

Why Frequency Matters for Unit Root Testing

H. Peter Boswijk

Department of Quantitative Economics, Universiteit van Amsterdam,

1018 WB Amsterdam, The Netherlands (H.P.Boswijk@uva.nl)

Franc Klaassen

Department of Economics, Universiteit van Amsterdam,

1018 WB Amsterdam, The Netherlands (F.J.G.M.Klaassen@uva.nl)

Abstract

It is generally believed that the power of unit root tests is determined only by the time span of observations, not by their sampling frequency. We show that the sampling frequency does matter for financial time series displaying fat tails and volatility clustering. Our claim builds on recent work on unit root testing based on non-Gaussian GARCH-based likelihood functions. Such methods yield power gains in the presence of fat tails and volatility clustering, and the strength of these features increases with the sampling frequency. This is illustrated using Monte Carlo simulations and an empirical application to real exchange rates.

KEY WORDS: Fat tails; GARCH; Mean reversion; Sampling frequency; Systematic sampling; Purchasing-power parity.

1 INTRODUCTION

Testing purchasing-power parity is one of the main applications of unit root and cointegration analysis. Although some researchers have tried to address this problem by checking whether nominal exchange rates and prices (or price differentials) are cointegrated in a multivariate framework, many others have focussed on the question whether real exchange rates contain a unit root. The fact that a unit root quite often cannot be rejected, so that no significant mean reversion in real exchange rates is found, is often not considered to be decisive evidence against purchasing-power parity. In contrast, the insignificant test result is typically explained by the notoriously low power of unit root tests, together with the fact that the tendency towards purchasing-power parity is so weak that it is not detected by conventional unit root tests. To solve this problem, researchers have tried to increase the power of these tests by obtaining more data; either by considering a longer time span (of a century or more) of data, or by considering a panel of exchange rates, exploiting cross-country restrictions, or a combination of these; see Frankel (1986), Frankel and Rose (1996), Lothian and Taylor (1996), and Taylor (2002).

Another way to obtain larger sample sizes is to consider the same time span, but using data observed at a higher frequency. However, it is generally believed that this route will not lead to more power, because the power of unit root tests is mainly affected by time span, and much less by sampling frequency. This result was first derived by Shiller and Perron (1985), and repeated in the influential review paper by Campbell and Perron (1991). At an intuitive level, finding significant mean reversion requires a realization of a process that indeed does pass its mean quite regularly within the sample; increasing the sampling frequency does not change this in-sample mean-reversion, whereas a longer history of the time series will demonstrate more instances of passing the mean.

At a more formal level, the theoretical basis for the result of Shiller and Perron (1985) is provided by a continuous-record asymptotic argument. Suppose that the data may be seen as discrete observations from the continuous-time Ornstein-Uhlenbeck process (i.e., a continuous-time Gaussian first-order autoregression). Then the power of a unit root test based on the discrete observations may be approximated by the asymptotic local power, which is essentially the power of a likelihood ratio test for reducing an Ornstein-Uhlenbeck

process to a Brownian motion process, and the latter power is solely determined by the time span and the mean-reversion parameter. Therefore, this approximate local power is the same, whether we consider 10 years of quarterly data or 10 years of daily data, say. Although in practice, the finite-sample power differs somewhat between these cases, this difference is negligible relative to the power gains that may be obtained from a longer time span.

The results mentioned so far refer to systematic sampling, where the low-frequency observations are obtained by skipping intermediate high-frequency observations. This type of sampling is relevant for stock variables, but also for asset prices, which are typically defined as end-of-day, end-of-week or end-of-month prices, so that their logarithmic first difference defines the return over the preceding day, week, or month, respectively. Choi (1992) shows that temporal aggregation of flow variables, where the low-frequency observations are defined by cumulating the intermediate high-frequency observations, can have an effect on the power of unit root tests, because of the necessity to account for higher-order dynamics in the aggregated data. However, this is only a finite-sample effect, which vanishes asymptotically, as can be derived from the theoretical results of Chambers (2004).

The present paper argues that the effect of systematic sampling on the asymptotic local power of unit root tests is not negligible when high-frequency innovations are fat-tailed and display volatility clustering, and these properties are accounted for in the construction of the unit root tests. Our claim is based on two results. First, recent research by Lucas (1995), Rothenberg and Stock (1997), Seo (1999), Boswijk (2001), and Li *et al.* (2002) has demonstrated that when the errors of an autoregressive process display these typical features of financial data (fat tails and persistent volatility clustering), then (quasi-) likelihood ratio tests within a model that takes these effects into account can be considerably more powerful than the conventional least-squares-based Dickey-Fuller tests, which do not explicitly incorporate these properties.

Secondly, these typical features tend to be more pronounced in high-frequency (in particular, daily) data than in lower-frequency (monthly or quarterly) data. This phenomenon is explained by the analysis of Drost and Nijman (1993), who show that systematic sampling and temporal aggregation of GARCH (generalized autoregressive conditional heteroskedasticity) processes decreases both the kurtosis of the conditional

distribution and the persistence in the volatility process.

Combining these two results implies that the possibilities of obtaining unit root tests with higher power increase when one moves from low-frequency to high-frequency observations. In this sense, frequency does matter. In a different model (allowing for regime-switching in the mean growth rate of a process), Klaassen (2005) found a similar effect of sampling frequency on the power of a test of the random walk null hypothesis. For the unit root test, we illustrate the power gain by a small-scale Monte Carlo experiment and an empirical application to real exchange rates in the post-Bretton Woods era.

The outline of the remainder of this paper is as follows. In Section 2, we review the most important results about testing for a unit root based on Gaussian and non-Gaussian likelihood functions, and we discuss the theoretical effect of sampling frequency. Section 3 discusses a stylized Monte Carlo experiment, showing that in a realistic situation, the possible power gains are considerable when moving from the monthly frequency to daily data. Section 4 discusses the empirical analysis. Here we show that for the real exchange rate of the German mark versus the British pound, the theoretically expected power gain from using higher frequency observations leads to a rejection of the unit root hypothesis by a non-Gaussian GARCH-based likelihood ratio test, whereas the Dickey-Fuller test remains insignificant. For the real US dollar exchange rates of the German mark and the British pound, on the other hand, none of the tests find evidence in favour of purchasing-power parity, regardless of the sampling frequency. The final section contains some concluding remarks.

2 LIKELIHOOD BASED UNIT ROOT TESTS

Consider the first-order autoregressive model for a time series $\{X_t\}_{t \geq 1}$:

$$X_t = \phi_0 + \phi X_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots \quad (1)$$

where the starting value X_0 is considered fixed, and where $\{\varepsilon_t\}_{t \geq 1}$ is a martingale difference sequence, i.e., a disturbance term with mean zero conditional on the past history of X_t . This implies that $\{\varepsilon_t\}_{t \geq 1}$ has mean zero and displays no autocorrelation. We focus on the first-order autoregression throughout this paper, but

the results may be generalized to higher-order autoregressions, which are often needed in practice to obtain uncorrelated disturbances. The main elements of this generalization are discussed in Appendix A.

Defining $\gamma = \phi - 1$, the model can be rewritten as

$$\Delta X_t = \phi_0 + \gamma X_{t-1} + \varepsilon_t. \quad (2)$$

The unit root hypothesis is $\mathcal{H}_0 : \gamma = 0$ ($\phi = 1$), to be tested against the alternative hypothesis $\mathcal{H}_1 : \gamma < 0$ ($\phi < 1$). Under \mathcal{H}_1 (and provided that $\phi > -1$), the process $\{X_t\}_{t \geq 1}$ is stationary with a constant mean $\mu = \phi_0 / (1 - \phi) = -\phi_0 / \gamma$. The parameter γ represents the strength of the tendency of the process to revert to this mean. Under \mathcal{H}_0 , the mean of $\{X_t\}_{t \geq 1}$ is not defined, and instead the process $\{\Delta X_t\}_{t \geq 1}$ is stationary with a constant mean ϕ_0 , which means that $\{X_t\}_{t \geq 1}$ displays a linear trend if $\phi_0 \neq 0$. In this paper we focus on the case where this trend under the null hypothesis is assumed to be zero, which is a realistic assumption for, e.g., real exchange rates and interest rates. A reparametrization that makes this assumption explicit is

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \varepsilon_t, \quad (3)$$

which shows that the constant $\phi_0 = -\gamma\mu$ drops out of the model under the null hypothesis $\gamma = 0$. Although we do not consider this explicitly, the conclusion of this paper continues to hold in generalizations of (1)–(3), allowing for a linear trend under both the null and the alternative hypothesis.

2.1 Dickey-Fuller Tests

Dickey and Fuller (1979) proposed testing \mathcal{H}_0 against \mathcal{H}_1 by rejecting the null hypothesis for large negative values of $\hat{\tau}_\mu$, the t -statistic for $\gamma = 0$ based on least-squares estimation of (2). Alternatively, Dickey and Fuller (1981) proposed to test the same null hypothesis, with the restriction on ϕ_0 implied by (3), by rejecting for large values of Φ_1 , the F -statistic for $\phi_0 = \gamma = 0$ in (2). The t -test is equivalent to a signed (one-tailed) likelihood ratio test for $\gamma = 0$ against $\gamma < 0$ in (2), under the assumption that $\{\varepsilon_t\}_{t \geq 1}$ is an independent and identically distributed (i.i.d.) $N(0, \sigma_\varepsilon^2)$ sequence. Under the same conditions, the F -test is equivalent to the likelihood ratio test for \mathcal{H}_0 against \mathcal{H}_1 in (3). However, the asymptotic properties of $\hat{\tau}_\mu$ and Φ_1 will continue to hold under much wider conditions, as considered below.

An essential ingredient in deriving the asymptotic properties is the functional central limit theorem, which states that if $\lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \text{var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$, and under otherwise fairly mild conditions (allowing for excess kurtosis and some forms of volatility clustering in $\{\varepsilon_t\}_{t \geq 1}$),

$$\frac{1}{\sigma_\varepsilon \sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \varepsilon_t \xrightarrow{\mathcal{L}} W(s), \quad s \in [0, 1],$$

where $\lfloor x \rfloor$ denotes the integer part of x , where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution as $n \rightarrow \infty$, and where $W(s)$ is a standard Brownian motion process on $[0, 1]$. This result, together with the continuous mapping theorem, implies that, under \mathcal{H}_0 and as the sample size n tends to infinity,

$$\hat{\tau}_\mu \xrightarrow{\mathcal{L}} \left[\int_0^1 \tilde{W}(s)^2 ds \right]^{-1/2} \int_0^1 \tilde{W}(s) dW(s), \quad (4)$$

$$2\Phi_1 \xrightarrow{\mathcal{L}} \int_0^1 dW(s) G(s)' \left[\int_0^1 G(s) G(s)' ds \right]^{-1} \int_0^1 G(s) dW(s), \quad (5)$$

with $\tilde{W}(s) = W(s) - \int_0^1 W(u) du$ and $G(s)' = [W(s), 1]$. The factor 2 on the left-hand side of (5) is included only to facilitate comparison with the limiting expression for the non-Gaussian-based likelihood ratio test considered below; the statistic $2\Phi_1$ is in fact the Wald test statistic for $\phi_0 = \gamma = 0$ in (2), and the distribution of the right-hand side expression of (5) is also the limiting null distribution of the likelihood ratio statistic (based on an i.i.d. Gaussian likelihood). The limiting distributions are tabulated by Dickey and Fuller (1979, 1981), based on simulation of a discretization of the relevant integrals.

Under a sequence of local alternatives $\mathcal{H}_n : \gamma = c/n$, where $c < 0$ is a non-centrality parameter, the results of Chan and Wei (1987) imply that as $n \rightarrow \infty$,

$$\hat{\tau}_\mu \xrightarrow{\mathcal{L}} \left[\int_0^1 \tilde{U}(s)^2 ds \right]^{-1/2} \int_0^1 \tilde{U}(s) [dW(s) + cU(s)ds], \quad (6)$$

$$2\Phi_1 \xrightarrow{\mathcal{L}} \int_0^1 [dW(s) + cU(s)ds] H(s)' \times \left[\int_0^1 H(s) H(s)' ds \right]^{-1} \int_0^1 H(s) [dW(s) + cU(s)ds], \quad (7)$$

where $\tilde{U}(s) = U(s) - \int_0^1 U(u) du$, $H(s) = [U(s), 1]'$, and where $U(s) = \int_0^s e^{c(s-u)} dW(u)$, an Ornstein-Uhlenbeck process, which is the solution to the stochastic differential equation $dU(s) = cU(s)ds + dW(s)$.

The probability that the right-hand side expression in (6) is less than the $100\alpha\%$ quantile of the null distribution in (4), or that the right-hand side of (7) exceeds the $100(1 - \alpha)\%$ quantile of the null distribution in (5), defines the asymptotic local power functions of the two tests. These provide an approximation to the actual power of $\hat{\tau}_\mu$ and Φ_1 for finite samples. As the non-centrality parameter c becomes larger in absolute value (i.e., more negative), the center of the distributions of $\hat{\tau}_\mu$ and Φ_1 will shift (to the left and to the right, respectively), leading to higher power.

These results have direct implications for the effect of sampling frequency on the power of the tests. For example, suppose that the model (1) applies to a sample consisting of 25 years of quarterly observations (hence $n = 100$), with $\phi = 0.95$ and hence mean reversion parameter $\gamma = -0.05$. Then the asymptotic local power of the Dickey-Fuller $\hat{\tau}_\mu$ test is the rejection probability of (6) for $c = n\gamma = -5$. This may be directly compared to the situation where we consider, for the same period of 25 years, the annual (end-of-year) observations $\{X_j^* = X_{4j}\}_{j=1}^{n^*}$, with $n^* = n/4 = 25$. Writing (1) with $\phi_0 = -\gamma\mu$ as $X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$, the implied model for the annual observations is

$$X_j^* - \mu = \phi^4(X_{j-1}^* - \mu) + \varepsilon_j^*,$$

where $\varepsilon_j^* = \sum_{i=0}^3 \phi^i \varepsilon_{4j-i}$, so that $\{\varepsilon_j^*\}_{j \geq 1}$ is a serially uncorrelated process. Therefore, the annual mean-reversion parameter is $\gamma^* = \phi^4 - 1 = -0.185$, so that the non-centrality parameter becomes $c^* = n^*\gamma^* = -4.637$, which corresponds to an asymptotic local power that is only marginally smaller than the local power relevant for the quarterly observations (see also Stock (1994), p. 2776). In other words, moving from 25 annual observations to 100 quarterly observations will increase the sample size by a factor 4, but will only marginally increase the local power: in this example, the power of the $\hat{\tau}_\mu$ test at the 5% significance level would increase from 10.9% to 11.8%. By comparison, moving from 25 annual observations to 100 annual observations leads to an increase of the non-centrality parameter by a factor 4, which corresponds to an increase in local power from 10.9% to 79%.

As indicated above, these asymptotic results also apply when $\{\varepsilon_t\}_{t \geq 1}$ is a stationary GARCH process, or when the distribution of $\{\varepsilon_t\}_{t \geq 1}$ has a larger kurtosis than the Gaussian distribution. However, when the actual data-generating process displays such deviations from the i.i.d. Gaussian assumption, then the

Dickey-Fuller test is not a likelihood ratio test, and hence it is no longer optimal. In that case, tests based on a likelihood function that captures this volatility clustering and distributional shape will have a larger asymptotic local power. This is considered next.

2.2 Likelihood Ratio Tests Based on a Non-Gaussian GARCH Likelihood

Unit root testing based on a non-Gaussian but i.i.d. likelihood is considered by Cox and Llatas (1991), Lucas (1995), and Rothenberg and Stock (1997). Tests based on a Gaussian GARCH likelihood are analyzed by Ling and Li (1998, 2003), Seo (1999), Boswijk (2001), and Li *et al.* (2002), *inter alia*. A combination of both approaches, i.e., unit root inference based on a non-Gaussian GARCH likelihood, can be found in Ling and McAleer (2003). We refer to these papers and the references therein for a full derivation of these results, and only mention the most important aspects here. Moreover, some technical results that are used here are derived in Appendix B.

Consider the model

$$\Delta X_t = \phi_0 + \gamma X_{t-1} + \varepsilon_t, \quad (8a)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (8b)$$

$$\eta_t = \frac{\varepsilon_t}{\sigma_t} \sim \text{i.i.d. } f(\eta_t), \quad (8c)$$

where $f(\eta_t)$ is a (possibly non-Gaussian) density with zero mean and unit variance, so that σ_t^2 is the conditional variance of ε_t . We assume that $\alpha + \beta < 1$, so that $\{\varepsilon_t\}$ has a finite unconditional variance $\sigma_\varepsilon^2 = \omega/(1 - \alpha - \beta)$. Let $\delta = (\gamma, \phi_0)'$ and $Z_t = (X_{t-1}, 1)'$, and let θ denote the full parameter vector, containing δ , the GARCH parameters (ω, α, β) , and possible additional parameters characterizing the distributional shape of η_t . Since the conditional density of ΔX_t given the past is $\sigma_t^{-1} f((\Delta X_t - \delta' Z_t)/\sigma_t)$, the log-likelihood function has the following form:

$$\ell(\theta) = \sum_{t=1}^n \ell_t(\theta) = \sum_{t=1}^n \left\{ -\frac{1}{2} \log \sigma_t^2 + \log f \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \right\}.$$

This leads to the likelihood ratio (LR) statistic, defined as usual:

$$LR = -2 \left[\ell(\hat{\theta}_0) - \ell(\hat{\theta}) \right], \quad (9)$$

where $\hat{\theta}$ is the unrestricted maximum likelihood (ML) estimator, and $\hat{\theta}_0$ is the ML estimator under the restriction $\delta = 0$. Furthermore, a quasi-maximum likelihood (QML) based t -statistic for $\gamma = 0$ is given by

$$t_{QML} = \frac{\hat{\gamma}}{\sqrt{\widehat{\text{var}}_{QML}(\hat{\gamma})}}, \quad (10)$$

where $\hat{\gamma}$ is the first component of $\hat{\theta}$, and $\widehat{\text{var}}_{QML}(\hat{\gamma})$ is the first diagonal element of

$$\widehat{\text{var}}_{QML}(\hat{\theta}) = \left(\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta \partial \theta'} \right)^{-1} \sum_{t=1}^n \frac{\partial \ell_t(\hat{\theta})}{\partial \theta} \frac{\partial \ell_t(\hat{\theta})}{\partial \theta'} \left(\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta \partial \theta'} \right)^{-1}. \quad (11)$$

As usual, the statistic LR is only a likelihood ratio statistic under correct specification of the model (8a)–(8c), and its asymptotic properties, sketched below, are obtained under that condition. In contrast, the t -statistic (10)–(11) is constructed to be robust against possible misspecification of f , and as such may be used as a quasi-likelihood-based statistic. The LR statistic (9) with f the standard Gaussian density was analyzed by Boswijk (2001); a version of the t -statistic derived from the same Gaussian likelihood can be found in Seo (1999), Ling and Li (2003), and Li *et al.* (2002), and a t -statistic for the non-Gaussian GARCH model (with f known or estimated non-parametrically) is given by Ling and McAleer (2003).

The asymptotic properties of the LR and t -statistic are derived most easily via a second-order Taylor series expansion of $\ell(\theta)$, which leads us to express LR , up to an asymptotically negligible term, as a quadratic form in the score vector. Defining the function $\psi(\eta) = -d \log f(\eta)/d\eta$, the partial derivative of the log-likelihood with respect to δ is given by

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \delta} &= \sum_{t=1}^n \left\{ \frac{1}{\sigma_t} Z_t \psi \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \right. \\ &\quad \left. + \frac{1}{2\sigma_t^2} \left[\psi \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \frac{\Delta X_t - \delta' Z_t}{\sigma_t} - 1 \right] \frac{\partial \sigma_t^2}{\partial \delta} \right\}, \end{aligned} \quad (12)$$

where the GARCH model implies that

$$\frac{\partial \sigma_t^2}{\partial \delta} = -2\alpha \sum_{i=1}^{t-1} \beta^{i-1} Z_{t-i} (\Delta X_{t-i} - \delta' Z_{t-i}).$$

Under the null hypothesis, the score vector (12) (divided by n) is asymptotically equivalent to

$$\frac{1}{n} \sum_{t=1}^n Z_t \left(\frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2} [\psi(\eta_t) \eta_t - 1] \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) = \frac{1}{n} \sum_{t=1}^n Z_t v_t, \quad (13)$$

where v_t is implicitly defined. It is shown in Appendix B that $\{v_t\}_{t \geq 1}$ is a martingale difference sequence with finite variance σ_v^2 , and that $\text{cov}(\varepsilon_t, v_t) = 1$. This leads to the following bivariate functional central limit theorem: as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \sigma_\varepsilon W(s) \\ \sigma_v B(s) \end{pmatrix}, \quad s \in [0, 1], \quad (14)$$

where $B(s)$ is a standard Brownian motion, such that $\text{cov}(\sigma_\varepsilon W(s), \sigma_v B(s)) = \text{cov}(\varepsilon_t, v_t) = 1$. Hence the correlation between $W(s)$ and $B(s)$ is given by $\rho = 1/(\sigma_\varepsilon \sigma_v)$. The sample moment (13) converges in distribution, as $n \rightarrow \infty$, to the stochastic integral $\sigma_\varepsilon \sigma_v \int_0^1 G(s) dB(s)$, and this in turn implies, under \mathcal{H}_0 and as $n \rightarrow \infty$,

$$t_{QML} \xrightarrow{\mathcal{L}} \left[\int_0^1 \tilde{W}(s)^2 ds \right]^{-1} \int_0^1 \tilde{W}(s) dB(s), \quad (15)$$

$$LR \xrightarrow{\mathcal{L}} \int_0^1 dB(s) G(s)' \left[\int_0^1 G(s) G(s)' ds \right]^{-1} \int_0^1 G(s) dB(s), \quad (16)$$

where $\tilde{W}(s)$ and $G(s)$ are the same as in (4)–(5).

For an i.i.d. Gaussian (quasi-) likelihood, it can be derived that $v_t = \varepsilon_t/\sigma^2$, so that $\rho = 1$. In general however, $0 < \rho \leq 1$, where smaller values of ρ are associated with a larger degree of variation in the volatility, and with fatter tails in the density $f(\eta)$. (The limit $\rho \rightarrow 0$ corresponds to the distribution of $\{\varepsilon_t\}_{t \geq 1}$ approaching an infinite-variance distribution; however, in that case (14) no longer applies and a different limit theory is needed.) This implies that the limiting null distributions of t_{QML} and LR depend on the nuisance parameter $\rho = \text{corr}(W(1), B(1))$. This parameter ρ may be estimated consistently by the sample correlation of $\hat{\varepsilon}_t$ and \hat{v}_t , and approximate quantiles of the asymptotic null distributions for a given value of ρ may be obtained from the normal approximation of Abadir and Lucas (2000) (for the t_{QML} statistic) and the Gamma approximation of Boswijk and Doornik (2005) (for the LR statistic).

Under local alternatives, we have as $n \rightarrow \infty$,

$$t_{QML} \xrightarrow{\mathcal{L}} \left[\int_0^1 \tilde{U}(s)^2 ds \right]^{-1} \int_0^1 \tilde{U}(s) \left[dB(s) + \frac{c}{\rho} U(s) ds \right], \quad (17)$$

$$\begin{aligned} LR \xrightarrow{\mathcal{L}} & \int_0^1 \left[dB(s) + \frac{c}{\rho} U(s) ds \right] H(s)' \left[\int_0^1 H(s) H(s)' ds \right]^{-1} \\ & \times \int_0^1 H(s) \left[dB(s) + \frac{c}{\rho} U(s) ds \right], \end{aligned} \quad (18)$$

where $U(s)$, $\tilde{U}(s)$ and $H(s)$ are the same as before. Comparing (17)–(18) with (6)–(7), we observe two differences. First, the term $dW(s)$ has been replaced by $dB(s)$ in the stochastic integrals, and secondly, the non-centrality parameter c has been replaced by c/ρ . The former has a relatively minor effect on the local power of the test, but the effective non-centrality parameter c/ρ implies that large power gains of t_{QML} and LR over $\hat{\tau}_\mu$ and Φ_1 , respectively, are obtained for cases where ρ is relatively small. As discussed above, this occurs when the volatility displays much variation (large values of the GARCH parameters α and β), and when the distribution of η_t has heavy tails.

These results again have implications for the effect of sampling frequency on power. As we move from low-frequency data to high-frequency data, maintaining the same time span, the parameter c hardly changes, for the same reasons as indicated in the previous sub-section. However, the results of Drost and Nijman (1993) imply that for the high-frequency data, the sum of the GARCH parameters $\alpha + \beta$ will be closer to 1 than for low-frequency observations, implying more persistent variation in the volatility, and that the kurtosis of $\{\varepsilon_t\}_{t \geq 1}$ will be higher. Both effects will cause ρ to be smaller for high-frequency data, so that the power of the test increases with sampling frequency. In the next section, we study how large this power increase is in a small Monte Carlo experiment.

3 A MONTE CARLO EXPERIMENT

As the data-generating process (DGP), we consider the AR(1)-GARCH(1,1)- $t(\nu)$ model

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \varepsilon_t, \quad (19a)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad (19b)$$

$$\eta_t = \frac{\varepsilon_t}{\sigma_t} \sim \text{i.i.d. } t(\nu), \quad t = 1, \dots, n. \quad (19c)$$

Here $t(\nu)$ denotes the *standardized* $t(\nu)$ distribution, such that $E(\eta_t^2) = 1$ and hence σ_t^2 is indeed the conditional variance of ε_t given the past. The initial values for the process are chosen as $\sigma_0^2 = \sigma_\varepsilon^2 = \omega/(1 - \alpha - \beta)$, and $X_0 = \mu + \sigma_0\eta_0$, where $\eta_0 \sim t(\nu)$, independent of $\{\eta_t\}_{t \geq 1}$.

For each replication, we first generate a high-frequency sample $\{X_t\}_{t=0}^n$, using parameter values that mimic properties of daily financial data, i.e.,

$$\alpha = 0.07, \quad \beta = 0.9, \quad \nu = 5.$$

These parameter values are such that the volatility is persistent ($\alpha + \beta = 0.97$), and both the conditional and the unconditional distribution have a high (but finite) kurtosis of

$$\kappa_\eta = 3\frac{\nu - 2}{\nu - 4} = 9 \quad \text{and} \quad \kappa_\varepsilon = \kappa_\eta \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - (\kappa_\eta - 1)\alpha^2} = 26.7,$$

respectively; see He and Teräsvirta (1999) for moments of GARCH processes. For the mean-reversion parameter we choose $\gamma = c/n$, where $c \in \{0, -5, -10, -15, -20, -25, -30\}$; thus we study both the size and the (size-corrected) power of various tests. Note that the analysis is invariant to μ and ω , which only determine the location and scale of X_t but not its dynamic properties or distributional shape.

To study the effect of systematic sampling, we then construct a low-frequency sample $\{X_j^*\}_{j=0}^{n^*}$ by a skip-sample version of X_t , i.e.,

$$X_j^* = X_{mj}, \quad j = 0, \dots, n^* = \frac{n}{m},$$

where we take $m = 20$. These may be thought of as the end-of-month versions of X_t , assuming 20 trading days in a month. We choose $n = 2000$ and hence $n^* = 100$; thus we mimic a sample of about 8 years of

daily data $\{X_t\}_{t=0}^n$, and the corresponding 8 years of monthly data $\{X_j^*\}_{j=0}^{n^*}$. We do not attempt to provide full characterization of the implied data-generating process for $\{X_j^*\}_{j=0}^{n^*}$. The main result for our purpose, which follows from the analysis of Drost and Nijman (1993), is that under the null hypothesis $\gamma = 0$, ΔX_j^* is a weak GARCH(1,1) process with $\alpha^* + \beta^* = (\alpha + \beta)^m = 0.54$ (so that the volatility displays less persistence), and that the kurtosis of the standardized errors is lower for low-frequency data. (Drost and Nijman (1993) define a process $\{\varepsilon_t\}$ to be weak GARCH if the linear projection of ε_t^2 on a constant and the past history of ε_t and ε_t^2 satisfies a GARCH specification.) Under the alternative $\gamma < 0$, the implied process is weak ARMA-GARCH of a higher order; however, because we consider very local alternatives, we suspect that an AR(1)-GARCH(1,1) model will also be adequate for the low-frequency data in these cases.

For both the low- and the high-frequency data we apply four unit root tests, each at the 5% nominal size:

- DF, the Dickey-Fuller $\hat{\tau}_\mu$ test for $\gamma = 0$ in the least-squares regression (2);
- QML-GARCH, the t_{QML} test for $\gamma = 0$ based on a Gaussian GARCH(1,1) likelihood (ignoring the conditional $t(\nu)$ -distribution of the DGP);
- QML- $t(\nu)$, the t_{QML} test for $\gamma = 0$ based on an i.i.d. $t(\nu)$ likelihood (ignoring the GARCH(1,1) aspect of the DGP);
- ML, the t_{QML} test for $\gamma = 0$ based on a GARCH(1,1)- $t(\nu)$ model.

The motivation for studying QML-GARCH and QML- $t(\nu)$ is to investigate how much power may be gained by taking into account only volatility clustering or fat-tailedness. These tests are based on a misspecified likelihood function, which is corrected in the construction of the standard error from (11). The test based on the GARCH- t likelihood is actually only a true maximum-likelihood based test for the high-frequency data; for the monthly observations, the actual DGP will not exactly be an AR(1)-GARCH(1,1)- $t(\nu)$ process, so that the test in this case is also a QML-based test. We only report results for the one-sided t -tests; further simulations reveal that the two-sided Φ_1 and LR tests generally have lower power than their one-sided counterparts, but otherwise display the same relative power behaviour.

We obtain p -values for each of the statistics using the normal approximation of Abadir and Lucas (2000), extended to allow for an intercept in the autoregression. The correlation parameter ρ is estimated simply as the sample correlation of the residuals $\hat{\varepsilon}_t$ and the “scores” \hat{v}_t , both evaluated at the unrestricted estimates. All results below are based on 10,000 replications, and have been obtained using Ox 6, see Doornik (2009).

– Insert Table 1 here –

Table 1 displays the actual size of the four tests, both for low-frequency and for high-frequency observations. Also indicated are the average (over 10,000 replications) estimates of the correlation parameter ρ . We observe that the (Q)ML tests display some over-rejection when based on 100 monthly observations. These distortions clearly decrease as we move to 2000 daily observations. The values of ρ indicate that both the largest power gains and the largest effect of sampling frequency on power may be expected from the ML test.

– Insert Figure 1 here –

The size-corrected power curves for the four tests are plotted against $-c$ in Figure 1. Recall that the mean-reversion parameter γ is given by c/n , so the range of c corresponds to $\gamma \in \{0, -0.0025, \dots, -0.015\}$. In each of the four panels, the vertical distance between the solid and the dashed curves indicates the gains from using high-frequency observations. To evaluate the power gain from using the (Q)ML tests applied to low-frequency observations, the low-frequency power of the Dickey-Fuller test is indicated by a dotted curve in the graphs for QML-GARCH, QML- $t(\nu)$, and ML.

As predicted from the values of ρ in Table 1, the largest power gains from high-frequency observations are obtained by the ML test, followed by the QML- $t(\nu)$ test. The QML-GARCH test, based on a Gaussian GARCH quasi-likelihood function, is the least able to exploit this power potential. This confirms earlier results, see Boswijk (2001), that for the type of GARCH parameter values typically encountered in practice, taking account of fat-tailedness has a bigger contribution to the power gain than taking account of volatility clustering. In the low-frequency case, the power of the (Q)ML tests is only slightly higher than the DF test

for low values of c ; for larger deviations from the null hypothesis, the DF has higher power. This indicates that in smaller samples of monthly observations, the gain from modelling the fat tails and volatility clustering is limited, and for larger values of c this gain is in fact more than off-set by the degrees-of-freedom loss from estimating more parameters. Finally, we see that as expected, the power of the Dickey-Fuller test is hardly affected by the sampling frequency.

The power curves also allow us to compare the effects of sampling frequency and time span on the power of the tests. Concentrating on the ML test, we see that for the daily frequency, a non-centrality parameter of $c = -10$ is sufficient to obtain a power of about 60%. When using monthly data, a non-centrality of about $c = -16$ is needed to obtain the same power. Suppose that $c = -10$ corresponds to 8 years of daily data ($n = 2000$), with daily mean-reversion parameter $\gamma = c/n = -0.005$. The corresponding monthly mean-reversion parameter is $\gamma^* = (1 - 0.005)^{20} - 1 = -0.095$, so that about 14 years of monthly observations ($\tilde{n} = 168$) are needed to obtain the desired non-centrality parameter $c = -16$. In other words, in this specific case 14 years of monthly data contain about the same information on the mean-reversion of the process as 8 years of daily data.

In summary, this Monte Carlo experiment confirms the theoretical predictions: sampling at a higher frequency increases the degree of fat-tailedness and volatility clustering in the data, and hence the possibility of obtaining more power from tests that take these properties into account. In the next section we explore whether this also has consequences for the empirical analysis of real exchange rates.

4 EMPIRICAL APPLICATION

In this section we apply the unit root tests studied in this paper to real exchange rate data in the post-Bretton Woods era (using data from April 1978 through October 2009). We focus on the three real exchange rates between the US, Germany and the UK (for the euro period, we use the fixed €/DM exchange rate to convert euro rates to D-mark rates). Daily nominal exchange rates have been converted to real exchange rates using daily producer price indices, constructed by log-linear interpolation from the monthly price indices taken from IMF's International Financial Statistics. From these daily real exchange rates, we define the end-of-

week, end-of-month and end-of-quarter series by systematic sampling. The question of interest is whether purchasing-power parity holds in the long run, i.e., whether a unit root in real exchange rates can be rejected.

– Insert Figure 2 here –

The daily log real exchange rates are depicted in Figure 2. In all three series we observe very slow (if any) mean-reversion: real exchange rates may persistently deviate from their mean for a large number of years. The graphs illustrate the common finding that it is hard to find evidence in favour of purchasing-power parity within the kind of time span considered here.

The first row of Figure 3 displays the daily exchange rate returns. The period from October 1990 to September 1992, when the British pound was pegged to the German mark within the European Monetary System, shows up as a period of low volatility in the mark-pound exchange rate returns. Because this is a relatively short period, we choose not to model this period separately from the rest of the sample.

– Insert Figure 3 here –

To investigate whether there is any scope for power improvement due to fat-tailedness and volatility clustering, Figure 3 also depicts some stylized properties of the daily returns, in particular their correlogram, the correlogram of squared returns, and the estimated density. We observe the typical characteristics of financial returns: very little serial correlation in the returns, positive and persistent correlation in the squared returns, and a relatively peaked and fat-tailed density of the returns, with a kurtosis given by 7.11, 5.89 and 6.76, respectively. Clearly, this characterizes only the unconditional distributions of the returns; whether the conditional distributions also display excess kurtosis will become evident from the estimated GARCH- $t(\nu)$ models for the returns.

– Insert Table 2 here –

Table 2 displays the p -values of the Dickey-Fuller $\hat{\tau}_\mu$ and Φ_1 tests, and the t_{QML} and LR tests based on (8) with $\eta_t \sim \text{i.i.d. } t(\nu)$, for the three different real exchange rates and four different frequencies. For ease of comparison, we only consider first-order autoregressive models. For the daily data, this sometimes leads

to some residual autocorrelation, but accounting for this by including lagged differences hardly affects the results in Table 2. The GARCH- $t(\nu)$ model has only been estimated when this model provides a significant improvement over the Gaussian i.i.d. model for the disturbances, which is the case for all series at the weekly and daily frequencies, and for two series at the monthly frequency. For the quarterly series, and the monthly series of the dollar-mark rate, the volatility clustering and leptokurtosis are not strong enough to cause significant deviations from the Gaussian i.i.d. model. All results in this section have been obtained using the Garch module within PcGive 13, see Doornik and Hendry (2009).

Consider first the p -values of the Dickey-Fuller tests. We observe that for all three series, these p -values display little variation over the different frequencies. None of these tests lead us to conclude that there is significant mean-reversion in the real exchange rates at the 5% significance level, for any sampling frequency. This may be seen as an illustration of Shiller and Perron (1985)'s conclusion that sampling frequency matters very little for the power of the Dickey-Fuller test.

For the two real exchange rates vis-à-vis the US dollar, we find that using the t_{QML} and LR tests does not lead to a different outcome: regardless of the sampling frequency, the unit root in these series cannot be rejected. In fact, the p -values of these tests are often higher than the Dickey-Fuller p -values, and generally do not decrease with the sampling frequency. This result holds despite the fact that the conditions for more powerful tests at higher frequencies are satisfied: as the sampling frequency increases, the estimated volatility persistence $\hat{\alpha} + \hat{\beta}$ increases, and the estimated degrees of freedom $\hat{\nu}$ decreases (as predicted by the analysis of Drost and Nijman (1993)), which together leads to a decrease in the estimated correlation $\hat{\rho}$.

For the real exchange rate of the German mark vis-à-vis the British pound, on the other hand, we do observe outcomes in line with the theoretical predictions. At the monthly frequency, the t_{QML} and LR tests have p -values that are already closer to 5% than the Dickey-Fuller tests, and as we move to weekly or daily observations, these p -values drop to 1% or even lower, indicating strong evidence against the unit root in this series. Note that for this series, the estimated mean-reversion parameter closely follows the theoretical predictions: for example, the daily mean-reversion parameter of -0.0013 would imply a weekly mean-reversion parameter of -0.0065 , quite close to the estimated -0.0069 , and the same applies to the monthly

and quarterly mean-reversion parameters implied by the weekly and monthly estimates, respectively. This highlights that the stronger evidence in favour of mean-reversion is not a consequence of a larger estimated effect, but of a more efficient estimate.

In summary, we conclude that in one out of three real exchange rates, the use of high-frequency data in combination with a GARCH- $t(\nu)$ based LR test changes the evidence against the unit root hypothesis from insignificant to significant, as predicted by the theoretical results and Monte Carlo evidence in the previous sections. An obvious explanation of the difference between the mark-pound rate and the other two rates is that the close geographical proximity of the UK and Germany, and their joint membership of the European Union, leads to more effective goods-market arbitrage between these two currencies. Another, more technical explanation is that the dollar exchange rates display more prominent swings in the first decade than the intra-European exchange rate. This might indicate that the possible mean-reverting behavior of the dollar-mark and dollar-pound real exchange rates over this period is not adequately described by a linear AR process, but requires, e.g., threshold effects or switching regimes. As the models considered here do not allow for such non-linear effects, this leads to an estimate of γ that is biased towards zero. Increasing the data frequency does not solve this problem, so that it is difficult to find significant evidence of purchasing-power parity in these cases. The results for the mark-pound real exchange rate, however, demonstrate that for series where the linear AR specification seems more appropriate, the power gains from increasing the frequency are empirically relevant.

5 CONCLUDING REMARKS

The main conclusion of this paper is that the common belief, that only time span matters for the power of unit root tests, is incorrect for financial data, where high-frequency observations display properties that may be exploited for obtaining tests with higher power. Clearly, the alternative approaches to obtaining more power, such as longer time series, panel data restrictions, or alternative treatments of the constant and trend are useful as well, and could be combined with the approach presented here. Similar power gains from non-Gaussian likelihood analysis may be obtained in a multivariate cointegration context, see Boswijk and Lucas

(2002). Although one can also apply GARCH likelihoods in a cointegration context (see Li *et al.* (2001) and Seo (2007)), the main problem here is to find a parsimoniously parametrized multivariate GARCH model that is reasonably well specified.

ACKNOWLEDGMENTS

Helpful comments from two anonymous referees are gratefully acknowledged.

APPENDIX A: HIGHER-ORDER AUTOREGRESSIONS

This appendix discusses how the theory in Section 2 may be extended to higher-order (but finite) autoregressive models. Consider the AR(p) model with mean μ :

$$X_t = \mu + \sum_{j=1}^p \phi_j (X_{t-j} - \mu) + \varepsilon_t.$$

The null and alternative hypotheses can be formulated in terms of the roots of the characteristic polynomial $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$. Under the null hypothesis $\mathcal{H}_0 : \phi(1) = 0$, the characteristic equation has a unit root and the process is integrated of order 1 (provided all other roots are outside the unit circle). Under the alternative $\mathcal{H}_1 : \phi(1) > 0$, and if all roots are outside the unit circle, the process is stationary. Defining $\gamma = -\phi(1)$ and $\gamma_j = -\sum_{i=j+1}^p \phi_i$, the model can be rewritten as

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \sum_{j=1}^{p-1} \gamma_j \Delta X_{t-j} + \varepsilon_t, \quad (\text{A.1})$$

which shows that γ may be interpreted as mean-reversion parameter, similar to the AR(1) case.

To define a sequence of local alternatives \mathcal{H}_n , we assume that one of the inverted roots of $\phi(z)$ is a real-valued sequence $\phi = 1 + c/n$, and all other roots are fixed, and outside the unit circle. This allows us to decompose $\phi(z)$ as

$$\phi(z) = (1 - \phi z)\psi(z),$$

where the roots of $\psi(z) = 1 - \sum_{j=1}^{p-1} \psi_j z^j$ are outside the unit circle. Note that this implies $\gamma = -\phi(1) =$

$\psi(1)c/n$. This leads to the representation

$$X_t - \mu = \phi(X_{t-1} - \mu) + u_t, \quad (\text{A.2})$$

with $u_t = \psi(L)^{-1}\varepsilon_t$, a stationary linear process with long-run variance $\sigma_u^2 = \sigma_\varepsilon^2/\psi(1)^2$. The process $\{u_t\}_{t \geq 1}$ satisfies the conditions of a functional central limit theorem, so that $(\sigma_u^2 n)^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} u_t \xrightarrow{\mathcal{L}} W(s)$, with $W(s)$ a standard Brownian motion. From this we find, under local alternatives \mathcal{H}_n , that $n^{-1/2} X_{\lfloor sn \rfloor} \xrightarrow{\mathcal{L}} \sigma_u U(s)$, with $U(s)$ an Ornstein-Uhlenbeck process, and that in turn can be used to show that the $\hat{\tau}_\mu$ and Φ_1 statistics defined from least-squares estimation of (A.1) have the same limiting distributions under the null and local alternatives as given in (4)–(7). Similarly, the non-Gaussian GARCH likelihood-based statistics t_{QML} and LR will have the same limiting distributions as given in (15)–(18).

If (A.2) applies to the high-frequency data $\{X_t\}_{t \geq 1}$, then we find, for the low-frequency (skip-sampled) observations $\{X_j^* = X_{mj}\}_{j \geq 1}$:

$$X_j^* - \mu = \phi^*(X_{j-1} - \mu) + u_j^*,$$

where $\{u_j^*\}_{j \geq 1}$ is now a stationary and invertible ARMA process, and $\phi^* = \phi^m$. Therefore, $\phi = 1 + c/n$ implies that

$$\phi^* - 1 = \left(1 + \frac{c}{n}\right)^m - 1 = \frac{mc}{n} + o(n^{-1}) = \frac{c}{n^*} + o(n^{-1}),$$

where $n^* = n/m$. Therefore, the same non-centrality parameter c applies to the high- and low-frequency observations.

APPENDIX B: NON-GAUSSIAN AND GARCH LIKELIHOOD ANALYSIS

In this appendix we provide some details about non-Gaussian GARCH-based likelihood analysis that are referred to in Section 2.2. We start from the score process $\{v_t\}_{t \geq 1}$, defined implicitly in (13) as

$$v_t = \frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2} [\psi(\eta_t)\eta_t - 1] \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i},$$

with $\psi(\eta) = -d \log f(\eta)/d\eta$. Let \mathcal{F}_t be the σ -field generated by $\{\varepsilon_{t-i}, \eta_{t-i}\}_{i \geq 0}$, so that σ_t is \mathcal{F}_{t-1} -measurable. We will show that $E[v_t | \mathcal{F}_{t-1}] = 0$, so that $\{v_t\}_{t \geq 1}$ is a martingale difference sequence with

respect to the filtration $\{\mathcal{F}_t\}_{t \geq 1}$; and that $\text{cov}(\varepsilon_t, v_t) = 1$, which implies that $\sigma_v^2 = \text{var}(v_t) \geq 1/\sigma_\varepsilon^2$, because $\rho^2 = \text{corr}(\varepsilon_t, v_t)^2 = 1/(\sigma_\varepsilon^2 \sigma_v^2) \leq 1$.

We start with

$$\begin{aligned} E[v_t | \mathcal{F}_{t-1}] &= E \left[\frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2} [\psi(\eta_t)\eta_t - 1] \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \middle| \mathcal{F}_{t-1} \right] \\ &= \frac{1}{\sigma_t} E[\psi(\eta_t) | \mathcal{F}_{t-1}] - \frac{\alpha}{\sigma_t^2} \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} E[\psi(\eta_t)\eta_t - 1 | \mathcal{F}_{t-1}] \\ &= \frac{1}{\sigma_t} E[\psi(\eta_t)] - \frac{\alpha}{\sigma_t^2} \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} E[\psi(\eta_t)\eta_t - 1], \end{aligned}$$

where we have used the ‘‘taking out what is known’’ rule for conditional expectations, and the fact that $\{\eta_t\}_{t \geq 1}$ is i.i.d., so that η_t is independent of \mathcal{F}_{t-1} .

For sufficiently regular densities, we have $E[\psi(\eta_t)] = 0$ and $E[\psi(\eta_t)\eta_t] = 0$. To see this, it will be useful to express $\psi(\eta)$ as the score of a location model $x = \mu + \eta$, with density $g(x, \mu) = f(x - \mu)$, evaluated at $\mu = 0$:

$$\psi(\eta) = \left. \frac{\partial \log g(x, \mu)}{\partial \mu} \right|_{\mu=0} = \frac{1}{f(x)} \left. \frac{\partial g(x, \mu)}{\partial \mu} \right|_{\mu=0} = -\frac{1}{f(\eta)} \frac{df(\eta)}{d\eta}.$$

This implies, assuming that the order of differentiation and integration may be changed:

$$E[\psi(\eta)] = \int_{\mathbb{R}} \frac{1}{f(x)} \left. \frac{\partial g(x, \mu)}{\partial \mu} \right|_{\mu=0} f(x) dx = \frac{d}{d\mu} \int_{\mathbb{R}} g(x, \mu) dx \Big|_{\mu=0} = 0,$$

because $\int_{\mathbb{R}} g(x, \mu) dx = 1$ for all μ . Furthermore,

$$E[\psi(\eta)\eta] = \int_{\mathbb{R}} \frac{1}{f(x)} \left. \frac{\partial g(x, \mu)}{\partial \mu} \right|_{\mu=0} x f(x) dx = \frac{d}{d\mu} \int_{\mathbb{R}} g(x, \mu) x dx \Big|_{\mu=0} = 1,$$

because $\int_{\mathbb{R}} g(x, \mu) x dx = \mu$. Combining these results leads to $E[v_t | \mathcal{F}_{t-1}] = 0$. Note that $\{v_t\}_{t \geq 1}$ (as well as $\{\varepsilon_t\}_{t \geq 1}$) is a serially dependent sequence, but this dependence does not arise in their first conditional moment.

Next, the martingale difference property for $\{v_t\}_{t \geq 1}$ and $\{\varepsilon_t\}_{t \geq 1}$ implies $E[\varepsilon_t] = E[v_t] = 0$, so that

$$\text{cov}(\varepsilon_t, v_t) = E[\varepsilon_t v_t] = E[E(\varepsilon_t v_t | \mathcal{F}_{t-1})],$$

using the law of iterated expectations. Next,

$$\begin{aligned} E[\varepsilon_t v_t | \mathcal{F}_{t-1}] &= E \left[\sigma_t \eta_t \left(\frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2} [\psi(\eta_t) \eta_t - 1] \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) \middle| \mathcal{F}_{t-1} \right] \\ &= E[\eta_t \psi(\eta_t)] - \frac{\alpha}{\sigma_t} \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} E[\psi(\eta_t) \eta_t^2 - \eta_t] = 1, \end{aligned}$$

because $E[\eta_t \psi(\eta_t)] = 1$, $E[\eta_t] = 0$, and

$$E[\psi(\eta) \eta^2] = \int_{\mathbb{R}} \frac{1}{f(x)} \frac{\partial g(x, \mu)}{\partial \mu} \bigg|_{\mu=0} x^2 f(x) dx = \frac{d}{d\mu} \int_{\mathbb{R}} g(x, \mu) x^2 dx \bigg|_{\mu=0} = 0,$$

since $\int_{\mathbb{R}} g(x, \mu) x^2 dx = 1 + \mu^2$ (the second uncentered moment of $x = \mu + \eta$), which has derivative 0 at $\mu = 0$.

References

- Abadir, K. M., and Lucas, A. (2000), “Quantiles for t -statistics Based on M -estimators of Unit Roots,” *Economics Letters*, 67, 131–137.
- Boswijk, H. P. (2001), “Testing for a Unit Root with Near-Integrated Volatility,” Tinbergen Institute Discussion Paper # 01-077/4 (www.tinbergen.nl/discussionpapers/01077.pdf).
- Boswijk, H. P., and Doornik, J. A. (2005), “Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors,” *Journal of Applied Econometrics*, 20, 797–810.
- Boswijk, H. P., and Lucas, A. (2002), “Semi-Nonparametric Cointegration Testing,” *Journal of Econometrics*, 108, 253–280.
- Campbell, J. Y., and Perron, P. (1991), “Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots,” in O. J. Blanchard and S. Fisher (eds.), *NBER Macroeconomics Annual* (Vol. 6), Cambridge, MA: The MIT Press.
- Chan, N. H., and Wei, C. Z. (1987), “Asymptotic Inference for Nearly Nonstationary AR(1) Processes,” *Annals of Statistics*, 15, 1050–1063.
- Chambers, M. J. (2004), “Testing for Unit Roots with Flow Data and Varying Sampling Frequency,” *Journal of Econometrics*, 119, 1–18.

- Choi, I. (1992), "Effects of Data Aggregation on the Power of Tests for a Unit Root", *Economics Letters*, 40, 397–401.
- Cox, D. D., and I. Llatas (1991), "Maximum Likelihood Type Estimation for Nearly Nonstationary Autoregressive Time Series," *Annals of Statistics*, 19, 1109–1128.
- Dickey, D. A. and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.
- Dickey, D. A. and Fuller, W. A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.
- Doornik, J. A. (2009), *An Object-oriented Matrix Programming Language - OxTM 6*, London: Timberlake Consultants Press and Oxford: www.doornik.com.
- Doornik, J. A., and Hendry, D. F. (2009), *Econometric Modelling - PcGiveTM 13* (Vol. III), London: Timberlake Consultants Press.
- Drost, F. C. and Nijman, Th. E. (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.
- Frankel, J. A. (1986), "International Capital Mobility and Crowding Out in the U.S. Economy: Imperfect Integration of Financial Markets or of Goods Markets?," in R. Hafer (ed.), *How Open is the U.S. Economy?* Lexington: Lexington Books.
- Frankel, J. A., and Rose, A. K. (1996), "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," *Journal of International Economics*, 40, 209–224.
- He, C., and Teräsvirta, T. (1999), "Properties of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 92, 173–192.
- Klaassen, F. (2005), "Long Swings in Exchange Rates: Are They Really in the Data?," *Journal of Business & Economic Statistics*, 23, 87–95.
- Li, S., Ling, W. K., and McAleer, M. (2002), "Recent Theoretical Results for Time Series Models with GARCH Errors," *Journal of Economic Surveys*, 16, 245–269.
- Li, S., Ling, W. K., and Wong, H. (2001), "Estimation for Partially Nonstationary Multivariate Autoregres-

- sive Models with Conditional Heteroskedasticity,” *Biometrika*, 88, 1135–1152.
- Ling, S., and Li, W. K. (1998), “Limiting Distributions of Maximum Likelihood Estimators for Unstable Autoregressive Moving-Average Time Series with General Autoregressive Heteroskedastic Errors,” *Annals of Statistics*, 26, 84–125.
- Ling, S., and Li, W. K. (2003), “Asymptotic Inference for Unit Root Processes with GARCH(1,1) Errors,” *Econometric Theory*, 19, 541–564.
- Ling, S., and McAleer, M. (2003), “On Adaptive Estimation in Nonstationary ARMA Models with GARCH Errors,” *Annals of Statistics*, 31, 642–674.
- Lothian, J., and Taylor, M. P. (1996), “Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries,” *Journal of Political Economy*, 104, 488–509.
- Lucas, A. (1995), “Unit Root Tests Based on M -Estimators,” *Econometric Theory*, 11, 331–346.
- Rothenberg, T. J., and J. H. Stock (1997), “Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations,” *Journal of Econometrics*, 80, 269–286.
- Seo, B. (1999), “Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity,” *Journal of Econometrics*, 91, 113–144.
- Seo, B. (2007), “Asymptotic Distribution of the Cointegrating Vector Estimator in Error Correction Models with Conditional Heteroskedasticity,” *Journal of Econometrics*, 137, 68–111.
- Shiller, R. J., and Perron, P. (1985), “Testing the Random Walk Hypothesis: Power versus Frequency of Observation,” *Economics Letters*, 18, 381–386.
- Stock, J. H. (1994), “Unit Roots, Structural Breaks and Trends,” in R. F. Engle and D. F. McFadden (eds.), *Handbook of Econometrics* (Vol. IV). Amsterdam: North Holland.
- Taylor, A. M. (2002), “A Century Of Purchasing-Power Parity,” *Review of Economics and Statistics*, 84, 139–150.

Table 1: Actual size (5% nominal level) and average correlation ρ of the four unit root tests.

	size		ρ	
	low freq.	high freq.	low freq.	high freq.
DF	0.058	0.056	1.000	1.000
QML-GARCH	0.071	0.046	0.966	0.851
QML- $t(\nu)$	0.076	0.049	0.926	0.842
ML	0.093	0.052	0.907	0.798

Note: Results based on 10,000 replications. The tests DF, QML-GARCH, QML- $t(\nu)$ and ML are defined in the text. “Size” refers to rejecton frequency at the 5% nominal significance level. ρ denotes the average sample correlation between between residuals $\hat{\varepsilon}_t$ and scores \hat{v}_t , where the latter are defined in (13).

Table 2: Unit root test p -values and estimated parameters for the real exchange rate returns.

sampling frequency	Dickey-Fuller		GARCH- $t(\nu)$		AR(1) and GARCH- $t(\nu)$ parameter estimates				
	$p(\hat{\tau}_\mu)$	$p(\Phi_1)$	$p(t_{QML})$	$p(LR)$	$\hat{\gamma} \times 10^2$	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}$	$\hat{\nu}$	$\hat{\rho}$
US dollar versus German mark									
quarterly	0.296	0.433	—	—	−6.327	—	—	—	—
monthly	0.336	0.488	—	—	−1.944	—	—	—	—
weekly	0.350	0.506	0.806	0.931	−0.189	0.096	0.925	11.17	0.920
daily	0.350	0.512	0.861	0.704	−0.016	0.055	0.990	5.640	0.841
US dollar versus British pound									
quarterly	0.155	0.225	—	—	−7.740	—	—	—	—
monthly	0.146	0.211	0.809	0.862	−1.395	0.037	0.987	7.716	0.926
weekly	0.167	0.242	0.853	0.865	−0.124	0.089	0.952	8.725	0.862
daily	0.187	0.271	0.860	0.359	−0.014	0.050	0.994	5.909	0.807
German mark versus British pound									
quarterly	0.082	0.121	—	—	−9.240	—	—	—	—
monthly	0.076	0.111	0.061	0.054	−2.854	0.193	0.906	5.599	0.780
weekly	0.092	0.133	0.020	0.005	−0.687	0.111	0.972	5.747	0.796
daily	0.103	0.146	0.012	0.000	−0.127	0.075	0.997	5.277	0.752

Note: Results based on the estimation period April 1978 through October 2009. $p(\cdot)$ denotes the p -value of the argument. The test statistics $\hat{\tau}_\mu$, Φ_1 , t_{QML} and LR are defined in Section 2. $\hat{\gamma}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\nu}$ are maximum likelihood estimators in the model (19a)-(19c), in cases where this model is estimated; in the other cases, $\hat{\gamma}$ is the least-squares estimate. $\hat{\rho}$ is the sample correlation between residuals $\hat{\varepsilon}_t$ and scores $\hat{\nu}_t$, where the latter are defined in (13).

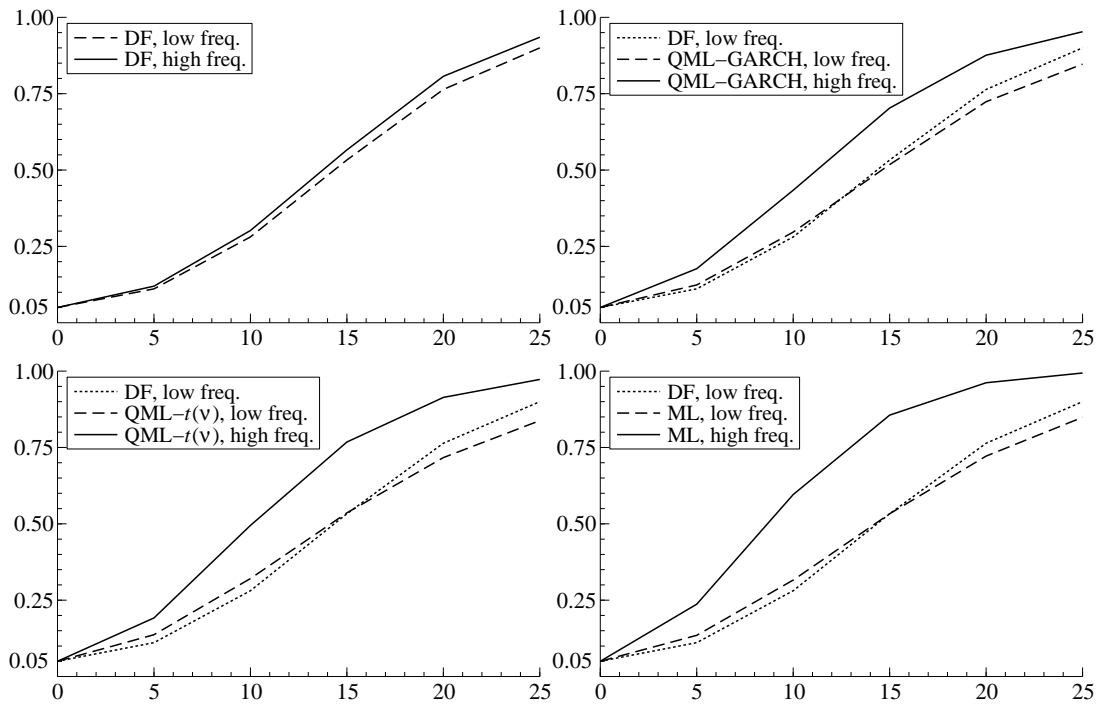


Figure 1: Power (as function of $-c$; 5% nominal level) of four unit root tests.

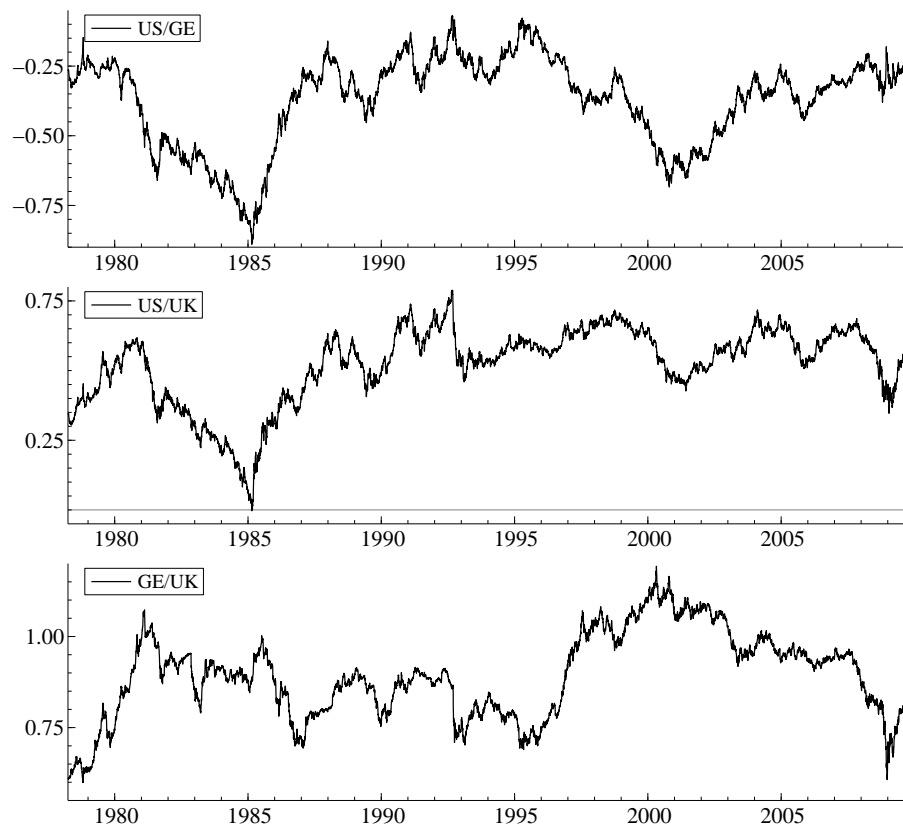


Figure 2: Daily log real exchange rates.

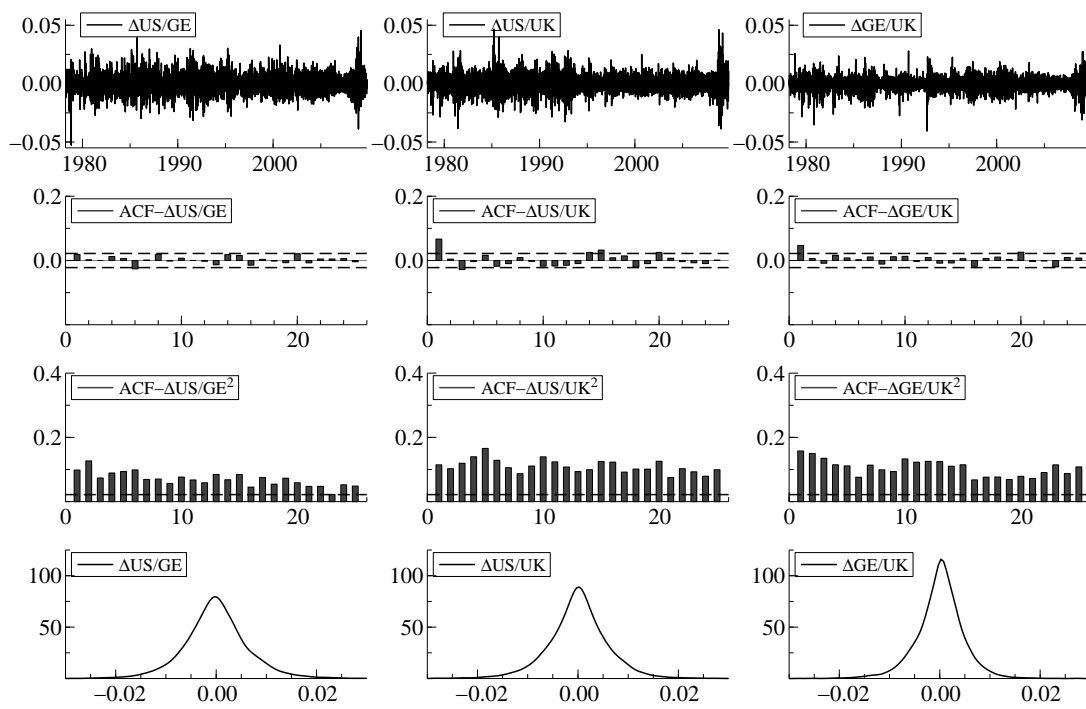


Figure 3: Graphs, correlograms and densities of daily (squared) real exchange rate returns.