The Long Range Dependence Paradigm for Macroeconomics and Finance

Marc Henry
Paolo Zaffaroni

Discussion Paper #:0102-19

Department of Economics
Columbia University
New York, NY 10027

March 2002
The Long Range Dependence Paradigm for Macroeconomics and Finance

Marc Henry, Paolo Zaffaroni

ABSTRACT The long range dependence paradigm appears to be a suitable description of the data generating process for many observed economic time series. This is mainly due to the fact that it naturally characterizes time series displaying a high degree of persistence, in the form of a long lasting effect of unanticipated shocks, yet exhibiting mean reversion. Whereas linear long range dependent time series models have been extensively used in macroeconomics, empirical evidence from financial time series prompted the development of nonlinear long range dependent time series models, in particular models of changing volatility. We discuss empirical evidence of long range dependence as well as the theoretical issues, both for economics and econometrics, such evidence has stimulated.

Keywords and phrases. Self-similarity, time series, FARIMA, nonlinear time series, ARCH, stochastic volatility, arbitrage.

1 Introduction

[48] first pointed out that nonparametrically estimated power spectra of many economic variables, such as industrial production and commodity price indexes, suggested overwhelming importance of the low frequency components. [84] observed a self-similar behaviour in the distribution of speculative prices, and proposed continuous and discrete time fractional models, such as fractional Brownian motion [85]. He then developed the Hurst rescaled range (hereafter R/S) analysis, originally introduced in [83], for social measurement purposes. However, initial sizeable empirical success of the long-range dependence (LRD) concept in economics is certainly related to the autoregressive fractionally integrated moving average model (hereafter ARFIMA), proposed simultaneously by [64] and [51], and the increased flexibility it brings to the Box-Jenkins linear time series methodology, in particular for modeling macroeconomic time series since [30]. The relative success of the LRD concept in economics may also be attributed to the development of a rationale for its presence in macro-level economic and financial systems based on the aggregation of micro units. This was originally proposed in [95] and further developed in [49], both in the context of contemporaneous aggregation of heterogeneous AR(1) micro-units.

Estimation of fully parametric long memory models such as ARFIMA appeared cumbersome, especially in the time domain, and, moreover, practical estimation
of these models indicated that the estimated values of the long memory parameter was largely affected by the parametric specification of the short run dynamics part of the model, i.e. the ARMA component. Hence, an important reason which gave further impetus to the importance of LRD in economics, was the introduction by [40] of a semiparametric estimator (hereafter GPH) for the estimation of the degree of LRD in a time series. In fact, GPH is robust to many forms of complicated short run dynamics insofar as it is based on a frequency domain ordinary least squares regression of periodogram ordinates in a shrinking band of frequencies around zero. Although not formally developed, this possibility was first suggested by [51] and [68]. Due to its simplicity of implementation (it is based on a simple univariate linear regression), this estimation method was extensively applied to macroeconomic and financial time series well before of any satisfactory theoretical analysis of its asymptotic distributional properties, finally provided by [102]. More efficient and robust semiparametric estimators of long memory have then been proposed by [69], [99], [62], [65], [89], [4], [105] and others.

Applications to macroeconomic time series, discussed in section 2, are intimately linked with the extensive unit root literature and the use of a fractional specification within the linear Box-Jenkins representation to account for high persistence of shocks coupled with possible mean reversion of the levels (i.e. the observations themselves) in series of income, consumption and prices (inflation and exchange rates).

Empirical finance research had a tremendous impact in emphasizing the importance of the LRD paradigm, both empirically and theoretically. On the one hand, the availability of very large time series of high-frequency data, allowed easier and more convincing detection of long memory. On the other hand, some empirical findings prompted the development of nonlinear time series models apt to fit the empirical distribution of asset returns, synthesized in a number of well known stylized facts, including slow decay of sample autocorrelations of squared returns. This has stimulated research aiming at establishing asymptotic theory for such models and deriving their implications in terms of asset pricing and risk management in arbitrage-free theoretical frameworks. These issues are the focus of section 3.

2 Applications of Long Range Dependence to Macroeconomic Time Series

2.1 The Fractional Model

For the sake of discussing macroeconomic applications, the LRD paradigm may be suitably summarized with the following nonparametric specification for the time series $Y_t$, $t = 0, \pm 1, \ldots$, where the usual notation for income is used:

$$ Y_t = \mu + (1 - L)^{-d} X_t $$

(2.1)

$$ X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} $$
where \( a_0 = 1 \) and \( \sum_{j=0}^{\infty} |a_j| < \infty \), and the \( \varepsilon_j \)'s follow a martingale difference sequence. This specification includes the ARFIMA model in the case where

\[
a(z) = \sum_{j=0}^{\infty} a_j z^j = \frac{b(z)}{a(z)}
\]

(2.2)

where \( a(z) \) and \( b(z) \) are finite order polynomials with zeros outside the unit circle in the complex plane. Other parametric specifications of LRD include fractional Gaussian noise and an extension of the Bloomfield exponential model as in 3.19 of [100]. Neither, however, has shared the empirical success of the ARFIMA model.

The process \( Y_t \) is denoted \( I(d) \) in reference to the degree of integration, and fundamental properties of the series \( Y_t \) can be described in terms of interval regions for the parameter \( d \). \( Y_t \) is covariance stationary for \( d < 1/2 \) and invertible for \( d > -1/2 \). In such cases, let \( f(\lambda) \) be the spectral density of the process satisfying

\[
\rho_k := \text{cov}(Y_t, Y_{t+k}) = \int_{-\pi}^{\pi} f(\lambda) \cos(k\lambda) d\lambda
\]

(2.3)

for \( k = 0, \pm 1, \ldots \). When \( d = 0 \), the spectral density is finite and positive at zero and autocorrelations are summable. Usual rules of inference apply, in particular relative to moment estimation. The case where \( d < 0 \), called antipersistence, is characterized by a shrinking spectral density at zero frequency, and it is empirically relevant to the extent that it describes the behaviour of overdifferenced series, i.e. the result of first differencing in series that were mistakenly believed to have a unit root (i.e. a degree of integration \( d = 1 \)). LRD proper arises when \( d > 0 \) and it is characterized by a singularity in the spectrum at frequency zero

\[
f(\lambda) \sim e^{\lambda^{-2d}} \quad \text{as} \quad \lambda \to 0^+
\]

(2.4)

and, correspondingly, nonsummable autocovariances following

\[
\rho_k \sim e^{k^{2d-1}} \quad \text{as} \quad k \to \infty.
\]

(2.5)

(2.4) and (2.5) can be used as alternative (though not equivalent, see [122]) definitions of LRD. When \( d \geq 1/2 \), the variance of the process is infinite, and a non-integrable pseudo-spectrum can be defined with power also concentrated at the origin; and the unit root case is embedded as a special case of specification 2.1. A particularly interesting region for macroeconomic applications is \( 1/2 < d < 1 \) where the time series \( Y_t \) has infinite variance, it displays a high degree of persistence as exemplified by 2.5 and yet mean reverts in the sense, for instance, that the impulse response function is slowly decaying. In particular, shocks do not have full persistence because for \( d < 1 \), \((1 - z)^{1-d}a(z)\) has a unit root, so that the long run impulse response of the system is zero.

### 2.2 Estimation of Long Range Dependence in Linear Models

Consider a sample of size \( n \) from the process \( Y_t \). A survey on asymptotic theory relevant (in particular) to parametric estimation of the ARFIMA model is
given in [100]. Exact maximum likelihood is efficient in the Fréchet-Darmois-Cramér-Rao (hereafter FDCR) sense when the $\varepsilon_t$'s are normally and identically distributed ([26]). The Whittle approximate likelihood is asymptotically efficient under Gaussianity ([38]) and remains $\sqrt{n}$-consistent and asymptotically normal for possibly non-normal identically and independently distributed $\varepsilon_t$'s ([46]). A result that is particularly relevant to macroeconomic time series is the apparently greater robustness properties of the Whittle estimate in small samples when the mean is unknown ([21]).

Parametric exact and approximate maximum likelihood estimation, however, relies heavily on a correct specification of the short run dynamics, i.e. the $a(z)$ and $b(z)$ polynomials in 2.2. The semiparametric approach advocated by [51] and [68], and developed by [40], [69], [99], [102], [101] and others, relies only on (2.1) or, more generally, on (2.4), which only specifies the spectral density in a neighbourhood of zero, the frequency of interest. Like the Whittle likelihood, they are periodogram based, but, in accordance with the local specification, only a small proportion of all periodogram ordinates are used in the estimation of $d$. The choice of bandwidth $m$, the number of periodogram ordinates $I(\lambda_j) = (1/2\pi n) |\sum_{t=1}^n Y_t \exp(i\lambda_j)|^2$ used in the estimation, with $\lambda_j = 2\pi j/n$ and $1/m + m/n \rightarrow 0$, is crucial, as it drives the efficiency of the $\sqrt{m}$-consistent estimates. See [58] for a discussion of optimal and feasible bandwidth choice in the stationary region for those of the above mentioned estimates that are $\sqrt{m}$-consistent and asymptotically normal.

Even though the local Whittle proposed by [69] is more efficient (see [101] for asymptotic theory and [77] for a multivariate extension), and more robust (see [104]), only the GPH has been extensively used by practitioners*. [30] and [31] apply it to the analysis of output and consumption, [20] to exchange rates, [56] to inflation rates. In the first two cases, $\sqrt{n}$ order bandwidths are used. This ad hoc bandwidth choice was suggested by [40] for the stationary region. [66] prove that the mean squared error minimizing bandwidth $m$ is of order $n^{4/5}$ which is the rate upper bound for its class of estimators as shown in [44]. However, this concerns the stationary region $d < 1/2$, and most macroeconomic time series seem to be nonstationary.* [56] report results with choices of $m$ of order of magnitude $n^\alpha$, $1/2 < \alpha < 1$, and they find little variation in the estimated parameter values.

As mentioned earlier, the parameter region $1/2 < d < 1$ where the process is infinite variance but mean reverting is of particular interest for the analysis of macroeconomic time series. [118] extended the GPH asymptotic normality result of [102] to that region for Gaussian series. However, the behaviour of the GPH across the unit root $d = 1$ value is problematic, as Monte Carlo results from [67] point to a possible inconsistency of the GPH for $d > 1$. [119] proposes a modified

---

* [105] propose a unifying framework for the GPH and the local Whittle estimates within a class of M-estimates, and improve efficiency with the use of higher-order kernels.

† In any case, orders of magnitude are poor guides in small samples when the constant is unknown. Indeed, [113] argues that the [30] result showing strong mean reversion in output is due to a mis specification of short run dynamics through too large a choice of bandwidth ($m = 11$ for a sample size of $n = 272$).
asymptotically normal version based on smoothing and tapering which robustifies
the estimate against linear trends -a particularly appealing feature given that
one of the major applications of the LRD paradigm has been to contribute to
the debate as to whether GNP is difference stationary or trend stationary. As
for small sample behaviour of the GPH and other semiparametric estimates as
compared to fully parametric ones, the crucial trade-off between robustness and
efficiency in the choice of bandwidth\(^1\) had been little documented in small sample
Monte Carlo investigations (see for instance [114], [58]).

2.3 Long Range Dependence in Aggregate Macroeconomic Series:
Evidence and Rationale

One of the major debates in macroeconomic research concerns the persistence
of economic shocks on income and the nature of the long cycles observed in US
output (Kondratiev, Kuznets or Juglar cycles) and documented in [70] and oth-
ers. [1] was the first to propose the use of the LRD paradigm to analyze such
long cycles and other low frequency characteristics of income series. Now, as [48]
noted, one of the more pervasive characteristics of macroeconomic time series is
a concentration of power (or variance) in low frequencies, which he dubbs “the
typical spectral shape of economic variables”. As [100] notes, such explosion of
spectral estimates at zero frequency would be consistent with the presence of a
unit root (and total persistence of shocks) in the levels of the series, if it were
not often associated, as is the case with the Beveridge Wheat Price Index ([11]),
with first differenced series with very low power around zero frequency, consist-
tent with \(I(1)\) behaviour, \(d < 0\). This would suggest that the original series were
integrated of order less than one. Besides, unlike the case of financial series with
efficient markets hypotheses, there are no theoretical underpinnings for an exact
unit root in macroeconomic time series as opposed to mean reverting fractional
alternatives. \(I(0)\) behaviour with \(0 < d < 1\) is also suggested by the conflicting
findings concerning aggregate price series from nine industrialized countries ana-
lyzed in [8], where the null of a unit root is rejected by Augmented Dickey-Fuller
procedures and the null of \(I(0)\) is rejected by KPSS tests.

An immediate explanation for evidence of LRD in aggregate price series is the
inheritance through agricultural prices of the dependence structure of geophysical
time series such as rainfall, riverflow and climactic series in which [83], [72], [63]
among others have found LRD to be pervasive. LRD shocks (possibly inherited
from underlying geophysical processes) in a real business cycle model of the
economy (see [71]) can account in the same way for the presence of LRD in
aggregate income series. However, a more appealing explanation for long memory
in income series is based on aggregation of output from a large number of sectors
each submitted to white noise shocks. [5] similarly justify the presence of LRD
in inflation series.\(^5\)

---

\(^1\)Reducing bandwidth increases the robustness of asymptotic results not only against
misspecification of short run dynamics, but also against departures from Gaussianity in
the \(z_t\)'s (see [104] and [19]).

\(^5\)In most existing literature on interest rates, the short rate process has the property
These justifications are based on an aggregation result in [95] and [49] with the more general implication that LRD may be found in the aggregate produced from a large number of heterogeneous autoregressive processes describing the microeconomic dynamics of each unit, e.g. the behaviour of each agent when heterogeneity among agents is allowed. In the generalized framework of [75], the results involving the memory can be summarized as follows. Focusing on AR(1) for simplicity's sake, let \( Y_{n,t} = \frac{1}{n} \sum_{i=1}^{n} y_{i,t} \) be the result of the aggregation of the \( n \) micro variables

\[
y_{i,t} = \alpha_i y_{i,t-1} + u_t + \epsilon_{i,t}, \quad i = 1, \ldots, n
\]

where \( u_t \) is a common shock and \( \epsilon_{i,t} \) is an idiosyncratic shock. The latter represents a source of time-varying heterogeneity. Assume that the shocks are i.i.d. white noise mutually independent. Heterogeneity across the coefficients \( \alpha_i \), assumed i.i.d. drawn from a nonnegative r.v., is defined by means of a mild semi-parametric specification of the latter's probability density function, say \( B(\alpha) \). As \( n \to \infty \), \( Y_{n,t} \) displays short range dependence, with exponentially decaying autocorrelation function (ACF), when the support of \( B(\alpha) \) is restricted to \( [0, \gamma) \), \( \gamma < 1 \). It may exhibit LRD when \( B(\alpha) \) has support \( [0, 1) \) and in particular, \( Y_{n,t} \) has, asymptotically, infinite variance LRD when \( B(\alpha) \) is concentrated enough in the neighbourhood of 1. Within an economic model with heterogeneous agents behaving according to the standard profit-maximization principle, [87] considers an economic model where agents face heterogeneous and time-varying costs of adjustment in setting their optimal level of output. In this context, he finds that cross-sectional distributions with the required shape in order to induce long memory in the aggregate might arise endogenously.

### 2.4 Long Range Dependence and Persistence

In the vast literature that followed [90], some consensus seems to have arisen on the fact that US postwar GNP is well represented by low order ARIMAs with a single unit root, mostly as a result of failure to reject null I(1) hypotheses against autoregressive alternatives. However, the non-standard asymptotics and lack of Pitman efficiency (as noted in [42]) as well as evidence in [31] that some of the traditional methods have low power against fractional alternatives, have prompted the use of procedures based on LRD specifications with continuous parameter values across the value of interest \( d = 1 \), such as [98], [30], [113] and [42] typically fail to discriminate between a unit root and mean reverting that correlation between short rates \( n \) periods apart goes to zero exponentially with \( n \). This is not supported by the data, as considerable residual variation is found in long yields. [5] partially resolve the puzzle with a long memory modeling of the short rate process and justify this assumption as either inherited from long memory in money growth, inflation and monetary policy in general, or as the effect of aggregation across agents with heterogeneous beliefs. They use a fully specified time domain ML procedure and cite [112].

See [93] for a survey on unit root testing.
fractional alternatives in a variety of US income series (including an extended version of the [90] data) due to the lack of data points available.

In spite of inconclusive inference results, the introduction of the LRD paradigm in the debate over the persistence of output has been instrumental in partially resolving empirical inconsistencies resulting from the unit root (and fully persistent) representation for output. First of all, conflicting evidence on persistence measures from unrestricted ARIMA and Unobserved Component models (in [18] and [120]) is discussed in a LRD framework by [30]. The second empirical inconsistency we mention a little more at length: it concerns the conflicting evidence of a unit root in income and the convergence of output to a steady state. Most empirical studies based on growth regressions point to a 2% uniform exponential rate of convergence to a long run steady state (see for instance [9] and [86]). This finding is inconsistent with full persistence of shocks implied by a unit root in income, but not with the type of persistence implied by infinite variance but mean reverting LRD income, as demonstrated by [88] in the framework of a Slow-Swan growth model. They show that an I(d) income with 1/2 < d < 1 could produce spurious evidence of uniform exponential convergence (as in the data considered by previous studies), while the actual hyperbolic rate of convergence would be too slow for standard unit root tests to reject the null of a unit root against covariance stationary alternatives.\footnote{Such a specification for income also accounts for the rejection of convergence in [10] based on cointegration tests, as, in case income is I(d), d < 1, the latter may be misspecified.}

The third empirical inconsistency we mention is the excessive smoothness observed in consumption series compared to the implications of the Permanent Income Hypothesis (hereafter PIH) based on a unit root specification for income and under rational expectations. This excess smoothness, or Deaton's Paradox, is presented in [27] who stresses that, with a fully persistent specification of income, the PIH under rational expectations implies that innovations to consumption should have larger variance than innovations to income. More precisely, if income follows model 2.1 with \( d = 1 \), then the PIH predicts (see for instance [53]) that consumption will react to innovations on income as \( \Delta C_t = \kappa_\infty \epsilon_t \), where \( \kappa_\infty = \sum_{j=0}^{\infty} \beta^j a_j \) is the long run discounted impulse response (\( \beta \) is the discount factor). [27], and [17] show that under a variety of ARIMA specifications for income and reasonable assumptions on the interest rate, \( \kappa_\infty \) is generally larger than 1, which is another way of stating Deaton’s Paradox. [32] point out that if \( d < 1 \), however, the impulse response function is decaying, which allows for a smoothed response of consumption to income shocks, which is more in line with the rationale of the PIH. [57] presents a similar resolution of the Deaton Paradox with postulated values of \( d \), and tests whether these paradox resolving values of \( d \) are consistent with observed income processes. However, he uses a test based on the [76] modified R/S statistic which has no clear optimality properties (see [100]), and seems to have undesirable small sample properties (see [117]).
2.5 Long Range Dependence and Co-persistence

Within the LRD paradigm, the standard definition of cointegration can be extended to allow for fractional integration in both the original variables and the cointegrated residuals, as proposed by [50], [99], [47]. This provides a generalized framework for the analysis of co-persistence which has been particularly fruitful in the description of exchange rates from the post Bretton Woods currency float, and in an empirical reappraisal of the purchasing power parity equilibrium (hereafter PPP). [29] challenge the view that deviations from purchasing power parity can be suitably modeled by martingale differences. They model exchange rates as ARFIMA series which they estimate with the Whittle quasi-likelihood, and find evidence of mean reversion in deviations from parity. A similar study is found in [22]**. [106] discuss parametric and semiparametric procedures for the estimation of various forms of fractional cointegration and establish their asymptotic distributional properties.

3 Long memory implications in finance

3.1 Long memory in asset prices

Following the mixed empirical success of the LRD paradigm in macroeconomics, a great deal of research has focused on possible evidence of LRD in financial asset returns defined as \( x_t = \log(P_t/P_{t-1}) \), where \( P_t \) is the price of the asset. [84] first suggested the possibility that asset prices could exhibit LRD. Using the classical R/S analysis, [52] uncovered significant statistical evidence of LRD in daily returns on securities listed at the New York Stock Exchange. These results were challenged by [76] who proposed a modified R/S statistic, robustifying for possible additional short memory dynamics by taking into account the first \( q \) lags of the autocovariance of the observables. He concluded that there is no such clear evidence of LRD in the levels of asset returns. [117], however, show that Lo’s modified procedure has dramatic power problems, depending on the assumed degree of short memory in the data. In particular, the test has rapidly decreasing power against LRD alternatives as \( q \) increases. [73] use more efficient semiparametric techniques supporting the possibility of LRD, but the results are critically re-examined in [78].

Hence, the empirical evidence concerning LRD in the levels of financial asset returns is far from being clear-cut. Note that most of modern finance theory is based on the martingale difference assumption for the levels of asset returns, in turn a by-product of a martingale assumption for log-prices and therefore the simple presence of autocorrelated returns data in levels, however weak, poses considerable problems. Indeed, there is some empirical evidence of significant autocorrelation at the first few lags often imputed to non-synchronous trading and other market microstructure effects; see e.g. [3] for forex returns and [34] for returns on the Standard & Poor 500 index (S&P 500).

**See also [23], [6], [20].
There is general agreement on the fact that asset returns exhibit more memory in the squares than in the levels. This stylized fact was refined to the presence of LRD in many nonlinear transformations of absolute returns (including squared and log-squared returns). Using daily data of the S&P 500 closing price index, [34] show that there is significant autocorrelation for lags up to 10 years (approximately 2,500 lags with daily data) when considering the squares return $x_t^2$ and more generally power transformations such as $|x_t|^\alpha$ for various positive values of $\alpha$. Considering a suitable $\alpha < 2$, the ACF of $|x_t|^\alpha$ would be well defined even when the $x_t$ have unbounded fourth moment. A slow rate of decay of shocks to the conditional variance of the S&P500 is also found in [39]. This pattern is consistent with theoretical autocovariances of the $x_t^2$ decaying hyperbolically.

This finding has been corroborated since the availability of high frequency tick-by-tick data. [59], [60], [3] and [2] consider the returns from using the foreign exchange spot rates DM/$ and Yen/$ and, using semiparametric estimators such as the GPH, the local Whittle and the average periodogram estimator of [99], find evidence of covariance stationary LRD in the squared, log-squared and absolute returns. [59] and [60] use the optimal bandwidth selection procedures discussed in [58].

Since [115], it has been thought that the existence of a stable relationship between stock returns and trading volumes could be exploited in forecasting speculative prices dynamics. Using daily data for the thirty stocks which contribute to the Dow Jones industrial average index, [79] provides empirical evidence of long memory in stock return volatility and trading volume, with the same degree of memory, although the hypothesis of fractional cointegration was rejected. A semiparametric estimator of the coherency was used.

Unlike macroeconomic time series, financial data (particularly high-frequency) makes it very natural to use semi and nonparametric estimators, as the usual efficiency loss is compensated by the possibility of using very long spans of data. This, in part, explains why fully parametric long memory models have not been advocated in order to detect, as a preliminary analysis, long memory in asset returns, using in turn the levels and the squares as observables. However, once the general statistical properties of the data are uncovered, precise and efficient estimation of the volatility of asset returns is crucial for valuation of derivative securities linked to such assets (such as options and futures). Furthermore, accurate forecasting of volatility of financial assets is a key ingredient for risk management. Nonparametric probability density estimators for long memory variates (see [96]) do extend to estimation of conditional moments but only for a finite number of lagged dependent variables. On the other hand, semiparametric long memory estimators, such as [102] and [101], focus precisely on long-horizon past information but ignore short run dynamics. Therefore, in order to model the intrinsically non-Markovian dynamics of LRD in conditional variance, and for practical pricing of financial assets, fully parametric long memory models of changing volatility must be considered.
3.2 Long memory volatility models

Financial asset returns display a number of regularities across assets and periods, in particular dynamic conditional heteroskedasticity. This clearly cannot be accounted for by linear models. Furthermore, as originally pointed out by [37] and [12], the data exhibit non-Gaussian kurtosis and dynamic asymmetries such as the ‘leverage effect’, expressed by a negative cross-correlation between current returns ($x_t$) and future squared returns ($x^2_{t+u}$ with $u > 0$). The ARCH model of [35] represents the most famous nonlinear time series model apt to account for such features, except for dynamic asymmetries. Denoting by $\mathcal{F}_t$ the $\sigma$-field of events generated by $\{x_s: s \leq t\}$ and setting $\sigma^2_t = \text{var}(x^2_t | \mathcal{F}_{t-1})$, the ARCH($p$) postulates that

$$\sigma^2_t = \omega + \alpha_1 x^2_{t-1} + \ldots + \alpha_p x^2_{t-p} \ a.s., \quad (3.1)$$

where $\omega > 0$, $\alpha_i > 0 (i = 1, \ldots, p)$. Therefore, the conditional variance is a symmetric function of past observations. A plethora of extensions have been developed in order to take other features into account, in particular dynamic asymmetries; see [13] for a recent survey.

[97] generalized (3.1) to

$$\sigma^2_t = \tau^2 + \sum_{j=1}^{\infty} \psi_j (x^2_{t-j} - \tau^2) \ a.s., \quad \sum_{i=1}^{\infty} \psi^2_j < \infty, \quad (3.2)$$

where $\tau^2 > 0$ and $\psi_j \geq 0$. This includes (3.1) when $\psi_j = 0$ for all $j > p$. [97] introduced the ARCH($\infty$) model (3.2) and

$$\sigma^2_t = \left( \omega + \sum_{j=1}^{\infty} \phi_j x_{t-j} \right)^2, \quad a.s., \quad \sum_{i=1}^{\infty} \phi^2_i < \infty, \quad (3.3)$$

as alternative hypotheses in deriving score tests of no-ARCH with optimal efficiency against such alternatives. Note that for (3.3) there is no need to impose nonnegativity on $\omega$ and on the $\phi_i$. Model (3.3), called linear ARCH (LARCH), was further considered in [45] who characterize the low-order statistical properties of the model. In particular, they establish sufficient conditions for LRD in the $x^2_t$ and weak convergence of their partial sums to fractional Brownian motion $B_H(t)$.

In order to impose martingale difference levels and autocorrelated squares, the data generating process for $x_t$ is routinely defined by

$$x_t = \epsilon_t \sigma_t, \quad (3.4)$$

where the so-called rescaled innovations $\epsilon_t$ are i.i.d. with $E(\epsilon_t) = 0$ and $E(\epsilon^2_t) = 1$.

[107] discuss LRD parameterizations for the $\psi_j$ in (3.2). An important parametric version is

$$1 - \sum_{j=1}^{\infty} \psi_j z^j = (1 - z)^d \frac{b(z)}{a(z)}, \quad (3.5)$$
where \( d \geq 0 \) and \( a(z) \) and \( b(z) \) are defined as in (2.2). Setting \( \psi(z) = 1 - \sum_{j=1}^{\infty} \psi_{\phi} z^{j} \) and \( \phi = \tau^{2} \psi(1) \), (3.2) can be rewritten as
\[
\sigma_{t}^{2} = \phi + (1 - \psi(L)) x_{t}^{2}.
\]
Note that in this case (3.5) implies \( \phi = 0 \).

Relaxing the condition \( E(\epsilon_{t}^{2}) = 1 \), [43] show that when
\[
(E(\epsilon_{t}^{4}))^{1/2} \sum_{i=1}^{\infty} \psi_{i} < 1,
\]
there exists a strictly stationary solution to (3.2) with bounded fourth moment. (3.6) allows for hyperbolically decaying \( \psi_{j} \), although parameterizations such as (3.5) are excluded when \( E(\epsilon_{t}^{2}) = 1 \). Moreover, they show that (3.6) implies summability of the ACF of the \( x_{t}^{2} \), ruling out LRD.

The ARCH(\( \infty \)) (3.2)-(3.4) with the parameterization (3.5) was considered in [7] but letting \( \phi \) be a free parameter. This implies that when both \( d, \phi > 0 \) (and \( E(\epsilon_{t}^{2}) = 1 \)) the \( x_{t}^{2} \) are not covariance stationary. They propose Gaussian pseudo-maximum likelihood estimation (PMLE) of the model. However, no formal asymptotic distribution theory yet exists for the PLME nor for exact MLE (making distributional assumptions on the rescaled innovations \( \epsilon_{t} \)). This holds irrespective of whether \( \phi \) is considered a free positive parameter or not and for any chosen finite parameterization of the \( \psi_{j} = \psi_{j}(\theta) \), \( \theta \) denoting a \( p \)-dimensional parameter, including (3.5).

Indeed, even when focusing on exponentially decaying \( \psi_{j} \), the asymptotic properties of the Gaussian PMLE are known only for the ARCH(\( p \)) (see [121]) and the GARCH(1,1) (see [74] and [80]), obtained when \( p = 2 \) and \( \psi_{j} = \alpha \beta^{j-1} \) for some nonnegative parameters \( \alpha, \beta \) satisfying the strict stationarity condition, \( E \log(\beta + \alpha \epsilon_{t}^{2}) < 0 \) (see [91]). In a linear long memory semiparametric framework, [104] show that the asymptotic properties of the local Whittle estimator [69] are unchanged when relaxing conditional homoskedasticity of the linear innovations to ARCH(\( \infty \)) dynamic conditional heteroskedasticity.

The ARCH(\( \infty \)), and more generally ARCH-type models, can be seen as nonlinear autoregressive models. One can also consider nonlinear moving average (MA) models. In order to take into account the ‘leverage effect’ and to relax nonnegativity constraints on the conditional variance parameters, [92] introduced the EGARCH, given by (3.4) and
\[
\log \sigma_{t}^{2} = \omega + \sum_{k=1}^{\infty} \beta_{k} g(\epsilon_{t-k}), \quad \sum_{k=1}^{\infty} \beta_{k}^{2} < \infty.
\]
The scalar function \( g(\cdot) \) defines the so-called news impact curve \( g(\epsilon_{t}) = \theta \epsilon_{t} + \gamma (| \epsilon_{t} | - E | \epsilon_{t} | ) \), for scalar parameters \( \gamma, \theta \). Dynamic asymmetry requires \( \theta \neq 0 \) and identifiability \( \beta_{1} = 1 \). [92] mentioned the possibility of long memory parameterizations for the \( \beta_{k} \), further developed in [14] who proposed estimation of the model by Gaussian PMLE. However, the asymptotic properties of the Gaussian PMLE are unknown both for short and long memory parameterization of the EGARCH model.
An alternative nonlinear MA has been proposed by [107]:

\begin{align}
  x_t &= \epsilon_t h_t, \\ h_t &= \rho + \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j}, \quad \sum_{j=1}^{\infty} \alpha_j^2 < \infty,
\end{align}

(3.8) (3.9)

generalizing the nonlinear MA(1) of [94]. This yields \( \sigma_t^2 = \text{var}(\epsilon_t) h_t^2 \). The choice of the \( \alpha_i \) defines the memory properties of the \( x_t^2 \), which are weaker or equal than the ones of the \( h_t \) depending on whether \( \rho \) is equal to zero or not. Focusing on a wide class of long memory parameterization, including the case where the \( \alpha_j \) are the moving average coefficients of the ARFIMA, [124] establishes \( n^{1/2} \)-consistency (\( n \) denotes sample size) and asymptotic normality of the Whittle estimator of the model, when fitting the spectral density of the squares \( x_t^2 \).

The nonlinear AR and MA models just described can be expressed as nonlinear transformations of present and lagged values of the \( \epsilon_t \). Therefore, these models could be defined as ‘one-shock’ models. In principle, one can invert the models and express the unobservable innovation, the \( \epsilon_t \), as a nonlinear function of present and lagged values of the observable, the \( x_t \), although formal proofs of invertibility may prove very difficult to establish.

Motivated by the fact that they seem to provide a closer approximation to continuous time pricing formulae of modern finance, Stochastic Volatility (SV) models, introduced by [116], replace (3.4) with

\[ x_t = \eta_\epsilon \sigma_t, \]

(3.10)

and set

\[ \log \sigma_t^2 = \mu_0 + \sum_{i=1}^{\infty} \mu_i \epsilon_{t-i}, \quad \sum_{i=0}^{\infty} \mu_i^2 < \infty, \]

where \( \{ \eta_i \} \) represents an i.i.d. sequence with \( E(\eta_i) = 0, \text{var}(\eta_i) = 1 \), different from the \( \epsilon_i \). Hence, SV-type models can be defined as ‘two-shock’ models. In most cases it is assumed that \( \eta_i \) and \( \epsilon_i \) are independent of one another, although this rules out the possibility to account for dynamic asymmetries. SV challenge ARCH-type models in modeling the empirical distribution of asset returns, see [41] for a complete survey. Although the decoupling of the two shocks makes SV more attractive for moment formulae, the latent nature of the conditional variance complicates the evaluation of the conditional variance through filtering and, more importantly, of the likelihood function itself.

A long memory SV model was introduced by [55], setting

\[ \log \sigma_t^2 - \mu_0 = (1 - L)^{-d} \epsilon_{t-1}, \quad 0 < d < 1/2, \]

and proposed estimation by a discrete Whittle estimator, fitting the model spectral density of \( x_t^2 \). [16] show consistency of this estimator and [28] show asymptotic normality of the GPH estimator for this model. [24] discuss statistical properties and temporal aggregation issues relative to a continuous time long memory SV model.
A ‘two-shock’ nonlinear MA was introduced in [108] which allows time-varying first and second conditional moments. It is defined by

\[ x_t = g_t + \eta_t h_t, \]

where the bivariate process \( g_t, h_t \) is independent of the \( \eta_t \), defined in (3.10). The memory properties of the \( x_t \) and \( x_t^2 \) are characterized for a general specification of the \( g_t, h_t \). The linear specification, with \( h_t \) given by (3.9) and \( g_t = \mu + \sum_{i=0}^{\infty} \beta_i \epsilon_{t-i} ; \sum_{i=0}^{\infty} \beta_i^2 < \infty \), seems the most appealing specification, prior to finite parameter modeling for estimation purposes.

Focusing on Gaussian \( \epsilon_t \) and \( \eta_t \), [103] shows how all these SV-type models can be nested in a general framework, setting \( x_t = f(\delta_t) \) where \( \delta_t \) is a \( q \)-dimensional stationary Gaussian process and \( f(\cdot) \) an arbitrary function \( f : R^q \rightarrow R \) such that \( E(f^2(\delta_t)) < \infty \). An asymptotic expansion for the ACF of the \( x_t \) so defined is provided, which allows to distinguish the impact of the nonlinear transformation \( f(\cdot) \) and of the memory of the \( \delta_t \), respectively, on the memory of the \( x_t \). In particular, assuming the \( \delta_t \) exhibit LRD, primitive conditions on the function \( f(\cdot) \) are established such that the \( x_t \) would exhibit the same or a weaker degree of LRD, respectively.

Despite the numerous LRD volatility models so far proposed, very little is known in terms of the possible sources of LRD in squared asset returns. Inspired by the corresponding linear time series analysis (cf. section 2.3), the contemporaneous aggregation mechanism seems to provide an important vehicle. [33] and [3] employ the linear aggregation results to a GARCH and SV setting, respectively. In both cases, the aggregate is not truly defined as the cross-sectional arithmetic average. In a strong GARCH setting, [123] establishes the probabilistic properties of the aggregate, defined as the cross-sectional arithmetic average of heterogeneous GARCH. In particular, it is shown that the aggregation mechanism never induces LRD in the conditional variance of the aggregate.

### 3.3 Pricing Implications

The principle of no arbitrage is the cornerstone of all pricing models of modern finance, since the seminal work of [54], [84] first analyzed the pricing implications of long memory. In a single asset framework, he assumed that the first difference of the fundamental follows a linear process, \( P_0(t) - P_0(t-1) = \sum_{i=0}^{\infty} \alpha_i N(t-i) \), for some i.i.d. white noise sequence \( \{ N(t) \} \) and \( \sum_{i=1}^{\infty} \alpha_i^2 < \infty \). Further, it was assumed that arbitrage would induce the speculative price of the asset, \( P(t) \), to be a linear function of the fundamental and to display the martingale property, implying \( P(t) - P(t-1) = N(t)(\sum_{i=0}^{\infty} \alpha_i) \) under risk neutrality and a zero risk-free interest rate. Hence, LRD of the differenced fundamental, i.e. non-summability of the \( \alpha_i \), directly implies the impossibility of implementing the no arbitrage principle.

Although with ambiguous results (cf. section 3.1), the empirical research that focused on the possibility of LRD in the levels of asset returns prompted the use of fractional Brownian motion as an alternative to the standard Brownian motion model. However, [81] first pointed out that fractional Brownian motion
of type $B_H(t)$ ($0 < H < 1, 0 \leq t \leq 1$) is not a feasible model for asset returns. In fact, $B_H(t)$ is not a semimartingale unless $H = 1/2$ for which $B_{1/2}(t) = B(t)$. For the case $1/2 \leq H < 1$ see also [25] and more recently [111], [109] extended the result to the case $0 < H < 1/2$.

Without the semimartingale property, the standard approach for pricing contingent claims, based on the equivalent martingale measure, fails. This result stems from the fact that the order-$q$ variation of $B_H(t)$ satisfies (cf. [109, eq. (2.2)]), as $n \to \infty$,

$$
\sum_{j=1}^{2^n} |B_H(j2^{-n}) - B_H((j-1)2^{-n})|^q \to_p \begin{cases} 
0, & qH > 1, \\
\infty, & qH < 1,
\end{cases}
$$

consistent with a semimartingale behaviour for $H = 1/2$ only. [109] shows that the failure of the semimartingale property characterizes all Gaussian processes $\{X(t)\}$ of the form

$$
X(t) = \int_{-\infty}^{t} \phi(t-s)dB(s) - \int_{-\infty}^{0} \phi(-s)dB(s), \quad t \in \mathbb{R}^+,
$$

with square integrable kernel $\phi(\cdot)$ satisfying $\phi(t) \sim ct^{H-1/2}$ as $t \to 0^+$. $B_H(t)$ represents a particular case setting $\phi(t) = t^{H-1/2}$. As a consequence, it follows that choosing $\phi(t)$ such that $\phi(0) = 1$, $\phi'(0) = 0$ and $0 < \phi''(t)t^{5/2-H} = O(1)$, $t \to \infty$, one can construct Gaussian processes which exhibit the same LRD as $B_H(t)$ while retaining the semimartingale property [109, section 5], [110] addressed the issue of arbitrage opportunities in a more general framework, with respect to the class of processes with continuous sample paths and bounded $q$-variation ($1 \leq q < 2$). [25] and [111] discuss fractional version of the Black-Sholes option pricing formula, based on $B_H(t)$. These developments, which use $B_H(t)$ for describing dynamics of asset prices, could still be useful, though, as the cornerstone assumption of zero transaction costs is likely to fail in practice.

When considering the possibility of LRD in squared returns with martingale difference levels (cf. section 3.2), the above mentioned implications in terms of the possibility of arbitrage do not apply. Therefore the standard risk-neutral valuation methods can be used. [24] discuss option pricing formulae based on a continuous time long memory SV model.

On the other hand, one must be aware of these implications when considering volatility models characterized by a time-varying conditional mean specified, since [36], as a linear function of the square-rooted conditional variance when the latter displays long memory. [82] and [61] proposed continuous time subordinated processes for financial asset returns, apt to display heavy tails and LRD in absolute-valued and squared returns.

Practical use of long memory asset pricing formulae in empirical applications is not so widespread yet. Using daily data of exchange traded long-term equity anticipation securities on the S&P 500 index, [15] propose an empirical applications of pricing long term options, parameterizing the conditional variance with the so-called fractionally integrated GARCH ((3.7) coupled with $\sum_{k=0}^{\infty} \beta^k L^k = \psi^{-1}(L)$ as from (3.5)).
Acknowledgments

The first author acknowledges support from a “Prix à Savant” of the Université Catholique de Louvain.

References


1. The Long Range Dependence Paradigm for Macroeconomics and Finance


1. The Long Range Dependence Paradigm for Macroeconomics and Finance


[80] Lumsdaine, R. (1996): “Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models,” *Econometrica*, 64, 575-596.


1. The Long Range Dependence Paradigm for Macroeconomics and Finance


Marc Henry, *Department of Economics, Columbia University, 420 W 118th Street, New York, NY 10027, U.S.A., e-mail: marc.henry@columbia.edu*

Paolo Zaffaroni, *Department of Applied Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DE, UK, e-mail: paolo.zaffaroni@econ.cam.ac.uk*