Robust testing of time trend and mean with unknown integration order errors*

Jiawen Xu[†]
Shanghai University of Finance and Economics
and Key Laboratory of Mathematical Economics

Pierre Perron[‡]
Boston University

March 13, 2013; Revised January 30, 2014

Abstract

We provide tests to perform inference on the coefficients of a linear trend assuming the noise to be a fractionally integrated process with memory parameter $d \in (-0.5, 1.5)$ by applying a quasi-GLS procedure using d-differences of the data. Doing so, the error term is short memory, the asymptotic distribution of the OLS estimators applied to quasi-differenced data and their t-statistics are unaffected by the value of d and standard procedures have a limit normal distribution. No truncation or pre-test is needed given the continuity with respect to d. To have feasible tests, we use the Exact Local Whittle estimator of Shimotsu (2010), valid for processes with a linear trend. The finite sample size and power of the tests are investigated via simulations. We also provide a comparison with the tests of Perron and Yabu (2009) valid for a noise component that is $I(0)$ or $I(1)$. The results are encouraging in that our test is valid under more general conditions, yet has similar power as those that apply to the dichotomous cases with d either 0 or 1.

JEL Classification: C22
Keywords: fractional integration, long-memory, linear time trend, inference, confidence intervals, quasi-GLS procedure.

---

*We are grateful to two referees and the editor for useful comments.

[†]School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China 200433, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China (xu.jiawen@mail.shufe.edu.cn).

[‡]Corresponding author: Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (perron@bu.edu).

1 Introduction

Many time series are well captured by a deterministic linear trend. With a logarithmic transformation, the slope of the trend function represents the average growth rate of the time series, a quantity of substantial interest. To be more precise, consider the following model for the time series process y_t:

$$y_t = \beta_1 + \beta_2 t + u_t, \tag{1}$$

where u_t are the deviations from the trend. The parameter β_2 is of primary interest. If $\beta_2 = 0$, then tests about β_1 pertains to the mean of the time series. Hypothesis testing on the slope of the trend function is important for many reasons. First, assessing whether a trend is present is of direct interest in many applications. Second, the correct specification of the trend function is important in other testing problems, such as assessing the nature of the noise component u_t. Third, tests for hypotheses about the values of β_1 and β_2 allow constructing confidence intervals via inversions. There is a large literature on issues pertaining to inference about the slope of a linear trend function, most related to the case where the noise component is stationary, i.e., integrated of order zero, $I(0)$. A classic result due to Grenander and Rosenblatt (1957) states that the estimate of β_2 obtained from a simple least-squares regression of the form (1) is asymptotically as efficient as that obtained from a Generalized Least Squares (GLS) regression when the process for u_t is correctly specified. However, when u_t has an autoregressive unit root, i.e., integrated of order one, $I(1)$, the estimate of the mean of the first-differenced series is efficient in large samples.

Several papers tackled the issue of constructing tests and confidence intervals for the parameter β_2 when it is not known a priori if u_t is $I(1)$ or $I(0)$. Sun and Pantula (1999) proposed a pre-test method which first applies a test of the unit root hypothesis and then chooses the critical value to be used according to the outcome of the test. Since the probability of using the critical values from the $I(0)$ case does not converge to zero when the errors are $I(1)$, the simulations reported accordingly show that substantial size distortions remain. Canjels and Watson (1997) considered various Feasible GLS methods. Their analysis is, however, restricted to the cases where u_t is either $I(1)$ or the autoregressive root is local to one. They do not allow $I(0)$ processes and, moreover, their method yields confidence intervals that are substantially conservative with common sample sizes. Roy et al. (2004) considered a test based on a one-step Gauss Newton regression but its limit distribution is not the same in the $I(1)$ and $I(0)$ cases (see Perron and Yabu, 2012). Vogelsang (1998), Bunzel and Vogelsang (2005) and Harvey et al. (2007) proposed tests valid with either $I(1)$ or $I(0)$ errors. Their approach, however, uses randomly scaled versions of tests for trends so that in finite samples the good properties of such tests are lost, at least to some extent. Perron and Yabu (2009) considered a Feasible Quasi GLS approach that uses a superefficient estimate of the sum of the autoregressive parameters α when $\alpha = 1$. The estimate of α is the OLS estimate obtained from an autoregression applied to detrended data and is truncated to take a value 1 when the estimate is in a $T^{-\delta}$ neighborhood of 1. This makes the estimate "super-efficient" when $\alpha = 1$ and implies that inference on the slope parameter can be performed using the standard normal or chi-square distribution whether $\alpha = 1$ or $|\alpha| < 1$.

Much of the literature focused on u_t being $I(0)$ or $I(1)$, special cases of fractionally integrated, $I(d)$, processes with memory parameter d. Since d can take any real value (within some interval), a long-memory process extends the classical dichotomy of $I(0)$ and $I(1)$ processes. Our aim is to provide tests to perform inference on the coefficients of a linear trend function assuming the noise component to be an $I(d)$ process with $d \in (-0.5, 1.5)$. The methodology is similar to that in Perron and Yabu (2009) and applies a quasi-GLS procedure using d-differences of the data. The error term is then short memory and the asymptotic distribution of the OLS estimators of (β_1, β_2) and their t-statistics are unaffected by the value of d and standard OLS procedures can be applied with the limit normal distribution. No truncation or pre-test is needed given the continuity with respect to d. To make our procedure feasible, we need an estimator of d valid with a fitted linear time trend and for a wide range of d. After experimenting with various possible estimators, we opted to use the Exact Local Whittle (ELW) estimator of Shimotsu (2010) who extended Shimotsu and Phillips (2005) to cover processes with a linear trend. It is valid for values of d in the range $(-.5, 1.5)$ and yields tests with good finite sample properties. A related paper is Iacone and al. (2013) who proposed a test for a break in the slope of a linear time trend when the order of integration is unknown, whose methodology is similar to ours. Also, Abadir, Distaso and Giraitis (2011) considered an $I(d)$ model with trend and cycles and derived the asymptotic distribution of the OLS estimate of the parameter of the slope of the trend.

This note is organized as follows. Section 2 describes the model and the test statistics, and Section 3 the estimate of d used to have feasible tests. Section 4 presents simulation results about the size and power of the tests in finite samples. We provide a comparison with the tests of Perron and Yabu (2009) valid when u_t is either $I(0)$ or $I(1)$. The results are encouraging in the sense that our test is valid under much more general conditions, yet has similar power as those that apply only to the dichotomous cases with d either 0 or 1. Section 5 provides brief conclusions and a mathematical appendix some technical derivations.

2 The model and test statistics

The data-generating process is assumed to be:

$$y_t = \beta_1 + \beta_2 t + u_t \tag{2}$$

for $t = 1, ..., T$, with u_t a fractionally integrated process satisfying the following assumptions.
•Assumption 1: The process u_t is generated by $\Delta^d u_t = (1-L)^d u_t = \varepsilon_t \P\{t \geq 1\}$, where Δ^d is the fractional difference operator and $\P\{A\}$ is the indicator function of the event A. Also, ε_t is a linear short memory process generated by $\varepsilon_t = A(L)v_t = \sum_{j=0}^{\infty} A_j v_{t-j}$ with $A(1)^2 > 0$, $\sum_{l=0}^{\infty} l|A_l| < \infty$, $v_t \sim i.i.d.\ (0, \sigma_v^2)$ and $E|v_t|^q < \infty$ with $q > \max(4, 2/(3-2d))$.
•Assumption 2: $f_\varepsilon(\lambda)$ is bounded for $\lambda \in [0, \pi]$; $f_\varepsilon(\lambda) \sim G_0 \in (0, \infty)$ as $\lambda \to 0_+$ and, for some $\beta \in (0, 2]$, $f_\varepsilon(\lambda) = G_0(1 + O(\lambda^\beta)$; in a neighborhood $(0, \delta)$ of the origin, $A(e^{i\lambda})$ is differentiable and $(d/d\lambda)A(e^{i\lambda}) = O(\lambda^{-1})$ as $\lambda \to 0_+$.

Assumptions 1-2 are mostly from Shimotsu (2010) and allow the estimate of d to be consistent and asymptotically normally distributed. Assumption 1 strengthens some of his conditions in order to have a functional central limit theorem for the partial sums of the u_t. The stated conditions for this to hold follow Marinucci and Robinson (2000). Applying a d-differencing transformation, the DGP can be written as:

$$y_t^d \equiv \Delta^d y_t = \beta_1 \Delta^d \P\{t \geq 1\} + \beta_2 \Delta^d t \P\{t \geq 1\} + \Delta^d u_t \P\{t \geq 1\}, \qquad (t = 1, ..., T)$$

Note that $\Delta^d u_t = \varepsilon_t$ and $\Delta^d y_1 = y_1$. We also define $X_t = [1, t]'$ and $X_t^d \equiv \Delta^d X_t = [\Delta^d \P\{t \geq 1\}, \Delta^d t \P\{t \geq 1\}]'$ with $\Delta^d X_1 = [1, 1]'$. Hence, the GLS transformed regression is:

$$y_t^d = X_t^{d\prime} \beta + \varepsilon_t, \qquad (t = 1, ..., T)$$

To obtain a feasible regression, we need to replace d by some consistent estimate $\hat{d}$ to be discussed in the next section. The tests will then be based on the regression

$$y_t^{\hat{d}} = X_t^{\hat{d}\prime} \beta + u_t^{\hat{d}}, \qquad (t = 1, ..., T) \tag{3}$$

where $u^{\hat{d}} = \Delta^{\hat{d}} u_t 1\{t \geq 1\}$. Let $\hat{\beta} = (X^{\hat{d}\prime} X^{\hat{d}})^{-1} X^{\hat{d}\prime} y^{\hat{d}}$ denote the OLS estimator of $[\beta_1, \beta_2]'$, where $X^{\hat{d}} = [X_1^{\hat{d}\prime}, ..., X_T^{\hat{d}\prime}]'$ and $y^{\hat{d}} = [y_1^{\hat{d}}, ..., y_T^{\hat{d}}]'$. The test statistic on the time trend coefficient β_2 for $H_0 : \beta_2 = \beta_2^0$ against $H_1 : \beta_2 \neq 0$, is constructed as the usual t-statistic:

$$t_{\hat{\beta}_2} = R(\hat{\beta} - \beta^0)/[\hat{\sigma}^2 R(X^{\hat{d}\prime} X^{\hat{d}})^{-1} R']^{1/2}$$

where $R = [0\ 1]$, $\beta^0 = (\beta_1^0, \beta_2^0)$ and $\hat{\sigma}^2$ is a consistent estimator of long-run variance $\sigma^2 = \sum_{j=-\infty}^{\infty} \Gamma(j)$ where $\Gamma(j) = E(\varepsilon_t \varepsilon_{t-j})$. Similarly, the test statistic on the constant term β_1 for $H_0 : \beta_1 = \beta_1^0$ can also be constructed as usual with:

$$t_{\hat{\beta}_1} = R_1(\hat{\beta} - \beta^0)/[\hat{\sigma}^2 R_1 (X^{\hat{d}'} X^{\hat{d}})^{-1} R_1']^{1/2}$$

where $R_1 = [1\ 0]$. The following theorem provides the limit distribution of the test statistics.

Theorem 1 *Let $\{y_t\}$ be generated by (2) under Assumptions 1-2. Suppose that we have estimates $\hat{d}$ and $\hat{\sigma}^2$ such that $\hat{d} - d = O_p(T^{-\kappa})$ for some $\kappa > 0$ and $\hat{\sigma}^2 - \sigma^2 = o_p(1)$. Then, a) under $H_0 : \beta_2 = \beta_2^0$, $t_{\hat{\beta}_2} \to^d N(0,1)$ for any $d \in (-0.5, 1.5)$; b) under $H_0 : \beta_1 = \beta_1^0$, $t_{\hat{\beta}_1} \to^d N(0,1)$ for any $d \in (-0.5, 0.5)$.*

A consistent estimate of σ^2 is readily available. Popular estimates are weighted sums of autocovariances of the form $\hat{\sigma}^2 = \hat{\Gamma}(0) + 2\sum_{j=1}^{T-1} \kappa(j,m)\hat{\Gamma}(j)$, where $\hat{\Gamma}(j) = T^{-1}\sum_{t=j+1}^{T} u_t^{\hat{d}} u_{t-j}^{\hat{d}}$ with $u_t^{\hat{d}}$ the OLS residuals from the regression (3) and $\kappa(\cdot)$ a kernel function with bandwidth m. In the simulations below, we use the Bartlett kernel and Andrews' (1991) data dependent method for selecting the bandwidth based on an AR(1) approximation. The choice of an appropriate estimate of d is more delicate and discussed in the next section.

3 Estimate of d

The Exact Local-Whittle (ELW) estimation procedure for the order of fractional integration of a process was studied by Shimotsu and Phillips (2005). It was subsequently extended by Shimotsu (2010) to cover the case with an unknown trend function, a needed feature in our context. It is also valid under a wide range of possible values for d including values greater than 1. Accordingly, we shall adopt it as the estimator of d to be used in constructing our test statistics. To describe the estimation procedure, define the discrete Fourier transform and the periodogram of y_t evaluated at the fundamental frequencies as

$$\omega_y(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t \exp(it\lambda_j), \qquad I_y(\lambda_j) = |\omega_y(\lambda_j)|^2$$

for $\lambda_j = (2\pi j/T)$, $j = 1, ..., T$. The ELW estimator of d is the minimizer of

$$Q_m(G, d) = \frac{1}{m} \sum_{j=1}^{m} [\log(G\lambda_j^{-2d}) + \frac{1}{G} I_{\Delta^d y}(\lambda_j)].$$

Concentrating $Q_m(G, d)$ with respect to G, the objective function is $R(d) = \log \hat{G}(d) - 2d(m^{-1})\sum_{j=1}^{m} \log(\lambda_j)$, where $\hat{G}(d) = m^{-1}\sum_{j=1}^{m} I_{\Delta^d y}(\lambda_j)$ and, within a pre-specified range to be defined below, the ELW estimator is $\widetilde{d} = \arg\min_{d\in[\Delta_1,\Delta_2]} R(d)$. Shimotsu (2010) extended the ELW estimation procedure to cover an unknown linear time trend via a two-step procedure applied to detrended data. The first step detrends the data by an OLS regression of y_t on $(1,t)$ with the residuals denoted $\hat{y}_t$. The modified objective function is then:

$$R_F(d) = \log \hat{G}_F(d) - 2d\frac{1}{m}\sum\nolimits_{j=1}^{m} \log(\lambda_j), \quad \hat{G}_F(d) = \frac{1}{m}\sum\nolimits_{j=1}^{m} I_{\Delta^d(\hat{y}-\varphi(d))}(\lambda_j)$$

where $\varphi(d) = (1-w(d))\hat{y}_1$ with $w(d)$ a twice continuous differentiable weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq 3/4$. As recommended by Shimotsu (2010), $w(d) = (1/2)[1+\cos(4\pi d)]$ for $d \in [1/2, 3/4]$. A two-step procedure is applied to ensure the global consistency of the estimate. In the first step, one uses the tapered local Whittle estimator of Velasco (1999) denoted $\hat{d}_T$, which is $\sqrt{m}$-consistent and invariant to a linear trend for $d \in (-1/2, 5/2)$. The second step estimator involves the following modification:

$$\hat{d}^{*}_{ELW} = \hat{d}_T - R'_F(\hat{d}_T)/R''_F(\hat{d}_T) \tag{4}$$

where $R'_F(\hat{d}_T)$ and $R''_F(\hat{d}_T)$ are the first and second derivatives of $R_F(d)$. Following Shimotsu (2010), we use $\max[R''_F(\hat{d}_T), 2]$ to improve the finite sample properties. The final estimator, denoted $\hat{d}_{ELW}$, is obtained iterating (4). Now consider the following assumption.

•Assumption 3: a) $-(1/2) < \Delta_1 < \Delta_2 \leq (7/4)$; b) as $T \to \infty$, $m^{-1} + m^{1+2\beta}(\log m)^2 T^{-2\beta} + m^{-\gamma}\log T \to 0$, for any $\gamma > 0$, with β as defined in Assumption 2.

From Shimotsu (2010), under Assumptions A1-A3, $\sqrt{m}(\hat{d}_{ELW} - d) \to^d N(0, 1/4)$. Hence, if our test statistics are constructed using $\hat{d}_{ELW}$, Theorem 1 continues to hold.

4 Simulation results

In this section, we consider the size and power of the test $t_{\hat{\beta}_2}$ for the slope of the trend via simulations, using 1,000 replications throughout (the results for the test $t_{\hat{\beta}_1}$ for the mean are qualitatively similar for the range $d \in (-0.5, 0.5)$). The data are generated by (2) with u_t an $ARFIMA(p,d,q)$ of the form $(1-L)^d u_t = \varepsilon_t \P\{t \geq 1\}$ with $A(L)\varepsilon_t = B(L)e_t$, where $A(L) = 1 - a_1 L - ... - a_p L^p$ and $B(L) = 1 + b_1 L + ... + b_q L^q$ are the autoregressive and moving average lag polynomials, respectively, and $e_t \sim i.i.d\ N(0,1)$. Assumptions A1 and A2 are satisfied if the roots of $A(L) = 0$ and $B(L) = 0$ are outside the unit circle. In all cases, we set $\beta_1 = \beta_2 = 0$ under the null hypothesis without loss of generality. Also, the estimate $\hat{d}_{ELW}$ is constructed with $m = T^{.65}$. We consider two-sided tests at the 5% significance level and for

$d = 0$ or 1, the results are compared to those obtained with the two versions of the Perron and Yabu (2009) tests, $t_{\beta}^{FS}(MU)$ or $t_{\beta}^{FS}(UB)$, which use different autoregressive estimates before applying the truncation (MU stands for Median Unbiased and UB for Upper Biased).

We start with the case of pure fractional processes with $A(L) = B(L) = 1$. We consider the range $d \in [-0.4, 1.4]$ and $T = 500$, 1000 and 2000. The results, presented in Table 1, show that the exact sizes of the test $t_{\hat{\beta}_2}$ are close to the nominal size in all cases. On the other hand, $t_{\beta}^{FS}(MU)$ and $t_{\beta}^{FS}(UB)$ show substantial size distortions unless $d = 0, 1$. When d is negative the tests are very conservative, while when $0 < d < 1$, the tests are liberal. The liberal size distortions are especially pronounced when $d = 1.4$. The power functions for a two-sided test of $\beta_2 = 0$ are presented in Figure 1 for $T = 500$. Given the size distortions of the Perron and Yabu (2009) tests when d is different from 0 and 1, we include them only for the case $d = 1$ (we return below to the case $d = 0$). When $d = 1$, $t_{\beta}^{FS}(MU)$ and $t_{\beta}^{FS}(UB)$ have higher power, as expected. This due to the fact that the Perron and Yabu (2009) tests apply a truncation to 1 when the autoregressive parameter is in a neighborhood of 1 leading to a smaller bias when $d = 1$. However, the differences are not large and decrease as T increases (from unreported simulations). As expected, the power of $t_{\hat{\beta}_2}$ is highest when d is small with the power decreasing monotonically as d increases (note the different scaling on the horizontal axis).

Table 2 presents results about the size of the tests for processes with short-run dynamics of the autoregressive form with an $AR(1)$ so that $A(L) = 1 - aL$ with $d = 0$, cases for which the Perron and Yabu tests were designed. We consider values of a ranging from 0 to 0.95. The results show that the exact size remains close to the nominal 5% level, unless a is close to 1, in which case the exact size of $t_{\hat{\beta}_2}$ is below nominal size. It is well known that in the presence of a short-run component that has strong correlation, most estimates of d are biased. Accordingly, it is of some comfort to see that our test retains decent size and exhibits no liberal size distortions. The power functions for a two-sided test of $\beta_2 = 0$ are presented in Figure 2 for $T = 500$. When $a = 0$, 0.3 or 0.5, all tests have essentially the same power. When $a = 0.7$ or 0.9, the Perron and Yabu tests have slightly higher power. When $a = 0.95$, $t_{\hat{\beta}_2}$ has much higher power, despite being conservative, unless the alternative is close to the null value.

We next consider the size and power of the tests using five different DGPs used in Qu (2011), which were motivated by financial applications of interest. These are given by:

DGP 1. ARFIMA$(1, d, 0)$: $(1 - a_1 L)(1 - L)^{0.4}\varepsilon_t = e_t$, where $a_1 = 0.4$ and -0.4.

DGP 2. ARFIMA$(0, d, 1)$: $(1 - L)^{0.4}\varepsilon_t = (1 + b_1 L)e_t$, where $b_1 = 0.4$ and -0.4.

DGP 3. ARFIMA$(2, d, 0)$: $(1 - a_1 L)(1 - a_2 L)(1 - L)^{0.4}\varepsilon_t = e_t$, with $a_1 = 0.3$, $a_2 = 0.5$.

DGP 4. $\varepsilon_t = z_t + \eta_t$, where $(1 - L)^{0.4} z_t = e_t$ and $\eta_t \sim i.i.d\ N(0, var(z_t))$.

DGP 5. $(1 - L)^{0.4}\varepsilon_t = \eta_t$ with $\eta_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.1\eta_{t-1}^2 + 0.85\sigma_{t-1}^2$.

In all cases, $e_t \sim i.i.d.\ N(0, 1)$. DGPs 1-3 are different cases of ARFIMA processes, DGP 4 is a fractionally integrated process with measurement errors and DGP 5 is a GARCH process. Note that DGPs 4 and 5 do not satisfy the conditions of Assumptions 1-2. We nevertheless include them to assess the robustness of the results given that conditional heteroskedasticity and measurement errors are prevalent features of many time series. Given the size distortions of the Perron and Yabu (2009) tests when d is different from 0 or 1, we only present results for the test $t_{\hat{\beta}_2}$.

Table 3 presents the exact sizes of the tests. In all cases, the exact size of $t_{\hat{\beta}_2}$ is near 5%, except for DGPs 2 and 5 for which the test has slight liberal size distortions when $T = 500$, which decrease as T increases. The power functions of the test for $T = 500$ are presented in Figure 3. In all cases, power increases rapidly to 1 as β_2 deviates from 0, with the exception perhaps of the case with GARCH errors. Comparing the results across DGPs, power decreases when additional short-run dynamics is present. The effect of measurement errors on the power is minor.

5 Conclusion

We provided tests to perform inference on the coefficients of a linear trend function assuming the noise to be a fractionally integrated process with memory parameter in the interval $(-0.5, 1.5)$. The results are encouraging in the sense that our test is valid under much more general conditions, yet has power similar to the Perron and Yabu (2009) tests that apply only to the dichotomous cases with d either 0 or 1. When d is different from 0 or 1, its exact size is close to the nominal size and power is good. Our procedure provides a useful tool for inference about the coefficients of a linear trend under general conditions on the noise component. Though we assumed the errors to follow a Type II long-memory process, we conjecture that our results remain valid with a Type I process as defined by Marinucci and Robinson (1999). First, as Shimotsu (2010) argues, his results remain valid for both types of processes. Also, the conditions for a functional central limit theorem for Type I processes are very similar, see e.g., Wang et al. (2003) and could be slightly modified accordingly.

References

Abadir, K. M., Distaso, W. and Giraitis, L. (2011) An I(d) model with trend and cycles. *Journal of Econometrics* **163**, 186-199.

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.

Bunzel, H. and Vogelsang, T. J. (2005) Powerful trend function tests that are robust to strong serial correlation with an application to the Prebish-Singer hypothesis. *Journal of Business and Economic Statistics* **23**, 381-394.

Canjels, E. and Watson, M. W. (1997) Estimating deterministic trends in the presence of serially correlated errors. *Review of Economics and Statistics* **79**, 184-200.

Grenander, U. and Rosenblatt, M. (1957) *Statistical Analysis of Stationary Time Series*, New York: John Wiley.

Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2007) A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics* **141**, 1302-1330.

Iacone, F., Leybourne, S. J. and Taylor, A. M. R. (2013) Testing for a break in trend when the order of integration is unknown. *Journal of Econometrics* **176**, 3045.

Marinucci, D. and Robinson, P. M. (1999) Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* **80**, 111-122.

Marinucci, D. and Robinson, P. M. (2000) Weak convergence of multivariate fractional Brownian motion. *Stochastic Processes and their Applications* **86**, 103-120.

Perron, P. and Yabu, T. (2009) Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics* **151**, 56-69.

Perron, P. and Yabu, T. (2012) Testing for trend in the presence of autoregressive error: a comment. *Journal of the American Statistical Association* **107**, 844.

Qu, Z. (2011) A test against spurious long memory. *Journal of Business & Economic Statistics* **29**, 423-438.

Robinson, P. M. (2005) Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* **33**, 1800-1842.

Robinson, P. M. and Iacone, F. (2005) Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* **129**, 263-298.

Roy, A., Falk, B. and Fuller, W. A. (2004) Testing for trend in the presence of autoregressive errors. *Journal of the American Statistical Association* **99**, 1082-1091.

Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* **26**, 501-540.

Shimotsu, K. and Phillips, P. C. B. (2005) Exact local Whittle estimation of fractional integration. *Annals of Statistics* **33**, 1890-1933.

Sun, H. and Pantula, S. G. (1999) Testing for trends in correlated data. *Statistics and Probability Letters* **41**, 87-95.

Velasco, C. (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20**, 87–127.

Vogelsang, T. J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica* **66**, 123-148.

Wang, Q., Lin, Y.-X. and Gulati, C. M. (2003) Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* **19**, 143-164.

Appendix

Proof of Theorem 1: We start by assuming that d and σ^2 are known and then show that the results remain the same under the condition stated. Consider first part (a). Let $K_T = diag\{T^{1/2-d}, T^{3/2-d}\}$, $X^d = [X_1^{d\prime}, ..., X_T^{d\prime}]'$, $X_t^d = [\mu_{0,t}\ \mu_{1,t}]'$, $\mu_{i,t} = \Delta^d t^i \P\{t \geq 1\}$ for $i = \{0, 1\}$ and $\varepsilon = [\varepsilon_1, ..., \varepsilon_T]'$ then:

$$t_{\hat{\beta}_2} = \frac{R(\hat{\beta} - \beta^0)}{[\sigma^2 R(X^{d\prime} X^d)^{-1} R']^{1/2}} = \frac{R(K_T^{-1} X^{d\prime} X^d K_T^{-1})^{-1}(K_T^{-1} X^{d\prime}\varepsilon)}{[\sigma^2 R(K_T^{-1} X^{d\prime} X^d K_T^{-1})^{-1} R']^{1/2}}$$

Lemma A.1 *(1) for* $-0.5 < d < 0.5$*:* $t_{\hat{\beta}_2} \xrightarrow{d} RC^{-1}L/[RC^{-1}R']^{1/2} \equiv A_1$*, where*

$$C = \begin{bmatrix} \frac{1}{\{\Gamma(1-d)\}^2(1-2d)} & \frac{1}{\Gamma(1-d)\Gamma(2-d)(2-2d)} \\ \frac{1}{\Gamma(1-d)\Gamma(2-d)(2-2d)} & \frac{1}{\{\Gamma(2-d)\}^2(3-2d)} \end{bmatrix}, \quad L = \begin{bmatrix} \frac{1}{\Gamma(1-d)} \int_0^1 r^{-d} dW(r) \\ \frac{1}{\Gamma(2-d)} \int_0^1 r^{1-d} dW(r) \end{bmatrix}$$

and

$$A_1 = \sqrt{3-2d}\left[(2-2d)\int_0^1 r^{1-d} dW(r) - (1-2d)\int_0^1 r^{-d} dW(r)\right]$$

(2) for $0.5 \leq d < 1.5$*:*

$$\begin{aligned} t_{\hat{\beta}_2} &= \frac{R(\widetilde{K}_T^{-1} X^{d\prime} X^d \widetilde{K}_T^{-1})^{-1}(\widetilde{K}_T^{-1} X^{d\prime}\varepsilon)}{[\sigma^2(\widetilde{K}_T^{-1} X^{d\prime} X^d \widetilde{K}_T^{-1})^{-1}]^{1/2}} \xrightarrow{d} \frac{\sigma R\widetilde{C}^{-1}\widetilde{L}}{[\sigma^2 R\widetilde{C}^{-1}R']^{1/2}} \\ &= C_{22}^{-1/2} L_2 = \sqrt{3-2d}\int_0^1 r^{1-d} dW(r) \equiv A_2 \end{aligned}$$

where $\widetilde{K}_T = diag\{1, T^{3/2-d}\}$ *with* C_{22} *and* L_2 *the relevant sub-matrices of* C *and* L*.*

Proof: From Lemma 1 of Robinson (2005), as $t \to \infty$, for $d \in (0, 1)$, $\Delta^d \P\{t \geq 0\} = \Gamma(1-d)^{-1} t^{-d} + O(t^{-1})$ and $\Delta^d t \P\{t \geq 0\} = \Gamma(2-d)^{-2} t^{1-d} + O(1)$. For $d \in (1, 1.5)$, $\Delta^d t \P\{t \geq 0\} = \Gamma(2-d)^{-1} t^{1-d} + O(t^{-1})$. From (A.34) of Robinson and Iacone (2005), for any $r \in (0, 1]$, we have a) for $d \in (0, 0.5)$: $T^{-d}\Delta^d \P\{[rT] \geq 0\} \to \Gamma(1-d)^{-1} r^{-d}$; b) for $d \in (0, 1.5)$, $T^{-d}\Delta^d [rT]\P\{[rT] \geq 0\} \to \Gamma(2-d)^{-1} r^{1-d}$. The facts that $K_T^{-1} X^{d\prime} X^d K_T^{-1} \to C$ and $K_T^{-1} X^{d\prime}\varepsilon \to L$ are proved in (A.36) of Robinson and Iacone (2005). For part (2), the result follows given that, when $0.5 \leq d < 1.5$,

$$\widetilde{K}_T^{-1} X^{d\prime} X^d \widetilde{K}_T^{-1} \to \begin{bmatrix} 0 & 0 \\ 0 & C_{22} \end{bmatrix} \quad \text{and} \quad \widetilde{K}_T^{-1} X^{d\prime}\varepsilon \to \widetilde{L} \equiv \begin{bmatrix} 0 \\ \sigma L_2 \end{bmatrix}$$

Lemma A.2 A_1 *and* A_2 *have a* $N(0, 1)$ *distribution.*

Note that A_2 can be approximated by the sum of functions of normal random variables $\sqrt{3-2d}T^{d-3/2}\sum_{j=1}^{T} j^{1-d}e_j$, with $e_j \sim i.i.d.\ N(0,1)$ so that $\sum_{j=1}^{T} j^{1-d}e_j$ is $N(0, \sum_{j=1}^{T} j^{2(1-d)})$. According to Faulhaber's formula, $\sum_{j=1}^{T} j^p = [(B+T)^{p+1} - B^{p+1}]/(p+1)$, with B the Bernoulli number. Also, $\sum_{j=1}^{T} j^{2(1-d)} \simeq T^{3-2d}/(3-2d)$, so that $\sqrt{3-2d}T^{d-3/2}\sum_{j=1}^{T} j^{1-d}e_j$ is $N(0,1)$. Similarly, A_1 can be approximated by

$$(2-2d)\sqrt{3-2d}T^{d-3/2}\sum_{j=1}^{T} j^{1-d}e_j - \sqrt{3-2d}(1-2d)T^{d-1/2}\sum_{j=1}^{T} j^{-d}e_j.$$

The first term is $N(0,(2-2d)^2)$ and the second is $N(0,(3-2d)(1-2d))$. The covariance of the two terms is $(3-2d)(1-2d)$, so that A_1 is $N(0,1)$, since $(2-2d)^2 + (3-2d)(1-2d)$ $-2(3-2d)(1-2d) = 1$.

For part (b), we have $t_{\hat{\beta}_1} \xrightarrow{d} \sigma R_1 C^{-1} L/[\sigma^2 R_1 C^{-1} R_1']^{1/2} \equiv B_1$, where

$$B_1 = \sqrt{1-2d}[(2-2d)\int_0^1 r^{-d}dW(r) - (3-2d)\int_0^1 r^{1-d}dW(r)].$$

Now, B_1 can be approximated by

$$(2-2d)\sqrt{1-2d}T^{d-1/2}\textstyle\sum_{j=1}^{T} j^{-d}e_j - \sqrt{1-2d}(3-2d)T^{d-3/2}\sum_{j=1}^{T} j^{1-d}e_j.$$

The first term is $N(0,(2-2d)^2)$ and the second is $N(0,(3-2d)(1-2d))$. Their covariance is $(3-2d)(1-2d)$, so that B_1 is $N(0,1)$, since $(2-2d)^2+(3-2d)(1-2d)-2(3-2d)(1-2d) = 1$.

It remains to show that the results remain the same with estimates of d and σ^2. The fact that the results remain the same when using a consistent estimate of σ^2 is trivial, hence we concentrate on using an estimate of d. We need to show that if $\hat{d} - d = O_p(T^{-\kappa})$ for any $\kappa > 0$, then (a) for $|d| < 0.5$,

$$K_T^{-1}X^{\hat{d}'}X^{\hat{d}}K_T^{-1} - K_T^{-1}X^{d'}X^{d}K_T^{-1} \xrightarrow{d} 0 \tag{A.1}$$

and

$$K_T^{-1}X^{\hat{d}'}u^{\hat{d}} - K_T^{-1}X^{d'}\varepsilon \xrightarrow{d} 0 \tag{A.2}$$

where $X^{\hat{d}} = [X_1^{\hat{d}'}, ..., X_T^{\hat{d}'}]'$, $u^{\hat{d}} = [u_1^{\hat{d}}, ..., u_T^{\hat{d}}]'$, $X_t^{\hat{d}} = [\hat{\mu}_{0,t}\ \hat{\mu}_{1,t}]'$, $u_t^{\hat{d}} = \Delta^{\hat{d}}u_t\P\{t \geq 1\}$, and $\hat{\mu}_{i,t} = \Delta^{\hat{d}}t^i\P\{t \geq 1\}$ for $i = \{0,1\}$; (b) for $0.5 \leq d < 1.5$,

$$\widetilde{K}_T^{-1}X^{\hat{d}'}X^{\hat{d}}\widetilde{K}_T^{-1} - \widetilde{K}_T^{-1}X^{d'}X^{d}\widetilde{K}_T^{-1} \xrightarrow{d} 0 \tag{A.3}$$

and

$$\widetilde{K}_T^{-1}X^{\hat{d}'}u^{\hat{d}} - \widetilde{K}_T^{-1}X^{d'}\varepsilon \xrightarrow{d} 0 \tag{A.4}$$

Note that from Iacone et al. (2013, A.22),

$$\hat{\mu}_{i,t} - \mu_{i,t} = o_p(t^{i-d}) \tag{A.5}$$

if $\hat{d} - d = O_p(T^{-\kappa})$ for some $\kappa > 0$. Consider first the case $|d| < 0.5$. We need to show that

$$T^{2d-1-i-j}(\sum_{t=1}^{T} \hat{\mu}_{i,t}\hat{\mu}_{j,t} - \sum_{t=1}^{T} \mu_{i,t}\mu_{j,t}) \xrightarrow{d} 0$$

for $i, j = \{0, 1\}$, or equivalently, that

$$T^{2d-1-i-j}[\sum_{t=1}^{T}(\hat{\mu}_{i,t} - \mu_{i,t})\mu_{j,t} + \sum_{t=1}^{T} \mu_{i,t}(\hat{\mu}_{j,t} - \mu_{j,t}) + \sum_{t=1}^{T}(\hat{\mu}_{i,t} - \mu_{i,t})(\hat{\mu}_{j,t} - \mu_{j,t})] \xrightarrow{d} 0 \quad \text{(A.6)}$$

We have

$$T^{2d-1-i-j}\sum_{t=1}^{T}(\hat{\mu}_{i,t} - \mu_{i,t})\mu_{j,t} \leq T^{2d-1-i-j}[\sum_{t=1}^{T}(\hat{\mu}_{i,t} - \mu_{i,t})^2 \sum_{t=1}^{T} \mu_{j,t}^2]^{1/2} \quad \text{(A.7)}$$

Using (A.5) and the fact that $|\mu_{i,t}| \leq Ct^{i-d}$ for $d \in (-0.5, 1.5)$ from Lemma 1 of Robinson (2005), (A.7) is $o_p(T^{2d-1-i-j+i-d+1/2+j-d+1/2}) = o_p(1)$. Using similar arguments, the other terms in (A.6) are $o_p(1)$, which establishes (A.1). For $d \in (-0.5, 1.5)$, we want to show that

$$T^{d-3/2}\sum_{t=1}^{T} \hat{\mu}_{1,t}u^{\hat{d}} - T^{d-3/2}\sum_{t=1}^{T} \mu_{1,t}\varepsilon_t \xrightarrow{d} 0$$

or, equivalently,

$$T^{d-3/2}[\sum_{t=1}^{T}(\hat{\mu}_{1,t} - \mu_{1,t})\varepsilon_t + \sum_{t=1}^{T} \mu_{1,t}(u^{\hat{d}} - \varepsilon_t) + \sum_{t=1}^{T}(\hat{\mu}_{1,t} - \mu_{1,t})(u^{\hat{d}} - \varepsilon_t)] \xrightarrow{d} 0$$

According to (A.29) and (A.30) in Iacone et al. (2013),

$$\begin{aligned} T^{d-3/2}\sum_{t=1}^{T} \mu_{1,t}(u^{\hat{d}} - \varepsilon_t) &= T^{d-3/2}\sum_{t=1}^{T} \mu_{1,t}\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d} - d)^r g^{(r)}(\varepsilon_t, 0) \\ &\quad + T^{d-3/2}\sum_{t=1}^{T} \mu_{1,t}\frac{1}{B!}(\hat{d} - d)^B g^{(B)}(\varepsilon_t, \widetilde{d} - d) \end{aligned}$$

where $|\widetilde{d} - d| \leq |\hat{d} - d|$, $g^{(r)}(\varepsilon_t, v) = \sum_{s=1}^{t-1} a_s^{(r)}(v)\varepsilon_{t-s}$ and $a_s^{(r)}(v) = \partial^r \Delta_s^{(v)}/\partial v^r$. Also $|\sum_{t=1}^{T} g^{(r)}(\varepsilon_t, 0)| = O_p(T^{1/2})$ and $|\mu_{1,t+1} - \mu_{1,t}| \leq Ct^{-d}$, hence

$$\begin{aligned} T^{d-3/2-r\kappa}\sum_{t=1}^{T} \mu_{1,t}g^{(r)}(\varepsilon_t, 0) &\leq T^{d-3/2-r\kappa}\sum_{t=1}^{T} |\mu_{1,t+1} - \mu_{1,t}||\sum_{s=t}^{T} g^{(r)}(\varepsilon_s, 0)| \\ &= o_p(T^{d-3/2-r\kappa-d+1/2+1}) = o_p(1) \end{aligned}$$

$$
\begin{aligned}
T^{d-3/2-B\kappa}\sum_{t=1}^{T}\mu_{1,t}g^{(B)}(\varepsilon_t,\widetilde{d}-d) &\leq T^{d-3/2-B\kappa}[\sum_{t=1}^{T}\mu_{1,t}^2\sum_{t=1}^{T}(g^{(B)}(\varepsilon_t,\widetilde{d}-d))^2]^{1/2} \\
&= O_p(T^{d-3/2-B\kappa+1-d+\frac{1}{2}+1}) = O_p(T^{1-B\kappa}) = o_p(1)
\end{aligned}
$$

for $B > 1/\kappa$. According to Lemma 4 of Robinson (2005),

$$
\hat{\mu}_{i,t}-\mu_{i,t} = \sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{i,t}^{(r)} + \frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{i,t}^{(B)} \tag{A.8}
$$

with $\mu_{i,t}^{(r)} = (\ln\Delta)^r\Delta^d t^i\P\{t\geq 1\}$ and $\widetilde{\mu}_{i,t}^{(B)} = (\ln\Delta)^B\Delta^{\widetilde{d}}t^i\P\{t\geq 1\}$. Hence,

$$
T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t}-\mu_{1,t})\varepsilon_t = T^{d-3/2}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{1,t}^{(r)} + T^{d-3/2}\sum_{t=1}^{T}\varepsilon_t\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{1,t}^{(B)}
$$

and

$$
T^{d-3/2-r\kappa}\sum_{t=1}^{T}\mu_{1,t}^{(r)}\varepsilon_t \leq T^{d-3/2-r\kappa}\sum_{t=1}^{T}|\mu_{1,t+1}^{(r)}-\mu_{1,t}^{(r)}||\sum_{s=t}^{T}\varepsilon_s| = o_p(1)
$$

since $|\sum_{s=t}^{T}\varepsilon_s| = O_p(T^{1/2})$ and $|\mu_{1,t+1}^{(r)}-\mu_{1,t}^{(r)}| = O((\ln T)^r t^{-d})$. Also, $T^{d-3/2-B\kappa}\sum_{t=1}^{T}\varepsilon_t\widetilde{\mu}_{1,t}^{(B)} = o_p(1)$ using similar arguments. Hence, $T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t}-\mu_{1,t})\varepsilon_t = o_p(1)$. Similarly, we can sow that $T^{d-3/2}\sum_{t=1}^{T}(\hat{\mu}_{1,t}-\mu_{1,t})(u^{\hat{d}}-\varepsilon_t) = o_p(1)$. It remains to show that for $d\in(-0.5, 0.5)$

$$
T^{d-1/2}\sum_{t=1}^{T}\hat{\mu}_{0,t}u^{\hat{d}} - T^{d-1/2}\sum_{t=1}^{T}\mu_{0,t}\varepsilon_t \xrightarrow{d} 0
$$

or, equivalently,

$$
T^{d-1/2}[\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t + \sum_{t=1}^{T}\mu_{0,t}(u^{\hat{d}}-\varepsilon_t) + \sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})(u^{\hat{d}}-\varepsilon_t)] \xrightarrow{d} 0 \tag{A.9}
$$

From (A.8),

$$
T^{d-1/2}\sum_{t=1}^{T}(\hat{\mu}_{0,t}-\mu_{0,t})\varepsilon_t = T^{d-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{B-1}\frac{1}{r!}(\hat{d}-d)^r\mu_{0,t}^{(r)} + T^{d-1/2}\sum_{t=1}^{T}\varepsilon_t\frac{1}{B!}(\hat{d}-d)^B\widetilde{\mu}_{0,t}^{(B)} = o_p(1)
$$

since,

$$
T^{d-1/2-r\kappa}\sum_{t=1}^{T}\varepsilon_t\mu_{0,t}^{(r)} \leq T^{d-1/2-r\kappa}|\mu_{0,t}^{(r)}||\sum_{s=1}^{T}\varepsilon_s| = O_p(T^{d-1/2-r\kappa-d+1/2}) = o_p(1)
$$

and $|\mu_{0,t}^{(r)}| = O((\ln t)^r(t^{-d}+(\ln t)^{-1}t^{-1}))$. Using similar arguments, the other two terms in (A.9) are also $o_p(1)$. This completes the proof of (A.1)-(A.4).

Table 1: Finite Sample Size; Pure Fractional Processes.

T	d	-0.4	0.2	0.4	0.8	1	1.4
500	ELW	0.054	0.056	0.057	0.047	0.062	0.051
	MU	0.000	0.098	0.157	0.155	0.053	0.462
	UB	0.000	0.093	0.139	0.100	0.052	0.462
1000	ELW	0.047	0.051	0.059	0.034	0.043	0.049
	MU	0.000	0.138	0.173	0.134	0.049	0.495
	UB	0.000	0.138	0.163	0.099	0.049	0.495
2000	ELW	0.054	0.047	0.044	0.051	0.043	0.039
	MU	0.000	0.178	0.277	0.108	0.046	0.559
	UB	0.000	0.178	0.277	0.097	0.046	0.559

Table 2: Finite Sample Size; AR(1) Processes with d=0.

T	AR	0	0.3	0.5	0.7	0.9	0.95
500	ELW	0.074	0.068	0.049	0.018	0.017	0.005
	MU	0.051	0.059	0.030	0.037	0.049	0.045
	UB	0.051	0.059	0.030	0.037	0.047	0.034
1000	ELW	0.085	0.068	0.057	0.008	0.017	0.005
	MU	0.057	0.031	0.046	0.046	0.042	0.053
	UB	0.057	0.031	0.046	0.046	0.042	0.051
2000	ELW	0.064	0.067	0.069	0.017	0.029	0.005
	MU	0.066	0.045	0.058	0.041	0.044	0.049
	UB	0.066	0.045	0.058	0.041	0.044	0.049

Table 3 Finite Sample Sizes; DGP 1-5 with d=0.4

	DGP-1	DGP-1	DGP-2	DGP-2	DGP-3	DGP-4	DGP-5
T	AR=0.4	AR=-0.4	MA=0.4	MA=-0.4	AR1=0.3, AR2=0.5	Measurement error	GARCH
500	0.061	0.059	0.084	0.107	0.033	0.052	0.084
1000	0.057	0.058	0.076	0.094	0.056	0.031	0.074
2000	0.048	0.065	0.08	0.069	0.045	0.052	0.069

Figure 1: Unadjusted power for pure fractional processes

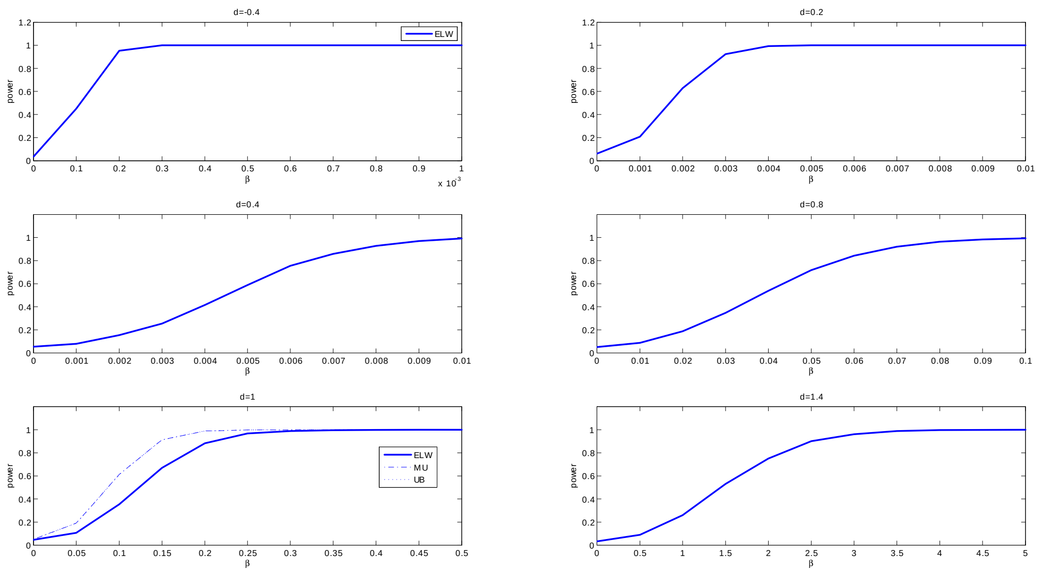

Figure 2: Unadjusted power for AR(1) processes

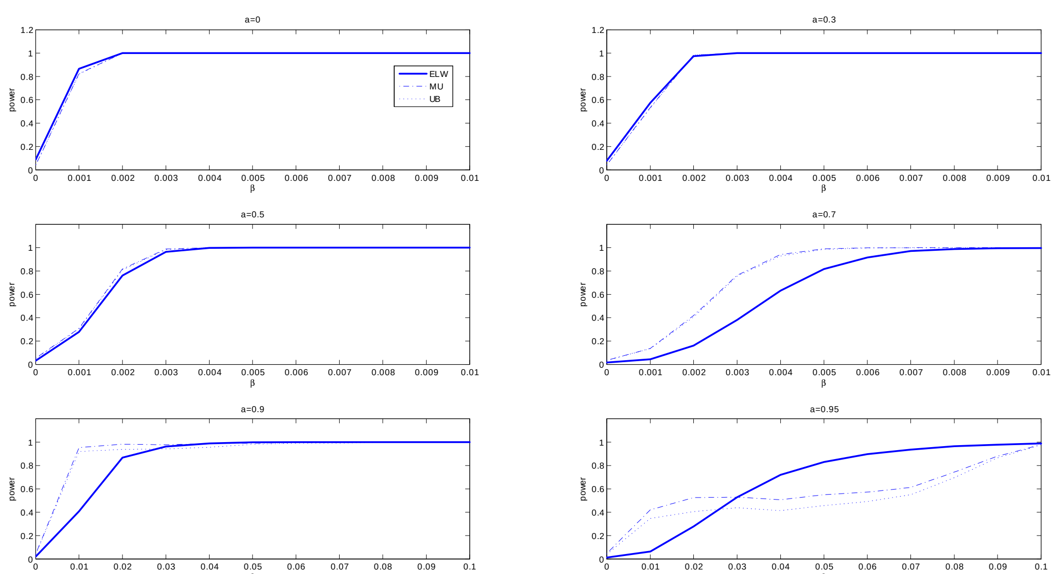

Figure 3: Unadjusted power for DGP 1-5