# The power of unit root tests under local-to-finite variance errors

Nunzio Cappuccio[a], Diego Lubian[b,*], Mirko Mistrorigo[c]

[a] *Department of Economics and Management "Marco Fanno", via del Santo 33, 35123 Padova, Italy*
[b] *Department of Economics, University of Verona, via dell'Artigliere 19, 37129 Verona, Italy*
[c] *Poste Italiane S.p.A.*

**Abstract**

We study the power of four popular unit root tests in the presence of a local-to-finite variance DGP. We characterize the asymptotic distribution of these tests under a sequence of local alternatives, considering both stationary and explosive ones . We supplement the theoretical analysis with a small simulation study to assess the finite sample power of the tests. Our results suggest that the finite sample power is affected by the $\alpha$-stable component for low values of $\alpha$ and that, in the presence of this component, the DW test has the highest power under stationary alternatives. We also document a rather peculiar behavior of the $DW$ test whose power, under the explosive alternative, suddenly falls from 1 to zero for very small changes in the autoregressive parameter suggesting a discontinuity in the power function of the $DW$ test.

*Keywords:* Unit root test, infinite variance, power, finite sample.

☆We would like to thank an anonymous referee for helpful comments. The usual disclaimers apply. All views expressed in this paper are those of the authors only and do not necessarily reflect those of Poste Italiane S.p.A.

*Corresponding author

*Email addresses:* nunzio.cappuccio@unipd.it (Nunzio Cappuccio), diego.lubian@univr.it (Diego Lubian), mirko.mistrorigo@gmail.com (Mirko Mistrorigo)

## 1. Introduction

This paper is concerned with the power of unit root tests under local departures from the maintained hypothesis of finite variance of the error term. Following the approach proposed by Amsler and Schmidt (2012), and used by Cappuccio and Lubian (2007), we provide expressions for the asymptotic distributions of unit root tests under a sequence of local alternatives from the unit root null hypothesis under the assumption that the error term of a driftless random walk belongs to the normal domain of attraction of a stable law in any finite sample but has finite variance in the limit as $T \uparrow \infty$.

A setup of local departures from finite variance is interesting because it allows us to investigate the behavior of unit root tests in borderline or near borderline cases between finite and infinite variance. This kind of robustness analysis may indeed be relevant in practical settings where the existence of the variance is dubious such as, for example, in the analysis of financial time series. It is well-known that the empirical distribution of financial asset returns is often characterized by fat tails suggesting the relevance of non-gaussian stable laws. However, the empirical evidence in favor of the stable model is not clear-cut (McCulloch, 1997) so that the local-to-finite variance approach can be useful for improving our understanding of the robustness of unit root and stationarity tests in these circumstances.

Amsler and Schmidt (2012) first proposed this approach and derived the null distribution of the KPSS test, Callegari et al. (2003) obtained the asymptotic distributions of $DF$ type tests of unit root, and Cappuccio and Lubian (2007) obtained the null distribution of additional stationarity and nonstationarity tests. In this paper we provide additional results both on the asymptotic distributions and the finite samples properties of unit root tests under a sequence of local alternatives and local-to-finite variance error term.

In the next section we derive the asymptotic distribution of four popular test of the unit root hypothesis under a sequence of stationary and explosive alternatives when the data generating process displays local-to-finite variance errors. In Section 3 we carry out a simulation experiment to study the finite sample power functions of the tests.

## 2. Asymptotic distributions under a sequence of local alternatives

Our modeling of the local-to-finite variance process follows the approach proposed in Amsler and Schmidt (2012) whereby the process has infinite variance in finite samples but collapses to the standard finite variance case asymptotically. The Data Generating Process for the error term $u_t$ is then given by

$$u_t = v_{1t} + \frac{\gamma}{aT^{1/\alpha-1/2}} v_{2t}. \tag{1}$$

where $v_{1t}$ is an i.i.d. process with zero mean and finite variance $\sigma^2$ and $v_{2t}$ is also an i.i.d. process, symmetrically distributed with distribution belonging to the normal domain of attraction of a stable law with characteristic exponent $\alpha$, with $\alpha \in (0;\ 2)$, denoted as $v_{2t} \in \mathcal{ND}(\alpha)$, and $a$ can be set equal to 1 as in Amsler and Schmidt (2012). It follows that that $u_t$ exhibits infinite variance in any finite sample size but finite variance in the limit as $T$ approaches infinity. The role played by the stable component decreases as the sample size increases even though this occurs at a slower rate as $\alpha$ increases. Thus, for a given $\gamma$, when $\alpha$ is close to 2 we need a large sample size to offset the stable component whereas for $\alpha < 1$ a relatively small sample size is required. By Donsker's theorem, it is well known that $T^{-1/2}\sum_{t=1}^{[Tr]} v_{1t} \Rightarrow \sigma W(r)$, where $\Rightarrow$ stands for the weak convergence of probability measures, and $W(r)$ is the standard Wiener process. Further, (see, for instance, Resnick, 1986; Phillips, 1990), for the partial sum process $a_T^{-1}\sum_{t=1}^{[Tr]} v_{2t}$ we have the following converegence results

$$\left( \frac{1}{a_T}\sum_{t=1}^{[Tr]} v_{2t}, \frac{1}{a_T^2}\sum_{t=1}^{[Tr]} v_{2t}^2 \right) \Rightarrow \left(U_\alpha(r), V(r)\right), \tag{2}$$

where $a_T = aT^{1/\alpha}$, $U_\alpha(r)$ is a Lévy $\alpha$-stable process on the space $D[0,1]$, $V(r)$ is its quadratic variation process $V(r) = [U_\alpha, U_\alpha]_r = U_\alpha^2(r) - 2\int_0^r U_\alpha^- \mathrm{d}U_\alpha$ (see Protter, 1990, pg. 58, Phillips, 1990, eq. (11)) and $U_\alpha^-(r)$ stands for the left limit of the process $U_\alpha(\cdot)$ in $r$. The process $V(r)$ is a Lévy $\alpha/2$-stable process appears frequently in the asymptotic distribution of unit root tests. For $\alpha \in (0,1)$, it is non a degenerate random variable, while for $\alpha = 2$ we have $V(1) = 1$.

The main convergence result used in the paper is the following

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_t \Rightarrow \sigma_1 W(r) + \gamma U_\alpha(r) \equiv Z_{\alpha,\gamma}(r)$$

whose proof follows directly from the above joint convergence and the continuous mapping theorem. A number of useful results on the limiting behavior of sample moments and partial sums of the local-to-finite variance error term have been provided by Cappuccio and Lubian (2007, Lemma 2.1).

We consider the baseline case of a driftless random walk and assume that $\{y_t\}$ is generated as

$$y_t = \rho y_{t-1} + u_t, \qquad t = 1, \ldots, T \tag{3}$$

where $\rho = 1$ and that the initial condition $y_0$ is any random variable whose distribution does not depend on $T$.

We consider four test statistics for testing the null hypothesis $H_{DS} : \rho = 1$ in (3) such as $T(\hat{\rho} - 1)$ and the $t$-ratio statistics, where $\hat{\rho}$ is the OLS estimator of $\rho$ in 3, proposed by Dickey and Fuller (1976), the Lagrange Multiplier test (hereafter $LM$) proposed by Ahn (1993), and the Durbin-Watson ($DW$) test. Formally, the $t$-ratio statistics , the $LM$ and $DW$ tests are given by

$$t_{\hat{\rho}} = \left( \sum_{t=2}^{T} y_{t-1}^2 \right)^{1/2} (\hat{\rho} - 1)/s \tag{4}$$

$$LM = \frac{\left( \sum_{t=2}^{T} (y_t - y_{t-1}) y_{t-1} \right)^2}{s^2 \sum_{t=2}^{T} y_{t-1}^2} \tag{5}$$

$$DW = \frac{\sum_{t=2}^{T} (y_t - y_{t-1})^2}{\sum_{t=2}^{T} y_{t-1}^2} \tag{6}$$

where $s^2 = \sum_{t=2}^{T} (y_t - y_{t-1})^2 / T$. As remarked by Paulauskas et al. (2011), the weak convergence to stochastic integrals for sample moments of i.i.d. random vectors in the domain of attraction of a multivariate stable law with an index $0 < \alpha < 2$ has been proved by Paulauskas and Rachev (1998). The limiting behavior of the above test statistics under DGP (1) and the null hypothesis $H_0 : \rho = 1$ has been provided by Cappuccio and Lubian (2007) and is reported here for completeness

$$T(\hat{\rho} - 1) \Rightarrow \frac{\int_0^1 Z_{\alpha,\gamma} \mathrm{d} Z_{\alpha,\gamma}}{\int_0^1 Z_{\alpha,\gamma}^2}, \qquad t_{\hat{\rho}} \Rightarrow \frac{\int_0^1 Z_{\alpha,\gamma} \mathrm{d} Z_{\alpha,\gamma}}{\left( K_\gamma(1) \gamma^2 V(1) \int_0^1 Z_{\alpha,\gamma}^2 \right)^{1/2}}$$

$$LM \Rightarrow \frac{\left( \int_0^1 Z_{\alpha,\gamma} \mathrm{d} Z_{\alpha,\gamma} \right)^2}{K_\gamma(1) \int_0^1 Z_{\alpha,\gamma}^2}, \qquad TDW \Rightarrow \frac{K_\gamma(1)}{\int_0^1 Z_{\alpha,\gamma}^2}$$

where $K_\gamma(1) = \sigma^2 + \gamma^2 V(1)$. The null distribution of the four test statistics for DGP (3) in the infinite variance case has been established by Ahn et al. (2001) and for DGPs with a constant or a constant and a drift by Callegari et al. (2003). Even though the process $u_t$ has finite variance, the limiting distributions of the unit root test statistics turns out to be a function of both the Wiener process $W(r)$ and the Lévy $\alpha$-stable process $U_\alpha(r)$, so that they depend on the maximal moment exponent $\alpha$ and the nuisance parameters $\sigma^2$ and $\gamma$. As expected, the weight of $U_\alpha(r)$ in the asymptotic distribution increasesas with $\gamma$.

To study the power of these tests, we consider local departures from the null hypothesis assuming that the data generating process is given by

$$y_t = \rho_T y_{t-1} + u_t \tag{7}$$

where $\rho_T = e^{c/T}$ with $c < 0$, a noncentrality parameter, to focus on the stationary alternatives. When $c = 0$ we are under the null hypothesis, while for $c > 0$ the process is explosive. When $c$ is negative but close to zero, the process (7) is said to be near-integrated or to display a local-to-unity root. Since $e^{c/T} = 1 + O(T^{-1})$, as well as for the OLS estimator under the null hypothesis, the asymptotic distribution of unit root tests will depend on $c$. The behavior induced by the parameter $c$ can be helpful to understand the effect on the asymptotic distribution of departures from the null hypothesis under a sequence of local to unity alternatives.

Under DGP (7), using standard results we have

$$T^{-1/2} y_{[Tr]} \Rightarrow \sigma J_c(r) + \gamma G_{c,\alpha}(r)$$

where $J_c(r) = \int_0^r e^{(r-s)c} \mathrm{d}W(s) = W(r) + c \int_0^r e^{(r-s)c} W(s)\mathrm{d}s$, and $G_{c,\alpha}(r) = \int_0^r e^{(r-s)c} \mathrm{d}U_\alpha(s) = U_\alpha(r) + c \int_0^r e^{(r-s)c} U_\alpha^-(s)\mathrm{d}s$. Applying the continuous mapping theorem we obtain immediately the following convergence for sample moments of the process (7). These are summarized in the following proposition.

Proposition 1. *Under DGP (7) and the local-finite variance assumption (1), as $T \uparrow \infty$, we have*

$$T^{-3/2} \sum_{y=1}^{T} y_j \Rightarrow \int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)] \mathrm{d}r \tag{8}$$

$$T^{-2} \sum_{y=1}^{T} y_j^2 \Rightarrow \int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r \tag{9}$$

$$T^{-1} \sum_{y=1}^{T} y_{j-1} u_j \Rightarrow \frac{1}{2}(\sigma J_c(1) + \gamma G_{c,\alpha}(1)) - c \int_0^1 ]\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r - \frac{1}{2}(\sigma^2 + \gamma^2 V(1)) \tag{10}$$

Given these results we can move to analyze the limiting distribution of the unit root tests under the sequence of local alternatives. We discuss each test at a time.

*Dickey-Fuller test $T(\hat{\rho} - 1)$*

Since $\rho_T \simeq 1 + c/T$, the limiting distribution of $T(\hat{\rho} - 1)$ is given by

$$T(\hat{\rho} - 1) \Rightarrow c + \frac{\frac{1}{2}[\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 - c \int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r - \frac{1}{2}(\sigma^2 + \gamma^2 V(1))}{\int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r}$$

In the standard case, which represents the reference settings when studying the behavior of the test, we have the $\gamma = 0$. The distribution under the sequence of local alternatives becomes

$$c + \frac{\int_0^1 J_c(r) \mathrm{d}W(r)}{\int_0^1 J_c^2(r) \mathrm{d}r}$$

which, using Lemma 2 in Phillips (1987) can be used to obtain the rejection regions of the null hypothesis $H_0 : c = 0$. Under the stationary alternative hypothesis $H_1 : c < 0$, we have that

$$c + \frac{(-2c)^{1/2} \overbrace{\left[\int_0^1 J_c(r)\mathrm{d}W(r)\right]}^{\Rightarrow N(0,1)}}{\underbrace{(-2c)\int_0^1 J_c^2(r)\mathrm{d}r}_{\to -1}}$$

so that, as $c \to -\infty$, the distribution is pushed to the left and its variance increases linearly with $c$ at the rate $\sqrt{2c}$. It follows that the critical region will be in the left tail of the distibution under the null. On the contrary, under the explosive alternative hypothesis $H_1 : c > 0$, using again Lemma 2 in Phillips (1987), we have

$$c + \frac{(2c)e^{-c} \overbrace{\left[(2c)e^{-c}\int_0^1 J_c(r)\mathrm{d}W(r)\right]}^{\Rightarrow \xi\eta}}{\underbrace{(2c)^2 e^{-2c}\int_0^1 J_c^2(r)\mathrm{d}r}_{\Rightarrow \eta}}$$

where $\xi$ and $\eta$ are independent $N(0, 1)$. In this case, as $c \to \infty$, the distribution moves to the right and, eventually, collapses on $c$ at the rate $e^{-c}$. Therefore, the rejection region will lie in the right tail of the null distribution.

*Dickey-Fuller* $t_{\hat{\rho}}$

From the limiting distribution of $T(\hat{\rho}_T - \rho_T)$ we have that $(\hat{\rho}_T - \rho_T) = O_p(T^{-1})$ and, therefore, $\hat{\rho}_T = \rho_T + O_p(T^{-1})$. Recalling that $\rho_T = e^{c/T}$, it follows that $\hat{\rho}_T \to 1$ a.s. as $T \uparrow \infty$. Since $s^2 \Rightarrow \sigma^2 + \gamma^2 V(1)$, we have

$$t_{\hat{\rho}} \Rightarrow c + \frac{\frac{1}{2}[\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 - -\frac{1}{2}(\sigma^2 + \gamma^2 V(1))}{\left((\sigma^2 + \gamma^2 V(1)) \int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r\right)^{1/2}}$$

Some algebraic rearrangements allow us to write the distribution under the finite variance case $\gamma = 0$ as

$$t_{\hat{\rho}} \Rightarrow c\left(\int_0^2 J_c(r)^2 \mathrm{d}r\right)^{1/2} + \frac{\int_0^2 J_c(r)\mathrm{d}W(r)}{\left(\int_0^2 J_c(r)^2 \mathrm{d}r\right)^{1/2}}$$

The second term of this expression is identical to the second term in the previous Dickey-Fuller test, while the first term differs for a multiplicative term. However, notice that

$$c\left(\int_0^2 J_c(r)^2 \mathrm{d}r\right)^{1/2} = c\frac{\overbrace{-2c\left(\int_0^2 J_c(r)^2 \mathrm{d}r\right)^{1/2}}^{\to 1}}{(-2c)^{1/2}}$$

so that, as $c \to -\infty$, the distribution is shifted to the left and the rejection region is given by the left tail of the null distribution. A similar reasoning applies to the sequence of explosive alternatives, the only difference with the first Dickey-Fuller test is that, given that in the first term $c$ is multiplied by a random variable, the distribution will not collapse to a single point but it will diverge to $\infty$ as $c$ increases. Hence, the rejection region is given by the right tail of the null distribution.

*LM test*

From (1) and the fact that $s^2 \Rightarrow \sigma^2 + \gamma^2 V(1)$, the distribution of the $LM$ test (6) is given by

$$LM \Rightarrow \frac{\left(\frac{1}{2}[\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 - \frac{1}{2}(\sigma^2 + \gamma^2 V(1))\right)^2}{(\sigma^2 + \gamma^2 V(1)) \int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r}$$

The critical region for this test statistics is easily found. Since the numerator is always positive, the distribution shifts to the right as $c$ diverges to $\pm\infty$ so that the rejection region always lies in the right tail. Formally, let us consider the case $\gamma = 0$:

$$\frac{\left(\frac{1}{2}\sigma J_c(1)^2 - \frac{1}{2}\sigma^2\right)}{\sigma^2 \int_0^1 \sigma J_c(r)^2 \mathrm{d}r}$$

The denominator converges to 0 as $c \to -\infty$ and diverges as $c \to \infty$. The numerator collapses to 1/4 as $c \to -\infty$,

$$\begin{aligned}
\left(\frac{1}{2}J_c(1)^2 - \frac{1}{2}\right)^2 &= \left[c\int_0^1 J_c(r)^2 \mathrm{d}r + \int_0^2 J_c(r) \mathrm{d}W(r)\right]^2 \\
&= \left[-\frac{1}{2}(-2c)\int_0^1 J_c(r)^2 \mathrm{d}r + \frac{1}{(-2c)^{1/2}}(-2c)^{1/2}\int_0^2 J_c(r)\mathrm{d}W(r)\right]^2 \\
&\to \left(-\frac{1}{2}\right)^2
\end{aligned}$$

and it diverges to $\infty$ for $c \to \infty$

$$\left[c\int_0^1 J_c(r)^2 \mathrm{d}r \int_0^2 J_c(r)\mathrm{d}W(r)\right]^2 =$$
$$\left[\frac{1}{c(2e^{-c})^2}(2ce^{-c})^2\int_0^1 J_c(r)^2 \mathrm{d}r + \frac{1}{(2ce^{-c})^{1/2}}(-2ce^{-c})\int_0^2 J_c(r)\mathrm{d}W(r)\right]^2$$

so that, in both cases, the distribution of the LM test shifts to the right.

*DW test*

The distribution of the $DW$ test is easy to deal with and its asymptotic distribution is given by

$$TDW \Rightarrow \frac{\sigma^2 + \gamma^2 V(1)}{\int_0^1 [\sigma J_c(r) + \gamma G_{c,\alpha}(r)]^2 \mathrm{d}r}$$

Whe $\gamma = 0$ we have that

$$TDW \Rightarrow \frac{\sigma^2}{\int_0^1 \sigma J_c(r)^2 \mathrm{d}r}$$

The numerator is a bounded and strictly positive quantity, while the denominator converges to 0 for $c \to -\infty$ and diverges for $c \to \infty$. The rejection region lies in the right tail of the null distribution under the stationary alternative and, on the contrary, it lies in the left tail under the explosive alternative. With regard to this latter case, we recall that the distribution under the null is positive and that, therefore, coherently the test goes to zero under the exposive alternative. This result is in line with the usual approximation of the DW test given by $DW \approx 2(1 - \hat{\rho})$ which implies that the test is approximately equal to $-2$ times the first DF test. Then, the behavior of this test is clearly the opposite of the behavior of the DF tests.

## 3. Finite sample properties

### *3.1. Montecarlo experiment*

The asymptotic null distributions depend on the unknown nuisance parameters $\alpha$ and $\gamma$ and, when analyzing the behavior under the null and under the alternative hypoteses,one should consider appropriate critical values for each combination of these two parameters. In practice, this approach seems to be difficult to follow in empirical settings where both $\alpha$ and $\gamma$ are unknown. Here, we follow a different approach by considering the critical values under the finite variance case $\gamma = 0$ and considering the local-to-finite variance component as a perturbation to the standard maintained null hypothesis of finite variance. Of course, we depart somehow from a rigorous theoretical analysis but we believe that our approach will be closer to the context of actual use of the test.

The power of the test statistic against stationary fixed alternatives, $T = 200$ and for $\alpha = 1$ and $\alpha = 1.5$ has been investigated by Ahn et al. (2001) in a simulation study, showing the consistency of the tests. In the MonteCarlo experiment we set the variance of $v_{1t}$ equal to 1, that is, $\sigma^2 = 1$. We consider 3 different sample sizes $T = 100, 1000, 10000$ and three different values for $\alpha = 0.5, 1, 1.5$. As in Cappuccio and Lubian (2007) and Amsler and Schmidt (2012) we consider the following grid for $\gamma$:

$$\gamma = \{0.1, 0.316, 1, 3.16, 10, 31.6\}$$

Table 1: Critical values of the test statistics at 5% and 95% for some values of $\gamma$

| | | 5% | | | 95% | | |
|---|---|---|---|---|---|---|---|
| $\gamma = 0$ | $T(\hat{\rho}-1)$ | $-8.04$ | | | 1.255 | | |
| | $t_{\hat{\rho}}$ | $-1.943$ | | | 1.256 | | |
| | $DW$ | 0.602 | | | 17.79 | | |
| | $LM$ | $-$ | | | 4.123 | | |
| | | $\alpha = 0.5$ | $\alpha = 1$ | $\alpha = 1.5$ | $\alpha = 0.5$ | $\alpha = 1$ | $\alpha = 1.5$ |
| $\gamma = 0.316$ | $T(\hat{\rho}-1)$ | $-7.039$ | $-7.429$ | $-7.608$ | 1.190 | 1.295 | 1.295 |
| | $t_{\hat{\rho}}$ | $-1.798$ | $-1.875$ | $-1.893$ | 1.130 | 1.235 | 1.283 |
| | $DW$ | 0.679 | 0.373 | 0.351 | 17.33 | 16.96 | 16.98 |
| | $LM$ | | | | 3.588 | 3.894 | 3.987 |
| $\gamma = 31.6$ | $T(\hat{\rho}-1)$ | $-4.663$ | $-6.340$ | $-7.261$ | 1.136 | 1.286 | 1.278 |
| | $t_{\hat{\rho}}$ | $-1.391$ | $-1.674$ | $-1.845$ | 0.820 | 1.034 | 1.158 |
| | $DW$ | 0.837 | 0.462 | 0.411 | 16.98 | 16.48 | 16.95 |
| | $LM$ | | | | 2.132 | 3.125 | 3.691 |

where $3.16 \approx \sqrt{10}$. As for $c$, we consider the following values

$$c = \{-18, -16, -14, -13, -12, -11, -10, -9, -8, -7, -6, -5, -4, \\ -3, -2, -1, -0.5, 0, 0.51, 2, 3, 4, 6, 8, 10\}$$

Overall, we have 3 sample sizes, 3 values for $\alpha$, 6 values for $\gamma$, and 26 different values for $c$, for a total of 1404 combinations of parameters/sample size. In addition, we consider the benchmark case $\gamma = 0$ which is used as a reference in the empirical research. For each combination of parameters and sample size we simulate 20000 realizations of the four test statistics. In practice, we proceed as follows: for each sample size $T$ we simulate the first-order autoregressive process with parameter $\rho_T = e^{c/T}$ and error term as in (1) for specific values of $\alpha$ and $\gamma$, and $\sigma = 1$. Then, we calculate the four test statistics. We repeat this procedure 20000 times. In addition, to approximate the limiting distribution under the null we compute 100000 simulation of the DGP under the null and then we tabulate the critical values of the test statistics. These critical values are reported in Table 1 for selectd values of $\gamma$.

We notice that the 5% critical values of the test statistics $T(\hat{\rho}-1)$, $t_{\hat{\rho}}$ and $DW$ move rightwards both as $\gamma$ increases holding $\alpha$ fixed and as $\alpha$ decreases holding $\gamma$ constant. Conversely, the 95% critical values of the $LM$ test move to the left. In all cases, this suggests that the researcher will reject the null too few times if she uses the standard critical values for $\alpha = 0$ when the $\alpha$-stable component is indeed present.

We begin by studying how the power of the test statistics changes with $c$. Using the critical values in Table 1 we compute the power of the tests by simply counting how many times the different test statistics lead to a rejection of the null hypothesis in our replications. We shall compute not only the non-adjusted power power, using the critical values under $\gamma = 0$, but also the adjusted power, computed using the true values of $\gamma$. Thus, we obtain a nonparametric estimate of the rejection probability under the null hypothesis, i.e., the power of the test. In Figure 1 and 3 we plot the size-unadjusted power curves of the four test statistics for two sample sizes, $T = 100$ (left column) and $T = 1000$ (right column), and three values of $\alpha$, namely, 0.5 (top), 1 (middle), and 1.5 (bottom) for $\gamma = 0.316$ and $\gamma = 31.6$, respectively.

As expected, for $c < 0$, i.e., for stationary alternatives, all tests are consistent and power increases and approaches unity as $c$ decreases. The power curves of the four test statistics are close to each other with the $DW$ test being more powerful as $c$ starts decreasing from zero and then being beated by the $t$-test as $c$ decreases. For $T = 100$, the $t$-ratio is slightly more powerful than the other tests and, on the other hand, the $LM$ test is less powerful, with noticeable differences in power for $T = 1000$ too. As $\gamma$ increases, i.e., as the $\alpha$-stable component becomes more important in finite samples, power tend to decrease. This occurs not only for $T = 100$ but also for $T = 1000$, a substantially large sample size. This phenomenon can be appreciated by comparing the behavior of the power functions reported in the top-left panels of Figure 1 and 3. When $\gamma = 31.6$, the power is much lower than when $\gamma = 0.316$ and it stay very low whenever $c$ is between 7 and 8, then it raises very sharply for large values of $\gamma$.

In Figures 2 and 4 we provide a zoom of the power function over the interesting interval $(-4, 4)$ for the parameter $c$ which corresponds to the interval $(0.96, 1.04)$ for the autoregressive parameter $\rho$. This allows us to examine the behavior of the test statistics in the vicinity of the null hypohesis and for values of the autoregressive parameter under the alternative which are likely to be relevant in empirical research. As expected from the theoretical analysis, even though power approaches unity as $c$ moves away from 0 in all cases and for all test statistics, quite a different empirical power is observed for negative $c$, stationary alternatives, and for positive $c$, explosive alternatives. In particular, power is quite low for stationary alternatives, ranging from about 20% when $\alpha = 1.5$ to around 10% when $\alpha = 0.5$. This result is in line with the well known fact that unit root tests have low power when the alternative hypothesis is very close to the null. Further, these Figures allows us to fully appreciate the highpower of the $DW$ test for small deviations from the null hypothesis in the stationary direction.

Under explosive alternatives and for all tests, apart from the $DW$ test for $T = 100$, power goes to unity quite fast (the $LM$ and the $DW$ are the slowest though) and this occurs for all values taken by $\gamma$. The test statistic $T(\hat{\rho} - 1)$ dominates the others and the $DW$ has the worst overall performance. In addition, the power of the $DW$ test exhibits a behavior in line with the theoretical

Figure 1: Non-adjusted power for $\gamma = 0.316$

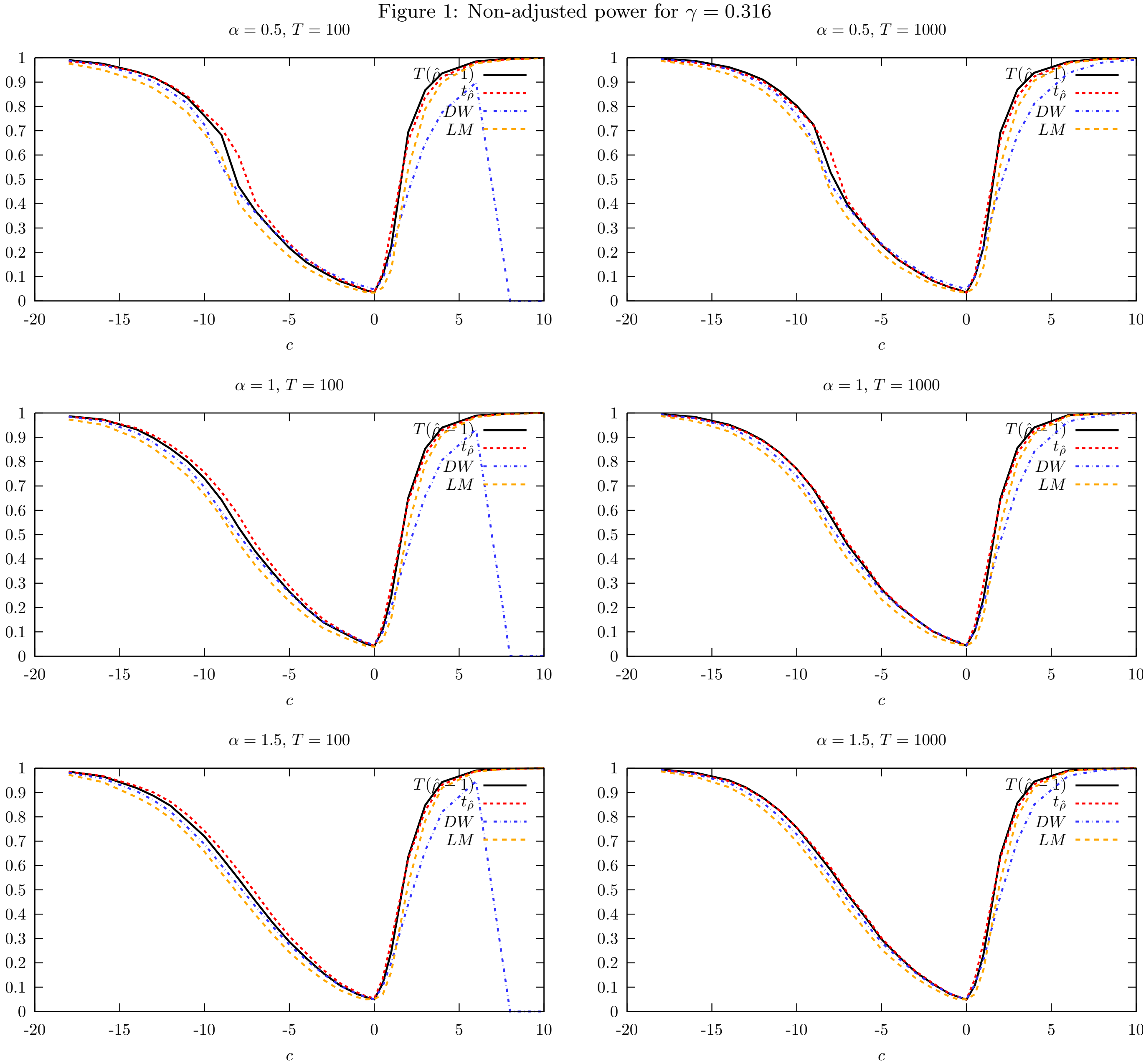


Figure 2: Non-adjusted power for $\gamma = 0.316$ (zoom)

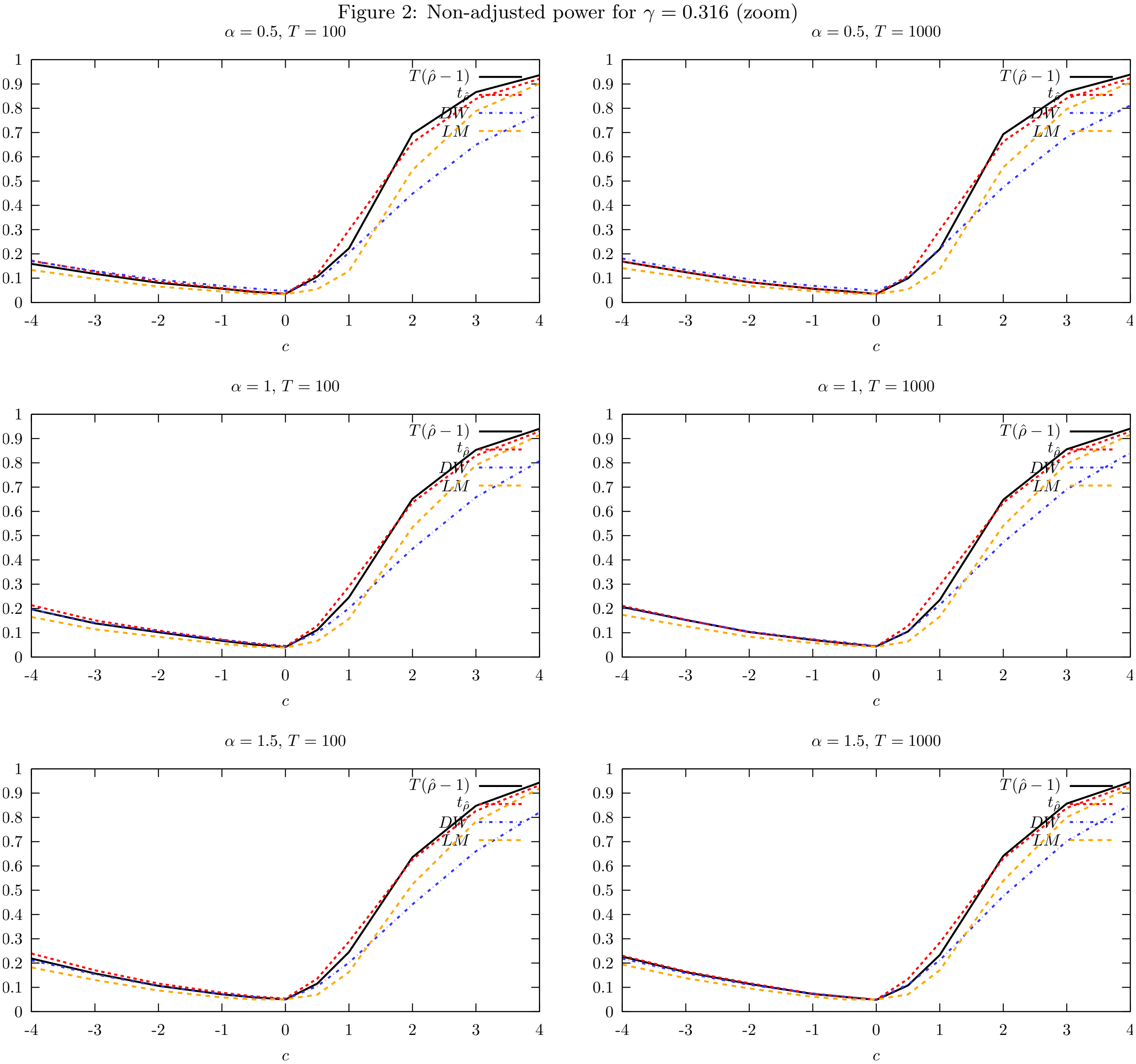


Figure 3: Non-adjusted power for $\gamma = 31.6$

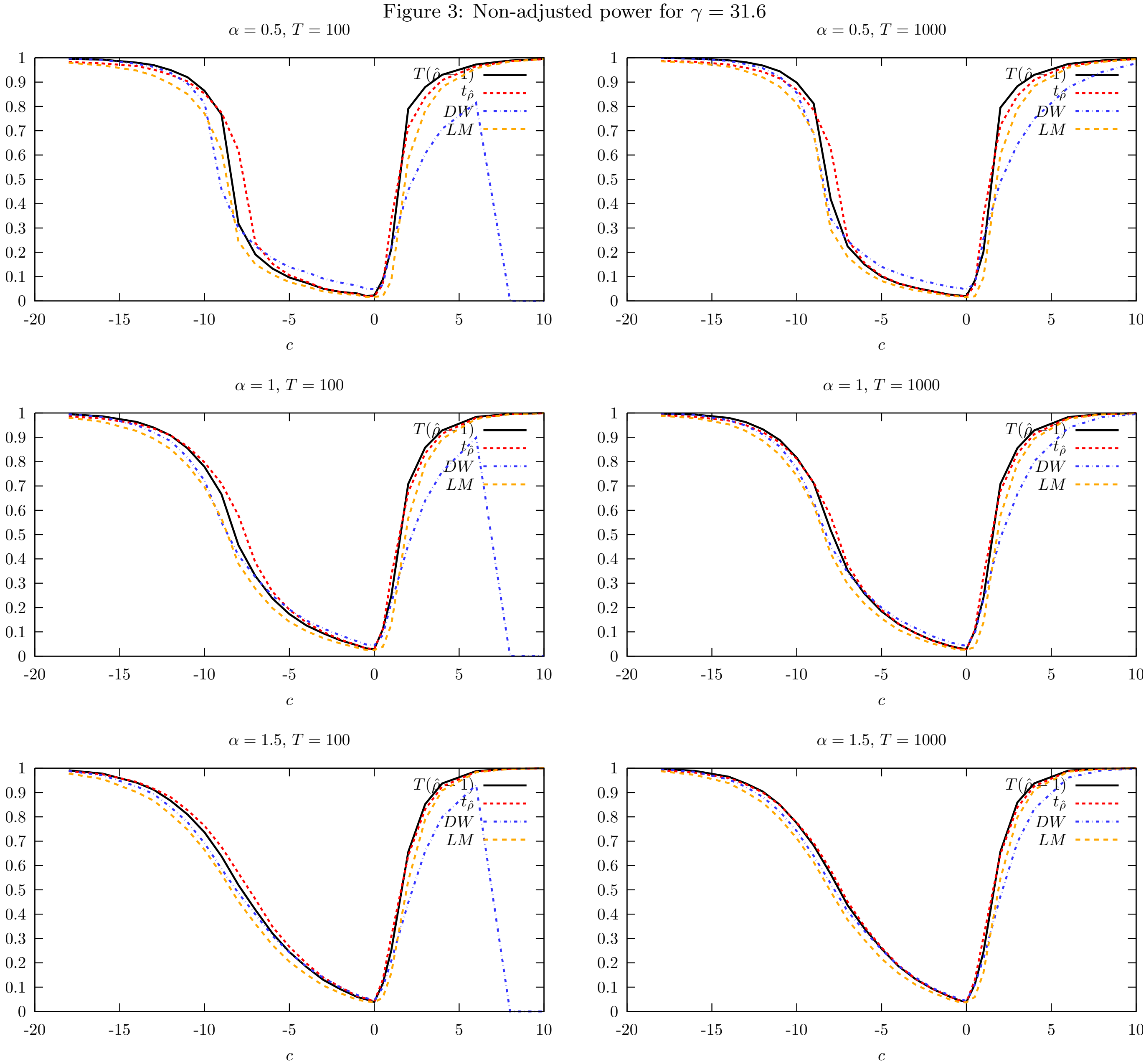


Figure 4: Non-adjusted power for $\gamma = 31.6$ (zoom)

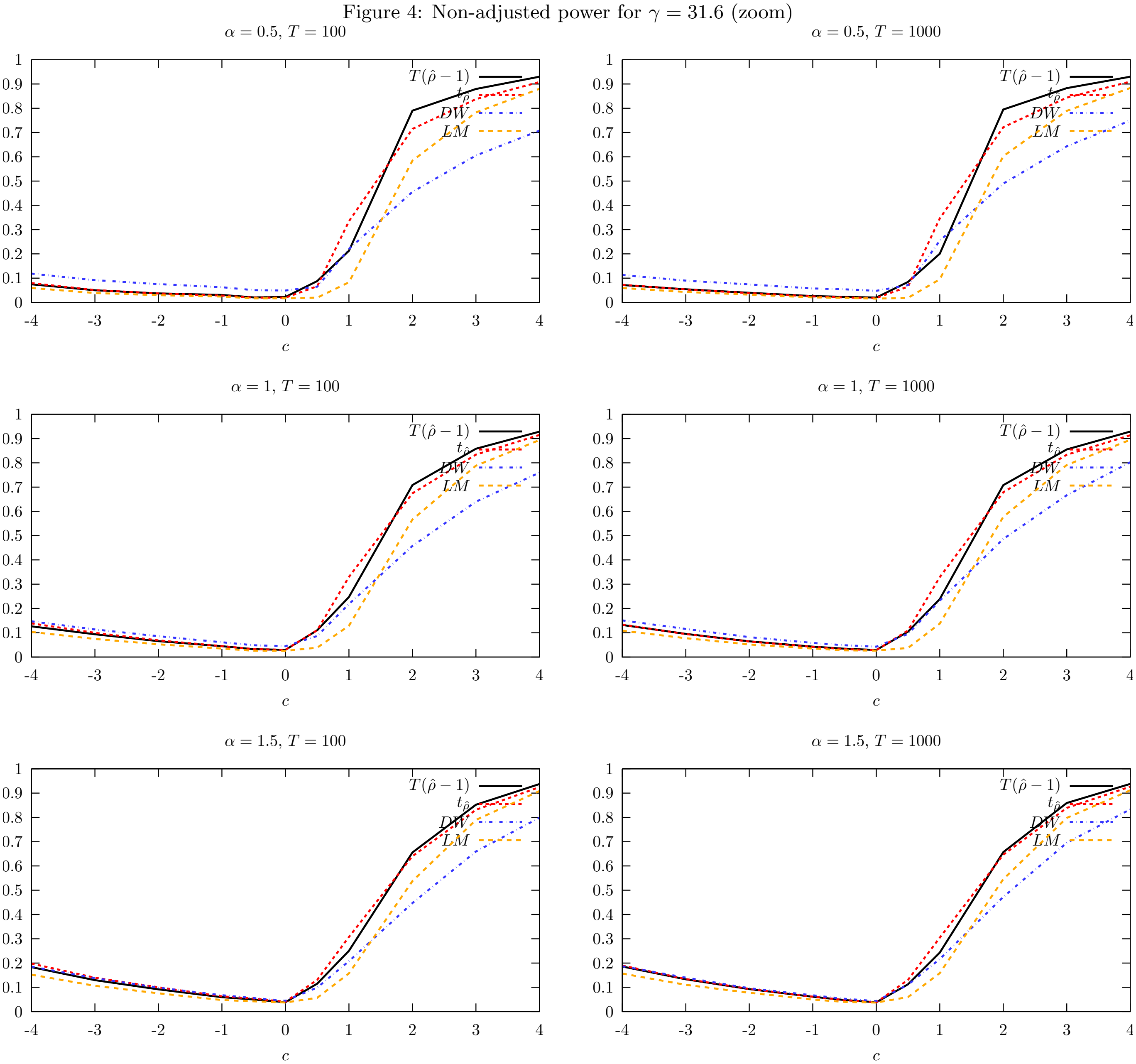


Table 2: Rejection rates $DW$ test at 5% of nominal size, $\alpha = 1$, $T = 1000$

| | | $\gamma$ | | | |
|---|---|---|---|---|---|
| | | 0.00 | 0.316 | 3.16 | 31.6 |
| $c$ | 20.0 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 21.0 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 22.0 | 1.00 | 0.99 | 1.00 | 1.00 |
| | 23.0 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 24.0 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 24.1 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 24.2 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 24.3 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 24.4 | 0.00 | 0.00 | 0.00 | 0.00 |

prediction since power falls suddenly from 1 to zero, suggesting the test might be inconsistent. This behavior seems to depend just on the sample size and not on the the values taken by $c$ and/or $\gamma$. In fact, if we consider the case $T = 100$, from Figures 1 we can see how power falls abruptly to zero when $c$ exceeds 7 whereas, for the same range of values of $c$ and $\alpha$, when $T = 1000$, see Figure 3, there is no fall in power.

Table 2 makes this point very clearly. For $T = 1000$ and $\alpha = 1$, and irrespective of $\gamma$, power is equal to 1 as $c$ increases from 20 to 24.2 and it collapses to zero for $c = 24.3$ ($c = 20$ corresponds to an autogressive parameter $\rho = 1.02$, $c = 24.2$ to $\rho = 1.0244$ and $c = 24.3$ to $\rho = 1.0245$). Thus, very small changes in the autoregressive parameter induce large changes in the power of $DW$ test suggesting the existence of a discontinuity in the power curve of the test.

It should be remarked that the behavior of the DW test does not depend on the $\alpha$-stable component or its weight in the DGP given it does not change with $\gamma$ or $\alpha$. In fact, from our analytical results we know that under the explosive alternative the DW test is inconsistent as it approaches zero as we move away from the null hypothesis. The theoretical result is obtained as $T \to \infty$ while in finite samples this may occur for a sample size as low as $T = 100$. A peculiar behavior of the $DW$ test has been documented elsewhere in the literature. Krämer (1985) has shown that in linear regression models without an intercept the power of the $DW$ is either 1 or zero, and Krämer and Zeisel (1990) have shown that for some choice of the regressors the power of the $DW$ test drops to zero as the autocorrelation among the disturbances increases. Bartels (1992) find that, when testing for no autocorrelation in the linear regression model, for some dataset the power of the $DW$ test goes to 0 as the autoregressive parameter goes to $\pm 1$. In a simulation study of the $DW$ test when the error term in linear regression models follows some strongly dependent process such as a fractionally integrated process, Kleiber and Krämer (2005) find that the power of the $DW$ test can drop to zero when the long memory parameter approaches the stationarity region.

This behavior of the $DW$ test is clearly a serious drawback of the test and the fact that for large sample sizes this behavior does not disappear raises serious concerns on its usefulness when considering explosive alternatives. On the other hand, we have just seen that, in the vicinity of the null hypothesis on the stationary side, the DW test has the best performance: little if any size distortion and higher power. The use of the DW test is therefore strongly recommended under those circumstances.

Figure 5 and 7 report the size-adjusted power of the tests, i.e., for an effective size of the test equal to 5%, for $\gamma = 0.316$ and $\gamma = 31.6$ and Figure 6 and 8 provide a zoom of the empirical power function over a range for $c$ associated to small departures from the null hypothesis. Adjusted-power curves are useful when the effective size of the tests, whose power properties are investigated, differ from the nominal size as it occurs sometimes in our case. When $\gamma$ is small, comparing Figure 1 and Figure 5, the adjusted power is close to the non-adjusted power. However, when $\gamma$ is high, looking at Figure 3 and Figure 7, we notice that the $DW$ test has quite low power against stationary alternatives while the other tests are much more powerful than in the size-unadjusted case. The low empirical power of the $DW$ test in size-adjusted case seems to shed some doubts on the overall usefulness of the $DW$ test as a nonstationarity test since (a) it has lower power under stationary alternatives and (b) it seems inconsistent under explosive alternatives. However, we would like to point out that the pratictioner would rarely be able to use the critical values generating an effective size equal to the nominal one, which would require the knowledge of $\alpha$ and $\gamma$, but it will rather use the standard critical values valid in the finite variance case. Under these circumstances, the high power of $DW$ test under stationary alternatives might be fully appreciated. Finally, Figure 6 and 8 allows to asses the heterogeneity in the power of the tests in the vicinity of the null hypothesis, in particular for explosive alternatives. The $t$-ratio has the best performance as $c$ increases, for any sample size and $\alpha$, but its power is rapidly reached by the $T(\hat{\rho} - 1)$ test with the $LM$ test lagging behind in all cases. As for the $DW$ test, as in the size-unadjusted case, we document the sudden fall in power from 1 to 0 as $c$ exceeds a given threshold, irrespective of the value taken by $\alpha$ and $\gamma$.

The same behavior can be observed for the adjusted power curves (Figure 5 and 7) only for the $DW$ test while for the other test statistics power seems to be higher when $\gamma$ is large.

Finally, in Figure (9), we graph the non-adjusted power curve of the four tests for $\alpha = 1$, $\gamma = 31.6$ and four increasing values for the sample sizes ($T = 50, 100, 1000, 10000$) to investigate its effect on power. If we consider stationary alternatives, i.e., $c$ negative, we notice that for the $t$-ratio power is not very sensitive to the sample size whereas for the other tests power increases with the sample size even thouh the beneficial effect of a larger sample size is mild inthe vicinity of the null hypothesis and it becomes stronger as we move further in the stationary region. Under explosive

Figure 5: Adjusted power for $\gamma = 0.316$

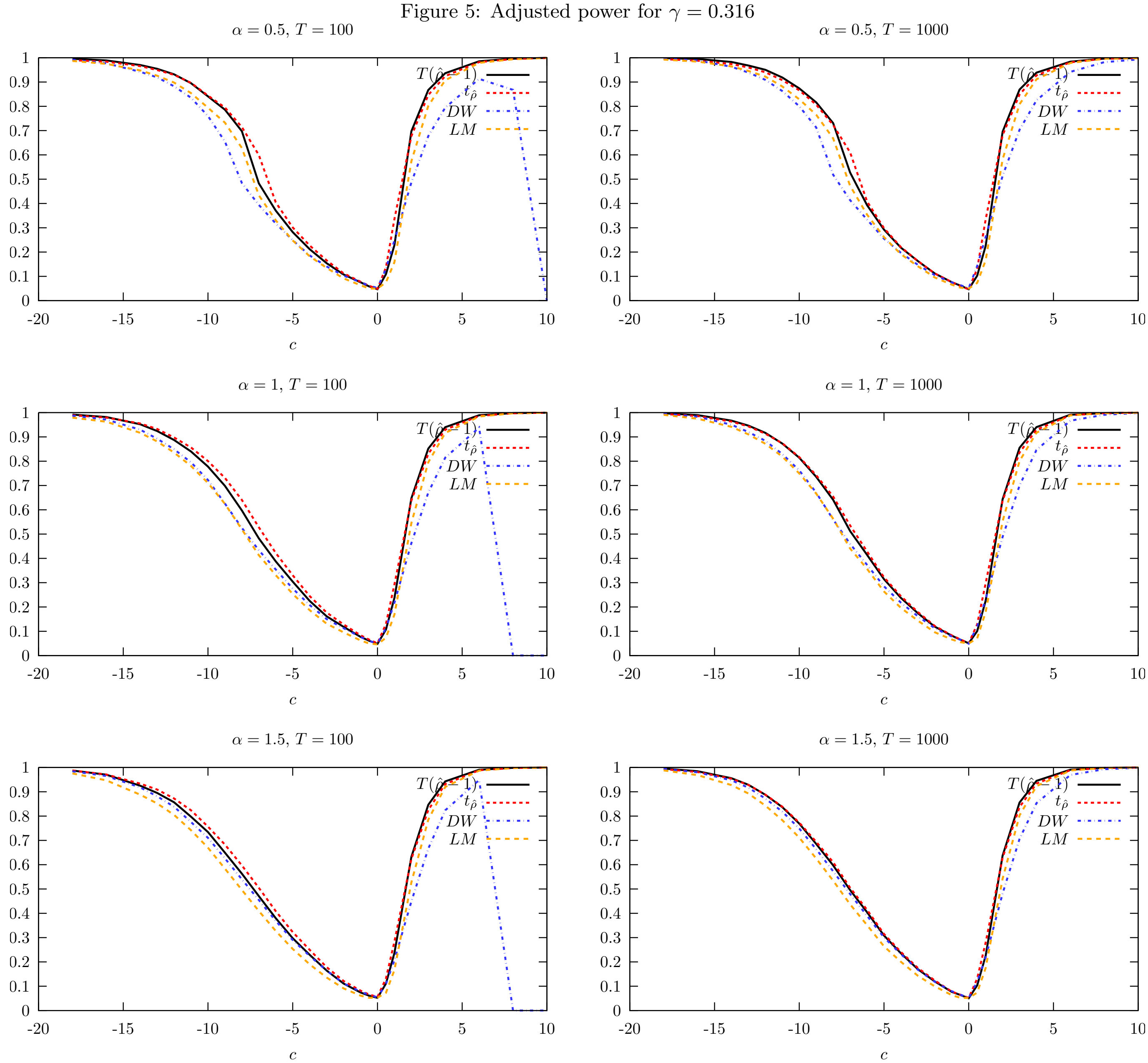


Figure 6: Adjusted power for $\gamma = 0.316$ (zoom)

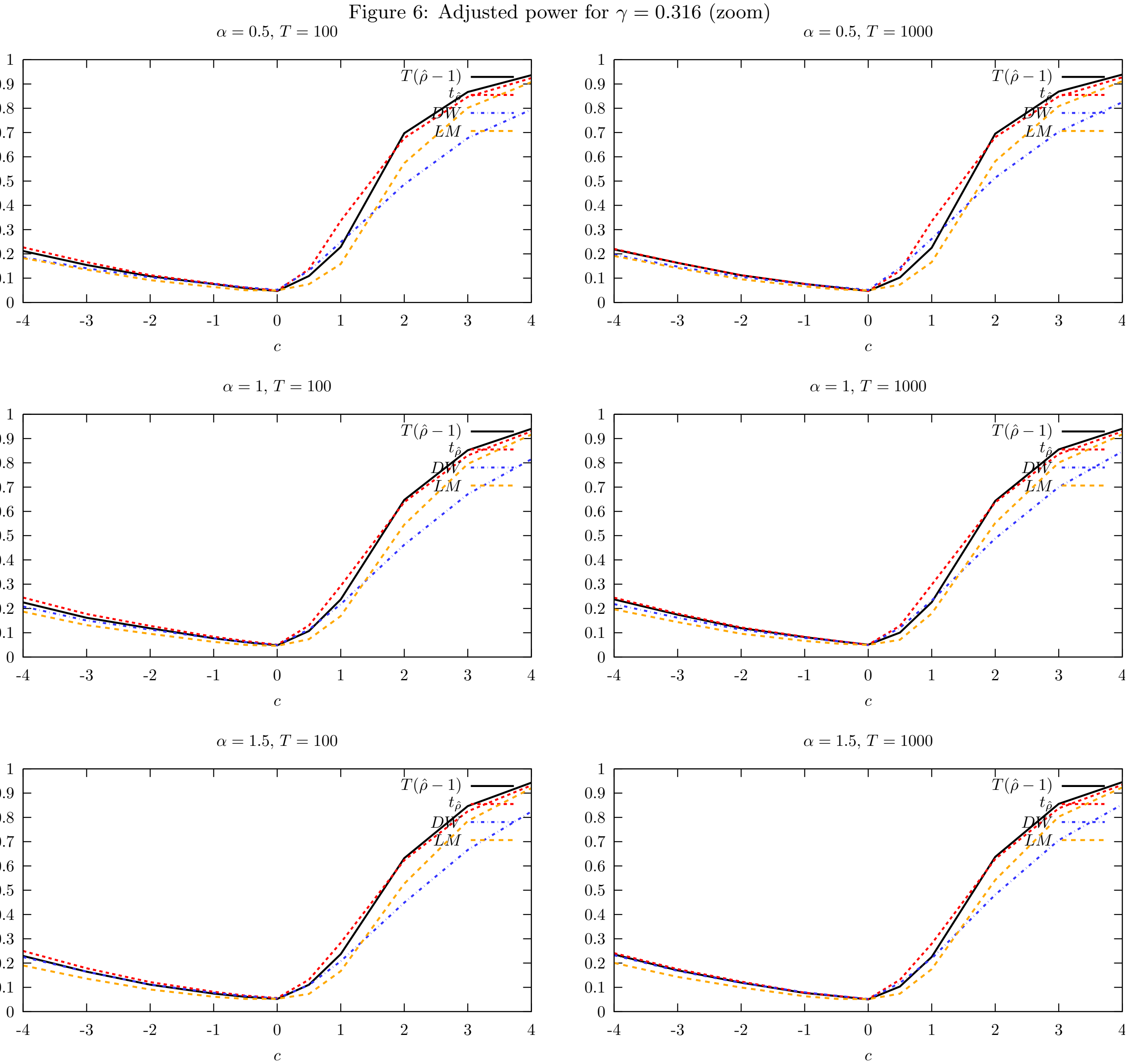


Figure 7: Adjusted power for $\gamma = 31.6$

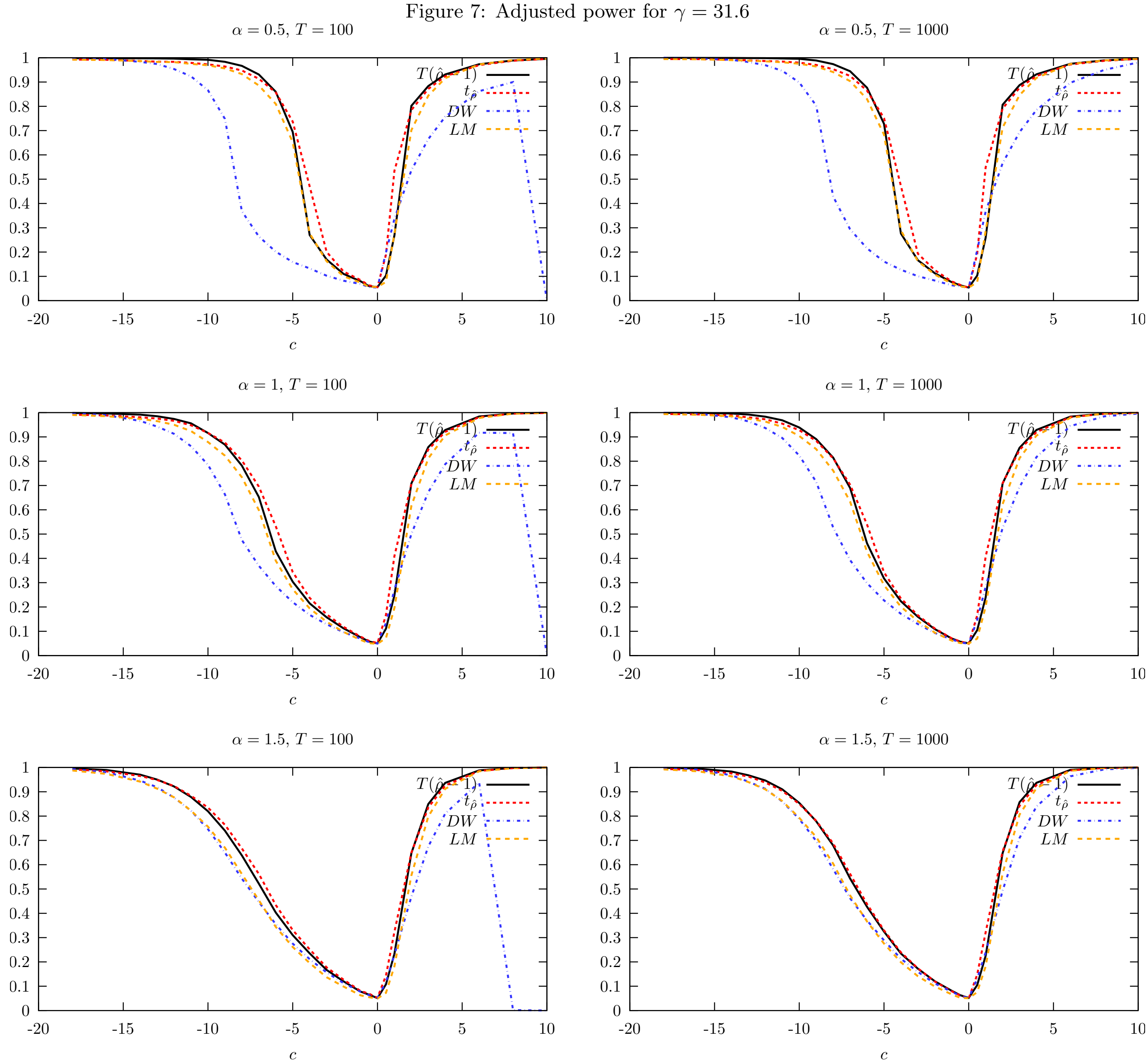


Figure 8: Adjusted power for $\gamma = 31.6$ (zoom)

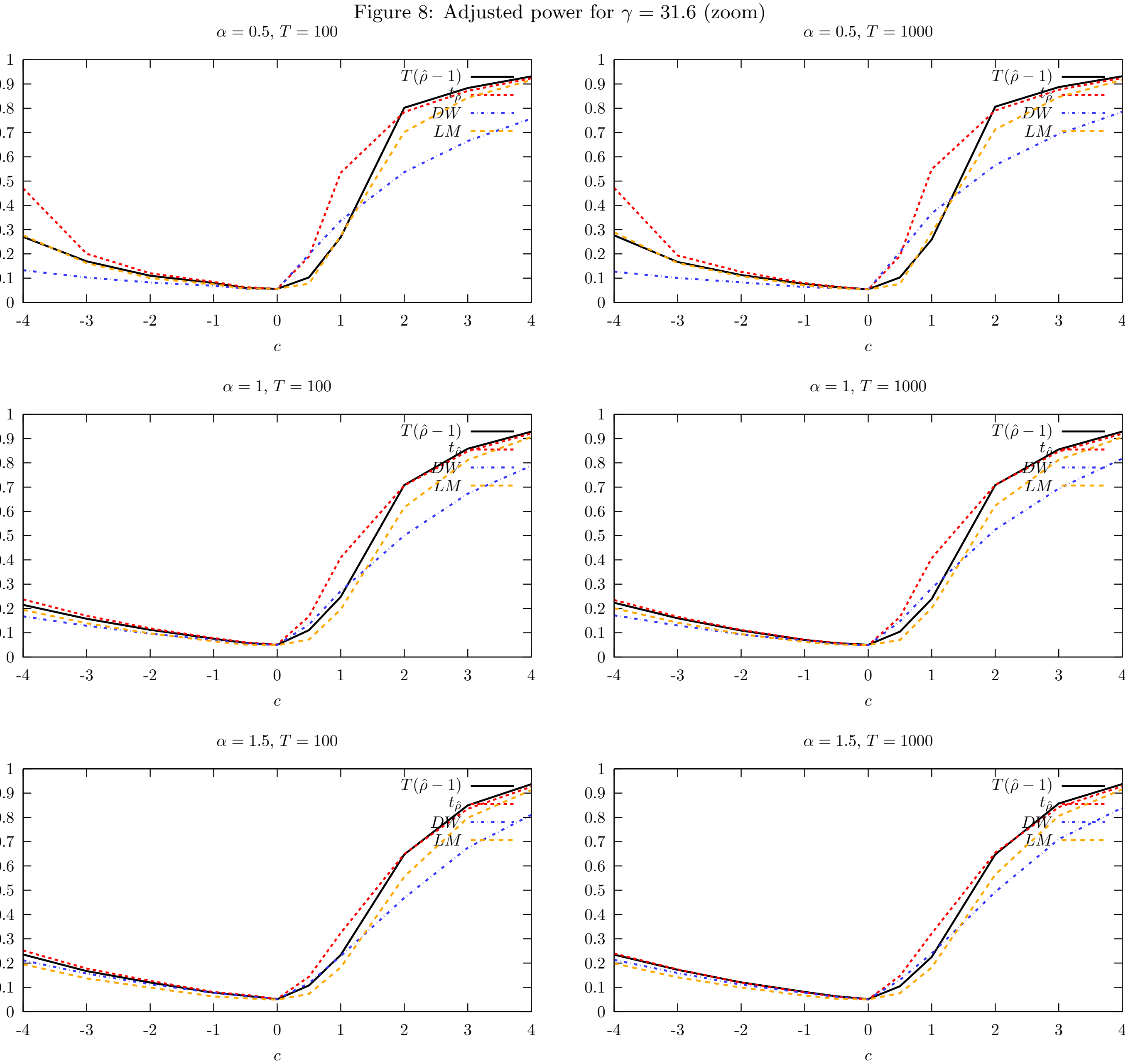
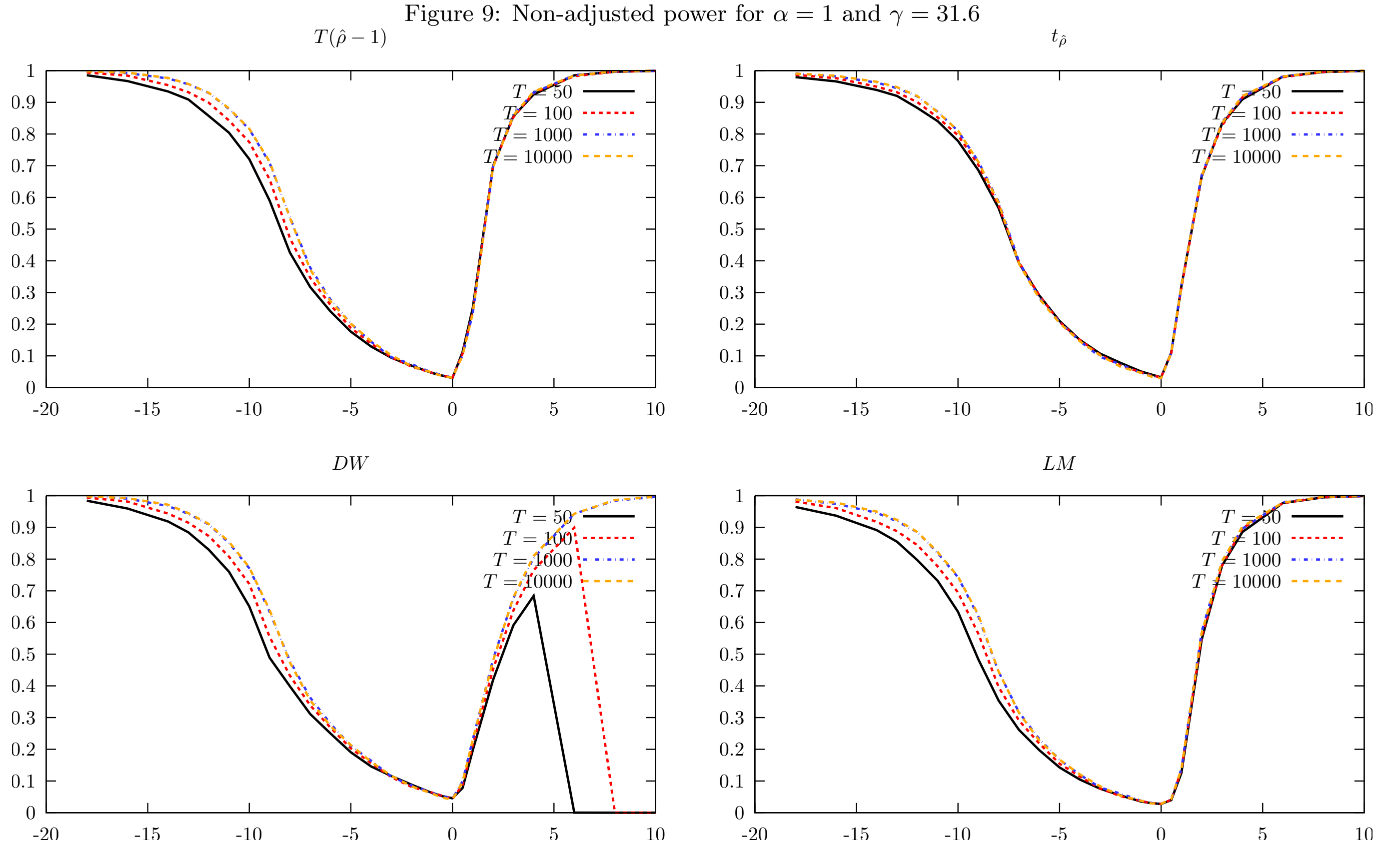


Figure 9: Non-adjusted power for $\alpha = 1$ and $\gamma = 31.6$




alternatives, the $DW$ test display, as discussed above, a sudden fall in power to 0 while the other tests have a similar behavior as a function of $t$.

## 4. Conclusions

In this paper we have studied the power of four popular unit root tests in the presence of a local-to-finite variance DGP. We have characterized the distribution of these tests under a sequence of local alternatives, considering both stationary and explosive alternatives . We carried out a small simulation study to assess the finite sample power of the tests. Our results suggest that the finite sample power is affected by the $\alpha$-stable component for low values of $\alpha$ and that, in the presence of this component, the DW test has the highest power under stationary alternatives. We also document a bizarre behavior of the $DW$ test which under the explosive alternative suddenly falls from 1 to zero for small changes in the alternative hypothesis suggesting a discontinuity in the power function of the $DW$ test.

## References


Ahn, S.K. (1993), 'Some tests for unit roots in autoregressive-integrated-moving average models with deterministic trends', *Biometrika* **80**, 855–868.

Ahn, S.K., S.B. Fotopoulos and L. He (2001), 'Unit root tests with infinite variance errors', *Econometric Reviews* **20**, 461–483.

Amsler, C. and P. Schmidt (2012), 'Tests of short memory with thick tails', *Journal of Business & Economic Statistics* **30**, 381–390.

Bartels, Robert (1992), 'On the power function of the durbin-watson test', *Journal of Econometrics* **51**(1-2), 101–112.

Callegari, F., N. Cappuccio and D. Lubian (2003), 'Asymptotic inference in time series regressions with a unit root and infinite variance errors', *Journal of Statistical Planning and Inference* **116**, 277–303.

Cappuccio, N. and D. Lubian (2007), 'Asymptotic null distributions of stationarity and nonstationarity tests under local-to-finite variance errors', *Annals of the Institute of Statistical Mathematics* **59**, 403–423.

Dickey, D.A. and W.A. Fuller (1976), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**, 427–431.

Kleiber, Christian and Walter Krämer (2005), 'Finite-sample power of the Durbin–Watson test against fractionally integrated disturbances', *Econometrics Journal* **8**(3), 406–417.

Krämer, Walter (1985), 'The power of the durbin-watson test for regressions without an intercept', *Journal of Econometrics* **28**(3), 363–370.

Krämer, Walter and Helmut Zeisel (1990), 'Finite sample power of linear regression autocorrelation tests', *Journal of Econometrics* **43**(3), 363–372.

McCulloch, J. H. (1997), 'Measuring tail thickness to estimate the stable index $\alpha$: a critique', *Journal of Business & Economic Statistics* **15**, 74–81.

Paulauskas, Vygantas and Svetlozar T. Rachev (1998), 'Cointegrated processes with infinite variance innovations', *Journal of Applied Probability* **8**(3), 775–792.

Paulauskas, Vygantas, Svetlozar T. Rachev and Frank. J. Fabozzi (2011), 'Comment on "Weak convergence to a matrix stochastic integral with stable processes"', *Econometric Theory* **27**, 907–911.

Phillips, P.C.B. (1987), 'Towards a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

Phillips, P.C.B. (1990), 'Time series regression with a unit root and infinite variance errors', *Econometric Theory* **6**, 44–62.

Protter, P. (1990), *Stochastic Integration and Differential Equations*, Springer and Verlag.

Resnick, S.I. (1986), 'Point processes regular variation and weak convergence', *Advances in Applied Probability* **18**, 66–138.