UvA ECONOMETRICS

Discussion Paper: 2003/12

Why frequency matters for unit root testing

H. Peter Boswijk and Franc Klaassen

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Why Frequency Matters for Unit Root Testing

H. Peter Boswijk*
Department of Quantitative Economics
Universiteit van Amsterdam

Franc Klaassen
Department of Economics
Universiteit van Amsterdam

February 2003; revised November 2004

Abstract

It is generally believed that for the power of unit root tests, only the time span and not the observation frequency matters. In this paper we show that the observation frequency does matter when the high-frequency data display fat tails and volatility clustering, as is typically the case for financial time series such as exchange rate returns. Our claim builds on recent work on unit root and cointegration testing based non-Gaussian likelihood functions. The essential idea is that such methods will yield power gains in the presence of fat tails and persistent volatility clustering, and the strength of these features (and hence the power gains) increases with the observation frequency. This is illustrated using both Monte Carlo simulations and empirical applications to real exchange rates.

Key words: Fat tails; GARCH; mean reversion; observation frequency; purchasing-power parity; unit roots.
JEL classification: C12, C22, F31.

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

*Corresponding author. Address: Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. E-mail: H.P.Boswijk@uva.nl.

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 (2000).

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 observation 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 observation 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.

This paper argues that the effect of observation frequency on the power of unit root tests is no longer 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 (1997), Ling and Li (1998, 2003), Seo (1999) and Boswijk (2001) 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 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 (2004) found a similar effect of observation 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 of the leading currencies vis-à-vis

the US dollar 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 observation 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 dollar exchange rate of the Japanese yen, the theoretically expected power gain from using higher frequency observations leads to a rejection of the unit root hypothesis by a GARCH-based likelihood ratio test, whereas the Dickey-Fuller test remains insignificant. For 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 observation frequency. The final section contains some concluding remarks.

2 Unit root testing based on Gaussian and non-Gaussian likelihoods

Consider an autoregressive model of order p for a time series $\{X_t\}$, formulated as

$$\Delta X_t = m + \gamma X_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta X_{t-i} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where the starting values $\{X_{1-p}, \ldots, X_0\}$ are considered fixed, and where ε_t is a disturbance term with mean zero. The number of lags p in the model should be such that the resulting disturbance ε_t displays no serial correlation; often p is chosen in practice using a model selection criterion (such as Akaike's or Schwartz's information criterion), in combination with a test for serial correlation.

The unit root hypothesis in this model is $\mathcal{H}_0 : \gamma = 0$, to be tested against the alternative hypothesis $\mathcal{H}_1 : \gamma < 0$. Under the alternative hypothesis, the process $\{X_t\}$ is stationary with a constant mean $\mu = -m/\gamma$. The parameter γ represents the strength of the tendency of the process to revert to this mean. Under the null hypothesis $\gamma = 0$, the mean of $\{X_t\}$ is not defined, and instead the process $\{\Delta X_t\}$ is stationary with a constant mean $m/\left(1 - \sum_{i=1}^{p-1} \phi_i\right)$, which means that $\{X_t\}$ displays a linear trend if $m \neq 0$. 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 useful reparametrization that makes this assumption explicit is

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \sum_{i=1}^{p-1} \phi_i \Delta X_{t-i} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{2}$$

which shows that the constant $m = -\gamma\mu$ drops out of the model under the null hypothesis $\gamma = 0$. Although we do not consider this explicitly, the main conclusion of this paper continues to hold in generalizations of (1)–(2), allowing for a linear trend under both the null and the alternative hypothesis.

2.1 The Dickey-Fuller test

Dickey and Fuller (1981) proposed to test $\mathcal{H}_0$ against $\mathcal{H}_1$ (with the restriction on m implied by (2)) by rejecting the null hypothesis for large values of the F-test statistic for $m = \gamma = 0$, denoted Φ_1, in

the regression model (1). This test is equivalent to the likelihood ratio test for $\mathcal{H}_0$ against $\mathcal{H}_1$ under the assumption that $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d.) $N(0, \sigma^2)$ sequence, but the asymptotic properties of Φ_1, considered in this section, will continue to hold under much wider conditions.

An essential ingredient in deriving the asymptotic properties is the functional central limit theorem, which states that under fairly mild conditions (allowing for excess kurtosis and some forms of volatility clustering in $\{\varepsilon_t\}$),

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \varepsilon_i \xrightarrow{\mathcal{L}} W(s), \qquad s \in [0, 1], \tag{3}$$

where $\lfloor x \rfloor$ denotes the integer part of x, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution as $n \to \infty$, 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 $n \to \infty$,

$$2\Phi_1 \xrightarrow{\mathcal{L}} \int_0^1 dW(s)G(s)' \left[\int_0^1 G(s)G(s)'ds\right]^{-1} \int_0^1 G(s)dW(s), \tag{4}$$

with $G(s)' = [W(s), 1]$. The factor 2 on the left-hand side 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 $m = \gamma = 0$ in (1), and the right-hand side expression also characterizes the limiting null distribution of the likelihood ratio statistic (based on an i.i.d. Gaussian likelihood). This limiting distribution is tabulated by Dickey and Fuller (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 $n \to \infty$,

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

where $H(s) = [U(s), 1]'$, and where $U(s) = \int_0^s e^{c(s-r)}dW(r)$, 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 (5) exceeds the $100(1 - \alpha)\%$ quantile of the null distribution in (4) defines the asymptotic local power function. This provides an approximation to the actual power of Φ_1 for finite samples. As the non-centrality parameter c becomes larger in absolute value (i.e., more negative), the distribution of Φ_1 will shift to the right, leading to higher power.

These results have direct implications for the effect of observation frequency on the power of the test. For example, when we have $n = 25$ annual observations on a time series and the true mean-reversion parameter is $\gamma = -0.2$, then the asymptotic local power is the rejection probability of (5) for $c = n\gamma = -5$. When we extend this sample to a century of data, such that $n = 100$, then the approximate power corresponds to $c = -20$, and will indeed be substantially larger. On the other hand, when we extend the sample by replacing 25 annual observations by $\tilde{n} = 100$ quarterly observations over the same 25 years, then this will not increase the asymptotic local power, because the quarterly mean-reversion parameter will be correspondingly smaller ($\tilde{\gamma} = -0.05$ on a quarterly basis), so that

the same non-centrality parameter $c = \tilde{n}\tilde{\gamma} = -5$ and hence the same asymptotic local power applies. Therefore, increasing the observation frequency has no effect on the local asymptotic power based on (5).

As indicated above, these asymptotic results also apply when $\{\varepsilon_t\}$ is a stationary GARCH process, or when the distribution of $\{\varepsilon_t\}$ has a larger kurtosis than the Gaussian distribution. The main assumption needed is that $\{\varepsilon_t\}$ has a finite unconditional variance. 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 likelihood

Unit root testing based on a non-Gaussian or GARCH likelihood is considered by Lucas (1997), Ling and Li (1998, 2003), Seo (1999), and Boswijk (2001), *inter alia*. We refer to these papers and the references therein for a full derivation of these results, and only mention the most important aspects here.

For notational convenience we consider a lag length of $p = 1$; the result may be generalized to higher-order autoregressions. Consider, therefore, the model

$$\Delta X_t = m + \gamma X_{t-1} + \varepsilon_t, \tag{6a}$$

$$\varepsilon_t = \sigma_t \eta_t, \tag{6b}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{6c}$$

$$\eta_t \sim \text{i.i.d. } g(\eta_t), \tag{6d}$$

where $g(\eta_t)$ is a (possibly non-Gaussian) density with zero mean and unit variance, such that σ_t^2 is the conditional variance of ε_t. We assume that $\alpha + \beta < 1$, such that $\{\varepsilon_t\}$ has a finite unconditional variance $\sigma^2 = \omega/(1 - \alpha - \beta)$. Let $\delta = (\gamma, m)'$ 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. Then the log-likelihood function has the following form:

$$\ell(\theta) = \sum_{t=1}^{n} \left\{ -\frac{1}{2} \log \sigma_t^2 + \log g \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \right\}. \tag{7}$$

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

$$LR = -2 \left[\ell(\tilde{\theta}) - \ell(\hat{\theta}) \right], \tag{8}$$

where $\hat{\theta}$ is the unrestricted maximum likelihood (ML) estimator, and $\tilde{\theta}$ is the ML estimator under the restriction $\delta = 0$.

The asymptotic properties of the LR 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 score function $\psi(\eta_t) = -\partial \log g(\eta_t)/\partial \eta_t$, the partial derivative of the log-likelihood with respect to δ is given by

$$\frac{\partial \ell(\theta)}{\partial \delta} = \sum_{t=1}^{n} \left\{ \frac{1}{\sigma_t} Z_t \psi\left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t}\right) + \frac{1}{2\sigma_t^2}\left[\psi\left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t}\right)\frac{\Delta X_t - \delta' Z_t}{\sigma_t} - 1\right]\frac{\partial \sigma_t^2}{\partial \delta}\right\}, \quad (9)$$

where the GARCH model implies that

$$\frac{\partial \sigma_t^2}{\partial \delta} = -2\alpha \sum_{i=1}^{t-1} \beta^{i-1} Z_{t-i}(\Delta X_{t-i} - \delta' Z_{t-i}). \quad (10)$$

Evaluating (9) in $\delta = 0$ gives the moment condition that is tested by the LR test.

Under the null hypothesis, this moment condition (divided by n) is asymptotically equivalent to

$$\frac{1}{n}\sum_{t=1}^{n} Z_t \left(\frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2}\left[\psi(\eta_t)\eta_t - 1\right]\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right) = \frac{1}{n}\sum_{t=1}^{n} Z_t \upsilon_t, \quad (11)$$

where υ_t is implicitly defined. It follows that $\{\upsilon_t\}$ is a martingale difference sequence with finite variance τ^2, and that $\mathrm{cov}(\varepsilon_t, \upsilon_t) = 1$. This leads to the following bivariate functional central limit theorem: as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor sn \rfloor}\begin{pmatrix} \varepsilon_i \\ \upsilon_i \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \sigma W(s) \\ \tau B(s) \end{pmatrix}, \qquad s \in [0,1], \quad (12)$$

where $B(s)$ is a standard Brownian motion, such that $\mathrm{cov}\left(\sigma W(s), \tau B(s)\right) = \mathrm{cov}(\varepsilon_t, \upsilon_t) = 1$. Hence the correlation between $W(s)$ and $B(s)$ is given by $\rho = 1/(\sigma\tau)$. The sample moment (11) converges in distribution, as $n \to \infty$, to the stochastic integral $\sigma\tau \int_0^1 G(s)dB(s)$, and this in turn implies, under $\mathcal{H}_0$ and as $n \to \infty$,

$$LR \xrightarrow{\mathcal{L}} \int_0^1 dB(s)G(s)' \left[\int_0^1 G(s)G(s)'ds\right]^{-1} \int_0^1 G(s)dB(s), \quad (13)$$

where $G(s)$ is the same as in (4).

For an i.i.d. Gaussian (quasi-) likelihood, it can be derived that $\upsilon_t = \varepsilon_t/\sigma^2$, 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 $g(\eta_t)$.[1] This implies that the limiting null distribution of LR depends on the nuisance parameter $\rho = \mathrm{corr}(W(1), B(1))$. This parameter ρ may be estimated consistently by the sample correlation of $\hat{\varepsilon}_t$ and $\hat{\upsilon}_t$, and approximate quantiles of the asymptotic null distribution for a given value of ρ may be obtained from the Gamma approximation discussed in Boswijk and Doornik (2004).

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

$$LR \xrightarrow{\mathcal{L}} \int_0^1 \left[dB(s) + \frac{c}{\rho}U(s)ds\right] H(s)' \left[\int_0^1 H(s)H(s)'ds\right]^{-1} \int_0^1 H(s)\left[dB(s) + \frac{c}{\rho}U(s)ds\right], \quad (14)$$

[1] The limit $\rho \to 0$ corresponds to the distribution of $\{\varepsilon_t\}$ approaching an infinite-variance distribution; however, in that case (12) no longer applies and a different limit theory is needed.

where $U(s)$ and $H(s)$ are the same as before. Comparing (14) with (5), 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 LR over Φ_1 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 observation frequency on power. As we move from low-frequency data to high-frequency data, maintaining the same time span, the parameter c does not change, for the same reasons as indicated in the previous sub-section. However, the results of Drost and Nijman (1993) on temporal aggregation of GARCH processes 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 the distribution of $\{\eta_t\}$ will be higher. Both effects will cause ρ to be smaller for high-frequency data, so that the power of the test increases with observation 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 model

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \varepsilon_t, \tag{15a}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{15b}$$

$$\eta_t := \frac{\varepsilon_t}{\sigma_t} \sim \text{i.i.d. } t(\nu), \qquad t = 1, \ldots, n. \tag{15c}$$

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 $\sigma_0^2 = \sigma^2 := \omega/(1 - \alpha - \beta)$, 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 $\{X_t, t = 0, \ldots, n\}$, 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

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

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 temporal aggregation, we then construct a low-frequency sample $\{\tilde{X}_\tau, \tau = 0, \ldots, \tilde{n}\}$ by a skip-sample version of X_t, i.e.,

$$\tilde{X}_\tau = X_{m\tau}, \qquad \tau = 0, \ldots, \tilde{n} = \frac{n}{m}, \tag{16}$$

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 $\tilde{n} = 100$; thus we mimic a sample of about 8 years of daily data $\{X_t, t = 1, \ldots, n\}$, and the corresponding 8 years of monthly data $\{\tilde{X}_\tau, \tau = 1, \ldots, \tilde{n}\}$. We do not attempt to provide full characterization of the implied data-generating process for $\{\tilde{X}_\tau\}$. The main result for our purpose, which follows from the analysis of Drost and Nijman (1993), is that under the null hypothesis $\gamma = 0$, $\Delta\tilde{X}_\tau$ is a weak[2] GARCH(1,1) process with $\tilde{\alpha} + \tilde{\beta} = (\alpha+\beta)^m = 0.54$ (so that the volatility displays less persistence), and that the kurtosis of the standardized errors decreases with temporal aggregation. 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 temporally aggregated 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:

- Φ_1, the Dickey-Fuller F-test for $\gamma = m = 0$ in the least-squares regression (1) with $p = 1$;
- QLR_G, the quasi-likelihood ratio (QLR) test for $\gamma = m = 0$, based on a Gaussian GARCH(1,1) model (ignoring the conditional t-distribution of the DGP);
- QLR_t, the QLR test for $\gamma = m = 0$, based on an i.i.d. $t(\nu)$ model for ε_t/σ (ignoring the GARCH(1,1) aspect of the DGP);
- LR, the LR test for $\gamma = m = 0$, based on a GARCH(1,1)-$t(\nu)$ model.

The motivation for studying QLR_G and QLR_t 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; although there is no formal proof available of their asymptotic validity, we suspect that by estimating the correlation parameter ρ from the data, we may diminish the effects of misspecification on the size of the tests. The LR test based on the GARCH-t likelihood is actually only a true likelihood ratio 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 QLR test.

We obtain p-values for each of the QLR and LR statistics using the Gamma approximation of Boswijk and Doornik (2004). The correlation parameter ρ is estimated simply as the sample corre-

[2]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.

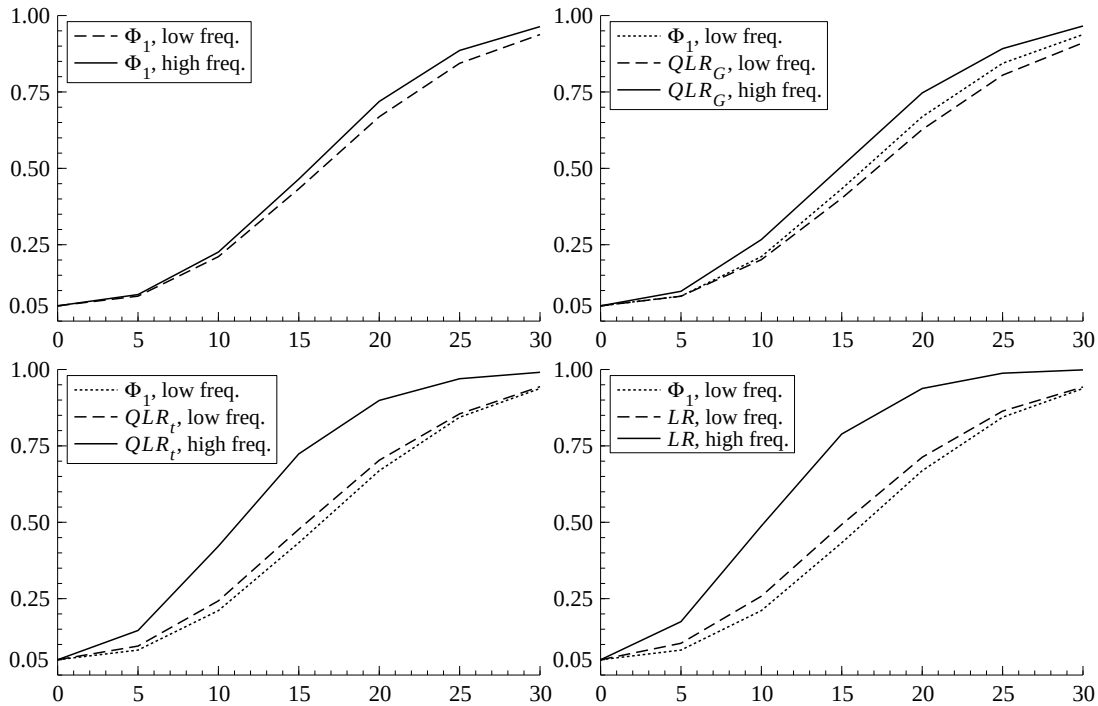
lation of the the residuals $\hat{\varepsilon}_t$ and the "scores" $\hat{\upsilon}_t$, both evaluated at the unrestricted estimates. All results below are based on 10000 replications, and have been obtained using Ox 3.3, see Doornik (2001).

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.
Φ_1	0.059	0.055	1.000	1.000
QLR_G	0.067	0.079	0.966	0.851
QLR_t	0.065	0.054	0.926	0.842
LR	0.065	0.054	0.907	0.798

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 10000 replications) estimates of the correlation parameter ρ. We observe that all tests display moderate over-rejection when based on 100 monthly observations. These distortions generally decrease as we move to 2000 daily observations, with the exception of the QLR_G test, where the over-rejection increases. Further simulations indicate that as n increases, the size distortion of QLR_G eventually decreases, albeit slowly. 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 LR test.

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



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 $\gamma \in \{0, -0.0025, \ldots,$

$-0.015\}$. 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)LR 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 QLR_G, QLR_t and LR.

As predicted from the values of ρ in Table 1, the largest power gains from high-frequency observations are obtained by the LR test, closely followed by the QLR_t test. The QLR_G test, based on a Gaussian GARCH quasi-likelihood function, is less 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 QLR_G test is even slightly less than that of the corresponding Dickey-Fuller test, which indicates that in such small samples, the gain from modelling the GARCH effect is more than off-set by the degrees-of-freedom loss from estimating more parameters. Finally, we see that as expected, the power of Φ_1 is hardly affected by the sampling frequency.

The power curves also allow us to compare the effects of observation frequency and time span on the power of the tests. Concentrating on the LR test, we see that for the daily frequency, a non-centrality parameter of $c = -10$ is sufficient to obtain a power of about 50%. When using monthly data, a non-centrality of about $c = -15$ 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 $\tilde{\gamma} = 20\gamma = -0.1$, so that over 12 year of monthly observations ($\tilde{n} = 150$) are needed to obtain the desired non-centrality parameter $c = -15$. In other words, in this specific case 12 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: temporal disaggregation 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 March 2002). We focus on the three main real exchange rates of the German mark[3], the British pound and the Japanese yen vis-à-vis the US dollar. We use end-of-quarter, end-of-month, end-of-week and daily nominal exchange rates (dollar prices of one unit of foreign currency). These are converted to real exchange rates using OECD producer price indices for manufacturing goods, where weekly and daily observations have been obtained by linear interpolation of the monthly log-prices. 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.

[3]For the euro period, we use the fixed €/DM exchange rate to convert dollar-euro rates to dollar-mark rates.

Figure 2: Daily log real exchange rates of three currencies against the US dollar.

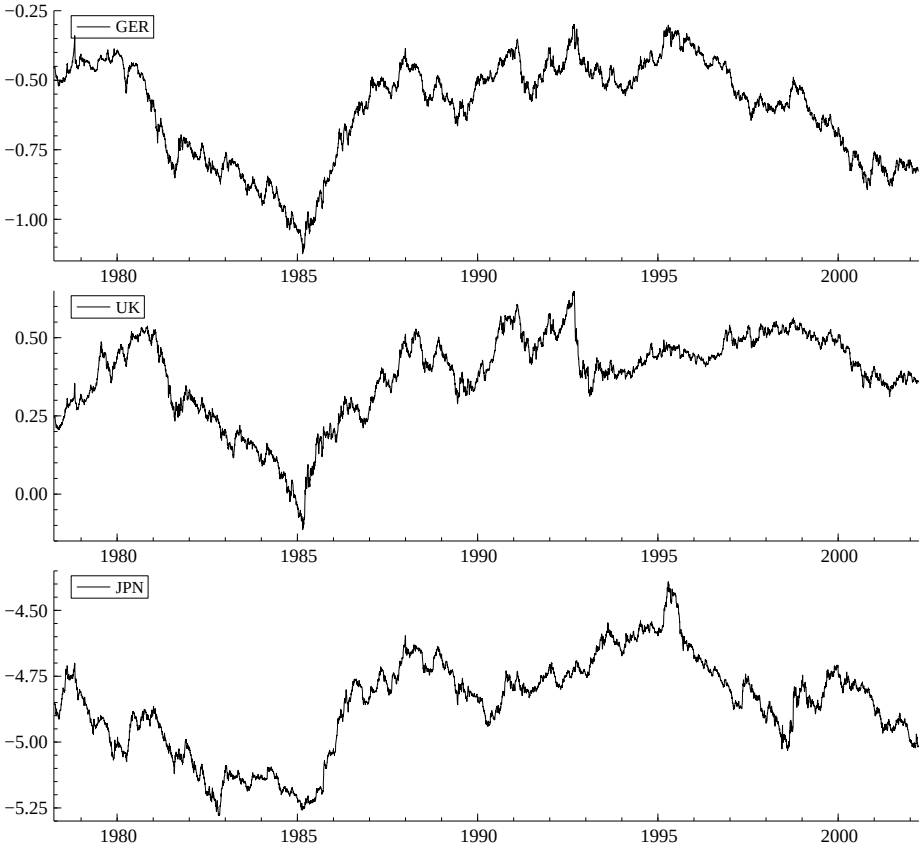
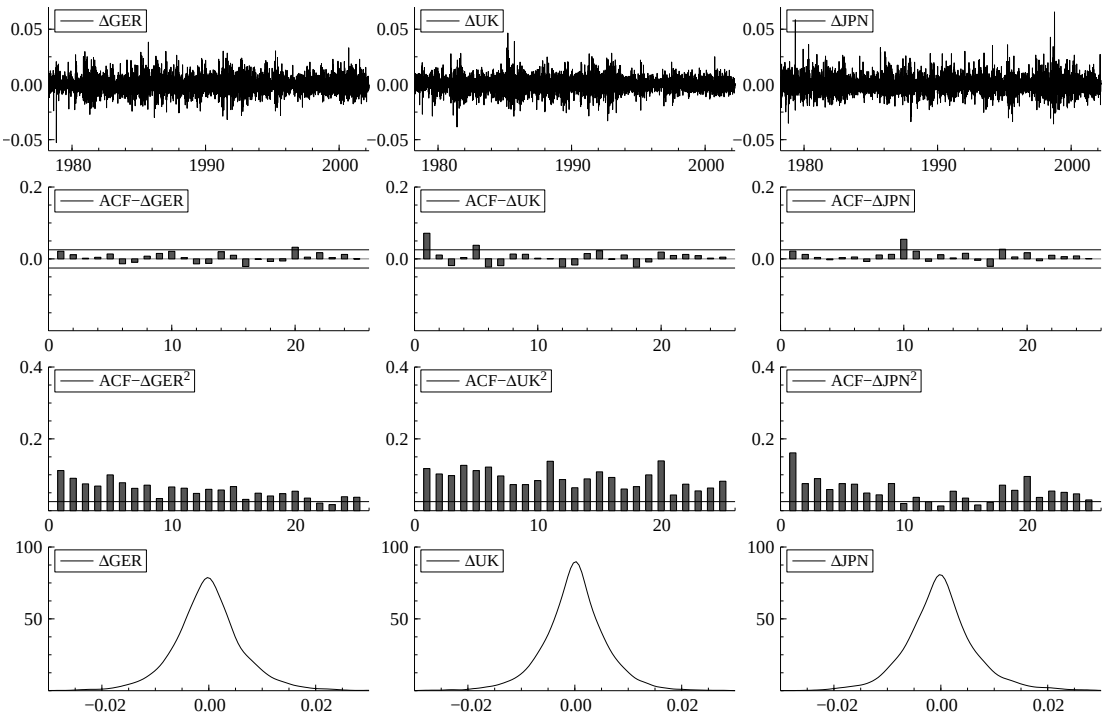


Figure 3: Graphs, correlograms and densities of the (squared) real exchange rate returns.



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.

To investigate whether there is any scope for power improvement due to fat-tailedness and volatility clustering, Figure 3 depicts some stylized properties of daily exchange rate 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 5.59, 6.23 and 7.73, 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 models for the returns.

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

observation	p-values		AR-GARCH-t parameter estimates			
frequency	$p(\Phi_1)$	$p(LR)$	$\hat{\gamma} \times 10^3$	$\hat{\alpha}$	$\hat{\alpha}+\hat{\beta}$	$\hat{\nu}$
German mark						
quarterly	0.641	–	–	–	–	–
monthly	0.720	–	–	–	–	–
weekly	0.720	0.673	0.035	0.114	0.854	10.13
daily	0.730	0.074	0.121	0.066	0.987	5.025
British pound						
quarterly	0.399	–	–	–	–	–
monthly	0.353	–	–	–	–	–
weekly	0.430	0.996	−0.353	0.058	0.975	7.342
daily	0.468	0.718	−0.038	0.054	0.994	5.261
Japanese yen						
quarterly	0.606	–	–	–	–	–
monthly	0.645	–	–	–	–	–
weekly	0.745	0.062	−1.424	0.061	0.944	5.364
daily	0.772	0.003	−0.048	0.066	0.981	4.183

Table 2 displays the p-values of the Dickey-Fuller Φ_1 test based on (1), and the LR test based on (6) with $\eta_t \sim$ i.i.d. $t(\nu)$, for the three different currencies and four different frequencies. For ease of comparison, we only consider first-order autoregressive models, i.e., $p = 1$. For the daily data, this often leads to some residual autocorrelation, but accounting for this by including lagged differences hardly affects the results in Table 2. The GARCH-t 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. For the monthly and quarterly series, the volatility clustering

and leptokurtosis are generally 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 10.1, see Doornik and Hendry (2001).

Consider first the p-values of the Dickey-Fuller Φ_1 test. We observe that for all three currencies, 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 conventional 5% level), and this conclusion is entirely unaffected by the choice of observation frequency. This may be seen as an illustration of Shiller and Perron's (1985) conclusion that observation frequency matters very little for the power of the Dickey-Fuller test.

The p-values of the LR test, on the other hand, display much more variation. For the German mark, the two tests have a very similar p-value at the weekly frequency, but when we move to daily data, some weak evidence against the unit root hypothesis seems to be indicated by the LR test, with a p-value of 7.4%. Closer inspection, however, reveals that the estimated mean-reversion parameter $\hat{\gamma}$ in this case has the wrong sign, such that this may not be interpreted as evidence in favour of purchasing-power parity.

For the British pound, we observe again considerable variation in the p-values of the LR test when we move from weekly to daily data, but this does not lead to more evidence against the unit root hypothesis. Note that for this currency, as well as for the other two, the estimated GARCH parameters $\hat{\alpha}$ and $\hat{\beta}$, and the estimated degrees-of-freedom parameter $\hat{\nu}$, indicate more persistent volatility clustering and stronger leptokurtosis as we increase the observation frequency, as predicted by the analysis of Drost and Nijman (1993).

Finally, we consider the results of the GARCH-t based LR test for the Japanese yen. Here we see that the evidence in favour of mean reversion in the real exchange rate does increase as we increase the observation frequency. For weekly observations, the LR test is significant at the 10% level (as opposed to the Dickey-Fuller test with a p-value of 75%), and when moving to daily data the LR test rejects the unit root even at the 1% significance level. Note that the estimated mean-reversion parameter is negative but very small in absolute value, especially at the daily frequency. Therefore, the mean-reversion tendency is very weak, but the LR test indicates that it is significant.

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 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. A possible explanation of the difference between the Japanese yen and the other two currencies is that the other two exchange rates display more prominent swings in the first decade than the dollar-yen rate. This might indicate that the possible mean-reverting behaviour 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, the models may interpret the swings as mean aversion, leading to an upward distortion of the estimate of γ. 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 Japanese yen, 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 could also try to apply GARCH likelihoods in a cointegration context, the main problem here is to find a parsimoniously parametrized multivariate GARCH model that is reasonably well specified. We leave this problem for future research.

References

Boswijk, H. P. (2001), "Testing for a Unit Root with Near-Integrated Volatility", Tinbergen Institute Discussion Paper # 01-077/4 (http://www.tinbergen.nl/discussionpapers/01077.pdf).

Boswijk, H. P. and J. A. Doornik (2004), "Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors", *Journal of Applied Econometrics*, forthcoming.

Boswijk, H. P. and A. Lucas (2002), "Semi-Nonparametric Cointegration Testing", *Journal of Econometrics*, 108, 253–280.

Campbell, J. Y. and P. Perron (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 C. Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics*, 15, 1050–1063.

Dickey, D. A. and W. A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 1057–1072.

Doornik, J. A. (2001), *Ox 3.0. An Object-Oriented Matrix Programming Language*. London: Timberlake Consultants Press and Oxford: http://www.nuff.ox.ac.uk/users/Doornik.

Doornik, J. A. and D. F. Hendry (2001), *Econometric Modelling Using PcGive 10, Volume III*. London: Timberlake Consultants Press.

Drost, F. C. and Th. E. Nijman (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 A. K. Rose (1996), "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries", *Journal of International Economics*, 40, 209–224.

Klaassen, F. (2004), "Long Swings in Exchange Rates: Are They Really in the Data?", *Journal of Business & Economic Statistics*, forthcoming.

He, C. and T. Teräsvirta (1999), "Properties of Moments of a Family of GARCH Processes", *Journal of Econometrics*, 92, 173–192.

Ling, S. and W. K. Li (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 W. K. Li (2003), "Asymptotic Inference for Unit Root Processes with GARCH(1,1) Errors", *Econometric Theory*, 19, 541–564.

Lothian, J. and M. P. Taylor (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. (1997), "Cointegration Testing Using Pseudo Likelihood Ratio Tests", *Econometric Theory*, 13, 149–169.

Seo, B. (1999), "Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity", *Journal of Econometrics*, 91, 113–144.

Shiller, R. J. and P. Perron (1985), "Testing the Random Walk Hypothesis: Power versus Frequency of Observation", *Economics Letters*, 18, 381–386.

Taylor, A. M. (2000), "A Century of Purchasing-Power Parity", NBER Working Paper Series # 8012 (http://www.nber.org/papers/w8012.pdf).