Why frequency matters for unit root testing

Boswijk, H.P.; Klaassen, F.

[Link to publication](#)

UvA ECONOMETRICS

Why frequency matters for unit root testing

H. Peter Boswijk and Franc Klaassen

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

Abstract

It is generally believed that for the power of unit root tests, only the time span and not the observation frequency matters. In particular this has lead researchers in the area of purchasing power parity to test for a unit root in real exchange rates based on data covering a time span of a century or more. 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 exhange 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 can yield substantial power gains in the presence of fat tails and persistent volatility clustering, and these feautures (and hence the power gains) typically decrease with temporal aggregation. 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.

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 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, researcher have tried 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 (2000).

---

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

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 test is mainly affected by time span, and much less by observation frequency. This result was first derived by Schiller 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 Schiller and Perron (1985) is provided by a continous record asymptotic argument. Suppose that the data may be seen as discrete observations from the continous-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. This approximate local power is the same, whether we consider 10 years of quarterly data or 10 years of daily data. Although the actual 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 result that only time span matters for the power of unit root tests is violated 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. Recent research by Lucas (1997), Ling and Li (1997, 1998), Boswijk (2001) and Klüppelberg *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. And because these properties tend to become less pronounced with temporal aggregation, the possibilities of high-powered unit root tests also decreases when one moves from high-frequency to low-frequency observations. In this sense, frequency does matter. We demonstrate this claim 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 non-Gaussian likelihood functions. Section 3 discusses a stylized Monte Carlo experiment, showing that in a realistic situation, the possible power gains at the daily frequency reduce considerably when moving to monthly data. Section 4 discusses the empirical analysis; here we show that for the real dollar exchange rate of the German mark, the British pound and the Japanese yen, the theoretically expected power gain indeed leeds to stronger evidence in favour of purchasing power parity. For a fourth exchange rate (of the Canadian versus the US dollar), we do not find any improvement over the Dickey-Fuller test. The final section contains some concluding remarks.

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

This section reviews the theory of testing for a unit root based on non-Gaussian and/or GARCH likelihood functions, building on recent work by Lucas (1997), Ling and Li (1997, 1998), Boswijk (2001), and Klüppelberg *et al.* (2002). To test for a unit root in a series $\{X_t, t = 1, \ldots, n\}$ which does not display any deterministic trend, we typically consider an autoregressive model of order p, rewritten as

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \sum_{i=1}^{p-1} \delta_i \Delta X_{t-i} + \varepsilon_t, \tag{1}$$

where the unit root hypothesis is $\mathcal{H}_0 : \gamma = 0$, to be tested against the stationarity alternative $\mathcal{H}_1 : \gamma < 0$. De model is usually linearized, replacing $\gamma(X_{t-1} - \mu)$ by $m + \gamma X_{t-1}$, where the restriction that $m = -\gamma\mu = 0$ under $\mathcal{H}_0$ is imposed either implicitly or explicitly, such that the process has a possibly non-zero mean but no linear trend under both the null and the alternative hypothesis. An explicit treatment of this restriction leads to the Φ_1 test of Dickey and Fuller (1981), i.e., the F-test for the exclusion of the constant term and X_{t-1}. Under the assumption that $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$, and with fixed starting values, this is a monotone transformation of the likelihood ratio test statistic for $\mathcal{H}_0$. The F-statistic does not have an asymptotic F-distribution under the null, but another asymptotic distribution, discussed below. The number of lags p in the model should be such that the resulting disturbance ε_t should display 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 asymptotic distribution of the likelihood ratio statistic for $\mathcal{H}_0$ is often characterized in terms of functionals of a Brownian motion process. An essential ingredient is the functional central limit theorem, which states that under fairly mild conditions on $\{\varepsilon_t\}$,

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \varepsilon_i \xrightarrow{\mathcal{L}} W(s), \qquad s \in [0, 1], \tag{2}$$

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

$$2\Phi_1 \xrightarrow{\mathcal{L}} \int_0^1 dW(s)F(s)' \left[\int_0^1 F(s)F(s)'ds \right]^{-1} \int_0^1 F(s)dW(s), \tag{3}$$

with $F(s)' = [W(s), 1]$. The distribution of the right-hand side expression is tabulated by Dickey and Fuller (1981), based on simulation of a discretization of the relevant integrals.

Under a sequence of local alternatives $\mathcal{H}_n : \gamma = c/n$, it can be shown that

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

where $U(s) = \int_0^s e^{c(s-r)}dW(r)$, an Ornstein-Uhlenback process, which is the solution to the stochastic differential equation $dU(s) = cU(s)ds + dW(s)$. Now $F(s)$ is defined as $[U(s), 1]'$. The probability

3

that the right-hand side expression in (4) exceeds the $100(1 - \alpha)\%$ quantile of the null distribution in (3) defines the asymptotic local power function. This provides an approximation to the actual power of Φ_1 for finite samples.

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 power function corresponds to the rejection probability of (4) 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 100 quarterly observations over the same 25 years, then this will not increase the power substantially, 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 $c = \tilde{\gamma}\tilde{n} = -5$ and hence the same asymptotic local power applies.

The results so far do not require ε_t to be i.i.d. Gaussian; the functional central limit theorem (2) also applies when ε_t is a stationary GARCH process, or when the distribution of ε_t has a bigger kurtosis than the Gaussian distribution. The main assumption needed is that ε_t has a finite unconditional variance, plus some fairly mild additional assumptions. However, when the actual data-generating process displays such deviations from the i.i.d. Gaussian assumption, then the Dickey-Fuller test, which is based on a Gaussian likelihood, is no longer optimal. In that case, tests based on a likelihood function that captures this volatility clustering and distributional shape will have a (sometimes substantially) bigger asymptotic local power, as shown by Lucas (1997) 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.

Suppose that the lag length $p = 1$ for convenience. Then the Dickey-Fuller test essentially checks whether the sample moment

$$\frac{1}{\sigma^2 n} \sum_{t=1}^{n} Z_t \Delta X_t = \frac{1}{\sigma^2 n} \sum_{t=1}^{n} \begin{pmatrix} 1 \\ X_{t-1} \end{pmatrix} \Delta X_t \tag{5}$$

differs significantly from 0. On the other hand, when we assume that $\eta_t := \varepsilon_t / \sigma_t \sim$ i.i.d. $g(\eta_t)$, where $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$, then the average log-likelihood function will look like

$$\bar{\ell}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ -\frac{1}{2} \log \sigma_t^2 + \log g \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \right\}, \tag{6}$$

where $\delta = (m, \gamma)'$, and where θ contains δ, the GARCH parameters (ω, α, β), and possible additional parameters characterizing the distributional shape of η_t. We assume that $\text{var}(\eta_t) = 1$, such that σ_t^2 is the conditional variance of ε_t. Defining the score function $\psi(\eta_t) = -\partial \log g(\eta_t)/\partial \eta_t$, the partial derivative of the average log-likelihood with respect to δ is given by

$$\frac{\partial \bar{\ell}(\theta)}{\partial \delta} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{1}{\sigma_t} Z_t \psi \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) + \frac{1}{2\sigma_t^2} \left[\psi \left(\frac{\Delta X_t - \delta' Z_t}{\sigma_t} \right) \frac{\Delta X_t - \delta' Z_t}{\sigma_t} - 1 \right] \frac{\partial \sigma_t^2}{\partial \delta} \right\}, \tag{7}$$

where the GARCH model implies that

$$\frac{\partial \sigma_t^2}{\partial \delta} = -2\alpha \sum_{i=1}^{t-1} \beta^{i-1} Z_{t-i} (\Delta X_{t-i} - \delta' Z_{t-i}). \tag{8}$$

Evaluating (7) in $\delta = 0$ gives the moment condition that is tested by the QLR test.

Under the null hypothesis, this moment condition is asymptotically equivalent to

$$\frac{1}{n}\sum_{t=1}^{n} Z_t \left(\frac{\psi(\eta_t)}{\sigma_t} - \frac{\alpha}{\sigma_t^2} \left[\psi(\eta_t)\eta_t - 1 \right] \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) = \frac{1}{n}\sum_{t=1}^{n} Z_t v_t, \tag{9}$$

which converges in distribution, as $n \to \infty$, to $\sigma\tau \int_0^1 F(s)dB(s)$, where $\sigma^2 = \omega/(1-\alpha-\beta) = \text{var}(\varepsilon_t)$, $\tau^2 = \text{var}(v_t)$, and where $B(s)$ is a standard Brownian motion, obtained as the limit in distribution of $(\tau^2 n)^{-1/2}\sum_{i=1}^{\lfloor sn \rfloor} v_i$; the process F is as defined before. Since $\text{cov}(v_t, \varepsilon_t) = 1$, it follows that $\rho = \text{corr}(v_t, \varepsilon_t) = 1/(\sigma\tau)$, and this is also the correlation between the two standard Brownian motions $W(s)$ and $B(s)$. For a Gaussian i.i.d. likelihood it can be checked that $v_t = \varepsilon_t/\sigma^2$, so that $\rho = 1$. In general however, $0 \leq \rho \leq 1$, where $\rho = 0$ corresponds to the limiting case of an infinite variance process for ε_t. The limiting distribution of the QLR statistic under local alternatives now may be expressed as

$$QLR \xrightarrow{\mathcal{L}} \int_0^1 \left[dB(s) + \frac{c}{\rho}U(s)ds \right] F(s)' \left[\int_0^1 F(s)F(s)'ds \right]^{-1} \int_0^1 F(s)] \left[dB(s) + \frac{c}{\rho}U(s)ds \right], \tag{10}$$

where $U(s) = \int_0^s e^{c(s-r)}dW(r)$ is the same Ornstein-Uhlenbeck process as before, and $F(s) = [U(s), 1]'$. Comparing (10) with (4), we see that the non-centrality parameter c now has been replaced by c/ρ, which explains why the largest power gains of QLR over Φ_1 are obtained for cases where ρ is relatively small (note that the limiting case $\rho = 0$ is not allowed, because an infinite-variance process does not satisfy a functional central limit theorem).

Under the null hypothesis $c = 0$, (10) reduces to $\int_0^1 dBF' \left[\int_0^1 FF'ds \right]^{-1} \int_0^1 FdB$, the distribution of which still depends on $\rho = \text{corr}(B(1), W(1))$. However, the parameter ρ may be estimated consistently by the sample correlation of $\hat{v}_t$ and $\hat{\varepsilon}_t$, 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 (1999).

3 A Monte Carlo experiment

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

$$\begin{aligned}
\Delta X_t &= \gamma(X_{t-1} - \mu) + \varepsilon_t, \\
\sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \\
\eta_t &:= \frac{\varepsilon_t}{\sigma_t} \sim \text{i.i.d. } t(\nu), \qquad t = 1, \ldots, n.
\end{aligned}$$

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 $\sigma_0^2 = \bar{\sigma}^2 := \omega/(1-\alpha-\beta)$, and $X_0 = \mu + \sigma_0\eta_0$.

For each replication, we first generate n observations of X_t, using parameter values that mimic properties of daily financial data, i.e.,

$$\alpha = 0.08, \quad \beta = 0.9, \quad \nu = 5,$$

and for γ we choose $\gamma = 1 + c/n$, where $c \in \{0, -5, -20\}$; thus we study both the size and the power of various tests, the latter for the case of weak ($c = -5$) and relatively strong ($c = -20$) mean reversion. 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 repeat the analysis with a skip-sample version of X_t, i.e.,

$$\tilde{X}_\tau = X_{m\tau}, \qquad \tau = 1, \ldots, \tilde{n} = \frac{n}{m},$$

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 $\tilde{n} = 100$; 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 results of Drost and Nijman (1993) imply that under the null hypothesis $\gamma = 0$, $\Delta \tilde{X}_\tau$ is a weak[1] GARCH(1,1) process with $\tilde{\alpha} + \tilde{\beta} = (\alpha + \beta)^m = 0.67$ (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 high- and the low-frequency data we apply four unit root tests:

- Φ_1, Dickey and Fuller's (1981) F-test for $\gamma = m = 0$ in the least-squares regression $\Delta X_t = m + \gamma X_{t-1} + \varepsilon_t$;

- QLR_{G-t}, the (Q)LR test for $\gamma = m = 0$, based on a GARCH(1,1)-$t(\nu)$ likelihood function with estimated ν; for the high-frequency data, this model is correctly specified, so this is the LR test;

- QLR_G, the QLR test for $\gamma = m = 0$, based on a Gaussian GARCH(1,1) likelihood function;

- QLR_t, the QLR test for $\gamma = m = 0$, based on an i.i.d. $t(\nu)$ model for ε_t/σ.

The motivation for studying QLR_G and QLR_t is to quantify the relative contribution of volatility clustering and fat-tailedness to the possible power gains.

We obtain p-values for each of the QLR statistics using the Gamma approximation of Boswijk and Doornik (1999). The correlation parameter ρ is estimated simply as the sample correlation of the "scores" $\hat{\upsilon}_t$ and the residuals $\hat{\varepsilon}_t$, both evaluated at the unrestricted estimates. All results below are based on 1000 replications, and have been obtained using Ox 3.2, see Doornik (2001).

---

[1]Drost and Nijman (1993) define a process ε_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.

Table 1: Rejection frequencies of the four unit root tests, high-frequency data, $n = 2000$.

c	Φ_1	QLR_{G-t}	QLR_G	QLR_t
0	0.057	0.057	0.096	0.062
-5	0.108	0.262	0.203	0.215
-20	0.714	0.955	0.867	0.895
ρ	1	0.749	0.791	0.809

Table 1 displays the rejection frequencies under the null and local alternatives of the four tests, for the high-frequency data. Also indicated are the average (over 1000 replications) estimates of the correlation parameter ρ. We observe that QLR_G has serious size distortions, whereas for the other three tests these size distortions are fairly moderate. Next, we see that all QLR tests have substantially better power than the Dickey-Fuller test, with the largest power gain for the QLR_{G-t} test (which is the likelihood ratio test in this case), and the smallest for the Gaussian GARCH-based test QLR_G. 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. The average values of the correlation parameter ρ is the smallest for QLR_{G-t}, corresponding also to the biggest power. From asymptotic theory, one would indeed expect the power to decrease with ρ; we see that this prediction is not satisfied by the relative power ordering of QLR_G and QLR_t, which presumably is a finite-sample effect.

Table 2: Rejection frequencies of the four unit root tests, low-frequency data, $\tilde{n} = 100$.

| c	Φ_1	QLR_{G-t}	QLR_G	QLR_t
0	0.067	0.073	0.082	0.069
-5	0.103	0.155	0.130	0.135
-20	0.693	0.804	0.691	0.788
ρ	1	0.864	0.935	0.894

Table 2 provides the corresponding figures for the low-frequency, temporally aggregated observations. As expected, the size distortions are somewhat larger for this smaller sample size. We see that the power of Φ_1 decreases only very little, confirming the basic idea that for Gaussian likelihood-based tests, the power of unit root tests is affected mainly by time span, and only very little by observation frequency. In contrast, we see a much more pronounced decrease in the power of the QLR tests, which now seem to approach the Φ_1 test (even though there is still some power superiority left). This is also reflected in the estimates of ρ, which are now closer to one.

In summary, this very limited Monte Carlo experiment confirms the theoretical predictions: temporal aggregation reduces 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 daily, weekly and monthly real exchange rate data in the post-Bretton Woods era (using data from April 1974 to March 2002). We focus on the real exchange rates of the German mark, British pound, Japanese yen and Canadian dollar vis-à-vis the US dollar. We use end-of-day, end-of-week and end-of-month nominal exchange rates. These are converted to real exchange rates using OECD producer price indices for manufacturing goods, where daily and weekly observations have been obtained by linear interpolation of the monthly log-prices.

Figure 1: Daily real exchange rates of the German mark, British pound, Japanese yen and Canadian dollar vis-à-vis the US dollar.
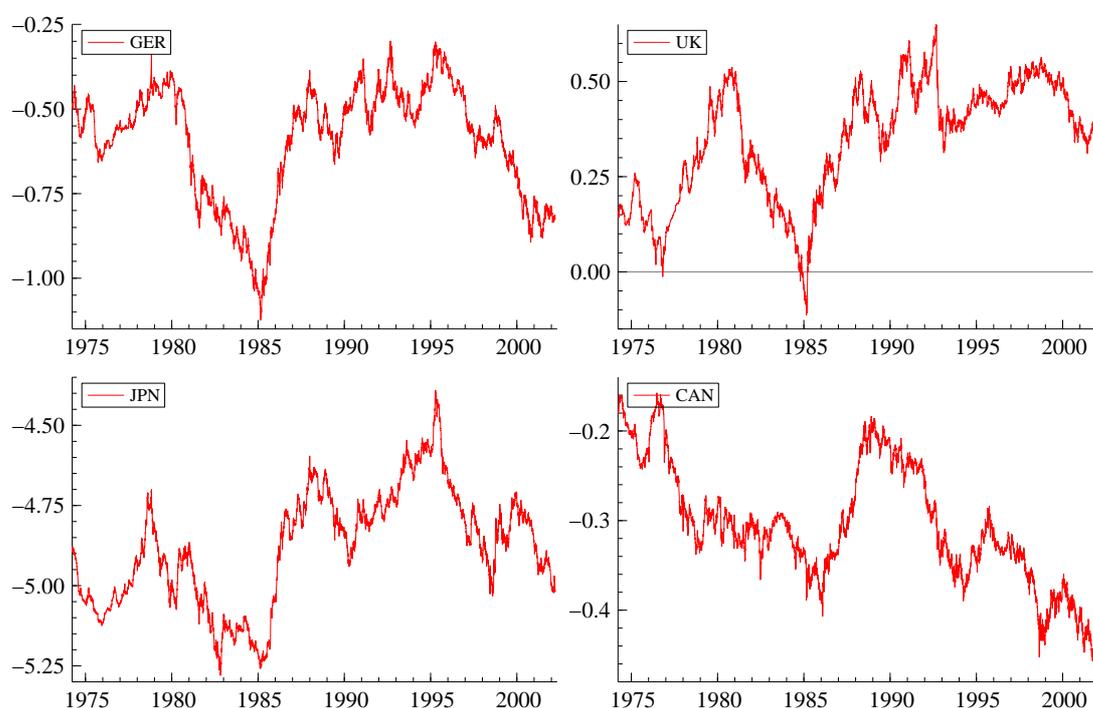

The log-real exchange rates are depicted in Figure 1. In all four series we observe very slow (if any) mean-reversion: real exchange rates may persistently deviate from their mean for more than a decade. This indicates that it is typically very hard to find evidence in favour of purchasing-power parity 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 2 depicts some stylized properties of daily exchange rate returns, in particular their estimated density, correlogram, and correlogram of squared returns. 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 density of the returns. The graphs do not directly indicate fat-tailedness, but the domain over which the densities are depicted indicates that extreme observations occur frequently. Note that for the Canadian dollar, the correlogram of squared returns seems to display

a different pattern; this might indicate that the GARCH(1,1) model is not adequate for this series, which will be investigated below.

Figure 2: Graphs, densities, correlograms and correlograms of squares for the returns on the real exchange rates depicted in Figure 1.
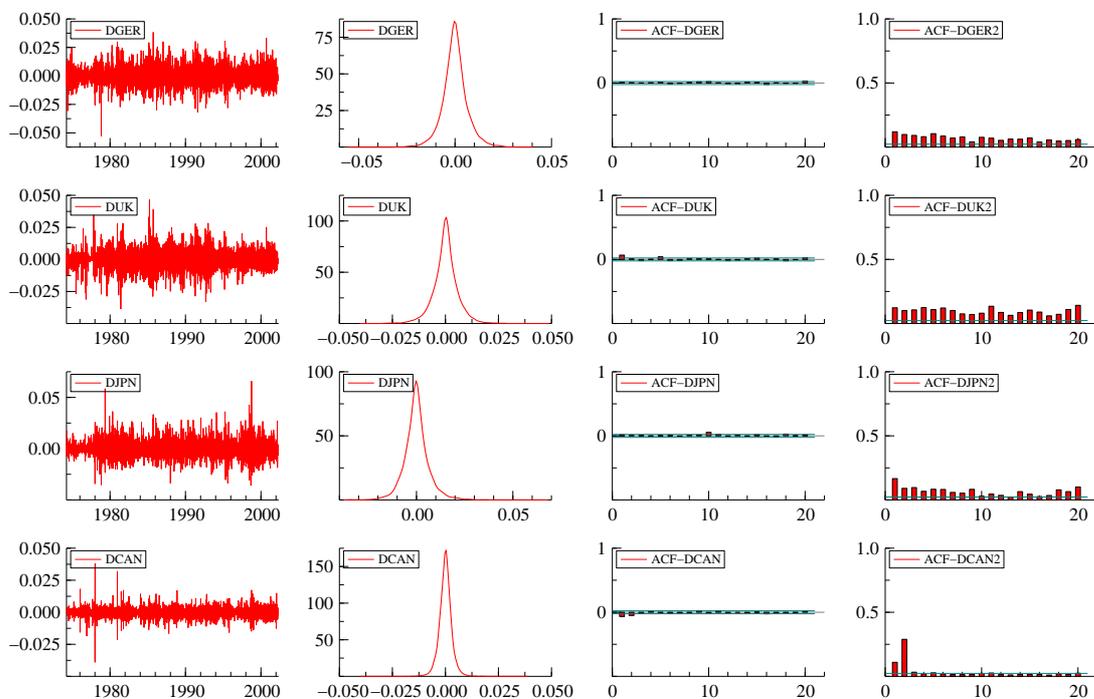
We consider first the German mark. Table 3 depicts the results for the four tests at the three different observation frequencies. Note that the table only gives the test statistics, not the p-values. However, the asymptotic theory indicates that the 5% critical value lies between 5.99 (the $\chi^2(2)$ critical value) and 9.17 (twice the asymptotic critical value for the Dickey-Fuller Φ_1 test, hence the appropriate asymptotic critical value for an LR version of this test), which suggests an easy bounds procedure. For ease of comparison, the Dickey-Fuller Φ_1 test statistic has been multiplied by two in the table.

Table 3: Unit root tests for the German mark real exchange rate.

frequency	$2\Phi_1$	QLR_{G-t}		QLR_G		QLR_t	
	test stat.	test stat.	$(\hat{\alpha} + \hat{\beta}, \hat{\nu})$	test stat.	$\hat{\alpha} + \hat{\beta}$	test stat.	$\hat{\nu}$
daily	2.351	4.772	$(1.000, 4.75)$	1.707	0.989	5.788	3.72
weekly	2.715	0.916	$(0.986, 8.47)$	1.668	0.981	2.258	5.98
monthly	2.979	1.105	$(0.955, 8.26)$	3.341	0.934	0.991	8.38

All results in this section have been obtained using the Garch module within PcGive 10.1, see Doornik and Hendry (2001). The GARCH models have been estimated under the restriction $\alpha + \beta \leq 1$, and in some cases the maximum of the likelihood function is found on the boundary, i.e., for $\hat{\alpha} + \hat{\beta} = 1$.

9

In all cases we have added a sufficient number of lagged differences, such that there is no residual autocorrelation, and also the GARCH(1,1) specification is not misspecified. The general impression from this table is that as we move from daily via weekly to monthly data, the Dickey-Fuller Φ_1 test is only moderately affected, whereas the GARCH-t- and t-based QLR test statistics clearly decrease, and hence convey less evidence of mean reversion. The Gaussian GARCH-based test does not fit into this picture, since the test statistic obtains a maximum value for monthly data. None of the tests are significant at the 5% level, but QLR_{G-t} and QLR_t do have a p-value less than 10% for the daily data, which indeed indicates that for this series, taking the fat-tailedness of the high-frequency data into account indeed helps to find more evidence in favour of purchasing-power parity, even though the evidence is still not overwhelming.

Table 4: Unit root tests for the British pound real exchange rate.

frequency	$2\Phi_1$	QLR_{G-t}		QLR_G		QLR_t	
	test stat.	test stat.	$(\hat{\alpha} + \hat{\beta}, \hat{\nu})$	test stat.	$\hat{\alpha} + \hat{\beta}$	test stat.	$\hat{\nu}$
daily	4.381	36.16	$(1.000, 4.92)$	9.179	0.999	4.162	3.19
weekly	3.993	4.132	$(0.985, 6.19)$	4.592	0.975	1.068	5.10
monthly	4.369	0.188	$(0.838, 9.53)$	7.909	0.843	0.940	7.43

The results for the British pound are given in Table 4. Here these results for the QLR_{G-t} are most remarkable: at the daily frequency there seems to be substantial evidence for mean reversion, which disappears as we move to weekly and monthly data. There is one caveat however, which is that the model with 5 lagged differences, which we use for the daily data, seems to display significant autocorrelation within the GARCH-t model. We have tried to remedy this using a bigger specification, but with no succes. A similar but less pronounced picture emerges from the QLR_G test. Using this test, the monthly data still seem to display significant mean reversion. In this particular case the test based on an i.i.d. $t(\nu)$ model does not lead to more evidence in favour of mean reversion.

Table 5: Unit root tests for the Japanese yen real exchange rate.

frequency	$2\Phi_1$	QLR_{G-t}		QLR_G		QLR_t	
	test stat.	test stat.	$(\hat{\alpha} + \hat{\beta}, \hat{\nu})$	test stat.	$\hat{\alpha} + \hat{\beta}$	test stat.	$\hat{\nu}$
daily	2.234	15.651	$(1.000, 4.14)$	3.531	0.997	14.35	3.08
weekly	2.963	7.283	$(1.000, 5.03)$	1.230	1.000	11.05	4.00
monthly	3.008	1.687	$(0.973, 5.60)$	2.504	0.912	2.521	6.66

The results for the Japanese yen, given in Table 5, again indicate that daily data provide more evidence for mean reversion, provided that a method is used that exploits the fat-tailedness. In this case taking the GARCH effect into account does not seem to yield much power gain. We see that even at the weekly frequency, there is still some evidence for purchasing-power parity based on QLR_{G-t} and QLR_t, but at the mothly frequency this evidence vanishes. As always, the Dickey-Fuller tests fail to detect any mean reversion at any observation frequency.

Table 6: Unit root tests for the Canadian dollar real exchange rate.

frequency	$2\Phi_1$	QLR_{G-t}		QLR_G		QLR_t	
	test stat.	test stat.	$(\hat{\alpha} + \hat{\beta}, \hat{\nu})$	test stat.	$\hat{\alpha} + \hat{\beta}$	test stat.	$\hat{\nu}$
daily	3.763	0.448	$(0.987, 5.35)$	5.767	0.982	0.219	4.07
weekly	2.963	3.892	$(0.935, 5.44)$	3.460	0.825	1.624	4.77
monthly	4.679	3.221	$(0.609, 5.10)$	5.623	0.50	2.650	5.19

Finally, for the Canadian dollar we observe in Table 6 that none of the tests give substantial evidence in favour of mean reversion, regardless of the obervation frequency. Note that although the correlogram of squared returns suggested otherwise, misspecification tests indicate that the GARCH(1,1) model does provide an adequate characterization of the volatility clustering. The parameter estimates clearly show that both the GARCH persistence and the fat-tailedness is the least pronounced for this series, which provides an explanation why the QLR tests do not help to detect mean reversion in this series.

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 test is incorrect for financial data, where high-frequency observations display properties that may be exploited for obtaining high-powered tests. 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 (1999), "Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors" , Tinbergen Institute Discussion Paper # 99-013/4, revised 2002 (http://www1.fee.uva.nl/ke/boswijk/gamma.pdf).

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.

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.

Klüppelberg, C., R. A. Maller, M. van de Vyver and D. Wee (2002), "Testing for Reduction to Random Walk in Autoregressive Conditional Heteroskedasticity Models", *Econometrics Journal*, 5, 387–416.

Ling, S. and W. K. Li (1997), "Estimating and Testing for Unit Root Processes with GARCH(1,1) Errors", Technical Report 140, Department of Statistics and Actuarial Science, The University of Hong Kong.

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.

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.

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).