Annals of Economics and Statistics, Number 145, March 2022

LONG RUN PREDICTIONS

CHRISTIAN GOURIEROUX[a] AND JOANN JASIAK[b]

This paper reexamines the modelling of long run risk in the econometric literature. We show that, if the macro or financial series are driven by the short and long run factors, then it is possible to identify all short run parameters, but not all long run parameters. We also develop techniques of evaluation of the long run estimation risk based on finite sample methods.

JEL Codes: C11, C22, C53, G18.
Keywords: Long Run Risk, Estimation Risk, Prediction, Local Level Model, Local-to-Unity Model, Ultra Long Run, Impossibility Theorem, Required Capital.

1. INTRODUCTION

This paper reexamines the modelling of long run risk in the econometric literature. We argue that the observed macro or financial series are driven by the short and long run factors and, if the number of observations over time is large, then under mild regularity conditions it is possible to identify and consistently estimate all the parameters characterizing the short run dynamics. Then it is possible to identify and consistently estimate some long run parameters as well, but not all of them. The uncertainty on these unidentified long run parameters can be evaluated by applying the "finite sample" techniques, which are either frequentist, or Bayesian. This difficulty arises even with a large number of observations. This has important consequences for the long run predictions, long run impulse responses, and for determining the required capital, or hedging against the long run risk. The reason is that the prediction intervals have to account for the "finite sample" uncertainty that is usually disregarded by the standard prediction softwares used by the practitioners. As mentioned in Hansen (2018) : "We (even) still see some academics or econometricians willingly deciding to take off uncertainty because policy makers are demanding precise answers".

Several articles report the confidence intervals for the long run parameters and large prediction intervals for the long run predictions based on Monte-Carlo studies : "The confidence intervals typically are wide" (Stock, 1991), or "These prediction sets may strike some readers as too large" (Müller and Watson, 2016), or Pástor and Stambaugh (2012) in stock volatility. However these wide prediction intervals are not yet taken into account by some financial and banking models, where the long run assets are assumed to be risk-free, thanks to total diversification over time. Typical examples are the banking theory in Diamond and Dybvig (1983), and the accounting practice where the long run interest rates are considered nonstochastic and constant over time.

The paper is organized as follows. The long run component models existing in the literature are discussed in Section 2. The difficulty to identify some long run parameters, i.e. the so-called impossibility theorem, is explained in Section 3. Section 4 analyzes the long run prediction problem and the importance of taking into account the estimation risk . Section 5 contributes to the debate on the (lack of) interest in testing the unit root

[a]University of Toronto, Toulouse School of Economics and CREST. christian.gourieroux@ensae.fr
[b]York University. nnaoj2012@gmail.com

hypothesis and on the trade-off between the statistical and strategic inferences. Section 6 concludes. The history of competing terminologies for the long run dynamic modelling is presented in Appendix A. This appendix distinguishes between the generic use and precise meaning. Appendix B discusses the limits of frequency analysis by projections. The acronyms used in the literature are gathered in Appendix C.

2. THE MODELLING OF LONG RUN DYNAMICS

For expository purpose, we consider a univariate series (y_t), which can be decomposed into two components as :

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

where $y_{s,t}$ is the short run component, usually a stationary ARMA or ARMA-GARCH process, and $y_{l,t}$ is a long run component, assumed independent of the short run component. The series (y_t) is observed at dates $t = 1, \ldots, T$, but the short and long run components are unobserved, leading to an unobserved component model. The assumption of independent components is the structural characteristic of this dynamic model. In this respect the analysis differs from the reduced form approaches based on frequency correlations (see Appendix B for the drawbacks of this latter approach).

In the simplest framework the long run component is assumed to follow an autoregressive process with a unit root, or has a close to unit root dynamics. These unobserved component models differ with respect to the type of trajectories they are intended to reproduce or approximate. They can feature nonstationarities with explosive patterns in their level or their variability, or display the occurrence of extreme values. The existence of such features largely depends on the variables of interest and their economic or financial interpretations. Whereas variables such as Gross Domestic Product (GDP), Consumer Price Index (CPI) feature trends, other variables have more stationary evolutions. These are the interest rates, unemployment rates, S&P realized volatility, productivity, some commodity prices. Other series become stationary after preliminary transformations, such as the GDP growth, inflation rate, deviation from the Interest Rate Parity (IRP), the Purchasing Power Parity (PPP), the spot-forward spreads, or the expenditure per capita that adjusts for demographical effects.

Due to technical issues, the theoretical literature has been generally focused on the univariate framework, and often without introducing the short run component. In practice, however, several series are considered jointly, some with stationary and others with nonstationary features, as in the cointegration models. Moreover the decomposition into the short and long run components is not necessarily additive, and the short and long run dynamics can be nonlinear.

Several dynamic models are discussed below including the close to unit root models. They have been introduced for two different, but related purposes. From a theoretical perspective the first objective is to define the local alternatives for the unit root hypothesis testing (see e.g. Elliott, Rothenberg, and Stock (1996) and the discussion in Section 5). From an applied perspective, the second objective is to understand why series with (almost) stationary features have empirical autocorrelation function (ACF) that decay slowly with the lag, at levels that do not correspond to the theoretical autocorrelation values.

2.1. *Random Walk (RW) plus Noise*

The basic model introduced in the literature assumes a strong white noise $y_{s,t}$ and a pure random walk dynamics of $y_{l,t}$:

$$y_{l,t} = \sum_{\tau=1}^{t} \varepsilon_{l,\tau}, \tag{2.2}$$

where $(\varepsilon_{l,t})$ is another strong white noise independent of $(y_{s,t})$ (see e.g. Harvey (1990) p45, and the references in Appendix A]. This model introduces a nonstationary long run component with trajectories featuring explosive behaviour over long random episodes. These behaviours are often incompatible with the economic interpretation of the variable of interest and may lead to unreasonable long run prediction intervals. "A random walk process hits any upper or lower bound with probability one and this is clearly an undesirable implication of this model" (Primiceri (2005), p825).

2.2. *Near-Integrated Models*

To partly circumvent this difficulty, stationary approximations of the random walk have been introduced. They usually consider an autoregressive parameter that depends on the number of observations, with a triangular array $y_{l,T,t}$, doubly indexed by t and T, instead of the standard time series $y_{l,t}$ for the long run component. A near integrated model is :

$$y_{l,T,t} = \rho_T y_{l,T,t-1} + \varepsilon_{l,t}, \tag{2.3}$$

$$\text{where } \rho_T \simeq \exp(-c/T), \text{ for } T \text{ large.} \tag{2.4}$$

These models, called near-nonstationary or near integrated (see Appendix A), have been widely used in the literature to analyze the asymptotic properties of historical autocorrelations and to perform tests of either the unit root, or of the martingale hypothesis (see Section 5). However, for T large, the near-integrated processes become close to a random walk (or a nonstationary martingale) and have the same drawback as the pure random walk for long run predictions.

2.3. *Local Level (LL) Model*

This model introduces an approximation of the long run dynamics that entails small changes in the long run component over the short run and significant changes in the long run. A LL model is :

$$y_{l,T,t} = \frac{1}{T} \sum_{\tau=1}^{t} \varepsilon_{l,\tau}. \tag{2.5}$$

Therefore, at any date t, $y_{l,T,t} - y_{l,T,t-1} = \varepsilon_{l,t}/T \simeq 0$, for large T. This justifies the idea of local-in-time changes. But, at any fixed T, the triangular array provides a random walk with the same drawback as a near-integrated process.

2.4. *Local-to-Unity (LTU)*

In the definition [1] proposed recently in Müller and Watson (2020), the LTU models combine a close to one autoregressive coefficient and a small innovation variance. An example is :

$$y_{l,T,t} = \exp(-c/T)y_{l,T,t-1} + \frac{1}{T}\varepsilon_{l,t}, c > 0, \tag{2.6}$$

(see Müller and Watson (2020), Section 2.2.). However, for large T, it is close to a non-stationary limit, although for T fixed it is stationary .

2.5. *Ultra Long Run (ULR) Model*

The idea is to introduce an instrumental long run component measured over a time scale with a time unit different from the short run component (Gourieroux and Jasiak, 2021, 2022). More precisely, let us consider a continuous time Ornstein-Uhlenbeck (OU) process x satisfying the diffusion equation :

$$dx_\tau = -cx_\tau d\tau - \sqrt{2c}\, dW_\tau, \tag{2.7}$$

where (W_τ) is a Brownian motion and $c > 0$. This is a stationary process with a stationary distribution $N(0,1)$, independent of c. Then the long run component is defined as :

$$y_{l,T,t} = \eta x_{t/T},\ \eta > 0. \tag{2.8}$$

Since the time discretized OU process is a Gaussian AR(1) process, it is easy to see that:

$$y_{l,T,t} = \rho_T y_{l,T,t-1} + \sqrt{1-\rho_T^2}\ \varepsilon_{l,t}, \tag{2.9}$$

where $\rho_T = \exp(-c/T)$ and the $\varepsilon'_{l,t}s$ are IIN $(0, \eta^2)$.

This specification differs from the LTU specification since :

$\sqrt{1-\rho_T^2} = \sqrt{1-\exp(-2c/T)}$ is of order $1/T^{1/2}$, instead of $1/T$ as in the LTU specification.

Moreover, for T large, the long run component tends to a "stationary" process such that $y_{l,t} = y_{l,0}$ is time independent, and takes a random value (stochastic level), because $y_{l,0} \sim N(0,1)$. This limiting process is a singular stationary process considered in Wold's representation theorem and also called "deterministic", or "purely predictable" in the literature.

3. STATISTICAL INFERENCE

Let us consider the parametric dynamic models of the short and long run dynamics. Vector α contains the parameters characterizing the short run dynamics and $c, \eta^2 = V(\varepsilon_{l,t})$ are the parameters characterizing the long run dynamics. These parameters are independent of the number of observations T. It is important to focus on such intrinsic parameters instead of the parameters, which depend on T such as ρ_T for instance, especially when

[1] According to initial definition the local-to-unity process is equivalent to the near integrated process (see Appendix A).

we are interested in performing long run predictions[2] (see the discussion in Section 5.2). As for the frequency of observations, the autoregressive coefficient at horizon 1 is either $\rho_T = \exp(-c/T)$, or 1, depending on the model. The variance of the innovation at horizon 1 is η^2 (for the near integrated model), η^2/T^2 for the local level model and the local-to-unity model, η^2/T for the ULR model.

Statistical inference requires high level mathematics, but the underlying principle is simple and is discussed below.

3.1. *The identification of short run parameters*

For all models where $y_{l,T,t} \simeq y_{l,T,t-1}$ for T large, the long run component is approximately constant in time over each episode of length H_T, with $H_T \to \infty, H_T/T$ tending to zero for large T. Therefore, except for the theoretical mean, all other short run parameters will be estimated consistently over such large, but "short run" episodes. The rate of convergence of the estimators of the short run parameters is expected to be of order $1/\sqrt{T}$. Since the theoretical mean of $y_{s,t}$ is not identifiable, later on, it is fixed to $Ey_{s,t} = 0$, for ease of exposition.

3.2. *The "identification" of the long run component*

Let us consider a historical average over H_T successive dates, $H_T \to \infty, H_T/T \to 0$, evaluated at distant dates kT, such as[3] :

$$\frac{1}{H_T}\sum_{t=kT+1}^{kT+H_T} y_T = \frac{1}{H_T}\sum_{t=kT+1}^{kT+H_T} y_{s,t} + \frac{1}{H_T}\sum_{t=kT+1}^{kT+H_T} y_{l,T,t}, 0 < k < 1.$$

By the Law of Large Numbers (LLN) applied to the short run component, the first historical average tends to zero (since $Ey_{s,t} = 0$). Therefore these averages eliminate the short run effects allowing us to focus on the long run component. Depending on the specification such as the near-integrated model (see Phillips (1987), Th. 3.1, b), local-level, local-to-unity (Müller and Watson, 2016), and ULR models (Gourieroux and Jasiak, 2021, 2022), such averages computed from a well-chosen H_T and possibly standardized by a power of T, converge either to a continuous time OU process (x_τ), up to scale η, or to integral transforms of such a process. [4]

In brief, by appropriate averaging, one can approximate the trajectory of an underlying O.U. process on $[0, 1]$. This approximation $\widehat{(\eta x_{\tau,T})}$ converges to the true process at a rate slower than $1/\sqrt{T}$, in general.

These local averaging do not imply a loss of information in the unobserved component model, but allow for filtering out the long run component. Indeed, to a dynamic model with short and long run time scales, the "likelihood principle" of Fisher known since the twenties (see Berger and Wolpert (1988)) does not apply. [5] Moreover, the likelihood

[2]Even if an estimator $\hat{c}_T$ of c is not consistent, $\hat{\rho}_T = \exp(-\hat{c}_T/T)$ will tend to 1, if $\hat{c}_T$ is of order 1.

[3]These averages are weighted to capture the variability at different frequencies in Müller and Watson (2016), eq. (2).

[4]In particular to frequency weighted averages in Müller and Watson (2016).

[5]The likelihood principle states that the "random" likelihood function is a sufficient statistic for the parameter.

principle should be modified to account for all unknown quantities, that are relevant to the statistical analysis. They include the parameters, and also the unobserved components that need to be filtered and predicted.

3.3. *The impossibility theorem*

From the discussion in Section 3.2, it follows that asymptotically, the approximation is as good as if the underlying continuous time process (ηx_τ) was observed (with errors) on $\tau \in [0.1]$. Unfortunately we come accross the impossibility theorem (Banon, 1978; Jiang and Knight, 1997) that applies to a continuous time diffusion process observed over a finite interval.

Impossibility Theorem :

From continuous time observations on (ηx_τ), we can estimate consistently the infinitesimal volatility $\sigma^2 = 2c\eta^2$. It is not possible to estimate consistently parameters c and η^2.

In practice this means , that we have approximate observations $\widehat{\eta x_{k_{1,T}}}, \ldots, \widehat{\eta x_{k_{J,T}}}$, with possibly $J = J_T$ large, that can be used to make inference on the long run parameters. Moreover this inference necessarily entails significant errors in η, c parameters even for very large T. Due to this impossibility theorem, we have to "mix" asymptotic inference for α, σ^2, with different rates of convergence, and "finite sample" inference [6] on parameter c.

The results above can be reinterpreted in the framework of stationary linear systems with independent shocks. When the number of inputs (sources) is strictly larger than the number of outputs (observations), the system is called undetermined since it is not possible to recover exactly the values of the inputs. Then the literature focuses on conditions that help identify the distribution of the sources, especially by applying dynamic independent component analysis. This practice has to be modified in a triangular model with parameters varying with the number of observations. We see that the sources can be asymptotically recovered, whereas the distribution of the long run component can never be identified.

3.4. *Finite sample/asymptotic inference*

Under a parametric specification, the parameters α, σ are consistently estimated and can be replaced by their estimates. Let us focus on the remaining parameter c. Two approaches can be followed.

i) Frequentist approach

Let us consider a given estimator $\hat{c}_T$ of c based on observations $y_1, \ldots, y_T$. The unobserved component model is often easy to simulate for any given value of c (c_0, say) and $\hat{\alpha}_T, \hat{\sigma}^2_T$ of other parameters. Therefore, it is easy to construct Monte-Carlo exact finite sample test procedures for testing $H_{c_0} = \{c = c_0\}$. Then exact finite sample confidence intervals for c based on $\hat{c}_T$ can be deduced by inverting the set of tests (see e.g.

[6] An alternative is to use panel data to solve this identifiability issue (see e.g. Moon and Phillips (2000), Gouriéroux, Phillips, and Yu (2010)). However such panel data are usually not available, especially for macro applications. It has also been proposed to modify the models by introducing the so-called block local-to-unity model (see Phillips, Moon, and Xiao (2001)). Such ad hoc adjustment of the models is to recover the consistency of the localizing parameter c. It is not in line with a prudential view of long run estimation risk.

Stock (1991), Chambers (2016), Gourieroux and Jasiak (2022) for examples of confidence belts). These belts and the theoretical analysis of the distribution of the estimator of the serial correlation close to unit root show that the finite sample distribution of this estimator is highly skewed. In particular this implies significant biases in the standard forecasts (Phillips, 1979; Boudoukh, Israel, and Richardson, 2021)

ii) Bayesian approach

Bayesian approaches are often used in structural unobserved component models (see e.g. Harrison and West (1991), West and Harrison (1997)), and have also been used in the local level, or local-to-unity framework (see Primiceri (2005), Schorfheide, Song, and Yaron (2018), Müller and Watson (2016, 2020)). Due to the identification issue concerning parameter c, the posterior of this parameter strongly depends on the selected prior and is likely not robust to a change of the prior.

4. LONG RUN PREDICTION INTERVAL

Similarly to the inference on parameter c, the predictions can be computed either in a Bayesian framework (Müller and Watson, 2016, 2020), or in a frequentist framework (Gourieroux and Jasiak, 2021, 2022). We consider prediction intervals, not just the plug-in pointwise predictions that are less informative, biased, and neglect the underlying risks. Let us first discuss the frequentist approach by clearly distinguishing between the theoretical and empirical predictions. Then we highlight a drawback of the Bayesian approach for long run prediction.

4.1. *Theoretical Predictions*

Under a frequentist approach the theoretical predictions are based on the distribution of y_{T+H} given $y_1, \ldots, y_T$ and parameters $\theta = (\alpha, \sigma, c)$, or equivalently on the distribution of:

$$y_{s,T+H} + y_{l,T,T+H}, \text{ given } y_{s,1} + y_{l,T,1}, \ldots, y_{s,T} + y_{l,T,T}.$$

As shown in Section 3, asymptotically the long run component can be filtered with an error close to zero. Then, it is equivalent to consider the information generated by $y_1, \ldots, y_T$, or the information generated by both $y_{s,1}, \ldots, y_{s,T}$ and $y_{l,T,1}, \ldots, y_{l,T,T}$. Equivalently the underlying values $y_{s,t}, y_{l,T,t}$ of the latent variables are known asymptotically.

Since the short and long run components are independent, the predictions of the short run and long run components can be performed separately. Moreover, if the prediction horizon H is large ($H \to \infty$, with T), the conditional distribution of $y_{s,T+H}$ at date T tends to a point mass at its mean equal to zero.

We conclude that, to make long run predictions, one can disregard the short run component and focus on the long run component only. Let us denote the pointwise theoretical prediction of $y_{l,T,T+H}$, at date T by:

$$\pi(H, T; y_{l,T,T}; \sigma, c), \tag{4.1}$$

which depends on $y_{l,T,T}$ and the long run parameters only.

4.2. *Empirical Pointwise Predictions*

The theoretical pointwise predictions given above depend on the unknown values $y_{l,T,T}, \sigma, c$. In the plug-in approach they are usually replaced by their approximations to get the com-

putable :

$$\pi(H, T; \hat{y}_{l,T,T}, \hat{\sigma}_T, \hat{c}_T), \tag{4.2}$$

where $\hat{y}_{l,T,T}$ is the consistent approximation of $y_{l,T,T}$ based on an appropriate average of y_t, and $\hat{\sigma}_T$, $\hat{c}_T$ are the estimators of σ, c, respectively. Due to the consistency properties, we get asymptotically :

$$\pi(H, T; \hat{y}_{l,T,T}; \hat{\sigma}_T, \hat{c}_T) \simeq \pi(H, T; y_{l,T,T}, \sigma, \hat{c}_T). \tag{4.3}$$

where $\hat{c}_T$ cannot be replaced by c due to the Impossibility Theorem.

4.3. *Prediction Errors*

The prediction error follows from equation (4.3) and is :

$$\begin{aligned} & y_{T,T+H} - \pi(H, T; \hat{y}_{l,T,T}; \hat{\sigma}_T, \hat{c}_T) \\ \simeq\ & y_{s,T+H} + [y_{l,T,T+H} - \pi(H, T; y_{l,T,T}; \sigma, c)] \\ & + [\pi(H, T; y_{l,T,T}; \sigma, c) - \pi(H, T; y_{l,T,T}; \sigma, \hat{c}_T)]. \end{aligned}$$

It involves two components linked to the long run factor, which are 1) the theoretical prediction error and 2) the effect of the estimation error on the point prediction.

Note that in all models of Section 2 the process $y_{l,T,t}$ is (asymptotically) a Gaussian AR(1) process. Therefore, for a given $y_{l,T,T}$, the future value $y_{l,T,T+H}$ and the past values $y_{l,T,1}, \ldots, y_{l,T,T-1}$ are independent. It follows that the components 1 and 2 in the decomposition of the empirical prediction error are also independent given $y_{l,T,T}$. Therefore the conditional distribution of the empirical prediction error is the convoluate of the conditional distribution of the theoretical prediction error and of the distribution of the second component, measuring the effect of the estimation risk in c. Usually, this second component is disregarded. **Yet, it becomes significant and even dominant in a long run framework**.

To illustrate this decomposition let us consider the ULR model. The empirical prediction error is decomposed as :

$$\begin{aligned} & y_{s,T+H} + [y_{l,T,T+H} - \exp(-cH/T) y_{l,T,T}] + [\exp(-cH/T) - \exp(-\hat{c}_T H/T)] y_{l,T,T} \\ =\ & y_{s,T(1+a)} + [y_{l,T,T(1+a)} - \exp(-ac) y_{l,T,T}] + [\exp(-ac) - \exp(-a\hat{c}_T)] y_{l,T,T}, \end{aligned}$$

over a long run horizon $H = H_T = aT$. Let us now discuss the magnitude of each random component. The short term component has a standard error equal to σ_s, that is of order 1. The second term has a standard error of order $\eta\sqrt{1 - \exp(-2ac)}$. It is also of order 1, like the last term since $\hat{c}_T$ is inconsistent. These three elements in the decomposition are of similar order. However the "magnitude" of the second term depends on the non identifiable parameter, and also has to be estimated. Therefore **there are asymptotically two related estimation risks to be taken into account**, that are the one concerning the theoretical magnitude of the mean square prediction error (second component) and the estimation error on the theoretical point prediction (third component).

A similar analysis can be done for a local-to-unity model. Since the variance of the long run innovation is of order $1/T^2$ instead of order $1/T$ for the ULR model, the second term

becomes asymptotically negligible like the estimation risk on that term. Thus we expect smaller long run prediction intervals for the LTU model than for the ULR model, just as a consequence of the model assumptions.

4.4. *Bayesian prediction*

In a Bayesian framework with parameter θ (including the short and long run parameters) and prior $p(\theta)$, say, the prediction interval at level $1-\alpha$ is constructed from the conditional distribution of y_{T+H} given $y_1, \ldots, y_T$:

$$(4.4) \qquad l(y_{T+H}|y_1, \ldots, y_T) = \int l(y_{T+H}|y_1, \ldots, y_T; \theta) p(\theta|y_1, \ldots, y_T) d\theta.$$

where $p(\theta|y_1, \ldots, y_T)$ denotes the posterior distribution of θ. Then a prediction interval (L_T, U_T) is such that $P[y_{T+H} \in (L_T, U_T)|y_1, \ldots, y_T] = 1 - \alpha$, for any $y_1, \ldots, y_T$ (see Müller and Watson (2020), Section 4.6, Müller, Stock, and Watson (2020)). It differs from a "classical" prediction intervals that has to satisfy :

$P[y_{T+H} \in (L_T, U_T)|y_1, \ldots, y_T; \theta] \geq 1 - \alpha$, for any $y_1, \ldots, y_T$, **and** θ. In fact the coverage condition introduced in the Bayesian approach is on average instead of being uniform in θ. It is less prudential and strongly dependent on the prior especially under the lack of identification issue, i.e. the impossibility theorem. The same difficulty arises while comparing confidence and credible sets for parameters (see Müller and Norets (2016), Müller and Watson (2020), Sections 6,7).

5. THE UNIT ROOT HYPOTHESIS

A large body of the long run econometric literature focuses on the test of the "unit root" hypothesis. The aim of this section is to clarify the following points :

i) What is the unit root hypothesis ? Does it admit a structural interpretation?

ii) What is the interest of testing for unit roots before performing long run predictions?

iii) What are the respective roles of the statistical and strategic inferences?

5.1. *The unit root hypothesis*

5.1.1. *The efficient market hypothesis (EMH)*

It follows from the debate between Sims (1988, 1991), and Phillips (1991), that the only structural interpretation of a unit root in Economics/Finance is through the notion of efficient market. Among the various definitions of efficiency that have been introduced such as the weak efficiency, strong efficiency, semi-strong efficiency (see e.g. Fama (1965, 1970)), only one is really relevant. Let us consider a stock market, denote by y_t^* the price of the stock, assume no dividend, and denote by y_t, the stock price discounted by the risk-free rate. The market is efficient if and only if :

$$(5.1) \qquad E_t(y_{t+1}) = y_t,$$

where E_t is the expectation conditional on the information available up to date t. This condition is interpreted as the impossibility of a perfect dynamic arbitrage between the

stock and the risk-free asset, as a consequence of Doob theorem (LeRoy, 1989).[7] The condition (5.1) is a martingale condition on the discounted stock price.

5.1.2. *The null hypothesis*

For a long time, economists have considered that this condition was equivalent either to the condition of random walk (RW) : $y_t = \sum_{i=1}^{t} \varepsilon_\tau$, where the (ε_t) are i.i.d. (see e.g. Sims (1988)), or to the condition of an integrated linear process where (ε_t) is a weak white noise. Both types of processes RW and I(1) are nonstationary.

However, there also exist stationary martingales. They include in particular :

i) noncausal autoregressive processes introduced to capture speculative bubbles, that are transitory explosions followed by crashes (Gouriéroux and Zakoïan, 2017);

ii) volatility induced mean-reverting processes, with a strong mean reversion effect when y_t takes extreme values (Conley, Hansen, Luttmer, and Scheinkman, 1997; Gouriéroux, Jasiak, and Monfort, 2020).

iii) pure predictable processes $y_t = y_0. \forall t$, with y_0 stochastic, that serves as the basis of modelling ULR processes (Gourieroux and Jasiak, 2022).

In this respect, the major part of the literature, except for the recent paper Gourieroux and Jasiak (2019), is not about testing the martingale hypothesis, but instead a null hypothesis of either RW, or I(1). In general, the testing procedures consist in testing a nonstationarity hypothesis against an alternative stationary process. This explains why the literature focuses on stationary local alternatives tending to nonstationary limits (see Section 2 and Appendix B).

5.2. *Pretesting*

The debate about the unit root hypothesis started from the early $90's$, with different conclusions obtained under Bayesian and classical approaches. The major arguments are the following ones :

i) **A Bayesian perspective :**

- There is no structural interpretation of the unit root hypothesis (see Section 5.1).
- Thus there is no reason to privilegiate this hypothesis, which is done by practitioners by applying the unit root (or cointegration) tests and using the unit root models when the unit root (resp. cointegration) hypothesis is not rejected. "There is no reason for distinguishing models (with unit roots) from their neighbors with roots close to the unit circle" (Sims (1988), p469).
- Classical statisticians have developed sophisticated test procedures for testing for unit root. In general, they do not evaluate the effect of accepting/rejecting the unit root hypothesis on the properties of estimators and predictors.
- A specification with parameters depending on T has no real meaning. In Bayesian terms the prior of a given individual does not depend on the number of observations.

ii) **A classical perspective.**

- In the long run framework, the Bayesian results strongly depend on the selected prior.

[7] Doob theorem says that a martingale is also such that $E_t(y_{t+v}) = y_t$, for any stopping time v. This means that there is no dynamic arbitrage opportunity for any strategic choice of the random time of reselling the risky stock.

- In particular, the choice of a uniform prior (or flat prior) on parameter ρ like in Sims and Uhlig (1991),Schorfheide, Song, and Yaron (2018), Müller and Watson (2018), p759, is questionable, since it underweights the nonstationary feature.
- The prior, that depends on the parametrization, may indirectly depend on the number of observations.

These arguments disregard the main purposes of long run modelling, which are the long run predictions and measures of long run impact of specific shocks. Even, if the estimation approach, tests, priors, and confidence or credible intervals can be used as intermediate tools for determining the long run prediction sets, their choices have to be based on a prediction criterion, rather than on an estimation criterion, or the power of a test. This renders all the (bayesian or classical) theoretical developments and debates about the unit root hypothesis less interesting.

Let us now discuss the argument against using the prior independent of T, that seems in contradiction with the triangular arrays introduced in the models of Section 2. In fact, the ULR model shows another structural interpretation of such dynamics, in which the observed series corresponds to an equilibrium with different agents acting on different time scales. Let us now discuss the effect of time scale in a simple, related case of high frequency data (HFD). Assume that observations on an O.U. process over one day are recorded at different scales every hour, 10mn, mn, second,... There are T observations, at every $1/T$ dates. A Bayesian statistician has to fix a common prior, independent of the choice of the observational frequency, that is a prior on the coefficients c, η of the underlying O.U. process. This will imply T dependent priors on the coefficients ρ_T, σ_T of the associated time discretized process. **In other words, fixing a uniform prior on ρ_T, independent of T, is meaningless, since it does not concern the right intrinsic structural parameter.**

5.3. *Statistical or strategic inference*[8]

As shown in Section 3, there is an identification issue in the long run analysis. This leads to wide prediction intervals for a given dynamic parametric model, and also to a lack of robustness with respect to the selection of the parametric dynamic model (including the choice of the prior in Bayesian analysis).

Thus, statistical inference based on a limited sampling period is not very informative about the long run. How to compensate this natural limitation of statistical analysis? It could be done by introducing additional strategic restrictions on the dynamics in order to influence the agent's behaviours. For instance, by considering a limited set of dynamic scenarios one can compute the long run required capital, fix by consensus an ultra long run interest rate for pricing the long run assets in the balance sheets and combine the priors of different agents interested in long run predictions(see Robinson (1977)). These subjective choices have to be clearly announced along with their justifications (as the retained balance between prudential and innovation friendly approaches, or the list of formal or informal identifying assumptions to yield substantive conclusions), and then regularly updated with the increasing number of observations.

[8]or decision making in Cox terminology (Cox, 1958)Cox (1958).

6. CONCLUDING REMARKS

Müller and Watson (2018) write that : "Proper long run inference depends critically on the exact form of persistence and there is limited sample information available to determine this form". As the analysis described in this paper shows, it is important to account for the asymptotic non identifiability of some long run parameters. This avoids underestimating the long run risk by neglecting the significant long run estimation risk and reveals the potential need for strategic inference and supervision. This is specially important when evaluating the required capital for the long run risk (see Gourieroux, Monfort, and Renne (2021)) and is a major step towards solving the so-called long run risk puzzle (Fama and French, 1988, 1989).

This paper considered a simple framework. It can be easily extended to the multivariate structural VAR framework (Müller and Watson, 2018; Gourieroux and Jasiak, 2022) predictive regressions when the regressors include both SR and LR component, to nonlinear dynamic models (Gourieroux et al. 2022), and to other types of parameters than a mean, for instance a scale parameter (see Shephard (1994) for local scale models), or a contagion parameter in epidemiological models (Gourieroux and Jasiak, 2021). The analysis of long run predictions is also the first step before deriving confidence intervals for impulse response functions at long run horizons (Pesavento and Rossi, 2006).

APPENDIX A: THE TERMINOLOGIES

In the context of modelling the dynamics of long run components, different terminologies have been introduced in the literature, with interpretations varying with time and authors. The aim of this appendix is to discuss the histories of some terms such as random walk plus noise model, local level (LL) model, near nonstationary model, near integrated model, local-to-unity (LTU) model, or ultra long run (ULR) component.

i) Random Walk plus Noise

This dynamic model is the most precisely defined in the literature (see e.g. Harvey (1990), p45). The model is :

$$\text{(a.1)} \qquad y_t = y_{s,t} + y_{l,t},$$

where $(y_{s,t})$ is a strong white noise, that is a sequence of i.i.d. variables with zero mean and common variance σ_s^2. (y_{lt}) is a random walk such that :

$$\text{(a.2)} \qquad y_{l,t} = y_{l,t-1} + \varepsilon_{l,t}.$$

where $(\varepsilon_{l,t})$ is another strong white noise with zero-mean, variance σ_l^2, and is independent of $(y_{s,t})$.

It is important to discuss the role of the initial value $y_{l,0}$.

(*) If $y_{l,0} = 0$, we have $y_{l,t} = \sum_{\tau=1}^{t} \varepsilon_{l,\tau}$. The process (y_t) has zero mean, and a variance equal to $V(y_t) = \sigma_s^2 + t\sigma_l^2$. This process is stationary in mean and nonstationary in variance.

(**) If $y_{l,0} \neq 0$, we have $y_{l,t} = y_{l,0} + \sum_{\tau=1}^{t} \varepsilon_{l,\tau}$. The process (y_t) has (conditional on $y_{l,0}$) a nonzero mean equal to $y_{l,0}$, and an explosive variance.

There is often a confusion in the applied literature about the notions of random walk, martingale, or integrated process of order 1, denoted I(1). In the I(1) framework, the errors $(\varepsilon_{l,t})$ are assumed zero-mean, with a constant variance and uncorrelated. In the martingale case, the $(\varepsilon_{l,t})$ is a martingale difference sequence (MDS), satisfying : $E(\varepsilon_{l,t}|\underline{\varepsilon_{l,t-1}}) = 0, \forall t, V(\varepsilon_{l,t}|\underline{\varepsilon_{l,t-1}}) = \sigma_l^2$, if it is conditionally homoscedastic (otherwise, the volatility is unconstrained) (see Section 5.1 for the Efficient Market Hypothesis).

ii) Local level (LL) model

Initially this terminology has been introduced in Harvey (1990), as an alternative for a random walk plus noise process with no drift ($y_{l,0} = 0$). Then the meaning has been modified in Harvey, Shephard (1990), to "a local level model, that has a deterministic level when the signal-to-noise ratio q is zero".

The Gaussian local level model, where the $\varepsilon_{l,t}$ are Gaussian, with σ_l fixed has been largely popularized by West and Harrison (1997), Commandeur and Koopman (2007), chapters 1-7, Durbin and Koopman (2012), chapter 2.

A modified version of the LL model where :

$$\text{(a.3)} \qquad y_{l,T,t} = \frac{1}{T}\sum_{\tau=1}^{t} \varepsilon_{l,\tau},$$

with a variance of $y_{l,T,t} - y_{l,T,t-1}$ tending to zero with the number of observations and $\varepsilon_{l,t}$ a MDS is considered in Müller and Watson (2020) eq. (4)-(5). Then, for each value of T, the long run component is a nonstationary process. This more recent definition is in line with Harvey (1990), since in a model $y_t = y_{st} + y_{l,T,t}$, the signal-to-noise ratio is $q_T = \sqrt{V(y_{l,T,t})/V(y_{s,t})} = \sigma_l\sqrt{t}/T\sigma_s \sim 0$, for T large and $t \leq T$. This is more compatible with the initial idea of local level, $y_{l,T,t} \sim y_{l,T,t-1}$, for T large (see e.g. Dahlhaus (2012)).

iv) Nearly Nonstationary Process

This terminology appears first in Ahtola and Tiao (1984) (see also Chan and Wei (1987)) for a long run component satisfying :

$$\text{(a.4)} \qquad y_{l,T,t} = \rho_T y_{l,T,t-1} + \varepsilon_{l,t},$$

where $\rho_T = \exp(c/T)$ and the $(\varepsilon_{l,t})$ a Gaussian white noise.

This long run component tends to a Gaussian random walk, when T tends to infinity. This convergence is through stationary AR(1) processes (resp. explosive AR(1) processes), if $c < 0$ (resp. $c > 0$).

v) Nearly-Integrated Process

This terminology has been introduced in Phillips (1987) for a model of the type (a.4) with a weaken assumption on the $(\varepsilon_{l,t})$, allowing for a martingale, or an I(1) process for the cumulated long run noise. This explains the name "integrated" (for I(1)). As, for the nearly nonstationary process, the associated long run component tends to a nonstationary process when T tends to infinity.

vi) Local-to-Unity (LTU) Process

This terminology has been used initially in Cavanagh (1986) (see also Stock (1991), p437, Müller and Watson (2016), p1729, Dou and Müller (2021), eq.1,) for ARMA models with a largest autoregressive root close to unity, i.e. such that $\rho_T \sim 1+c/T$, and fixed variance of innovation. In this sense, it covers processes similar to near-integrated processes. As these processes, it converges to a nonstationary ARIMA process, when T tends to infinity.

This explains why some authors come back on the alternative terminology of nearly-integrated (see e.g. Cavanagh, Elliott, and Stock (1995)).

However, it is important to note that the definition of LTU process has been modified in the recent survey by Müller and Watson (2020), p11, eq.2, where $\rho_T \simeq 1 + c/T \simeq \exp(c/T)$, and the variance of the innovation is not fixed, but tends to zero at speed $1/T^2$. This ensures smooth changes in the trajectory of the process, but automatically eliminates the uncertainty for large T.

vii) Ultra Long Run (ULR) Process

The ULR processes are Gaussian autoregressive processes defined over an ultra large time scale (see Gourieroux and Jasiak (2022) and Section 2.5 of the text). They differ from the LTU model by the speed of convergence to zero of the innovation variance. The appropriate choice of the rates of convergence of ρ_T to 1 and σ_T to 0 allows for a limiting **stationary unit root process**, that is an example of the singular component in the Wold decomposition of a stationary series. The rates of $\rho_T - 1$ and σ_T are of order $1/T$ and $1/\sqrt{T}$, respectively. Similar rates have been considered in the "dampered unit root" model of Gospodinov (2009), Gospodinov, Maynard, and Pesavento (2021), but without interpretation in terms of time scale.

APPENDIX B: THE LIMITS OF FREQUENCY ANALYSIS BY PROJECTION

In linear analysis of time series, the knowledge of the autocovariance function is equivalent to the knowledge of its spectral density, i.e. of its Fourier transform. Then the linear analysis can be performed either in the time domain (from the autocovariances), or in the frequency domain (from its spectral density). In the

frequency domain, it is common to assimilate the long run to the frequency close to zero, and to find an approximation of the long run component by projecting the series on low frequencies (see e.g. Engle (1974), Müller and Watson (2016, 2020) for a description of this approach, Müller and Watson (2018), Lunsford and West (2019), Section 11, for the use of low frequency correlation, or Andersen and Varneskov (2021), Section 3, for a local spectrum approach of predictive regressions). This practice has several drawbacks.

i) It is well-known that a Gaussian white noise, say, has a constant spectral density. Therefore its projection on low frequency is not zero. Then, with this approach a long run effect arises spuriously as the Gaussian white noise is the typical example of a pure short run component.

ii) The approach relying on unobserved components with different time scales is more advantageous, as it allows us to separate the components and to shock them separately, that is evaluate the impulse responses to shocks specific to the long run component (the so-called long run systemic shocks). This is not possible when impulses are applied directly to the aggregate time series, even frequency by frequency.

iii) The Fourier transformation is a linear approach, that cannot be easily extended to nonlinear dynamics (see Velasco and Lobato (2018); Velasco (2020) for recent attempts to analyze higher order dynamics in the frequency domain).

APPENDIX C: ACRONYMS

ACF : Auto-Correlation Function
CPI : Consumer Price Index
EMH : Efficient Market Hypothesis
HFD : High Frequency Data
I(1) : Integrated at order 1
IRP : Interest Rate Parity
LL : Local Level
LLN : Law of Large Numbers
LTU : Local-to-Unity
MDS : Martingale Difference Sequence
PPP : Purchasing Power Parity
RW : Random Walk
ULR : Ultra Long Run

REFERENCES

AHTOLA, J., AND G. C. TIAO (1984): "Parameter inference for a nearly nonstationary first-order autoregressive model," *Biometrika*, 71(2), 263–272.

ANDERSEN, T. G., AND R. T. VARNESKOV (2021): "Consistent inference for predictive regressions in persistent economic systems," *Journal of Econometrics*, 224(1), 215–244.

BANON, G. (1978): "Nonparametric identification for diffusion processes," *SIAM Journal on Control and Optimization*, 16(3), 380–395.

BERGER, J. O., AND R. L. WOLPERT (1988): "The likelihood principle," IMS.

BOUDOUKH, J., R. ISRAEL, AND M. RICHARDSON (2021): "Biases in long-horizon predictive regressions," *Journal of Financial Economics*.

CAVANAGH, C. L. (1986): *Roots local to unity*. Harvard Institute of Economic Research.

CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): "Inference in models with nearly integrated regressors," *Econometric theory*, 11(5), 1131–1147.

CHAMBERS, M. J. (2016): "The estimation of continuous time models with mixed frequency data," *Journal of Econometrics*, 193(2), 390–404.

CHAN, N. H., AND C.-Z. WEI (1987): "Asymptotic inference for nearly nonstationary AR (1) processes," *The Annals of Statistics*, pp. 1050–1063.

COMMANDEUR, J. J., AND S. J. KOOPMAN (2007): *An introduction to state space time series analysis*. Oxford University Press.

CONLEY, T. G., L. P. HANSEN, E. G. LUTTMER, AND J. A. SCHEINKMAN (1997): "Short-term interest rates as subordinated diffusions," *The Review of Financial Studies*, 10(3), 525–577.

COX, D. R. (1958): "Some problems connected with statistical inference," *Ann. Math. Statist.*, 29(2), 357–372.

DAHLHAUS, R. (2012): "Locally stationary processes," in *Handbook of statistics*, vol. 30, pp. 351–413. Elsevier.

DIAMOND, D. W., AND P. H. DYBVIG (1983): "Bank runs, deposit insurance, and liquidity," *Journal of political economy*, 91(3), 401–419.

DOU, L., AND U. K. MÜLLER (2021): "Generalized Local-to-Unity Models," *Econometrica*, 89(4), 1825–1854.

DURBIN, J., AND S. J. KOOPMAN (2012): *Time series analysis by state space methods*, vol. 38. OUP Oxford.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

ENGLE, R. F. (1974): "Band spectrum regression," *International Economic Review*, pp. 1–11.

FAMA, E. F. (1965): "The behavior of stock-market prices," *The journal of Business*, 38(1), 34–105.

——— (1970): "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, 25(2), 383–417.

FAMA, E. F., AND K. R. FRENCH (1988): "Permanent and temporary components of stock prices," *Journal of political Economy*, 96(2), 246–273.

——— (1989): "Business conditions and expected returns on stocks and bonds," *Journal of financial economics*, 25(1), 23–49.

GOSPODINOV, N. (2009): "A new look at the forward premium puzzle," *Journal of Financial Econometrics*, 7(3), 312–338.

GOSPODINOV, N., A. MAYNARD, AND E. PESAVENTO (2021): "Inference in Conditional Vector Error-Correction Models with Small Signal-to-Noise Ratio," .

GOURIEROUX, C., AND J. JASIAK (2019): "Robust analysis of the martingale hypothesis," *Econometrics and statistics*, 9, 17–41.

——— (2021): "Temporally Local Maximum Likelihood with Application to SIS Model," *arXiv preprint arXiv:2107.06971*.

——— (2022): "Long Run Risk in Stationary Structural Vector Autoregressive Models," *arXiv preprint arXiv:2202.09473*.

GOURIÉROUX, C., J. JASIAK, AND A. MONFORT (2020): "Stationary bubble equilibria in rational expectation models," *Journal of Econometrics*, 218(2), 714–735.

GOURIEROUX, C., A. MONFORT, AND J.-P. RENNE (2021): "Required Capital for Long-run Risks," *Available at SSRN 3907847*.

GOURIÉROUX, C., P. C. PHILLIPS, AND J. YU (2010): "Indirect inference for dynamic panel models,"

Journal of Econometrics, 157(1), 68–77.

GOURIÉROUX, C., AND J.-M. ZAKOÏAN (2017): "Local explosion modelling by non-causal process," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 737–756.

HARRISON, J., AND M. WEST (1991): "Dynamic linear model diagnostics," *Biometrika*, 78(4), 797–808.

HARVEY, A. C. (1990): "Forecasting, structural time series models and the Kalman filter," .

JIANG, G. J., AND J. L. KNIGHT (1997): "A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model," *Econometric Theory*, 13(5), 615–645.

LEROY, S. F. (1989): "Efficient capital markets and martingales," *Journal of Economic literature*, 27(4), 1583–1621.

LUNSFORD, K. G., AND K. D. WEST (2019): "Some evidence on secular drivers of US safe real rates," *American Economic Journal: Macroeconomics*, 11(4), 113–39.

MOON, H. R., AND P. C. PHILLIPS (2000): "Estimation of autoregressive roots near unity using panel data," *Econometric Theory*, 16(6), 927–997.

MÜLLER, U. K., AND A. NORETS (2016): "Credibility of confidence sets in nonstandard econometric problems," *Econometrica*, 84(6), 2183–2213.

MÜLLER, U. K., J. H. STOCK, AND M. W. WATSON (2020): "An Econometric Model of International Growth Dynamics for Long-horizon Forecasting," *The Review of Economics and Statistics*, pp. 1–47.

MÜLLER, U. K., AND M. W. WATSON (2016): "Measuring uncertainty about long-run predictions," *Review of Economic Studies*, 83(4), 1711–1740.

——— (2018): "Long-run covariability," *Econometrica*, 86(3), 775–804.

——— (2020): "Low-frequency analysis of economic time series," *preparation. In Handbook of Econometrics. Elsevier.*

PÁSTOR, L., AND R. F. STAMBAUGH (2012): "Are stocks really less volatile in the long run?," *The Journal of Finance*, 67(2), 431–478.

PESAVENTO, E., AND B. ROSSI (2006): "Small-sample confidence intervals for multivariate impulse response functions at long horizons," *Journal of Applied Econometrics*, 21(8), 1135–1155.

PHILLIPS, P. C. (1979): "The sampling distribution of forecasts from a first-order autoregression," *Journal of Econometrics*, 9(3), 241–261.

——— (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74(3), 535–547.

——— (1991): "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6(4), 333–364.

PHILLIPS, P. C., H. R. MOON, AND Z. XIAO (2001): "How to estimate autoregressive roots near unity," *Econometric Theory*, 17(1), 29–69.

PRIMICERI, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72(3), 821–852.

ROBINSON, G. K. (1977): "Conservative statistical inference," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3), 381–386.

SCHORFHEIDE, F., D. SONG, AND A. YARON (2018): "Identifying long-run risks: A Bayesian mixed-frequency approach," *Econometrica*, 86(2), 617–654.

SHEPHARD, N. (1994): "Local scale models: State space alternative to integrated GARCH processes," *Journal of Econometrics*, 60(1-2), 181–202.

SIMS, C. A. (1988): "Bayesian skepticism on unit root econometrics," *Journal of Economic dynamics and Control*, 12(2-3), 463–474.

——— (1991): "Comment by Christopher A. Sims onTo Criticize the Critics', by Peter CB Phillips," *Journal of Applied Econometrics*, pp. 423–434.

SIMS, C. A., AND H. UHLIG (1991): "Understanding unit rooters: A helicopter tour," *Econometrica: Journal of the Econometric Society*, pp. 1591–1599.

STOCK, J. H. (1991): "Confidence intervals for the largest autoregressive root in US macroeconomic time series," *Journal of monetary economics*, 28(3), 435–459.

VELASCO, C. (2020): "Identification and estimation of Structural VARMA models using higher order dynamics," *arXiv preprint arXiv:2009.04428*.

VELASCO, C., AND I. N. LOBATO (2018): "Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models," *The Annals of Statistics*, 46(2), 555–579.

WEST, M., AND J. HARRISON (1997): *Bayesian forecasting and dynamic models*. 2nd edition, Springer Verlag, New-York.