Phillips-Perron-type unit root tests in the nonlinear ESTAR framework

By Christoph Rothe and Philipp Sibbertsen *

Summary: In this paper, we propose Phillips-Perron type, semi-parametric testing procedures to distinguish a unit root process from a mean-reverting exponential smooth transition autoregressive one. The limiting nonstandard distributions are derived under very general conditions and simulation evidence shows that the tests perform better than the standard Phillips-Perron or Dickey-Fuller tests in the region of the null.

Keywords: Exponential smooth transition autoregressive model, unit roots, Monte Carlo simulations, purchasing power parity. JEL C12, C32.

1. Introduction

The application of unit root tests, such as the Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP), has become standard in applied time series econometrics. In some areas of economics, however, their frequent inability to reject the null of nonstationarity contradicts well-established theoretical expectations.

One example is the purchasing power parity (PPP) hypothesis. In its absolute form, it states that a common basket of goods, when quoted in the same currency, should cost the same in all countries. Due to several factors like taxes, tariffs, trade restrictions and other market imperfections, the relationship may only be expected to hold in the long run. In empirical studies, however, the null hypothesis of a unit root in the deviations from parity can often not be ruled out using ordinary testing procedures.

As a possible explanation, Dumas (1992), Sercu et al. (1995) and Michael et al. (1997) consider models of real exchange rate determination which take transaction costs into account. Their theory suggests that the larger the deviation from PPP, the stronger the tendency to move back to equilibrium. The result is a nonlinear, mean-reverting stochastic process. They claim that standard unit root tests might lack power in such cases, and thus the evidence against PPP in the literature might be the result of a type II error.

It has been demonstrated by Michael et al. (1997), Sarno (2000), Taylor et al. (2001) and Baum et al. (2001), amongst others, that the exponential smooth transition autoregressive (ESTAR) model is an effective way to describe such processes. It is part of the general class of STAR models, which constitute a broad and flexible family of nonlinear time series models, that has recently been applied in various fields of economics. For an excellent survey, see van Dijk et al. (2002).

The problem is that these authors assume stationarity of deviations from parity a priori, without using formal inference. To overcome this problem,

Kapetanios *et al.* (2003) construct an ADF type test called t_{NL} in that framework, and confirm its superiority over its linear counterpart in some situations. Our paper extends their work by proposing two new test statistics using a semi-parametric approach along the lines of Phillips (1987) and Phillips and Perron (1988). We give analytical expressions for the asymptotic distributions, which are free of nuisance parameters. By simulation, we also study the performance in small samples and find some improvements over the t_{NL}.

The plan of the paper is as follows. Section 2 describes the theoretical framework and derives some important asymptotic results. In Section 3, the new tests are developed along with their limit distributions. Finite sample properties are assessed in Section 4, while a small empirical application is presented in Section 5. Finally, Section 6 concludes. A more technical theorem is given in the Appendix.

Some words on notation. Throughout the paper, $\|X\|_r = (E(|X|^r))^{1/r}$ denotes the L_r norm, W denotes a standard Wiener process on $[0, 1]$, and '⇒' denotes weak convergence.

2. The nonlinear ESTAR framework

Consider the case of a simple STAR(1) model given by

$$y_t = \phi_1 y_{t-1}(1 - G(y_{t-1}, \theta)) + \phi_2 y_{t-1} G(y_{t-1}, \theta) + v_t, \qquad (1)$$

where $G(y_{t-1}, \theta)$ is called the transition function, is continuous and bounded between 0 and 1. The easiest way to think of such a nonlinear model is as a weighted average of two AR(1) processes, with dynamically changing weights determined by the level of the process in the previous period. The error term $\{v_t\}$ is also assumed to obey some assumptions, which are discussed in more detail below. To begin with, we will assume that y_t is a mean zero stochastic process. Other cases will be discussed in Section 3.

While there are various transition functions being considered in the literature, our focus will be on the exponential one, which is of the form

$$G(y_{t-1}, \theta) = 1 - \exp(-\theta y_{t-1}^2), \qquad (2)$$

with $\theta \geq 0$. This function is symmetrically U-shaped around zero and tends to 1 as y_{t-1} gets large in absolute value, provided that θ is positive. In the case of $\theta = 0$, however, the function is zero everywhere. So effectively, θ determines the speed of the transition between the two regimes.

For our particular application, we follow Kapetanios *et al.* (2003) and impose $\phi_1 = 1$, and $-1 < \phi_2 < 1$. In other words, the two regimes of our STAR model are assumed to correspond to a unit root process and a stable AR(1), respectively.

Combining these restrictions with (1) and (2), we can write our model in the following, reparameterized form:

$$\Delta y_t = \gamma y_{t-1}(1 - \exp(-\theta y_{t-1}^2)) + v_t, \qquad (3)$$

where $\gamma = \phi_2 - 1$. When $-2 < \gamma < 0$, which we assume holds, and $\theta > 0$, the process is globally mean-reverting towards the long-run level zero, while it locally acts like a unit root process when being close to that same level. We will say that the time series has a partial unit root in this case. For $\theta = 0$, there is no mean-reverting behaviour and the process has a full unit root.

It seems likely that standard linear unit root tests such as the ADF or PP have difficulties to distinguish these two cases, i. e. they lack power against the nonlinear alternative described above. For the former test, this was shown by Kapetanios et al. (2003), while simulation evidence on the latter one is presented in Section 5 of this paper. We will therefore now describe an approach to test for a unit root that takes our specific alternative into account.

The hypothesis we want to test is

$$H_0 : \theta = 0 \quad vs. \quad H_1 : \theta > 0 \,. \tag{4}$$

Note that the parameter γ is not identified under the null. Loosely speaking, this means that nothing can be learned about γ from the data if H_0 is actually true. This highly affects standard inferential procedures (see Davies, 1987), and thus a direct testing procedure is not feasible. However, following Luukkonen et al. (1988), the problem can be overcome by using a first order Taylor series approximation of the exponential function around zero, leading to the auxiliary regression

$$\Delta y_t = \delta y_{t-1}^3 + u_t, \tag{5}$$

where $\delta = \theta\gamma$, $u_t = v_t + \gamma y_{t-1}R(y_{t-1}, \theta)$ and $R(y_{t-1}, \theta)$ is the remainder of the Taylor approximation. Here the identification problem is no longer present.

To get an idea of the goodness of the approximation, Figure 1 depicts the undisturbed relationship (i. e. without the error term) between consecutive observations implied by the respective regressions. As one can see, differences are only getting obvious when y_{t-1} becomes large in absolute value. Auxiliary regression (5) therefore captures the dynamics of the true process quite well.

In the regression (5), our hypotheses stated above are equivalent to

$$H_0 : \delta = 0 \quad vs. \quad H_1 : \delta < 0 \,. \tag{6}$$

The idea now is to base inference on estimates of δ and the respective t ratios. Note that although our regression is only an approximation, under the null hypothesis this does not influence the properties of the error term, since $R(y_{t-1}, \theta) \equiv 0$ and thus $u_t = v_t$ when H_0 is true. Consequently, it does not affect the asymptotic distribution theory under H_0.

Denote the ordinary least squares (OLS) estimate of δ by

$$\hat{\delta} = \frac{\sum_{t=1}^T y_{t-1}^3 \Delta y_t}{\sum_{t=1}^T y_{t-1}^6} \,. \tag{7}$$
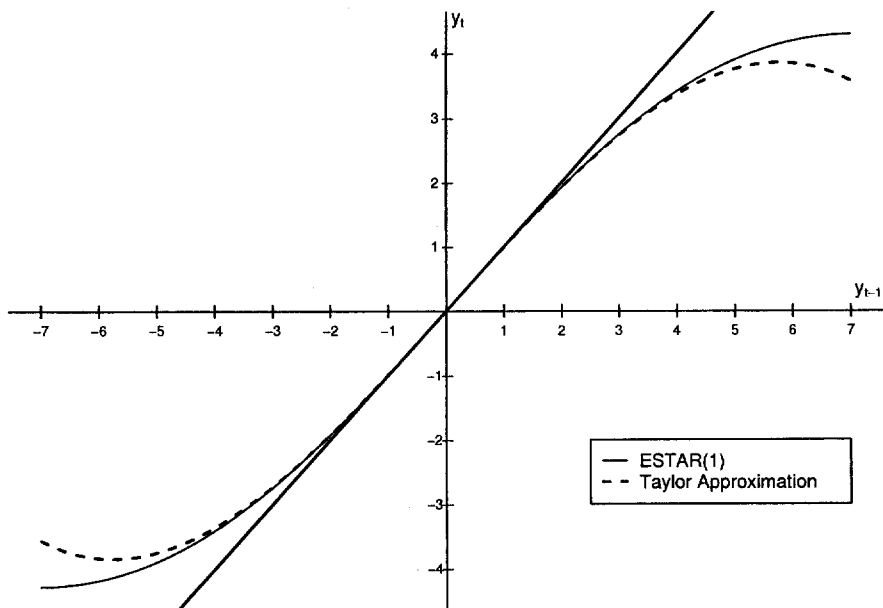
FIGURE 1. Undisturbed relationship between y_t and y_{t-1} for the ESTAR(1) model (3) and corresponding Taylor approximation (5), using $\gamma = -1$, $\theta = 0.01$. The bisecting line represents a unit root process.

Then the corresponding conventional t statistic is given by

$$
t_{\hat{\delta}} = \frac{\hat{\delta}}{\sqrt{\text{Var}(\hat{\delta})}} = \frac{\sum_{t=1}^{T} y_{t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} y_{t-1}^6}} , \tag{8}
$$

where

$$
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (\Delta y_t - \hat{\delta} y_{t-1}^3)^2 \tag{9}
$$

is the usual estimator of the error variance.

   To derive the asymptotics of $\hat{\delta}$ and $t_{\hat{\delta}}$, we have to be precise about the structure we are going to allow for the error term $\{v_t\}$. Kapetanios *et al.* (2003) only consider the t statistic (which they call t_{NL}) and show that for an i.i.d. sequence with finite second moments the limit distribution is given by

$$
t_{\hat{\delta}} \Rightarrow \frac{\frac{1}{4} W(1)^4 - \frac{3}{2} \int_0^1 W(r)^2 dr}{\sqrt{\int_0^1 W(r)^6 dr}} . \tag{10}
$$

They also demonstrate that the asymptotic distribution is the same in the case of an autoregressive process of order p, when the auxiliary regression (5) is extended with lagged first differences, as in Said and Dickey (1984), to

$$
\Delta y_t = \delta y_{t-1}^3 + \sum_{i=1}^{p} \rho_j \Delta y_{t-i} + \epsilon_t , \tag{11}
$$

where ϵ_t is i.i.d. $(0, \sigma^2)$ under the null. Given that p increases with the sample size in a suitable way, it can be shown that this also holds for general stationary and invertible ARMA processes.

However, it seems unintuitive to restrict the attention to linear error sequences in the general context of nonlinear time series. A flexible structure that allows for various forms of temporal dependence and also heteroscedasticity is given by the following assumption:

ASSUMPTION 1 (Hansen, 1992) *For some $p > \beta > 2$, $\{v_t\}$ is a zero mean, strong mixing sequence with mixing coefficients α_m of size $-p\beta/(p - \beta)$ and $\sup_{i \geq 1} \|v_i\|_p = C < \infty$. In addition, $(1/T)E\left(\left(\sum_1^T v_i\right)^2\right) \to \lambda^2 > 0$ as $T \to \infty$.*

The basic idea behind it is a trade-off between the existence of moments and the mixing decay rate, which prevents single observations from being too influential. For example, if extreme realizations occur relatively often (as will be the case when relatively few moments are finite), then the effect of such observations should wear off at a faster rate (requiring a larger size). Assumption 1 thus assures that the functional central limit theorem (FCLT) and results regarding the convergence to stochastic intergrals will apply to certain normalized partial sums of $\{v_t\}$.

The parameter λ^2 is the long-run variance. Note that while the second moments of $\{v_t\}$ are not assumed to be constant over time, a strong law of large numbers (McLeish, 1975) asserts that

$$\frac{1}{T} \sum_{i=1}^{T} v_i^2 \xrightarrow{a.s.} \sigma^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} E(v_i^2). \tag{12}$$

Hence there is a parameter σ^2 that can be interpreted as the *average* error variance. When $\{v_t\}$ is a sequence of i.i.d. variables, we obviously have $\lambda^2 = \sigma^2$.

Having defined the theoretical framework, we can now establish some convergence results that will be essential for the further asymptotic development.

LEMMA 1 *Let Assumption 1 hold with $\beta = 6$ and H_0 be true. Then, as $T \to \infty$, we have:*

$$(a) \quad T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \lambda^2 \int_0^1 W(r)^2 dr,$$

$$(b) \quad T^{-4} \sum_{t=1}^{T} y_{t-1}^6 \Rightarrow \lambda^6 \int_0^1 W(r)^6 dr,$$

$$(c) \quad T^{-2} \sum_{t=1}^{T} y_{t-1}^3 v_t \Rightarrow \frac{1}{4}\lambda^4 W(1)^4 - \frac{3}{2}\sigma^2\lambda^2 \int_0^1 W(r)^2 dr.$$

PROOF Both a) and b) are standard results that follow from the FCLT and the continuous mapping theorem (CMT), whereas c) is a result of Theorem 4 in the appendix, its subsequent remark, and the Itô formula.

Now we can start to discuss the properties of the quantities in (7)-(9). Our first concern is whether the estimates are consistent. This is ensured by the following theorem.

THEOREM 1 *Under the conditions of Lemma 1, we have*

$$a) \ \hat{\delta} \xrightarrow{p} 0 \ and \ \hat{\delta} = O_p(T^{-2}),$$

$$b) \ \hat{\sigma}^2 \xrightarrow{p} \sigma^2 \ and \ \hat{\sigma}^2 - \sigma^2 = O_p(T^{-1}).$$

PROOF Part (a) can be shown by a simple comparison of the convergence rates of the nominator and denominator in (7), which are $O_p(T^2)$ and $O_p(T^4)$, respectively. To prove (b), write

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} v_t^2 - \frac{2\hat{\delta}}{T} \sum_{t=1}^{T} y_{t-1}^3 v_t + \frac{\hat{\delta}^2}{T} \sum_{t=1}^{T} y_{t-1}^6 \tag{13}$$

and note that the first term converges to σ^2 in probability by (12), and the last two terms are both $O_p(T^{-1})$. Hence, the result follows.

Part (a) of the theorem implies that $\hat{\delta}$ converges to its true value zero under H_0 at rate T^2, which is even faster than the usual superconsistent rate T encountered in ordinary time series regressions with a unit root. Using this and the previous results, we can now derive the asymptotic distributions of the appropriately scaled version of $\hat{\delta}$ and of $t_{\hat{\delta}}$.

THEOREM 2 *Under the conditions of Lemma 1, we have*

$$a) \ T^2 \hat{\delta} \Rightarrow \frac{\frac{1}{4}\lambda^4 W(1)^4 - \frac{3}{2}\sigma^2\lambda^2 \int_0^1 W(r)^2 dr}{\lambda^6 \int_0^1 W(r)^6 dr},$$

$$b) \ t_{\hat{\delta}} \Rightarrow \frac{\frac{1}{4}\lambda^4 W(1)^4 - \frac{3}{2}\sigma^2\lambda^2 \int_0^1 W(r)^2 dr}{\sqrt{\sigma^2\lambda^6 \int_0^1 W(r)^6 dr}}.$$

PROOF Part (a) follows directly from Lemma 1 (b) and (c) and the continuous mapping theorem (CMT), whereas part (b) additionally requires Theorem 1 (b).

3. THE Z_{NL} TEST

Unfortunately, the limiting distributions derived in Theorem 2 depend on unknown nuisance parameters, and hence the respective test statistics cannot directly be used. We solve the problem by the semi-parametric approach proposed in Phillips (1987) and Phillips and Perron (1988). Without specifying an explicit model for $\{v_t\}$, we use consistent estimates of λ^2 and σ^2 to modify the test statistics in such a way that the parameters cancel each other asymptotically.

Let us assume for the moment that we have two such estimators named $\hat{\lambda}^2$ and $\hat{\sigma}^2$, without worrying about their exact functional form. These consistent estimates can now be used to develop two Phillips-Perron-type test statistics, whose limiting distributions do not depend on nuisance parameters.

Define

$$
Z_{NL}(\delta) = \hat{\lambda}^2 T^2 \left(\hat{\delta} - \frac{3}{2} \sum_{t=1}^{T} y_{t-1}^2 \left(\hat{\lambda}^2 - \hat{\sigma}^2 \right) \left(\sum_{t=1}^{T} y_{t-1}^6 \right)^{-1} \right) \tag{14}
$$

and

$$
Z_{NL}(t) = \frac{\hat{\sigma}}{\hat{\lambda}} t_{\hat{\delta}} - \frac{3}{2} \sum_{t=1}^{T} y_{t-1}^2 \left(\hat{\lambda}^2 - \hat{\sigma}^2 \right) \left(\hat{\lambda}^2 \sum_{t=1}^{T} y_{t-1}^6 \right)^{-1/2}, \tag{15}
$$

where the first term is a transformation of the OLS estimate $\hat{\delta}$ and the second term one of the corresponding t-statistic $t_{\hat{\delta}}$. The respective asymptotic distributions of the two quantities are given by the following theorem:

THEOREM 3 *Under the conditions of Lemma 1, we have*

$$
a)\ Z_{NL}(\delta) \Rightarrow \frac{\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr}{\int_0^1 W(r)^6 dr},
$$

$$
b)\ Z_{NL}(t) \Rightarrow \frac{\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(r)^2 dr}{\sqrt{\int_0^1 W(r)^6 dr}}.
$$

PROOF Both (a) and (b) are a direct consequence of Theorem 2, Lemma 1 and the CMT.

Since their limit distributions are free of nuisance parameters, these two statistics can now be used to test the unit root hypothesis. Figure 2 depicts the respective density functions under H_0. It turns out that the one of $Z_{NL}(\delta)$ is heavily left skewed, whereas the one of $Z_{NL}(t)$ is much more concentrated and relatively mildly skewed.
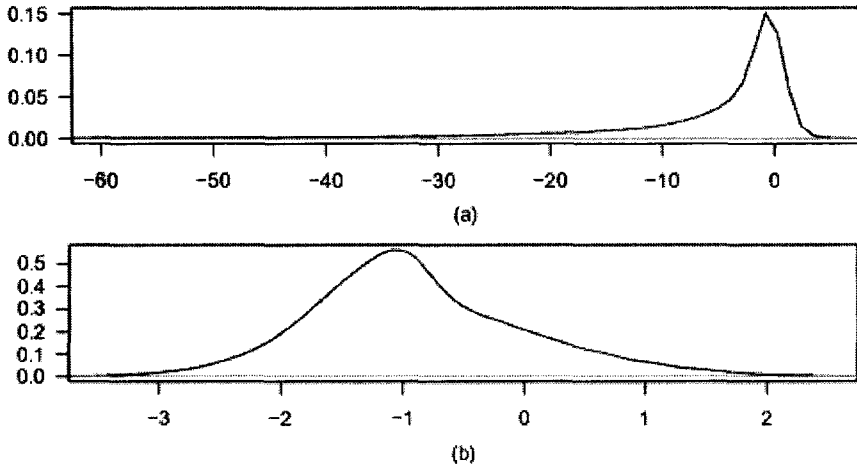
FIGURE 2. (a) Asymptotic probability density function of the $Z_{NL}(\delta)$ statistic.
(b) Asymptotic probability density function of the $Z_{NL}(t)$ statistic.

Up to now, we have confined our analysis to stochastic processes with mean zero, which certainly limits the usefulness of the results for empirical applications. To accommodate for nonzero means and/or deterministic linear trends, we use the same minor modifications as in Kapetanios et al. (2003). In case of a nonzero mean, i. e. when our model is $x_t = \mu + y_t$, we use the demeaned data, $\hat{y}_t = x_t - \bar{x}$, where $\bar{x}$ is the sample mean. The new limiting distributions are basically the same as in Theorem 3, but with $W(r)$ replaced by $W^*(r)$, a standard demeaned Wiener process on $[0, 1]$. Similarly, when we assume both a nonzero mean and a linear trend, i. e. when our model is $x_t = \mu + \beta t + y_t$, we will use the demeaned and detrended data, $\hat{y}_t = x_t - \hat{\mu} - \hat{\beta} t$, where $\hat{\mu}$ and $\hat{\beta}$ are the simple OLS estimates. In this case, $W(r)$ is replaced by $W^{\#}(r)$, a demeaned and detrended standard Wiener process defined on $[0, 1]$.

Asymptotic critical values for these three constellations, termed Case 1, Case 2 and Case 3, are given in Table 1. They were obtained by simulating the respective statistics with $T = 1\,000$ observations and $100\,000$ replications. Note that the right panel corresponds to the critical values for the t_{NL} test in Kapetanios et al. (2003) since their statistic and the $Z_{NL}(t)$ have the same limit distribution under the null.

Finally, a word on the estimation of the nuisance parameters. In the case of σ^2, the usual estimator $\hat{\sigma}^2$ has already shown to be consistent under the null hypothesis in Theorem 1 (b). There is also a large number of long-run variance estimators proposed in the literature (see den Haan and Levin, 1997, for a comprehensive review). While roughly speaking any consistent estimator will yield the same asymptotic results, the exact functional form may of course influence finite sample porperties. For the purpose of our simulation study in the following section, we choose the combination of a 'prewhitening' AR(1) filter and a kernel based estimator as proposed by

Fractile (%)	$Z_{NL}(\delta)$			$Z_{NL}(t)$		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
1	-130.069	-257.140	-503.641	-2.801	-3.471	-3.944
5	-51.611	-122.133	-279.296	-2.208	-2.929	-3.402
10	-29.359	-79.556	-200.274	-1.920	-2.653	-3.123

Cases refer to raw, demeaned, and demeaned and detrended data, respectively.

TABLE 1. Asymptotic critical values of the $Z_{NL}(\delta)$ and $Z_{NL}(t)$ statistics.

Andrews and Monahan (1992). More specifically, we estimate the regression $\hat{u}_t = \tilde{\alpha}\hat{u}_{t-1} + \tilde{e}_t$ by OLS and use the corresponding residuals to calculate $\hat{\lambda}_e^2 = \hat{\Gamma}(0) + 2\sum_{i=1}^{T-1} w\left(i/l_T\right)\hat{\Gamma}(i)$, where $\hat{\Gamma}(\tau) = T^{-1}\sum_{t=\tau+1}^{T}\tilde{e}_t\tilde{e}_{t-\tau}$, $w(\cdot)$ is the Bartlett kernel and the lag truncation parameter l_T is set at $\lfloor 4(T/100)^{2/9}\rfloor$. The final estimate of the long-run variance is then given by

$$\hat{\lambda}^2 = \frac{\hat{\lambda}_e^2}{(1-\hat{\alpha})^2}. \tag{16}$$

It should be noted that although this procedure performed well in informal simulation experiments, our particular choice is rather an ad hoc decision than based on theoretical considerations. The question whether other approaches might be more favourable in our context is beyond the scope of this paper.

4. SIMULATION RESULTS

In order to assess the properties of the new test statistics and to compare their performance with the procedures proposed by Kapetanios *et al.* (2003), Phillips and Perron (1988) and Said and Dickey (1984), we conduct a series of simulation experiments. For brevity, we only consider the first of the three cases described above, with the data being neither demeaned nor detrended. Although the power is lower for the last two cases, the performances of the tests relative to each other are similiar.

Our first concern is the size of the tests in finite samples, and thus we construct a null model with moving average errors by

$$y_t = y_{t-1} + v_t \quad \text{with} \quad v_t = e_t + \rho e_{t-1}, \tag{17}$$

where the e_t are i.i.d $N(0,1)$, and we consider $\rho \in \{0, \pm0.2, \pm0.5, \pm0.8\}$.

To evaluate the power of the two testing procedures against a globally mean-reverting ESTAR(1), we generate the data by

$$\Delta y_t = \gamma y_{t-1}(1 - \exp(-\theta y_{t-1}^2)) + v_t \quad \text{with} \quad v_t = e_t + \rho e_{t-1}, \tag{18}$$

and consider $\theta \in \{0.01, 0.05, 0.1, 1\}$ and $\gamma \in \{-1.5, -1, -0.5, -0.1\}$. We also used various specifications of the error process, but for brevity only the

case where $\rho = 0.5$ and the e_t are i.i.d $N(0,1)$ is reported here. Furthermore, we considered other types of error structures - including conditional heteroscedasticity, nonlinear dependence and heavy tails - but found no qualitative differences in the tests' performances.

For both size and power simulations, the nominal level of the tests is 5%, the sample size being considered is $T \in \{50, 100\}$, and the number of replications is set at $10\,000$.

In contrast to Kapetanios et al. (2003), we do not assume that the DGP is known when applying the tests to the data. In particular, we determine the number of augmentations used for the t_{NL} and ADF tests by minimizing the Schwarz Information Criterion SIC in the respective regressions. This should certainly give a more realistic impression of the properties of the tests than just using the 'true' order.

Table 2 shows the results on the size of the various tests. For $\rho \in \{0, 0.2, 0.5\}$ all sizes are generally close to the nominal level of 5%. We also observe the tendency to overreject the null somewhat when $\rho \in \{-0.2, 0.8\}$. For large negative values of ρ, however, size distortions are extreme for all of the six tests, with the ADF type test being a little less affected. This is a well known result for the usual unit root tests. The problem seems to persist in the nonlinear case, and remedies proposed e. g. by Perron and Ng (1996) do not easily generalize, so that none of the tests is recommendable when the error process almost has a negative moving average unit root.

Error Process	$Z_{NL}(\delta)$	$Z_{NL}(t)$	t_{NL}	$Z(\alpha)$	$Z(t)$	ADF
T=50						
$\rho = 0.8$	0.090	0.094	0.078	0.085	0.080	0.072
$\rho = 0.5$	0.069	0.069	0.076	0.067	0.063	0.068
$\rho = 0.2$	0.040	0.046	0.044	0.047	0.047	0.050
$\rho = 0$	0.036	0.052	0.052	0.049	0.048	0.050
$\rho = -0.2$	0.053	0.094	0.107	0.077	0.083	0.106
$\rho = -0.5$	0.145	0.278	0.206	0.256	0.281	0.230
$\rho = -0.8$	0.495	0.712	0.379	0.651	0.706	0.466
T=100						
$\rho = 0.8$	0.083	0.084	0.056	0.089	0.083	0.051
$\rho = 0.5$	0.064	0.063	0.068	0.066	0.060	0.064
$\rho = 0.2$	0.046	0.045	0.046	0.048	0.047	0.046
$\rho = 0$	0.042	0.051	0.048	0.052	0.051	0.050
$\rho = -0.2$	0.069	0.091	0.108	0.084	0.087	0.104
$\rho = -0.5$	0.201	0.280	0.134	0.237	0.252	0.152
$\rho = -0.8$	0.717	0.808	0.305	0.683	0.722	0.385

Note: Data under the null is generated by (19).

TABLE 2. The size of alternative test.

	θ = 0.01						θ = 0.05					
	$Z_{NL}(\delta)$	$Z_{NL}(t)$	t_{NL}	$Z(\alpha)$	$Z(t)$	ADF	$Z_{NL}(\delta)$	$Z_{NL}(t)$	t_{NL}	$Z(\alpha)$	$Z(t)$	ADF
$T = 50$												
$\gamma = -1.5$	0.624	0.790	0.724	0.547	0.584	0.581	0.971	0.985	0.967	0.945	0.952	0.943
$\gamma = -1.0$	0.472	0.663	0.625	0.423	0.464	0.484	0.902	0.954	0.907	0.857	0.875	0.839
$\gamma = -0.5$	0.272	0.416	0.422	0.252	0.294	0.309	0.683	0.793	0.731	0.636	0.668	0.651
$\gamma = -0.1$	0.078	0.139	0.155	0.086	0.117	0.123	0.166	0.263	0.276	0.179	0.224	0.242
$T = 100$												
$\gamma = -1.5$	0.996	0.995	0.976	0.974	0.976	0.937	1.000	1.000	1.000	1.000	1.000	0.999
$\gamma = -1.0$	0.977	0.976	0.940	0.926	0.931	0.875	1.000	1.000	0.997	0.999	0.999	0.995
$\gamma = -0.5$	0.826	0.879	0.817	0.721	0.740	0.715	0.997	0.993	0.972	0.991	0.992	0.961
$\gamma = -0.1$	0.194	0.307	0.321	0.191	0.220	0.237	0.519	0.574	0.539	0.511	0.544	0.512

	θ = 0.1						θ = 1.0					
	$Z_{NL}(\delta)$	$Z_{NL}(t)$	t_{NL}	$Z(\alpha)$	$Z(t)$	ADF	$Z_{NL}(\delta)$	$Z_{NL}(t)$	t_{NL}	$Z(\alpha)$	$Z(t)$	ADF
$T = 50$												
$\gamma = -1.5$	0.995	0.996	0.992	0.995	0.996	0.990	0.999	0.998	0.981	1.000	1.000	0.997
$\gamma = -1.0$	0.973	0.983	0.964	0.963	0.974	0.957	0.987	0.976	0.957	1.000	1.000	0.995
$\gamma = -0.5$	0.827	0.877	0.811	0.817	0.839	0.783	0.898	0.832	0.772	0.987	0.986	0.943
$\gamma = -0.1$	0.228	0.309	0.312	0.244	0.291	0.305	0.280	0.325	0.327	0.334	0.372	0.372
$T = 100$												
$\gamma = -1.5$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	1.000
$\gamma = -1.0$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.985	1.000	1.000	1.000
$\gamma = -0.5$	0.999	0.997	0.988	0.999	0.999	0.992	0.981	0.981	0.944	1.000	1.000	1.000
$\gamma = -0.1$	0.604	0.606	0.570	0.628	0.655	0.618	0.661	0.588	0.546	0.754	0.770	0.724

Note: Data under the alternative is generated by (20) with $\rho = 0.5$.

TABLE 1. The power of alternative test with MA(1) errors.

Results of the power simulations are presented in Table 3. The finding of Kapetanios *et al.* (2003) was that their t_{NL} test performed better than the ADF for small values of θ. In that case, the process is highly persistent. When θ grows, however, the average values of the transition function $G(y_{t-1}, \theta)$ get closer to one and the process more and more resembles a stable, linear AR(1). So eventually, the power of the ADF dominates.[1] Our simulations confirm these findings and show that the same holds for the two usual Phillips-Perron tests $Z(\alpha)$ and $Z(t)$. Their performance turns out to be marginally worse than the ADF's one, but they beat the t_{NL} provided that θ is sufficiently large. All of the six tests have in common that their power is close to one for small values of γ, corresponding to little persistence of the process.

Regarding our two new tests, we make the following observations. First, comparing $Z_{NL}(\delta)$ and $Z_{NL}(t)$, we note that the power of the latter dominates the one of the former, particularly when θ is small. In addition, $Z_{NL}(\delta)$ performs similar or worse than all its linear competitors, leading to the conclusion that this test should not be applied in practice. This comes somewhat unexpectedly, since unit root tests based directly on regression coefficients are usually found to be more powerful than the ones based on t-ratios (Phillips and Xiao, 1998).

Secondly, it emerges from Table 3 that $Z_{NL}(t)$ is somewhat more powerful than t_{NL} when the error sequence is an MA(1). While the differences are admittedly not huge, they certainly justify the use of our procedure in empirical applications.

5. EMPIRICAL APPLICATION: PPP

As an empirical application of the unit root test developed in the previous chapter, consider the case of real exchange rates, defined as

$$Q_t = S_t \cdot \frac{P_t^*}{P_t}, \tag{19}$$

where S_t is the nominal exchange rate and P_t^*, P_t are the foreign and domestic price levels at time t, respectively. It is common, however, to use the logarithmic form

$$q_t = s_t + p_t^* - p_t, \tag{20}$$

where all lower case variables are the log of the respective upper case counterpart. As outlined in the Introduction, the theory of purchasing power parity, or PPP, predicts that the time series q_t should be mean reverting, since short run deviations from parity should be arbitraged away by market forces in the long run. Note that this is equivalent to saying that s_t, p_t^* and p_t are cointegrated with known cointegration vector $(1, 1, -1)'$. Standard

---

[1] Note that θ is not a scale free parameter and that thus it is difficult to say which values are actually 'small' or 'large'.
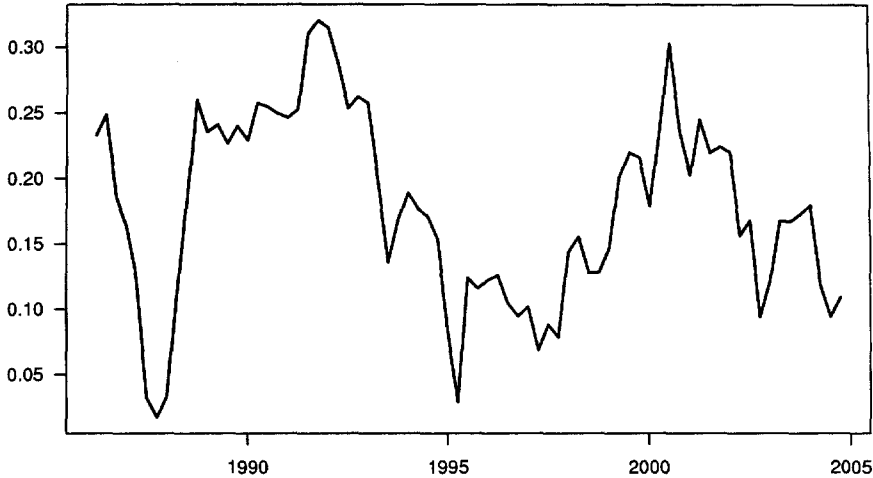
FIGURE 3. Quarterly (log) real bilateral exchange rate of the Australian and New Zealand Dollar.

unit root tests, however, are often not able to reject the null of a unit root in q_t, which has been attributed to possible nonlinearities in the process. We expect our $Z_{NL}(t)$ test (and the similarly constructed t_{NL}) to be more powerful than the usual procedures in this case.

For our illustration, we consider the case of Australia and New Zealand. Due to their geographical situation and common history as British colonies, the countries have strong economic ties. Furthermore, with the 1983 Closer Economic Relations Trade Agreement (known as CER), most trade restrictions have successively been removed. All this makes it more likely to find evidence in support of PPP.

Our dataset, obtained from the IMF's *International Financial Statistics Online* service, contains the nominal exchange rate of the Australian to the New Zealand Dollar, together with the consumer price indices (CPI) of the respective countries as proxies of the price level. Observations were sampled at quarterly intervals, starting in 1986, after which both currencies were floating freely, and running to 2004. The resulting real exchange rate is displayed in Figure 3. It can be observed that the two countries' series has roughly stayed constant over the inspected time period, although there have been substantial fluctuations.

We applied four unit root tests to each of the time series: the usual ADF and PP tests based on the t-statistic, the t_{NL} from Kapetanios *et al.* (2003), and our $Z_{NL}(t)$ test. Other procedures discussed earlier were omitted since the simulation results from the previous section suggest that their finite sample properties are at most similar, if not inferior, compared to the others. Since there is no theoretical justification for a linear trend in real exchange rates, we included only an intercept in the regression for the former two tests, and used demeaned data for the two latter ones. The

	ADF	$Z(t)$	t_{NL}	$Z_{NL}(t)$
Statistic	-2.336	-2.584	-3.362	-3.186
p-value	0.160	0.095	0.015	0.025

TABLE 4. Test statistics and corresponding p-values.

resulting test statistics and corresponding simulated p-values (computed by the same means as the critical values in Table 1) are presented in Table 4.

As one can see, the evidence in favour of PPP from the ADF and PP tests is not compelling. In contrast, the t_{NL} and $Z_{NL}(t)$ strongly reject the hypothesis of a unit root in the real exchange rate, in line with economic theory.

In addition to that, we estimated the ESTAR model from (3) for $y_t - \mu$ by nonlinear least squares (NLS), yielding

$$\Delta y_t = \underset{(0.102)}{-0.247}(y_{t-1}\underset{(0.017)}{-0.188})(1-\exp(\underset{(295.846)}{-207.617}(y_{t-1}\underset{(0.017)}{-0.188})^2)) + \hat{v}_t \,, \quad (21)$$

where the numbers in parentheses are the corresponding standard errors. A further analysis of the residuals shows no sign of serial dependence, which demonstrates that the model is well specified. It is also strongly favoured over a linear AR model with intercept by both the AIC and SIC.

As pointed out by van Dijk *et al.* (2002), the large standard error of the estimate of θ should not be interpreted as evidence against nonlinearity, since the asymptotic distribution is nonstandard when $\theta = 0$ due to the identification problem. In fact, the reason is that the residual sum of squares function is rather flat with respect to θ, and thus a precise estimate is hard to obtain.

6. CONCLUDING REMARKS

In various fields of economics, theory suggests processes to be stationary or at least mean-reverting. A prime example is the theory of purchasing power parity, which we use to motivate our study. The frequent inability of standard testing procedures to reject the null of a unit root in this cases has long caused discomfort amongst applied researchers. Recently, this has been attributed to a lack of power of those tests against nonlinear alternatives.

This paper contributes to the literature in this field by proposing two new test statistics in the framework of the exponential smooth transition autoregressive (ESTAR) model. We derive their non-standard limit distributions analytically and assess the finite sample properties through a simulation study. As a result, one of the tests turns out to have slightly favourable properties over a procedure suggested by Kapetanios *et al.* (2003), the only one available in this context so far. In a small empirical application to real exchange rates we demonstrate that the approach is at least a promising one to solve the puzzle of purchasing power parity.

As always, there is some room for improvements. First, the model considered under the alternative is still somewhat restrictive. One might want to allow for a more complex structure of the model under the alternative, involving a higher order of lagged variables, or a non-constant mean or time trend, changing smoothly with the regimes. Other transition functions, such as the logistic one, would be interesting to consider as well, as they allow for asymmetric adjustment.

Second, solving the identification problem under the null hypothesis by other means than linearization could be considered. The issue is discussed, *inter alia*, by Andrews and Ploberger (1994) and Hansen (1996). In a much more general context than the one presented in this paper, they suggest calculating the test statistics as a function of the unidentified nuisance parameters, and then use the supremum or average over all possible values to conduct inference. The respective limit distributions will typically be non-standard and critical values are again obtained via simulation. Theoretical considerations suggest that these procedures will yield a better test. However, our approach presented in this paper has the advantage of greater conceptual and computational simplicity.

REFERENCES

ANDREWS, D. W. K., MONAHAN, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60** 953–966.

ANDREWS, D. W. K., PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383–1414.

BAUM, C. F., BARKOULAS, J. T., CAGLAYAN, M. (2001). Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. *Journal of International Money and Finance* **20** 379–399.

DAVIES, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43.

DEN HAAN, W., LEVIN, A. (1997). A practitioner's guide to robust covariance matrix estimation. In *Handbook of Statistics 15* (G. Maddala and C. Rao, eds.), 299–342. Elsevier, Amsterdam.

DUMAS, B. (1992). Dynamic equilibrium and the real exchange rate in a spatially separated world. *Review of Financial Studies* **5** 153–180.

HANSEN, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8** 489–500.

HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430.

KAPETANIOS, G., SHIN, Y. AND SNELL, A. (2003). Testing for a unit root in the nonlinear star framework. *Journal of Econometrics* **112** 359–379.

LUUKKONEN, R., SAIKKONEN, P., TERÄSVIRTA, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* **75** 491–499.

MCLEISH, D. L. (1975). A maximal inequality and dependent strong laws. *Annals of Probability* **3** 829–839.

MICHAEL, P., NOBAY, A. R., PEEL, D. A. (1997). Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy* **105** 862–879.

PERRON, P. AND NG, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* **63** 435–463.

PHILLIPS, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301.

PHILLIPS, P. C. B., PERRON, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75** 335–346.

PHILLIPS, P. C. B., XIAO, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys* **12** 423–469.

SAID, S., DICKEY, D. (1984). Testing for unit roots in autoregressive moving average of unknown order. *Biometrika* **71** 559–607.

SARNO, L. (2000). Real exchange rate behavior in the middle east: A reexamination. *Economics Letters* **66** 127–136.

SERCU, P., UPPAL, R., VAN HULLE, C. (1995). The exchange rate in the presence of transaction costs: Implications for tests of purchasing power parity. *Journal of Finance* **50** 1309–1319.

TAYLOR, M. P., PEEL, D., SARNO, L. (2001). Nonlinear mean-reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles. *International Economic Review* **42** 1015–1042.

VAN DIJK, D., TERÄSVIRTA, T., FRANSES, P. H. (2002). Smooth transition autoregressive models - A survey of recent developments. *Econometric Reviews* **21** 1–47.

Christoph Rothe
Institute of Statistics
Department of Economics
University of Mannheim
L7, 3-5 D-68131 Mannheim
crothe@rumms.uni-mannheim.de

Philipp Sibbertsen
Institute of Statistics
Faculty of Economics
University of Hannover
Königsworther Platz 1
D-30167 Hannover
sibbertsen@statistik.uni-hannover.de

A. MATHEMATICAL APPENDIX

For the following convergence theorem, consider a time series of the form

$$y_i = y_{i-1} + v_i, \quad i = 1 \ldots n, \tag{22}$$

and define $V_t = \sum_{i=1}^{t} v_i$ and $V_n(s) = n^{-1/2} V_{\lfloor ns \rfloor}$ for $s \in [0, 1]$, where $\lfloor x \rfloor$ denotes the integer part of x.

THEOREM 4 *Let* $\Lambda = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=i+1}^{\infty} E(v_i v_j)$. *If Assumption 1 holds with* $\beta = 6$, *then*

$$\int_0^s V_n(r)^3 dV_n(r) \Rightarrow \int_0^s B(r)^3 dB(r) + 3\Lambda \int_0^s B^2(r) dr,$$

where $B = \lambda W$ is a Brownian motion with variance λ^2.

PROOF To begin with, recall that by the FCLT $V_n(\cdot) \Rightarrow B(\cdot)$ and hence $V_n^3(\cdot) \Rightarrow B^3(\cdot)$ by the CMT. The following derivation extensively uses Theorem 3.1 to 3.3 from Hansen (1992), which will be refered to as H1 to H3, respectively.

Let $\mathcal{F}_t = \sigma(v_i : i \leq t)$ be the smallest sigma-field containing the past history of $\{v_t\}$, and define

$$\epsilon_i = \sum_{k=0}^{\infty} \left(E(v_{i+k}|\mathcal{F}_i) - E(v_{i+k}|\mathcal{F}_{i-1}) \right),$$

$$z_i = \sum_{k=1}^{\infty} E(v_{i+k}|\mathcal{F}_i).$$

Then $\epsilon_i + z_{i-1} - z_i = E(v_i|\mathcal{F}_i) = v_i$ and it is easy to show that $\{\epsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence. With $Y_n(s) = n^{-1/2} \sum_{i=1}^{\lfloor ns \rfloor} \epsilon_i$, one directly obtains the equality

$$\int_0^s V_n^3(r)dV_n(r) = \int_0^s V_n^3(r)dY_n(r) + \Lambda_n^*(s),$$

where

$$\Lambda_n^*(s) = \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} (V_i^3 - V_{i-1}^3)z_i - \frac{1}{n^2} V_{\lfloor ns \rfloor}^3 z_{\lfloor ns \rfloor+1}.$$

Then, by H1, we obtain

$$\int_0^s V_n^3(r)dY_n(r) \Rightarrow \int_0^s B^3(r)dB(r).$$

Now consider the bias term

$$\Lambda_n^*(s) = \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} (V_i^3 - V_{i-1}^3)z_i - \frac{1}{n^2} V_{\lfloor ns \rfloor}^3 z_{\lfloor ns \rfloor+1}$$

$$= \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} (3V_{i-1}^2 v_i + 3V_{i-1}v_i^2 + v_i^3)z_i + o_p(1)$$

$$= \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1}^2 v_i z_i + \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1} v_i^2 z_i + \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} v_i^3 z_i + o_p(1).$$

First, by the triangle inequality and Hölder's inequality,

$$E \sup_{0 \leq s \leq 1} \left| \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} v_i^3 z_i \right| \leq E \sup_{0 \leq s \leq 1} \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \left| v_i^3 z_i \right|$$

$$\leq \frac{1}{n^2} \sum_{i=1}^{n} E \left| v_i^3 z_i \right|$$

$$\leq \frac{1}{n^2} \sum_{i=1}^{n} \left\| v_i^3 \right\|_{\frac{4}{3}} \left\| z_i \right\|_4$$

$$\leq \frac{1}{n^2} \sum_{i=1}^{n} \left\| v_i \right\|_4^3 \left\| z_i \right\|_4 \to 0$$

as $n \to \infty$, since by Assumption 1 and the proof of H1, respectively, both $\|v_i\|_4$ and $\|z_i\|_4$ are bounded.

Second, set $e_i = v_i z_i - \Lambda$, which by H2 satisfies the conditions of H3, yielding

$$\sup_{0 \le s \le 1} \left| \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1}^2 e_i \right| \xrightarrow{p} 0$$

and hence we have

$$\frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1}^2 v_i z_i = \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1}^2 \Lambda + \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1}^2 e_i$$

$$\Rightarrow 3\Lambda \int_0^s B^2(r) dr.$$

Finally,

$$E \sup_{0 \le s \le 1} \left| \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} V_{i-1} v_i^2 z_i \right| \le E \sup_{0 \le s \le 1} \frac{3}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \left| V_{i-1} v_i^2 z_i \right|$$

$$\le \frac{3}{n^2} \sum_{i=1}^{n} E \left| V_{i-1} v_i^2 z_i \right|$$

$$\le \frac{3}{n^2} \sum_{i=1}^{n} \|V_{i-1}\|_2 \|v_i^2 z_i\|_2 \to 0$$

as $n \to \infty$. To see this, note that by Hölder's inequality

$$\|v_i^2 z_i\|_2 = (E|v_i^4 z_i^2|)^{1/2} \le \|v_i\|_6^2 \|z_i\|_6$$

and thus is bounded. Furthermore,

$$\frac{3}{n^2} \sum_{i=1}^{n} \|V_{i-1}\|_2 = \frac{3}{n^2} \sum_{i=1}^{n} \left(E(V_{i-1}^2) \right)^{1/2}$$

$$= \frac{3}{n^{3/2}} \sum_{i=1}^{n} \left(\frac{1}{n} E(V_{i-1}^2) \right)^{1/2}$$

$$\le \frac{3}{n^{3/2}} \sum_{i=1}^{n} \left(\frac{1}{n} E(V_n^2) \right)^{1/2} \to 0$$

as $n \to \infty$, since $n^{-1} E(V_n^2)$ is $O(1)$ and thus bounded.

REMARK: Note that since $E(V_n) = 0$ for all n,

$$\lambda^2 = \lim_{n \to \infty} \text{Var}\left(\frac{1}{\sqrt{n}} V_n \right)$$

$$= \lim_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} E(v_i^2) + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(v_i v_j) \right)$$

$$= \sigma^2 + 2\Lambda.$$