bias due to the replacement of ρ by its estimate.

A priori, these biases can impact the shape of the second-characteristic function as well as its level, slope, or curvature. In this respect the Gaussian case is very special. Indeed let us consider the asymptotic behaviour of the sample second-characteristic function for ρ fixed and T large, which converges to :

$$\begin{aligned} \psi(u,v) &= \log E[\exp(iuy_t + ivy_{t-1})] \\ &= -\frac{1}{2}(u,v)\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}, \end{aligned}$$

which implies : $\dfrac{\partial \psi(u,v)}{\partial v} = -[\rho(1)u + v]$.

Therefore, when applying the estimation method of Section 5.1, with an estimator $\hat{\rho}$ of ρ that converges to ρ^* (not necessarily equal to ρ), we get the pseudo second-characteristic functions b_s^*, c_l^* such that :

$$\begin{aligned} \frac{db_s^*(u)}{du} &= \frac{\partial \psi(-u/\rho^*, u)}{\partial v} = -u\left[1 - \frac{\rho(1)}{\rho^*}\right], \\ \frac{dc_l^*(u)}{du} &= \frac{\partial \psi(0,u)}{\partial v} - \frac{\partial \psi}{\partial v}(-u/\rho^*, u) = -u\rho(1)/\rho^*. \end{aligned}$$

For instance, if the estimator $\hat{\rho}$ is consistent : $\rho^* = \rho$, we get :

$$\frac{db_s^*(u)}{du} = -u\left(1 - \frac{\rho(1)}{\rho}\right) = -up_s, \frac{dc_l^*(u)}{du} = -up_l.$$

This means that an inconsistent $\hat{\rho}$ has no visible impact on the shape of the imaginary and real components : the imaginary components remain equal to zero, and the real components linear in u. However, we can observe changes in the slopes of these linear functions.

It is more difficult to describe theoretically the direct bias, when ρ_T tends to 1 when T tends to infinity. In particular its impact on the shapes of the derivatives of the second characteristic functions is difficult to derive.

The plots in Figure 8 show that, for Gaussian processes, some biases in the shapes of second-order derivatives of the second-characteristic functions can be detected. For $T = 5000$ the biases affect the imaginary component of the noise. For $T = 2000$ the biases are more pronounced for all shapes. By considering the expansion of a second characteristic function in a neighbourhood[6] of $u = 0$, we see that the shape of a polynomial of degree 3 in Figure 8.a (x real), is capturing a kurtosis effect, and the parabolic shapes for the imaginary component a skewness effect. These are due to biases in the marginal and joint distributions of y_t, y_{t-1} [see the figures in Appendix 6] .

This section demonstrated that in the presence of a L.R. component the standard asymptotic theory does not apply even for a rather large number of observations ($T = 2000$, i.e. 8 years, $T = 5000$, i.e. 20 years). The next section provides a convenient framework for this non-standard analysis

[6]The expansion of the second-characteristic function is :

$$\psi(u) = \log E\exp(iuX) \simeq 1 + iuK_1 - \frac{u^2}{2}K_2 - \frac{iu^3}{6}K_3 + \frac{u^4}{24}K_4 + o(u^4),$$

where K_j denotes the cumulant of order j. Therefore the expansions of its real and imaginary part are :

$$Re\psi(u) = 1 - \frac{u^2}{2}K_2 + \frac{u^4}{24}K_4 + o(u^4),$$

$$Im\psi(u) = uK_1 - \frac{u^3}{6}K_3 + o(u^4).$$

The interpretations follow by considering the derivatives of these components with respect to u.

5.4 Near Unit Root/Small Sigma asymptotic inference

This section provides a heuristic description of near unit root/small sigma asymptotic inference. This is related to the literature on near unit root with fixed sigma [see e.g. Chan, Wei (1987), Phillips (1987)].

However, our approach is different and resembles the high frequency data (HFD) analysis from an underlying virtual diffusion model. In this framework it is known that we cannot expect to estimate consistently the drift of the diffusion equation from the observations of the diffusion over a finite interval [see e.g. Jiang, Knight (1997), Theorem 2, for high frequency data, and Banon (1978) for continuous observations]. This explains the asymptotic result below obtained for the estimator of coefficient ρ. Let us assume that the shocks $\varepsilon_{s,t}, \varepsilon_{l,t}$ are standard Gaussian.[7]

5.4.1 Asymptotic behaviour of the sample mean.

Proposition 2 : If the shocks $\varepsilon_{s,t}, \varepsilon_{l,t}$ are standard Gaussian :

i) the conditional distribution of y_{t+h} given $y_{s,t}, y_{l,t}$ is Gaussian with mean :

$$E\left(y_{t+h}|y_{s,t}, y_{l,t}\right) = \sqrt{p_l}\rho^h y_{l,t},$$

and variance :

$$V\left(y_{t+h}|y_{s,t}, y_{l,t}\right) = (1-p_l) + p_l(1-\rho^{2h}), h \geq 1.$$

ii) The conditional distribution of $\frac{1}{h}(y_{t+1} + \ldots + y_{t+h}) \equiv y_{t+1}(h)$ given $y_{s,t}, y_{l,t}$ is Gaussian, with mean :

$$E(y_{t+1}(h)|y_{s,t}, y_{l,t}) = \frac{1}{h}\sqrt{p_l}\rho\frac{1-\rho^h}{1-\rho}y_{l,t},$$

and variance :

[7] The asymptotic results can be extended to models with non Gaussian shocks by introducing conditions of weak convergence of partial sums to functionals of Brownian motion [see e.g. Davidson (1994), Berkes, Weber (2007), Magdalinos, Phillips (2007) for conventional regularity conditions]. In the Gaussian framework the analysis can be performed directly as shown in this section.

$$V(y_{t+1}(h)|y_{s,t}, y_{l,t}) = \frac{p_s}{h} + \frac{p_l(1+\rho)}{h(1-\rho)} - \frac{2(1+\rho)\rho p_l}{h^2(1-\rho)^2}(1-\rho^h) + \frac{p_l\rho^2}{h^2(1-\rho)^2}(1-\rho^{2h}).$$

Proof : See Appendix 4.

Let us now consider the case when h is large, ρ is close to 1, and p_l is held fixed. Since $\rho^h \simeq \exp(h \log \rho) \simeq \exp[-h(1-\rho)]$, the conditional mean is equivalent to :

$$E[y_{t+1}(h)|y_{s,t}, y_{l,t}] \simeq \sqrt{p_l}\frac{1-\exp(-h(1-\rho))}{h(1-\rho)}y_{l,t}.$$

The conditional variance is equivalent to :

$$V[y_{t+1}(h)|y_{s,t}, y_{l,t}]$$

$$\simeq \frac{2p_l}{h(1-\rho)} - \frac{4p_l}{h^2(1-\rho)^2}(1-\exp[-h(1-\rho)]) + \frac{p_l}{h^2(1-\rho)^2}[1-\exp(-2h(1-\rho))].$$

Therefore, the asymptotic behaviour of these conditional moments depends on the asymptotic behaviour of $h(1-\rho)$, when T increases.

Proposition 3 : If the shocks $\varepsilon_{s,t}, \varepsilon_{l,t}$ are standard Gaussian, we have the following equivalences for the first and second-order moments :

i) If $\lim_{T\to\infty} h(1-\rho) = 0$,

$$E[y_{t+1}(h)|y_{s,t}, y_{l,t}] \simeq \sqrt{p_l}y_{l,t},$$

$$V[y_{t+1}(h)|y_{s,t}, y_{l,t}] = o(1).$$

ii) If $\lim_{T\to\infty} h(1-\rho) = \gamma$, then

$$E[y_{t+1}(h)|y_{s,t}, y_{l,t}] = \sqrt{p_l}\frac{1-\exp(-\gamma)}{\gamma}y_{l,t}, +o(1)$$

$$V[y_{t+1}(h)|y_{s,t}, y_{l,t}] = \frac{2p_l}{\gamma} - \frac{4p_l}{\gamma^2}[1-\exp(-\gamma)] + \frac{p_l}{\gamma^2}[1-\exp(-2\gamma)] + o(1).$$

iii) If $\lim_{T\to\infty} h(1-\rho) = \infty$,

$$E[y_{t+1}(h)|y_{s,t}, y_{l,t}] \simeq \frac{\sqrt{p_l}}{h(1-\rho)} y_{l,t}, = o(1),$$

$$V[y_{t+1}(h)|y_{s,t}, y_{l,t}] \simeq \frac{2p_l}{h(1-\rho)} = o(1).$$

The asymptotic results of Proposition 3 show different behaviours when the forecasting horizon grows with the sample size (i.e. with ρ)[8].

i) If h increases slowly, the Law of Large Numbers can be applied to the short term component, whose average tends to zero, whereas the L.R. component is approximately constant and equal to $\sqrt{p_l} y_{l,t}$. Then $y_{t+1}(h)$ tends to a constant.

ii) If h increases at a speed that offsets the speed of convergence of ρ to 1, then $y_{t+1}(h)$ is conditionally Gaussian and nondegenerate.

Hence, for large h the sample average of y_t does not provide a consistent estimator of the theoretical mean of the process, which is equal to zero.

iii) If h increases very quickly, the conditional mean and variance tend to zero at equal rates.

Therefore, it is important to consider the entire term structure of predictions, rather than the short term predictions only [as in Hirano, Wright (2017) for instance].

Also note that, since in case ii) the estimator of the mean is not close to the true value, a bootstrap for evaluating the confidence interval on such a parameter does not satisfy the standard regularity conditions needed to approximate finite sample distribution and is misleading [see e.g. Bansal, Kiku, Yaron (2012), Section 4.1 for using bootstrap in a LRR model].

[8] See Valkanov (2003), Hjalmarsson (2008) for similar analysis in a framework of regression model.

5.4.2 Asymptotic behaviour of the sample autocorrelations

More asymptotic results can be derived from extensions of Donsker Theorem [see e.g. Berkes, Weber (2007)]. For expository purpose, we assume that the available daily observations on the L.R. component are [9] :

$$y_{l,t} = \sqrt{p_l}\tilde{x}_{l,t/T}, t = 1, \ldots, T, \tag{5.5}$$

where $(\tilde{x}_c)$ is a continuous Ornstein-Uhlenbeck process satisfying the diffusion equation :

$$d\tilde{x}_c = -k\tilde{x}_c dc + \sqrt{2k}\ d\tilde{W}_c,\ k > 0, \tag{5.6}$$

where the virtual time unit in diffusion equation (5.6) is infinitely large.

Since the frequency of observations is fixed, (the day), the observations are defined in (5.5) by specifying an initial time unit of T, such as $T = 250$ (opening days) for one year, $T = 25000$ for a century, $T = 250000$ for a millenium,.. In this framework, we have : $\rho_T = \exp(-k/T) \simeq 1 - k/T$, that tends to 1 when T tends to infinity and then $\sqrt{1-\rho_T^2} \simeq \sqrt{2k/T}$ [10].

Let us now reconsider Proposition 3 ii) with $h_T = T$ and $\lim_{T\to\infty} h_T(1 - \rho_T) = k$. We have :

$$y_{l,1}(T) = \frac{1}{T}(y_{l,1} + \ldots + y_{l,T}) = \frac{\sqrt{p_l}}{T}(\tilde{x}_{1/T} + \ldots + \tilde{x}_{T/T}).$$

This is a Riemann sum that convergences in distribution to the associated stochastic integral :[11]

$$y_{l,1}(T) \underset{T\to\infty}{\longrightarrow} \sqrt{p_l} \int_0^1 \tilde{x}_c dc. \tag{5.7}$$

The Ornstein-Uhlenbeck diffusion equation (5.6) can be solved in closed form. We have :

---

[9] A better notation would be $y_{l,t,T}$ to account for a triangular array. We omit index T in the notation for expository purpose.

[10] Compare with the conditions in Gospodinov (2009), Assumption B, Gospodinov et al. (2017), Assumption B. In our framework the constants c, λ of Assumption B are linked by $\lambda = \sqrt{2c}$ to get the interpretation in terms of time deformation of process $y_{l,t}/\sqrt{p_l}$. They become unconstrained when process $y_{l,t}$ is considered.

[11] This is the local asymptotic normality (LAN) set up in the Le Cam sense, where a sequence of "experiments" converges to a limiting Gaussian experiment [see Roussas, Bhattacharya (2011) for a discussion of LAN and its extensions.]

$$\tilde{x}_c = \tilde{x}_0 \exp(-kc) + \sqrt{2k} \int_0^c \exp[-k(c-s)] d\tilde{W}_s. \tag{5.8}$$

Next we deduce by integration :

$$\int_0^1 \tilde{x}_c dc = \tilde{x}_0 \frac{1-\exp(-k)}{k} + \sqrt{\frac{2}{k}} \int_0^1 (1-\exp[-k(1-s)]) d\tilde{W}_s. \tag{5.9}$$

The expressions of the conditional first and second-order moments in Proposition 3 ii) are easily derived from (4.9). For instance the variance is equal to :

$$V[y_{l,1}(T)|y_{l,1}] = \frac{2p_l}{k} \int_0^1 (1-\exp[-k(1-s)])^2 ds. \tag{5.10}$$

Let us now apply the HFD technique used above to derive also the asymptotic distribution of the sample autocorrelation. We have (see Appendix 5) :

Proposition 4 : Let us consider the sample autoregressive coefficient :

$\hat{\rho}_T(1) = \sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1.}^2$ Under (2.1)-(2.4) and (4.1)-(4.2),

$$\hat{\rho}_T(1) \underset{T\to\infty}{\longrightarrow} \frac{-\frac{p_l}{2}\tilde{x}_1^2 + p_l \int_0^1 \tilde{x}_c^2 dc}{p_s + p_l \int_0^1 \tilde{x}_c^2 dc},$$

where $\longrightarrow$ denotes the convergence in distribution.

For prediction purpose, the standard point prediction is deduced from the theoretical prediction by replacing ρ by $\hat{\rho}_T$. Very often the uncertainty on coefficient ρ is disregarded in the computation of the prediction interval by assuming implicitly that $\hat{\rho}_T$ is consistent. In our setting, it is not the case and even for large T there remains uncertainty on $\hat{\rho}_T$ and considerable caution should be exercised in interpreting evidence regarding long run predictions.

5.4.3 Finite sample properties

To give some insights on the (large) finite sample and limiting distributions of $\hat{\rho}_T(1)$ that depend on parameters ρ and p_l, we provide in Table 3, the probability that $\hat{\rho}_T(1)$ is outside the interval $[-2/\sqrt{T}, +2/\sqrt{T}]$ (corresponding to the standard bounds for the ACF), that is the probability of detecting the L.R. component. This exercice is similar to the exercise in Shephard, Harvey (1990), when they try to detect a small nonstationary deterministic component.

[Insert Table 3 : Probability of Detecting the L.R. component]

This probability is computed for sample sizes $T = 250$ (one year), $T = 500$ (2 years), $T = 1000$ (4 years). The values of parameter ρ are set to $\rho = 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 0.999$, and the value of p_l are $0.1, 0.2; 0.3, 0.4, 0.5$. By considering the ratio $\sqrt{p_s/p_l} = \sqrt{(1-p_l)/p_l}$, the short run component is three times the L.R. component for $p_l = 0.1$, twice the L.R. component for $p_l = 0.2$. The combination is equiweighted for $p_l = 0.5$. The probability of detecting the L.R. component can be very small. This probability diminishes when p_l decreases, or when ρ increases.

The finite sample confidence intervals at 95% for the first-order correlation are provided in Table 4 :

[Insert Table 4 : Confidence Interval for $\rho(1)$]

These confidence intervals (CI) differ significantly from the standard confidence intervals ($\pm 2/\sqrt{T}$). We observe both a drift effect, since the correct confidence interval are systematically centred at a higher level, and a variance effect, due to larger uncertainty on $\hat{\rho}(1)$, resulting in wider CI.

Table 5 shows the finite sample confidence intervals for the Yule-Walker estimator of ρ.

[Insert Table 5 : Confidence Interval for ρ]

The confidence intervals appear very wide, even for rather large number of observations. This is a consequence of the sample counterpart of formula (3.1) : $\hat{\rho} = \hat{\gamma}(2)/\hat{\gamma}(1)$, and the fact that $\hat{\gamma}(1), \hat{\gamma}(2)$ take random values close to zero.

5.5 Bayesian analysis

In a recent paper Bayesian methods are used by Schorfheide, Song, Yaron (2018) [12] to estimate a long run risk model under the parametric form (5.4). The difficulties encountered to derive reasonable asymptotic distributions of classical estimators arise also in the Bayesian framework.[13]

How to select a prior ? There is no clear rule except one : if a parameter θ has a unit and if we change the unit of θ to get θ^*, then the priors have to be changed accordingly. In model (4.4), with the interpretation of ρ as a change of time unit, this minimal coherency is expected. In other words, Bayesian asymptotic results similar to classical asymptotic results in Section 4.4 will be obtained, if the prior is introduced on the underlying parameter k of the associated diffusion. Then the prior on ρ_T will be deduced from the prior on k by applying the transformation $k \to \exp(-k/T)$, and will depend on T. This modifies the standard Bayesian asymptotics, since the prior now depends on T. It will be more concentrated in a neighbourhood of $\rho = 1$ to account for the asymptotic lack of identification [see the discussion in Poirier (1998)].

6 Concluding Remarks

A long run component has often been modelled as a near nonstationary process, as for instance a near unit root process with fixed sigma. This modelling approach creates trajectories with globally explosive trends, that are often not compatible with the expected long run behaviour of the macroeconomic or financial series of interest such as GDP growth, spot, or forward exchange rates, commodity prices. In this paper we explored the alternative of a stationary L.R. process with both near unit root and small sigma. It accounts for more "local" than "global" persistence effects. We explain how such a modelling of the L.R. component observed with noise can help solve some (long run) predictability puzzles mentioned in the literature.

The presence of a long run component implies that we cannot expect to estimate consistently the theoretical mean and the autoregressive coefficient of the long run dynamics. The asymptotic uncertainty on these estimated parameters can be derived and has to be taken into account when performing

[12] see also Stambaugh (1999).

[13] Chosen uniform on (-1, 1) in Schorfheide et al. (2018).

long run predictions or testing the hypothesis of market efficiency for long run adjusted portfolios.

While detecting and/or accommodating such long run component in the series of interest, standard impulse response analysis [see e.g. Bansal, Yaron (2004), Gourieroux, Jasiak (2019)], cointegration analysis [Ng, Perron (1997), Bansal, Dittmar, Kiku (2009), Gospodinov et al. (2017)], predictive regression with long run predictor [Torous, Valkanov (2000), Hjalmarsson (2008), (2011)] need to be modified and the pricing formulas for derivatives (futures and call options) with long time-to-maturity need to be adjusted. This also concerns the modelling of the long run term structure of interest rate. Since the long run prediction results are very fragile, this justifies robustness analysis based on competing scenarios (i.e. competing calibration of ρ) [see e.g. Bidder, Dew-Becker (2016)]. This also provides a new setting for fixing the associated technical standard for long run discounting by the supervisory Authority in Solvency 2, for knowing "how much (will) you pay to resolve long run risk" [Epstein et al. (2017)] and for pricing long run products such as life insurances, pensions, or the demographic and climate risks existing in the balance sheets of banks and insurance companies [see Hansen, Sargent (2010), Hansen, Scheinkman (2012), Gourieroux, Monfort, Renne (2019)]. Other types of L.R. processes can require an equivalent asymptotic analysis to robustify the long run prediction, as rare structural change [Hirano, Wright (2019)], or rare disasters [Wachter (2013), Barro, Liao (2016), Gourieroux et al. (2019)].

References

[1] Bandi, F., and B., Perron (2008) : "Long Run Risk-Return Trade Offs", Journal of Econometrics, 143, 349-374.

[2] Bandi, F., Perron, B., Tamoni, A., and C., Tebaldi (2018) : "The Scale of Predictability", Journal of Econometrics, 208, 120-140.

[3] Banon, G. (1978) : "Nonparametric Identification for Diffusion Processes", SIAM Journal of Control and Optimization, 16, 380-395.

[4] Bansal, R., Dittmar, R., and D., Kiku (2009) : "Cointegration and Consumption Risk in Equity Returns", Review of Financial Studies, 22, 1343-1375.

[5] Bansal, R., Kiku, D., and A., Yaron (2007) : "Risks For the Long Run : Estimation and Inference", DP Duke Univ.

[6] Bansal, R., Kiku, D., and A., Yaron (2010) : "Long Run Risks, the Macroeconomy and Asset Prices", American Economic Review, 100, 542-546.

[7] Bansal, R., Kiku, D., and A., Yaron (2012) : "Empirical Evaluation of the Long-Run Risks Model for Asset Prices", Critical Finance Review, 1, 183-221.

[8] Bansal, R., and A., Yaron (2004) : "Risk for the Long Run : A Potential Resolution of Asset Pricing Puzzle", Journal of Finance, 59, 1481-1509.

[9] Barro, R., and G., Liao (2016) : "Options Pricing Formula with Disaster Risk", NBER DP 21888.

[10] Beeler, J., and J., Campbell (2012) : "The Long Run Risks Model and Aggregate Asset Prices : An Empirical Assessment", Critical Finance Review, 1, 141-182.

[11] Berkes, I., and M., Weber (2007) : "On Complete Convergence of Triangular Arrays of Independent Random Variables", Stat. and Prob. Letters, 77, 952-965.

[12] Beveridge, S., and C., Nelson (1981) : "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle", Journal of Monetary Economics, 7, 151-174.

[13] Bidder, R., and I., Dew-Becker (2016) : "Long-Run Risk is the Worst Case Scenario", American Economic Review, 106, 2494-2527.

[14] Bonomo, M., Garcia, R., Meddahi, N., and R., Tedongap (2015) : "The Long and the Short of the Risk Return Trade-Off", Journal of Econometrics, 187, 580-592.

[15] Bradley, W., and W., Cook (2012) : "Two Proofs of the Existence and Uniqueness of Partial Fraction Decomposition", International Mathematical Forum, 7, 1517-1535.

[16] Breiman, L. (1996) : "Bagging Predictors", Machine Learning, 36, 105-139.

[17] Calvet, L., and A., Fisher (2007) : "Multifrequency News and Stock Returns", Journal of Financial Economics, 86, 178-212.

[18] Chan, N., and C., Wei (1987) : "Asymptotic Inference of Nearly Nonstationary AR(1) Processes", Annals of Statistics, 15, 1050-1063.

[19] Croce, M., Lettau, M., and S., Ludvigson (2014) : "Investor Information, Long Run Risk and the Term Structure of Equity", Review of Financial Studies, 28, 706-742.

[20] Davidson, J. (1994) : "Stochastic Limit Theory", Oxford Univ. Press.

[21] Deng, A. (2014) : "Understanding Spurious Regressions in Financial Economics", Journal of Financial Econometrics, 12, 122-150.

[22] Efron, B. (2014) : "Estimation and Accuracy After Model Selection", Journal of the American Statistical Association, 109, 991-1007.

[23] Epstein, L., Farhi, E., and T., Strzalecki (2017) : "How Much Would You Pay to Resolve Long Run Risk", American Economic Review, 104, 2680-2697.

[24] Fama, E., and K., French (1988)a : "Permanent and Temporary Components of Stock Prices", Journal of Political Economy, 96, 246-275.

[25] Fama, E., and K., French (1988)b : "Dividend Yields and Expected Stock Returns", Journal of Financial Economics, 25, 23-49.

[26] Fama, E., and K., French (1989) : "Business Conditions and Expected Returns on Stocks and Bonds", Journal of Financial Economics, 25, 23-49.

[27] Ferson, W., Sarkissian, S., and T., Simon (2003) : "Is Stock Return Predictability Spurious?", Journal of Investment Management, 1, 1-10.

[28] Gollier, C. (2016) : "Evaluation of Long Dated Assets : The Role of Parameter Uncertainty", Journal of Monetary Economics, 84, 66-89.

[29] Gospodinov, N. (2009) : "A New Look at the Forward Premium Puzzle", Journal of Financial Econometrics, 7, 312-338.

[30] Gospodinov, N., Maynard, A., and E., Pesavento (2017) : "Conditional Inference in Nearly Cointegrated Vector Error Correction Models and Small Signal-to-Noise Ratio", Fed. Reserve Atlanta, DP.

[31] Friedman, M. (1957) : "The Permanent Income Hypothesis : A Theory of the Consumption Function", Princeton University Press.

[32] Gourieroux, C., and J., Jasiak (2019) : "Dynamic Deconvolution of Independent Autoregressive Sources", CREST-DP.

[33] Gourieroux, C., and A., Monfort (1997) : "Time Series and Dynamic Models", Cambridge Univ. Press.

[34] Gourieroux, C., Monfort, A., Mouabbi, S., and J.P., Renne (2019) : "Disastrous Defaults", CREST DP.

[35] Gourieroux, C., Monfort, A., and J.P., Renne (2019) : "A New Modelling of Ultra Long Run Risk, with Application to the Term Structure of Interest Rates ", CREST-DP.

[36] Grammig, J., and E.M., Kuchlin (2018) : "A Two-Step Indirect Inference Approach to Estimate the Long-Run Risk Asset Pricing Model", Journal of Econometrics, 205, 6-33.

[37] Granger, C., and P., Newbold (1974) : "Spurious Regressions in Econometrics", Journal of Econometrics, 74, 111-120.

[38] Hansen, L., Heaton, J., and N., Li (2008) : "Consumption Strikes Back ? Measuring Long Run Risk", Journal of Political Economy, 116, 260-302.

[39] Hansen, L., and T., Sargent (2010) : "Fragile Beliefs and the Price of Uncertainty", Quantitative Economics, 1, 129-162.

[40] Hansen, L., and J., Scheinkman (2012) : "Pricing Growth Rate Risk", Finance and Stochastics, 16, 1-15.

[41] Hirano, K., and J., Wright (2017) : "Forecasting with Model Uncertainty : Representations and Risk Reduction", Econometrica, 85, 617-643.

[42] Hirano, K., and J., Wright (2019) : "Analyzing Cross-Validation for Forecasting with Structural Instability", Penn. State Univ. DP.

[43] Hjalmarsson, E. (2008) : "Inference in Long Horizon Regression", DP 853, Board of Governors of the Federal Reserve System.

[44] Hjalmarsson, E. (2011) : "New Methods for Inference in Long-Horizon Regressions", Journal of Financial and Quantitative Analysis 46, 815-839.

[45] Jiang, G., and J., Knight (1997) : "A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short Term Interest Rate Model", Econometric Theory, 13, 615-645.

[46] Kadane, J. (1971) : "Comparison of k-Class Estimators when the Disturbances are Small", Econometrica, 39, 729-737.

[47] Magdalinos, T., and P., Phillips (2007) : "Limit Theory for Moderate Deviations from Unity", Journal of Econometrics, 136, 115-130.

[48] Malloy, C., Moskowitz, T., and A., Vissing-Jorgensen (2009) : "Long-Run Stockholder Consumption Risk and Asset Returns", Journal of Finance, 6, 2427-2479.

[49] Maravall, A. (1979) : "Identification in Dynamic Shock-Error Model", Lecture Notes in Econometrics and Mathematical Systems, 165, Springer.

[50] Moon, R., Rubia, R., and R., Valkanov (2004) : "Long Horizon Regressions when the Predictor is Slowly Varying", DP Univ. Southern California.

[51] Nabeya, S., and P., Perron (1994) : "Local Asymptotic Distribution Related to the AR(1) Model with Dependent Errors", Journal of Econometrics, 62, 229-264.

[52] Ng, S., and P., Perron (1997) : "Estimation and Inference in Nearly Unbalanced Nearly Cointegrated Systems", Journal of Econometrics, 79, 53-81.

[53] Ortu, F., Tamoni, A., and C., Tebaldi (2013) : "Long Run Risk and the Persistence of Consumption Shocks", Review of Financial Studies, 26, 2876-2915.

[54] Phillips, P. (1979) : "The Sampling Distribution of Forecasts from a First-Order Autoregression", Journal of Econometrics, 9, 241-261.

[55] Phillips, P. (1986) : "Understanding Spurious Regressions in Econometrics", Journal of Econometrics, 33, 311-340.

[56] Phillips, P. (1987) : "Towards a Unified Asymptotic Theory for Autoregression", Biometrika, 74, 535-547.

[57] Phillips, P., (1998) : "New Tools for Understanding Spurious Regressions", Econometrica, 66, 1299-1326.

[58] Pohl, W., Schmedders, K., and O., Wilms (2018) : "Asset Pricing with Heterogenous Agents and Long-Run Risk", DP Tilburg Univ.

[59] Poirier, D. (1998) : "Revising Beliefs in Nonidentified Models", Econometric Theory, 14, 483-509.

[60] Roussas, G., and D., Bhattacharya (2011) : "Revisiting Local Asymptotic Normality (LAN) and Passing on to Local Asymptotic Mixed Normality (LAMN) and Local Asymptotic Quadratic (LAQ) Experiments", in Advances in Directional and Linear Statistics, M., Wells, and A., Sengupta, eds., 253-280, Physica-Verlag, Springer, Heidelberg.

[61] Schorfheide, F., Song, D., and A., Yaron (2018) : "Identifying Long-Run Risks : A Bayesian Mixed-Frequency Approach", Econometrica, 86, 617-654.

[62] Shephard, N., and A., Harvey (1990) : "On the Probability of Estimating a Deterministic Component in the Local Level Model", Journal of Time Series Analysis, 11, 339-347.

[63] Stambaugh, R. (1999) : "Predictive Regressions", Journal of Financial Economics, 54, 375-421.

[64] Torous, W., and R., Valkanov (2000) : "Boundaries of Predictability: Noisy Predictive Regressions", DP UCLA.

[65] Valkanov, R. (2003) : "Long Horizon Regressions : Theoretical Results and Applications", The Journal of Financial Economics, 68, 201-232.

[66] Vasicek, O. (1977) : "An Equilibrium Characterization of the Term Structure", Journal of Financial Economics, 5, 177-184.

[67] Wachter, J. (2013) : "Can Time Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility ?, Journal of Finance, 68, 987-1035.

[68] Yule, G. (1926) : "Why Do We Sometimes Get Nonsense-Correlations Between Time Series ? A Study in Sampling and the Nature of Time Series", Presidential Address, Journal of Royal Statistical Society, 89, 1-69.

Appendix 1

Proof of Proposition 1

We have :

$$
\begin{aligned}
& E[\exp(iuy_t + ivy_{t-1})] \\
= \; & E[\exp(iuy_{s,t} + ivy_{s,t-1})]E[\exp(iuy_{l,t} + ivy_{l,t-1})] \\
= \; & E[\exp(iuy_{s,t})]E[\exp(ivy_{s,t-1})] \\
& E[\exp[i(u\rho + v)y_{l,t-1}]E[\exp(iu\tilde{\varepsilon}_{l,t})].
\end{aligned}
$$

By taking the logarithm of both sides of the equality, we deduce that :

$$
\begin{aligned}
\psi(u,v) &= c_l(u\rho + v) + b_l(u) + b_s(u) + b_s(v) \\
&= c_l(u\rho + v) + c_l(u) - c_l(\rho u) + b_s(u) + b_s(v),
\end{aligned} \tag{a.1}
$$

by Lemma 1.

Therefore we get :

$$
\frac{\partial\psi(u,v)}{\partial v} = \frac{dc_l}{du}(u\rho + v) + \frac{db_s}{du}(v), \forall u, v. \tag{a.2}
$$

Let us now choose appropriate values for u and v.

i) If $u = 0$, we get :

$$
\frac{\partial\psi(0,v)}{\partial v} = \frac{dc_l}{du}(v) + \frac{db_s}{du}(v). \tag{a.3}
$$

ii) If $u = -v/\rho$, we have :

$$
\begin{aligned}
\frac{\partial\psi(-v/\rho,v)}{\partial v} &= \frac{dc_l}{du}(0) + \frac{db_s(v)}{dv} \\
&= \frac{db_s(v)}{dv},
\end{aligned} \tag{a.4}
$$

since $\dfrac{dc_l}{du}(0) = E(y_{l,t}) = 0$.

Proposition 1 follows from (a.3)-(a.4).

Appendix 2

Theoretical Regressions

In this appendix, we consider the general framework of variables Y_t related to sources X_t by $Y_t = AX_t$, where the K sources are independent autoregressive processes:

$$x_{jt} = \rho_j x_{j,t-1} + \varepsilon_{j,t}, j = 1, \ldots, J,$$

with $E\varepsilon_{jt} = 0, V\varepsilon_{jt} = 1 - \rho_j^2$. Then we consider the regression :

$$y_{1,t+1,t+h} = \beta_{hk} y_{2,t-k+1} + w_{hkt},$$

with $y_{1,t} = \sum_{j=1}^{J} a_{1j} x_{jt}, y_{2t} = \sum_{j=1}^{J} a_{2j} x_{jt}$.

The formulas of Section 4 are direct consequences of the results below.

2.1 Expressions of regression coefficients

Proposition A.1: The theoretical regression coefficient is :

$$\beta_{hk} = \frac{\sum_{j=1}^{K} a_{1j} a_{2j} \gamma(h, k, \rho_j)}{\sum_{j=1}^{K} a_{2j}^2 \gamma(h, \rho_j)},$$

and the theoretical correlation between the dependent and explanatory variables is:

$$R_{hk} = \frac{\sum_{j=1}^{K} a_{1j} a_{2j} \gamma(h, k, \rho_j)}{\left[\sum_{j=1}^{K} a_{1j}^2 \gamma(h, \rho_j)\right]^{1/2} \left[\sum_{j=1}^{K} a_{2j}^2 \gamma(h, \rho_j)\right]^{1/2}},$$

where :

$$\gamma(h, k, \rho) = \frac{\rho(1-\rho^k)(1-\rho^h)}{(1-\rho)^2},$$

$$\gamma(h, \rho) = \frac{1+\rho}{1-\rho} h - \frac{2\rho}{(1-\rho)^2}(1-\rho^h).$$

Proof:

a) **Expression of** $\gamma(h, k; \rho)$

The covariance matrix between $(x_t, x_{t-1}, \ldots, x_{t-k+1})'$ and $(x_{t+1}, x_{t+2}, \ldots x_{t+h})$ is equal to :

$$\Gamma(k, h) = \begin{pmatrix} \rho & \rho^2 & \ldots & \rho^h \\ \rho^2 & \rho^3 & & \rho^{h+1} \\ \vdots & & & \vdots \\ \rho^k & \rho^{k+1} & & \rho^{k+h} \end{pmatrix}.$$

We have : $\gamma(h, k; \rho) = e'_k \Gamma(k, h) e_h$, where e_h is the h-dimensional vector with unitary components. Therefore :

$$\begin{aligned} \gamma(h, k; \rho) &= [\rho + \rho^2 + \ldots + \rho^k] + \ldots + [\rho^h + \rho^{h+1} + \ldots + \rho^{h+k}] \\ &= [1 + \rho + \ldots + \rho^{k-1}][\rho + \ldots + \rho^h] \\ &= \rho \frac{(1-\rho^k)(1-\rho^h)}{(1-\rho)^2}. \end{aligned}$$

b) **Expression of** $\gamma(h; \rho)$

The variance-covariance matrix of $(x_{t+1}, \ldots, x_{t+h})'$ is :

$$
\Gamma(h) = \begin{bmatrix} 1 & \rho & \dots & \rho^{h-1} \\ \rho & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho^{h-1} & \dots & & 1 \end{bmatrix}.
$$

Therefore :

$$
\begin{aligned}
\gamma(h;\rho) &= h + 2\left\{(\rho + \ldots + \rho^{h-1}) + (\rho + \ldots + \rho^{h-2}) + \ldots + \rho\right\} \\
&= h + 2\rho\left\{\frac{1-\rho^{h-1}}{1-\rho} + \frac{1-\rho^{h-2}}{1-\rho} + \ldots + \frac{1-\rho}{1-\rho}\right\} \\
&= h + \frac{2\rho}{1-\rho}\left[h - 1 - [\rho + \rho^2 + \ldots + \rho^{h-1}]\right] \\
&= \frac{1+\rho}{1-\rho}h - \frac{2\rho}{1-\rho}\left[1 + \frac{\rho}{1-\rho}(1-\rho^{h-1})\right] \\
&= \frac{1+\rho}{1-\rho}h - \frac{2\rho}{(1-\rho)^2}[1-\rho^h].
\end{aligned}
$$

QED

We are interested in the patterns of β_{hk}, R_{kk}, for varying h, k. The corollary below follows directly from Proposition A.1.

Corollary A.1 : $\lim_{h,k\to\infty} \beta_{hk} = \lim_{h,k\to\infty} R_{hk} = 0$.

Proof: This is due to the term in $\gamma(h,\rho)$, which is linear in h. In particular, the convergence to 0 is at a hyperbolic speed (not geometric), and therefore rather slow.

2.2) **Close to unit root behaviour**

When $\rho = 1 - \delta/h, \delta > 0$, we get : $1 - \rho^h = 1 - \exp[h\log(1-\delta/h)] \sim 1 - \exp(-\delta)$, for h large. It follows that :

$$
\begin{aligned}
R(h, 1-\delta/h, \sigma^2) &\sim \frac{(1-\exp(-\delta))^2}{\delta^2} h^2 / \left[\frac{2h^2}{\delta} + h\sigma^2 - 2(1-\exp(-\delta)) \frac{h^2}{\delta^2} \right] \\
&\sim \frac{(1-\exp(-\delta))^2}{\delta^2} / \left\{ \frac{2}{\delta} - 2\frac{(1-\exp(-\delta))}{\delta^2} \right\} \\
&= \frac{(1-\exp(-\delta))^2}{2[\exp(-\delta) - 1 + \delta]} \equiv R_\infty(\delta), \text{ for } h \to \infty.
\end{aligned}
$$

Lemma : The function $R_\infty(\delta)$ is a decreasing function of δ, such that: $R_\infty(0) = 1, R_\infty(\infty) = 0.$

Proof : $\dfrac{dR_\infty(\delta)}{d\delta}$ has the same sign as the function :

$$
\begin{aligned}
a(\delta) &= 2(1-\exp(-\delta))[\exp(-\delta) - 1 + \delta] - (1-\exp(-\delta))^2[-\exp(-\delta) + 1] \\
&= (1-\exp(-\delta))\{2[\exp(-\delta) - 1 + \delta] - (1-\exp(-\delta))^2] \\
&= [1-\exp(-\delta)]2[\exp(-\delta) - 1 + \delta](1 - R_\infty(\delta)),
\end{aligned}
$$

and it is easy to see that the second factor on the right hand side is always negative.

QED

Appendix 3

Real and Imaginary Part of $\dfrac{\partial \hat{\psi}}{\partial v}(u,v)$

We have from (4.2) :

$$
\frac{\partial \hat{\psi}}{\partial v}(u,v) = \frac{-\Sigma_t[y_{t-1}\sin(uy_t + vy_{t-1})] + i\Sigma_t[y_{t-1}\cos(uy_t + vy_{t-1})]}{\Sigma_t \cos(uy_t + vy_{t-1}) + i\Sigma_t \sin(uy_t + vy_{t-1})}.
$$

By multiplying the numerator and denominator by the conjugate of the denominator, we get.

$$\widehat{\frac{\partial \psi}{\partial v}}(u,v) = \frac{1}{\Delta}\{-\Sigma_t[y_{t-1}\sin(uy_t+vy_{t-1})]+i\Sigma_t[y_{t-1}\cos(uy_t+vy_{t-1})]\}$$
$$\{\Sigma_t \cos(uy_t+vy_{t-1}) - i\Sigma_t \sin(uy_t+vy_{t-1})\},$$

where :

$$\Delta = [\Sigma_t \cos(uy_t+vy_{t-1})]^2 + [\Sigma_t \sin(uy_t+vy_{t-1})]^2. \quad \text{(a.5)}$$

We deduce :

$$\begin{aligned} Re\frac{\partial \hat{\psi}}{\partial v}(u,v) &= \frac{1}{\Delta}\{\Sigma_t[y_{t-1}\cos(uy_t+vy_{t-1})]\Sigma_t[\sin(uy_t+vy_{t-1})] \\ &\quad -\Sigma_t[y_{t-1}\sin(uy_t+vy_{t-1})]\Sigma_t[\cos(uy_t+vy_{t-1})]\}, \quad \text{(a.6)} \end{aligned}$$

$$\begin{aligned} Im\frac{\partial \hat{\psi}}{\partial v}(u,v) &= \frac{1}{\Delta}\{\Sigma_t[y_{t-1}\sin(uy_t+vy_{t-1})]\Sigma_t[\sin(uy_t+vy_{t-1})] \\ &\quad +\Sigma_t[y_{t-1}\cos(uy_t+vy_{t-1})]\Sigma_t[\cos(uy_t+vy_{t-1})]\}. \quad \text{(a.7)} \end{aligned}$$

Appendix 4

Proof of Proposition 2

We have : $y_{t+1}(h) = y_{s,t+1}(h) + y_{l,t+1}(h)$,

i) Conditional mean

We get :

$$E[y_{t+1}(h)|y_{s,t}, y_{l,t}] = E[y_{s,t+1}(h)|y_{s,t}] + E[y_{l,t+1}(h)|y_{l,t}]$$

$$= \frac{1}{h}\sqrt{p_l}[\rho + \ldots + \rho^h]y_{l,t}$$

$$= \frac{1}{h}\sqrt{p_l}\rho\frac{1-\rho^h}{1-\rho}y_{l,t}.$$

ii) Conditional variance

We have

$$
\begin{aligned}
& V[y_{t+1}(h)|y_{s,t}, y_{l,t}] \\
= \ & V[y_{s,t+1}(h)|y_{s,t}] + V[y_{l,t+1}(h)|y_{l,t}] \\
= \ & \frac{1}{h}p_s + p_l V[x_{l,t+1}(h)|y_{l,t}] \\
= \ & \frac{1}{h}p_s + \frac{1}{h^2}p_l(1-\rho^2)V[(1+\rho+\ldots+\rho^{h-1})\varepsilon_{l,t+1} + (1+\rho+\ldots+\rho^{h-2})\varepsilon_{l,t+2} \\
& +\ldots+\varepsilon_{l,t+h}] \\
= \ & \frac{1}{h}p_s + \frac{1}{h^2}p_l(1-\rho^2)\left\{\frac{(1-\rho^h)^2}{(1-\rho)^2} + \frac{(1-\rho^{h-1})^2}{(1-\rho)^2} + \ldots \frac{(1-\rho^2)}{(1-\rho)^2}\right\} \\
= \ & \frac{1}{h}p_s + \frac{1}{h^2}p_l\frac{1+\rho}{1-\rho}\sum_{k=1}^{h}[1-2\rho^k+\rho^{2k}] \\
= \ & \frac{1}{h}p_s + \frac{1}{h^2}p_l\frac{1+\rho}{1-\rho}\left\{h - 2\rho\frac{1-\rho^h}{1-\rho} + \rho^2\frac{1-\rho^{2h}}{1-\rho^2}\right\} \\
= \ & \frac{p_s}{h} + \frac{p_l(1+\rho)}{h(1-\rho)} - \frac{2(1+\rho)\rho p_l}{h^2(1-\rho)^2}(1-\rho^h) + \frac{p_l\rho^2}{h^2(1-\rho)^2}(1-\rho^{2h}).
\end{aligned}
$$

Appendix 5

Proof of Proposition 4

a) Let us first consider the denominator. We have :

$$\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2 = \frac{1}{T}\sum_{t=2}^{T} y_{s,t-1}^2 + \frac{2}{T}\sum_{t=2}^{T} y_{s,t-1}y_{l,t-1} + \frac{1}{T}\sum_{t=2}^{T} y_{l,t-1}^2.$$

i) By the LLN, we get : $\frac{1}{T}\sum_{t=2}^{T} y_{s,t-1}^2 \to p_s$.

ii) $\frac{1}{T}\sum_{t=2}^{T} y_{s,t-1}y_{l,t-1} \to 0$, since $E(y_{s,t}y_{l,t}) = Ey_{s,t}Ey_{l,t} = 0$ and the series is normally convergent :
$\frac{1}{T}\sum_{t=1}^{T} E|y_{s,t-1}y_{l,t-1}| = \frac{1}{T}\sum_{t=1}^{T}(E|y_{s,t-1}|E|y_{l,t-1}|) = E[y_{s,t-1}|E|y_{l,t-1}| < \infty$,
by stationarity.

iii) The last term is :

$$\frac{1}{T}\sum_{t=1}^{T-1} y_{l,t}^2 = \frac{p_l}{T}\sum_{t=1}^{T-1} \tilde{x}_{t/T}^2 \longrightarrow p_l \int_0^1 \tilde{x}_c^2 dc,$$

by the convergence of Riemann sum.

b) Similarly we can decompose the numerator as :

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T} y_t y_{t-1} &= \frac{1}{T}\sum_{t=2}^{T} y_{s,t}y_{s,t-1} + \frac{1}{T}\sum_{t=2}^{T} y_{s,t}y_{l,t-1} + \frac{1}{T}\sum_{t=2}^{T} y_{l,t}y_{s,t-1} \\ &+ \frac{1}{T}\sum_{t=2}^{T} y_{l,t}y_{l,t-1}.\end{aligned}$$

It is easily checked that the three first terms tend to zero. Let us now consider the last term. We have :

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T} y_{l,t}y_{l,t-1} &= \frac{1}{T}\sum_{t=2}^{T} y_{l,t-1}^2 + \frac{1}{T}\sum_{t=2}^{T}[y_{l,t-1}(y_{l,t} - y_{l,t-1})] \\ &= p_l\left[\frac{1}{T}\sum_{t=2}^{T} \tilde{x}_{t/T}^2 + \frac{1}{T}\sum_{t=2}^{T}[\tilde{x}_{t-1/T}(\tilde{x}_{t/T} - \tilde{x}_{(t-1)/T}))\right].\end{aligned}$$

This quantity converges in distribution to :

$$
\begin{aligned}
& p_l\left[\int_0^1 \tilde{x}_c^2 dc + \int_0^1 \tilde{x}_c d\tilde{x}_c\right] \\
= \quad & p_l(\int_0^1 \tilde{x}_c^2 dc + \frac{1}{2}\tilde{x}_1^2).
\end{aligned}
$$

The result follows.

Appendix 6

Marginal and Joint Distributions of y_t, y_{t-1}

Figure 9a : Marginal Distribution, Equiweighted
$(T = 2000,\ \text{dot line } T = 5000,\ \text{black line}))$

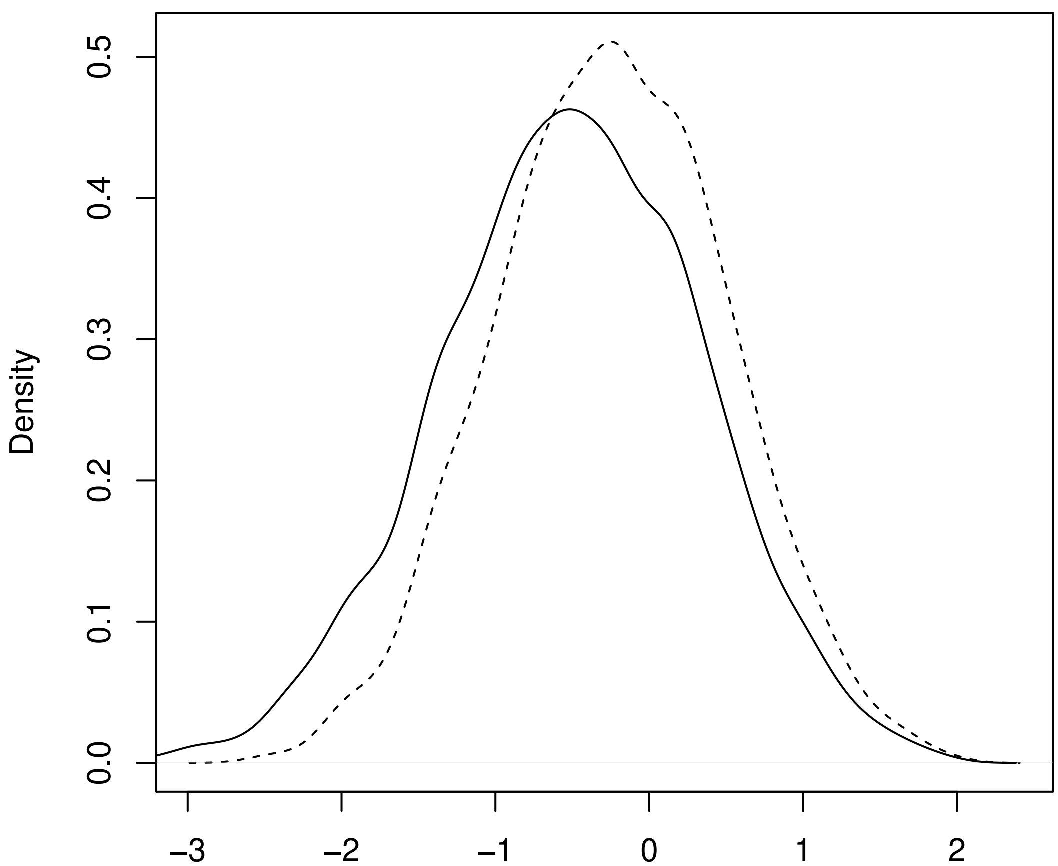

Figure 9b : Marginal Distribution, Large Noise
$(T = 2000,\ \text{dot line}\ T = 5000,\ \text{black line}))$

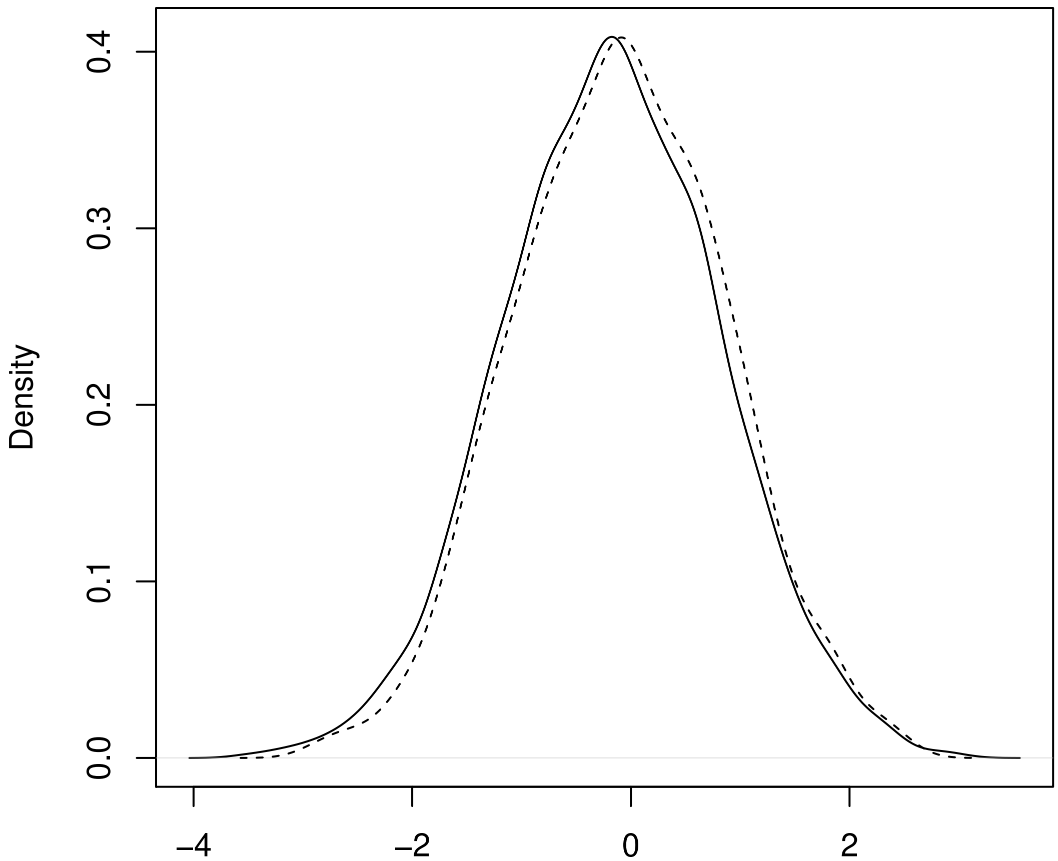

Figure 10 : Joint Distribution of y_t, y_{t-1}
(equiweighted : series 1, large noise : series 2, $(T = 2000, 5000)$)

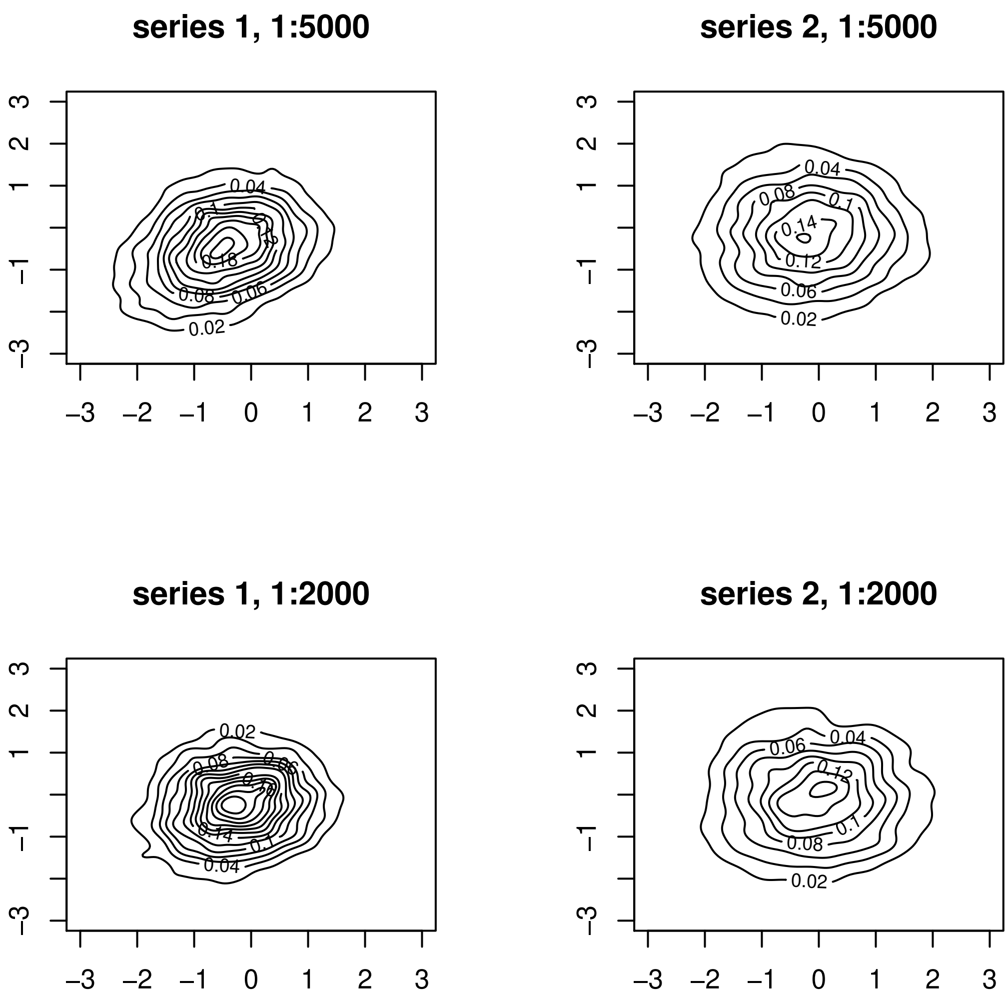

Table 1a : Summary Statistics : Sample Mean

number of observations	250	500	1000	2000	5000
white noise	-0.026	-0.017	-0.010	-0.016	-0.017
equiweighted	-0.027	-0.036	-0.109	-0.244	-0.510
large noise	-0.028	-0.024	-0.042	-0.089	-0.174

Table 1b : Summary Statistics : Sample Variance

number of observations	250	500	1000	2000	5000
white noise	1.030	0.978	0.996	0.944	0.992
equiweighted	0.563	0.533	0.548	0.557	0.716
large noise	0.993	0.942	0.959	0.913	0.975

Table 1c : Summary Statistics : 5%-quantile

number of observations	250	500	1000	2000	5000
white noise	-1.723	-1.627	-1.560	-1.560	-1.630
equiweighted	-1.365	-1.229	-1.303	-1.455	-1.956
large noise	-1.737	-1.629	-1.604	-1.641	-1.765

Table 1d : Summary Statistics : 95%-Quantile

number of observations	250	500	1000	2000	5000
white noise	1.499	1.545	1.641	1.572	1.628
equiweighted	1.172	1.147	1.108	0.992	0.864
large noise	1.477	1.527	1.568	1.494	1.460

Table 1e : Summary Statistics : First-Order Correlation

number of observations	250	500	1000	2000	5000
white noise	-0.004	-0.031	-0.027	0.011	-0.002
equiweighted	0.038	0.011	0.024	0.124	0.274
large noise	0.000	-0.028	-0.024	0.019	0.021

Table 2 : Estimation of ρ

number of observations	250	500	1000	2000	5000
equiweighted	0.105	2.727	2.500	1.008	1.180
large noise	X	0.250	-0.625	1.105	1.333

Table 3 : Probability of Detecting the L.R. Component

T= 250

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	23.50	66.00	92.75	98.75	100.00
0.875	24.25	63.25	92.75	99.50	100.00
0.9	20.50	66.50	90.50	99.00	100.00
0.925	26.00	65.25	89.25	98.50	99.75
0.95	22.25	62.50	88.50	97.25	99.50
0.975	17.50	53.75	77.00	92.75	97.50
0.999	6.50	6.00	7.50	14.00	20.75

T= 500

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	43.50	91.50	99.75	100.00	100
0.875	47.50	90.50	99.75	100.00	100
0.9	42.75	92.50	100.00	100.00	100
0.925	46.25	92.25	99.75	100.00	100
0.95	46.00	90.50	99.25	100.00	100
0.975	40.00	85.75	97.75	100.00	100
0.999	6.00	9.50	22.00	40.25	55

T= 1000

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	72.00	99.50	100	100.0	100.00
0.875	70.00	99.25	100	100.0	100.00
0.9	74.50	99.50	100	100.0	100.00
0.925	75.50	100.00	100	100.0	100.00
0.95	76.75	100.00	100	100.0	100.00
0.975	73.00	99.00	100	100.0	100.00
0.999	14.25	36.50	59	78.5	90.25

Table 4.a : Confidence Interval for $\hat{\rho}(1), T = 250$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$-0.058, 0.205$	$0.005, 0.288$	$0.086, 0.394$	$0.162, 0.468$	$0.162, 0.468$
0.875	$-0.049, 0.197$	$0.005, 0.303$	$0.085, 0.385$	$0.164, 0.470$	$0.230, 0.555$
0.9	$-0.054, 0.215$	$-0.001, 0.312$	$0.060, 0.404$	$0.153, 0.490$	$0.250, 0.589$
0.925	$-0.078, 0.204$	$0.009, 0.323$	$0.068, 0.410$	$0.135, 0.485$	$0.232, 0.596$
0.95	$-0.068, 0.209$	$0.006, 0.297$	$0.047, 0.430$	$0.113, 0.530$	$0.189, 0.590$
0.975	$-0.076, 0.213$	$-0.026, 0.314$	$0.020, 0.397$	$0.061, 0.504$	$0.125, 0.637$
0.999	$-0.139, 0.129$	$-0.119, 0.136$	$-0.101, 0.151$	$-0.106, 0.177$	$-0.079, 0.240$

Table 4.b : Confidence Interval for $\hat{\rho}(1), T = 500$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$-0.010, 0.167$	$0.057, 0.257$	$0.139, 0.349$	$0.207, 0.424$	$0.305, 0.528$
0.875	$-0.011, 0.182$	$0.049, 0.267$	$0.142, 0.360$	$0.211, 0.456$	$0.211, 0.456$
0.9	$-0.011, 0.179$	$0.050, 0.279$	$0.139, 0.372$	$0.213, 0.465$	$0.302, 0.551$
0.925	$-0.011, 0.182$	$0.065, 0.271$	$0.130, 0.369$	$0.204, 0.466$	$0.303, 0.576$
0.95	$-0.029, 0.194$	$0.053, 0.286$	$0.128, 0.398$	$0.199, 0.512$	$0.283, 0.598$
0.975	$-0.024, 0.194$	$0.042, 0.325$	$0.097, 0.399$	$0.144, 0.525$	$0.224, 0.631$
0.999	$-0.083, 0.097$	$-0.066, 0.126$	$-0.052, 0.182$	$-0.047, 0.257$	$-0.022, 0.308$

Table 4.c : Confidence Interval for $\hat{\rho}(1), T = 1000$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$0.020, 0.150$	$0.096, 0.236$	$0.165, 0.328$	$0.257, 0.405$	$0.346, 0.498$
0.875	$0.023, 0.155$	$0.083, 0.248$	$0.179, 0.343$	$0.250, 0.423$	$0.342, 0.509$
0.9	$0.016, 0.157$	$0.091, 0.246$	$0.182, 0.349$	$0.255, 0.438$	$0.352, 0.521$
0.925	$0.014, 0.157$	$0.088, 0.260$	$0.170, 0.354$	$0.257, 0.457$	$0.348, 0.547$
0.95	$0.017, 0.158$	$0.093, 0.272$	$0.173, 0.374$	$0.253, 0.472$	$0.340, 0.556$
0.975	$0.013, 0.167$	$0.082, 0.279$	$0.152, 0.386$	$0.222, 0.501$	$0.292, 0.612$
0.999	$-0.057, 0.106$	$-0.027, 0.150$	$-0.025, 0.238$	$0.011, 0.309$	$0.030, 0.395$

Table 5.a : Confidence Interval for $\hat{\rho}, T = 250$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$[-9.481, 6.739]$	$[-0.174, 4.393]$	$[0.272, 1.572]$	$[0.449, 1.258]$	$[0.559, 1.128]$
0.875	$[-8.443, 6.594]$	$[-0.649, 5.023]$	$[0.294, 1.738]$	$[0.444, 1.366]$	$[0.519, 1.139]$
0.90	$[-4.774, 9.938]$	$[-0.608, 4.531]$	$[0.350, 1.839]$	$[0.531, 1.455]$	$[0.536, 1.128]$
0.925	$[-16.306, 6.225]$	$[-0.379, 3.277]$	$[0.334, 1.983]$	$[0.476, 1.486]$	$[0.574, 1.192]$
0.95	$[-8.663, 6.123]$	$[-0.574, 3.718]$	$[0.234, 2.369]$	$[0.547, 1.511]$	$[0.680, 1.342]$
0.975	$[-9.305, 11.923]$	$[-3.916, 8.978]$	$[-0.195, 2.640]$	$[0.321, 2.481]$	$[0.553, 1.555]$
0.999	$[-11.724, 7.754]$	$[-17.211, 16.444]$	$[-9.454, 9.727]$	$[-13.367, 11.458]$	$[-22.019, 9.758]$

Table 5.b : Confidence Interval for $\hat{\rho}, T = 500$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$[-1.127, 6.404]$	$[0.244, 1.859]$	$[0.519, 1.332]$	$[0.559, 1.127]$	$[0.646, 1.024]$
0.875	$[-2.801, 10.940]$	$[0.320, 2.511]$	$[0.504, 1.279]$	$[0.582, 1.154]$	$[0.676, 1.048]$
0.9	$[-3.190, 8.188]$	$[0.336, 2.157]$	$[0.533, 1.348]$	$[0.650, 1.150]$	$[0.685, 1.058]$
0.925	$[-7.091, 8.991]$	$[0.373, 2.015]$	$[0.583, 1.336]$	$[0.675, 1.192]$	$[0.686, 1.077]$
0.95	$[-3.496, 6.863]$	$[0.340, 1.919]$	$[0.594, 1.480]$	$[0.720, 1.224]$	$[0.774, 1.143]$
0.975	$[-10.320, 6.803]$	$[0.295, 2.451]$	$[0.487, 1.551]$	$[0.659, 1.298]$	$[0.774, 1.177]$
0.999	$[-12.403, 15.602]$	$[-11.252, 8.338]$	$[-17.463, 10.186]$	$[-3.906, 6.869]$	$[-8.299, 4.300]$

Table 5.c : Confidence Interval for $\hat{\rho}, T = 1000$

ρ/p_l	0.1	0.2	0.3	0.4	0.5
0.85	$[-0.071, 3.516]$	$[0.465, 1.487]$	$[0.612, 1.106]$	$[0.667, 1.023]$	$[0.716, 0.961]$
0.875	$[0.064, 3.034]$	$[0.518, 1.530]$	$[0.612, 1.156]$	$[0.712, 1.084]$	$[0.736, 0.979]$
0.90	$[0.055, 3.578]$	$[0.522, 1.444]$	$[0.673, 1.184]$	$[0.730, 1.053]$	$[0.769, 1.005]$
0.925	$[0.099, 3.362]$	$[0.512, 1.464]$	$[0.700, 1.184]$	$[0.770, 1.100]$	$[0.804, 1.039]$
0.95	$[0.041, 3.037]$	$[0.526, 1.464]$	$[0.705, 1.215]$	$[0.767, 1.100]$	$[0.822, 1.058]$
0.975	$[0.058, 2.772]$	$[0.582, 1.545]$	$[0.716, 1.317]$	$[0.823, 1.176]$	$[0.852, 1.087]$
0.999	$[-8.570, 13.681]$	$[-8.832, 6.932]$	$[-3.405, 3.895]$	$[-0.014, 2.862]$	$[0.336, 2.202]$

Figure 1a : Trajectories of Gaussian AR(1) process

(small ρ, in increasing order)

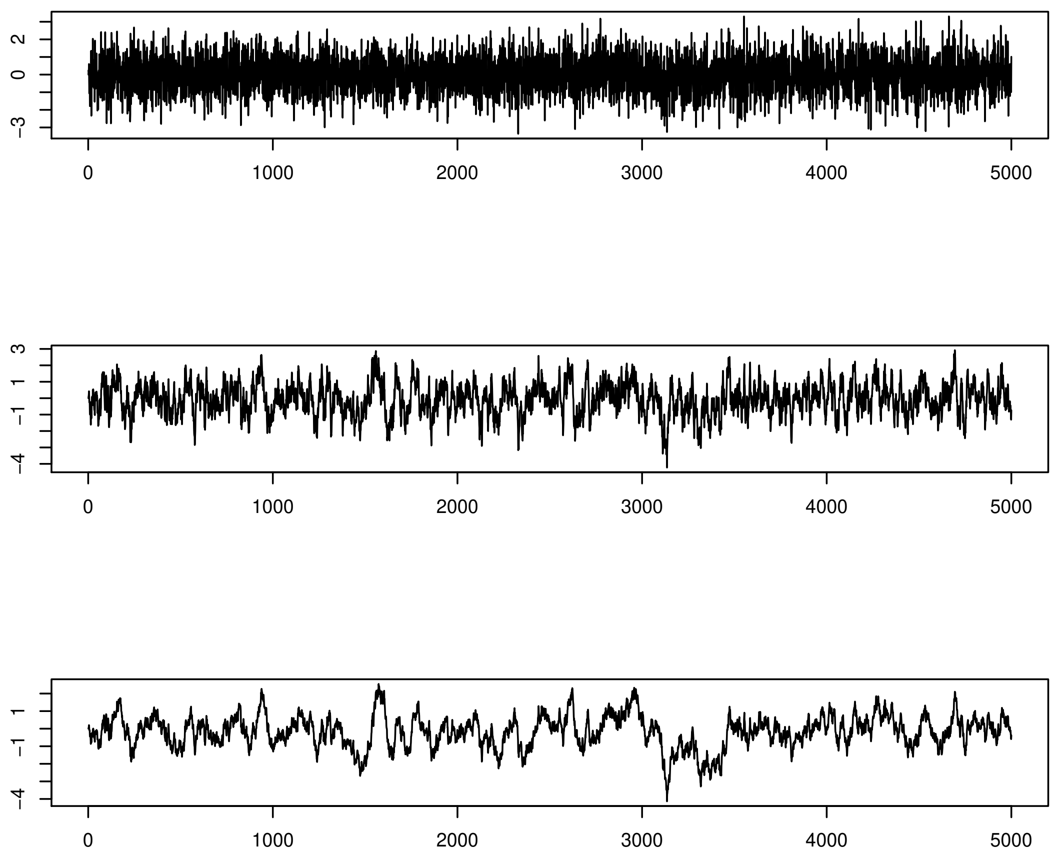

Figure 1b : Trajectories of Gaussian AR(1) process

(large ρ, in increasing order)

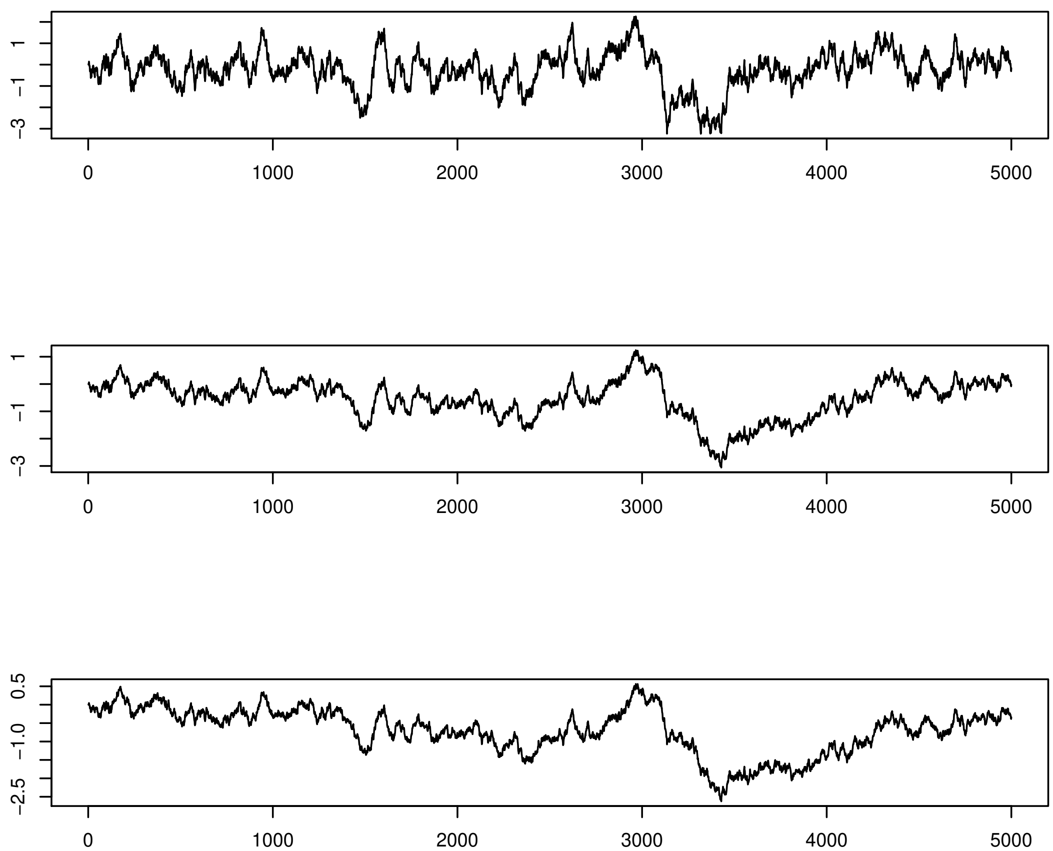

Figure 2 : Trajectories of Noisy Long Run Components

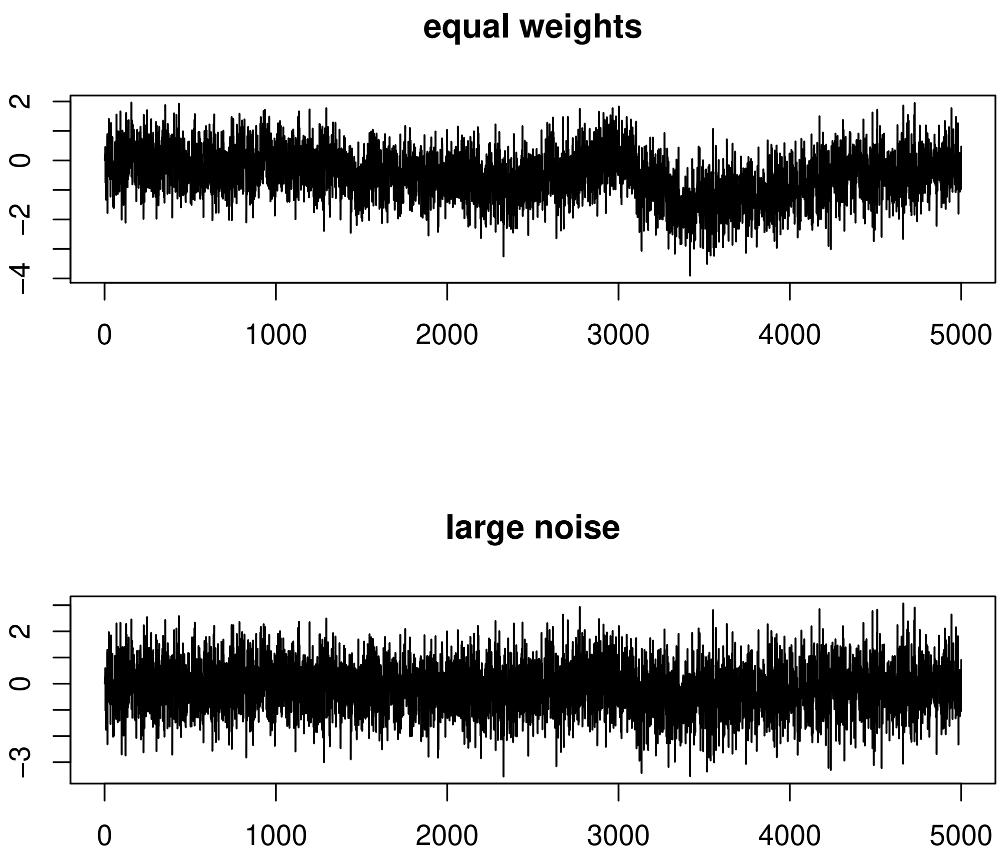

Figure 3 : Component Spectral Density, $\rho = 0.3, 0.8$.

(solid line $\rho = 0.8$, dashed line : $\rho = 0.3$).

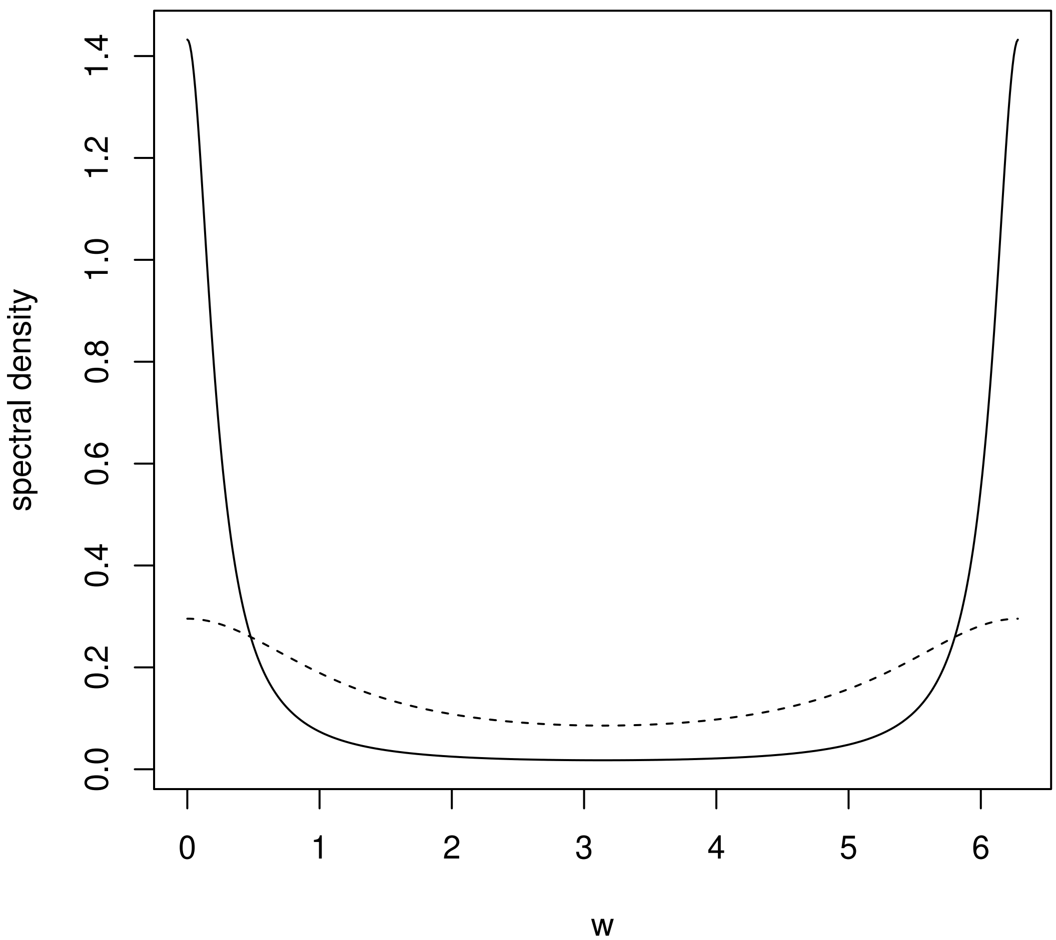

Figure 4 : $\beta_{hh} = R_{hh}$, function of h.

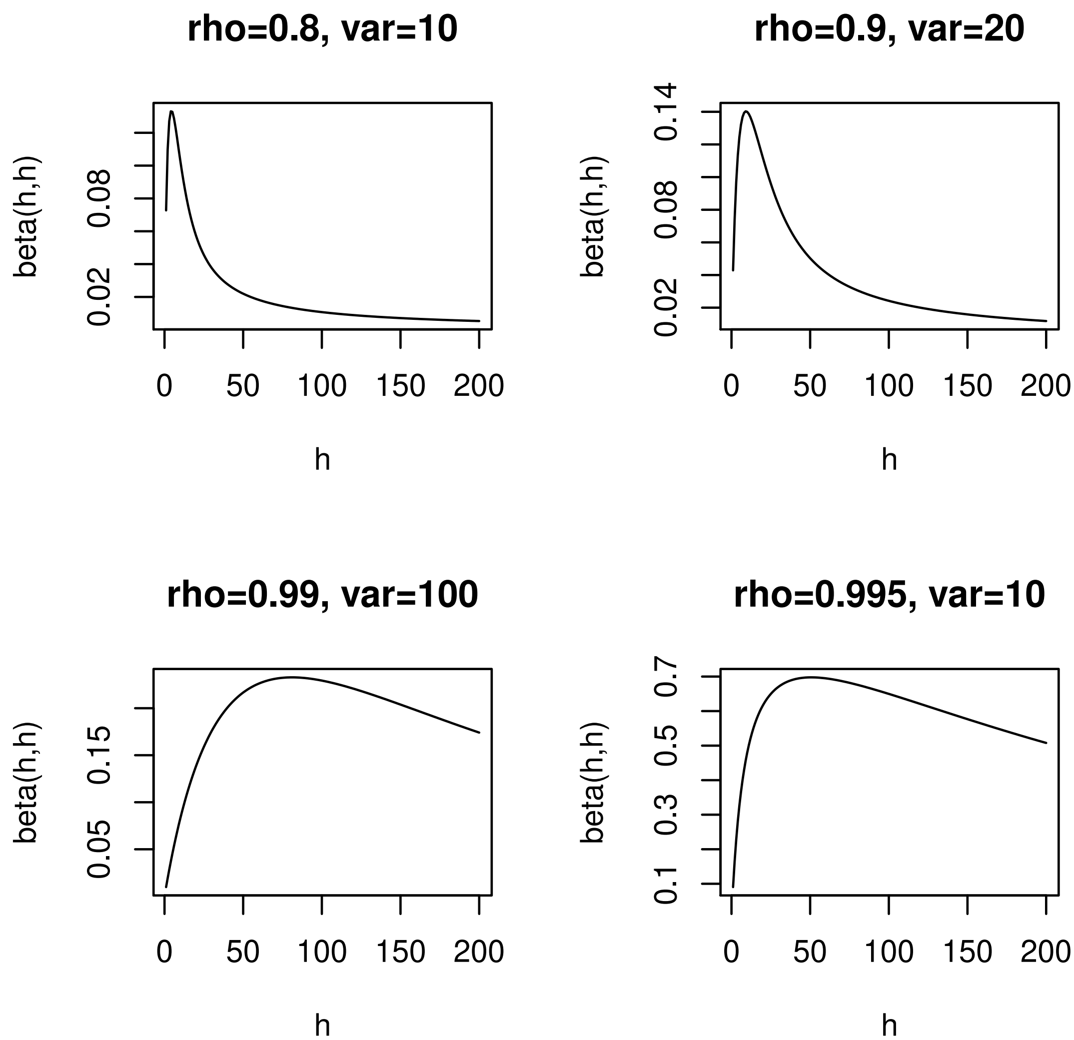

Figure 5 : Simulated Path

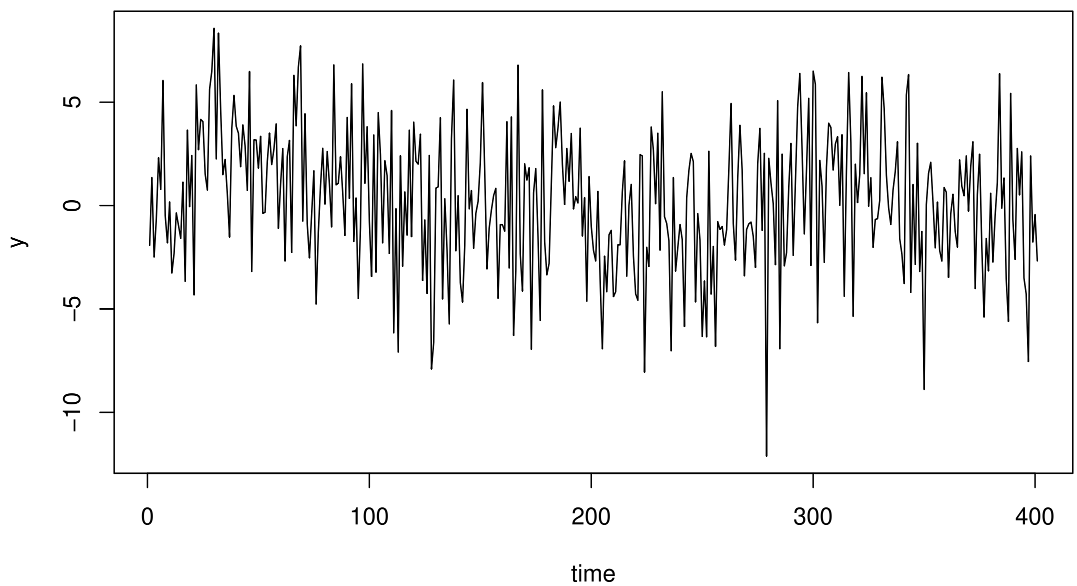

Figure 6 : True and Estimated Coefficients

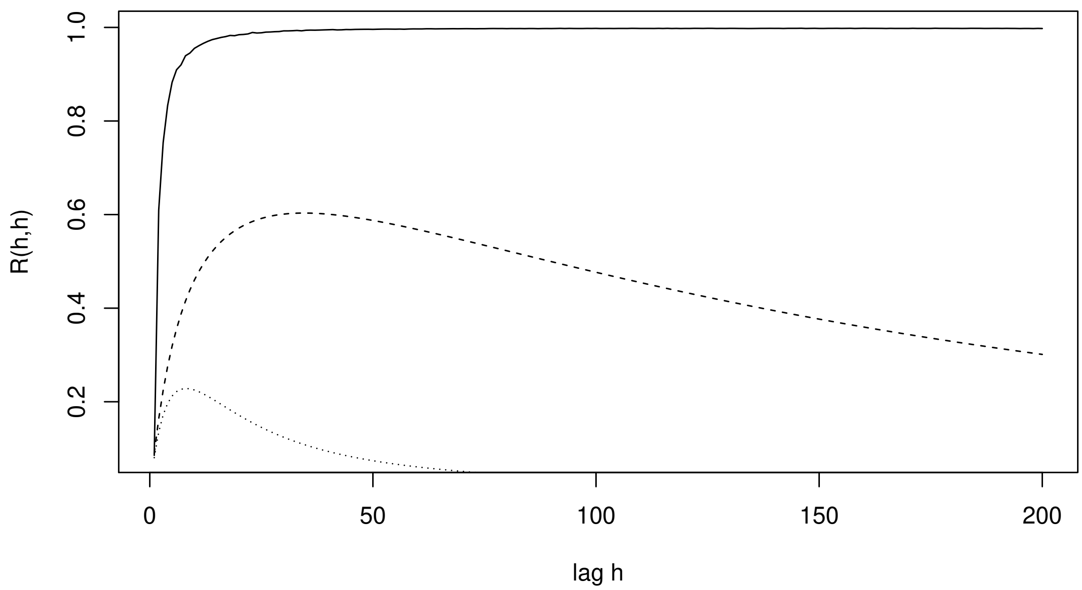

Figure 7a : ACF of Noisy Long Run Component : Equiweighted

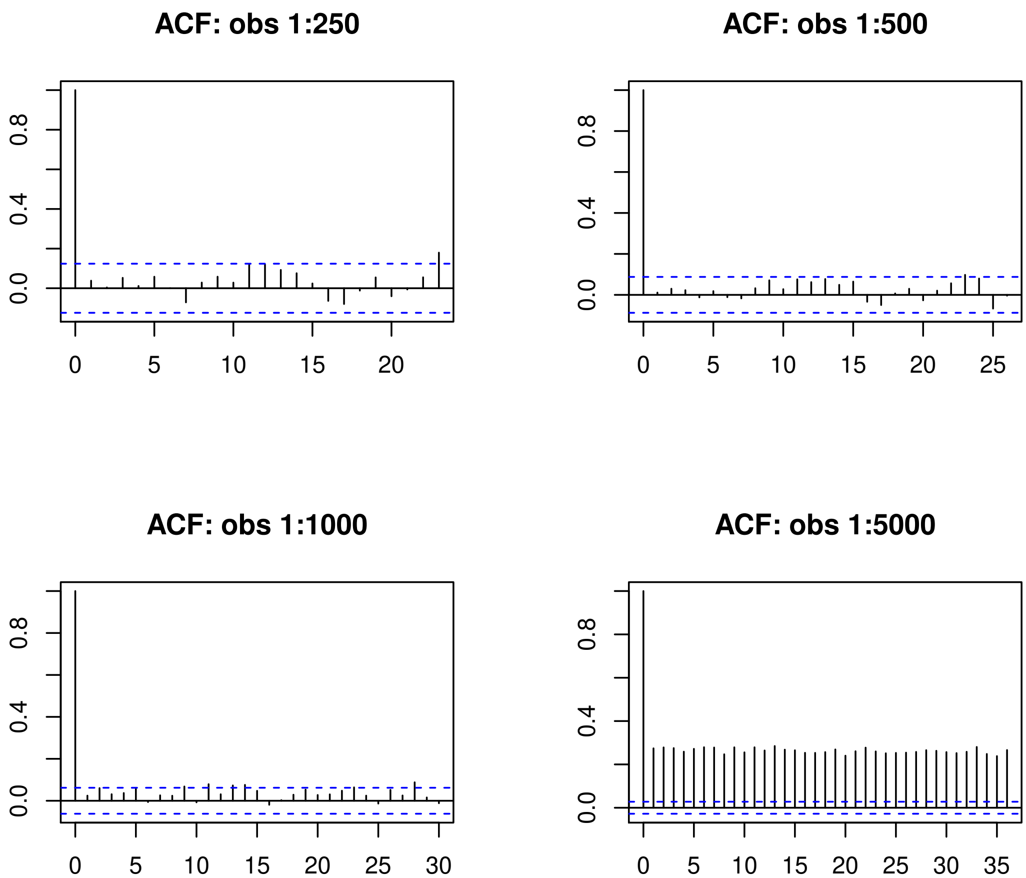

Figure 7b : ACF of Noisy Long Run Component : Large Noise

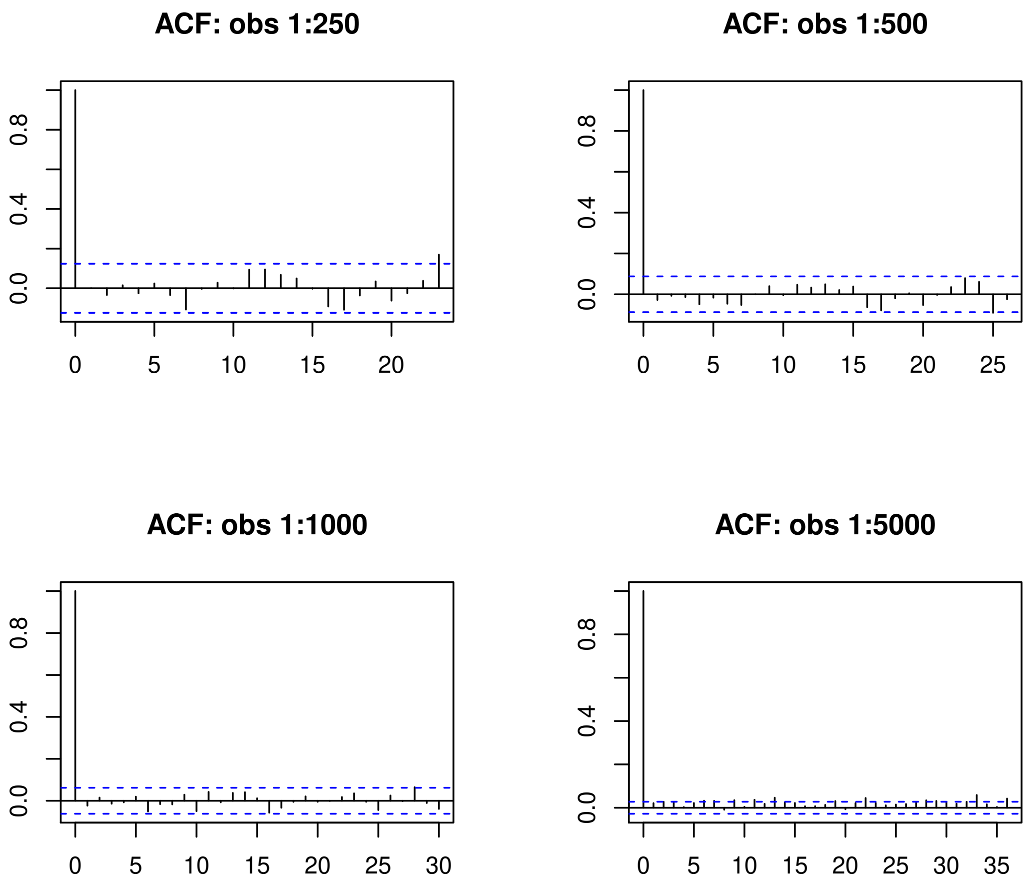

Figure 8a : Functional Estimates of the Distributions
(equiweighted, $T = 2000$)

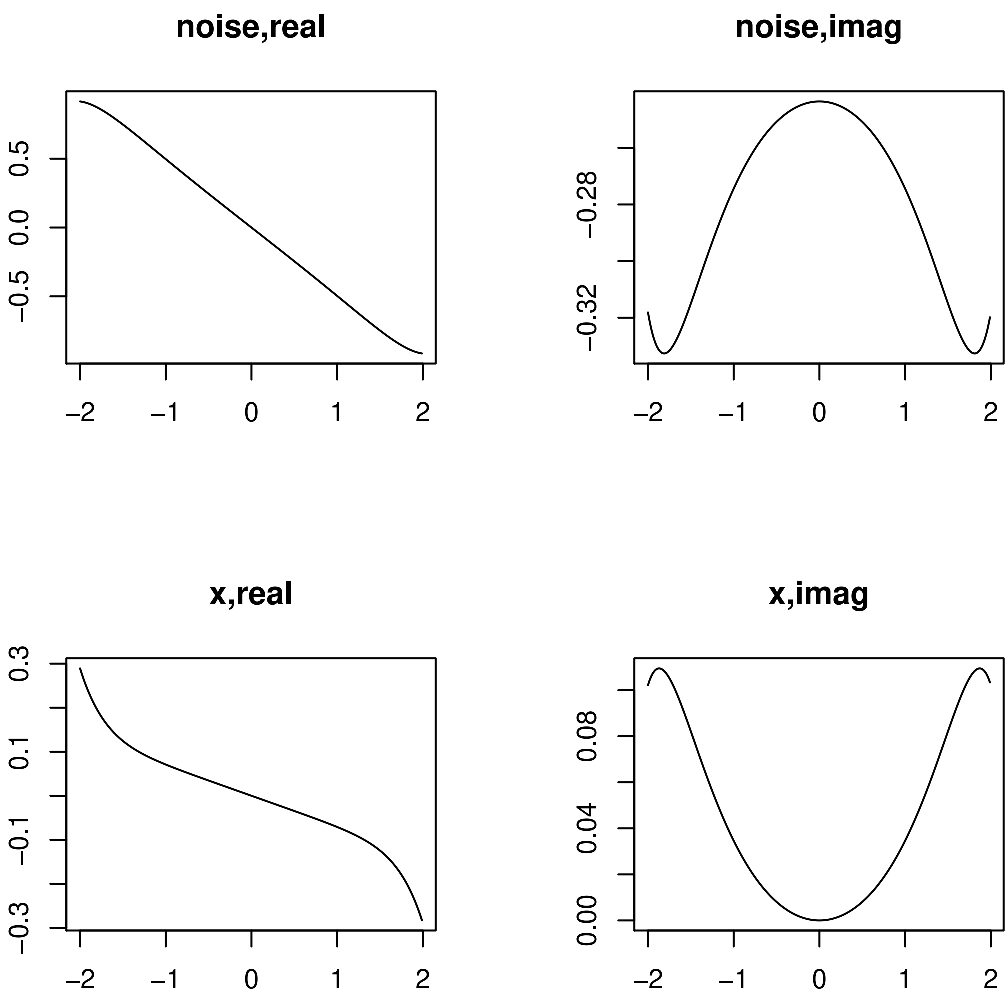

Figure 8b : Functional Estimates of the Distributions
(equiweighted, $T = 5000$)