

Finite sample power of the Durbin-Watson test against fractionally integrated disturbances

Christian Kleiber and Walter Krämer

Fachbereich Statistik, Universität Dortmund,
D-44221 Dortmund, Germany

E-mail: {kleiber, walterk}@statistik.uni-dortmund.de

Version January 2004

Abstract

We consider the finite sample power of various tests against serial correlation in the disturbances of a linear regression when these disturbances follow a stationary long memory process. It emerges that the power depends on the form of the regressor matrix and that, for the Durbin-Watson test and many other tests that can be written as ratios of quadratic forms in the disturbances, the power can drop to zero for certain regressors. We also provide a means to detect this zero-power trap. Our results depend solely on the correlation structure and allow for fairly arbitrary nonlinearities.

JEL classification: C22.

Keywords: Durbin-Watson test, power, autocorrelation, long memory.

1 Introduction and Summary

We consider the standard linear regression model

$$y = X\beta + u,$$

where y is $T \times 1$, X is $T \times k$ (nonstochastic, of rank $k < T$), β is a $k \times 1$ vector of regression coefficients to be estimated, and u is a $T \times 1$ disturbance vector whose components follow a

stationary fractionally integrated ARMA process (ARFIMA(p, d, q))

$$\phi(B)(1 - B)^d u_t = \theta(B)\varepsilon_t, \quad (1)$$

where $\phi(B)$ and $\theta(B)$ are polynomials in the backshift operator B of orders p and q , respectively, and $\{\varepsilon_t\}$ is a weak white noise sequence. This process is stationary and causal if and only if $d < 0.5$ and the AR polynomial $\phi(z)$ satisfies the usual stationarity and causality conditions for autoregressive models. Long memory corresponds to $d > 0$.

The present paper is concerned with the exact finite sample power of various autocorrelation tests against fractional alternatives, extending Krämer (1985), Zeisel (1989), Krämer and Zeisel (1990), Bartels (1992) and Löbus and Ritter (2000), all of whom confine themselves to the classical case of AR(1) alternatives. For concreteness, we focus on the Durbin-Watson test, one of the most intensely studied statistics in all of econometrics. Although it was originally designed as a test against AR(1) disturbances, it is well-known to be (approximately) locally optimal against a wide range of short-memory and spatial alternatives (Kariya, 1988, King and Evans, 1988). More recently, the properties of a modification of the Durbin-Watson test, due to Nabeya and Tanaka (1990), as a unit root test against short-range dependent alternatives have been studied by Hisamatsu and Maekawa (1994), and against long-range dependent alternatives by Tsay (1998). Nakamura and Tanaguchi (1999) investigate the asymptotics of a standardized Durbin-Watson statistic as a test for independence against fractionally integrated alternatives.

The test statistic is

$$DW = u^\top Q_1 u / u^\top Q_2 u,$$

where $Q_1 = M_X A M_X$, $Q_2 = M_X = I - X(X^\top X)^{-1}X^\top$, and A is the Toeplitz-like matrix

$$A = \begin{bmatrix} 1 & -1 & \cdots & 0 & 0 \\ -1 & 2 & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & 2 & -1 \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

More generally, our results carry over to all autocorrelation tests which are ratios of quadratic forms in the disturbances. Examples of such tests are briefly discussed in section 4. All of them were designed for particular short-memory alternatives, but short memory is not distinguishable from long memory in finite samples, and therefore it is of interest to know about the power of such tests when the disturbances are generated by (1). In particular, one might expect that, in the case of a strongly dependent process of the fractionally integrated type, an autocorrelation test will easily detect dependence, perhaps more easily than in the classical AR(1) case. However, this is not true; in fact, the power of the test can be arbitrarily low in the vicinity of 0.5, the boundary of the stationarity region. Specifically, we shall present regressor matrices X for which, as $d \rightarrow \frac{1}{2}$, (i) the power approaches one, (ii) the power approaches zero, and (iii) the power approaches a constant $C \in (0, 1)$.

For our main results in section 2, we only require the innovations $\{\varepsilon_t\}$ in (1) to be weak white noise, hence we can allow for fairly arbitrary nonlinearities of e.g. the GARCH type. This is in contrast to the more common large-sample asymptotics in the long-memory area, where $\{u_t\}$ is usually assumed to be a linear process. Our exact finite sample power computations in section 3 also require the innovations ε_t to be normal i.i.d.

2 High correlation asymptotics

Long memory is a concept of strong positive dependence, hence we shall focus on the one-sided Durbin-Watson test against positive autocorrelation which rejects for small values of DW . The rejection probability can then be written as

$$\begin{aligned} P\left(u^\top(Q_1 - c_\alpha Q_2)u < 0\right) &= P\left(\eta^\top R^{\frac{1}{2}}(Q_1 - c_\alpha Q_2)R^{\frac{1}{2}}\eta < 0\right) \\ &= P\left(\sum_{t=1}^T \lambda_t \eta_t^2 < 0\right), \end{aligned} \quad (2)$$

where $R \equiv R(d) = \frac{1}{\sigma_u^2} E(uu^\top)$, with entries $r_{t,t+h}$, $0 \leq |h| < T - t$, is the correlation matrix of the disturbances, $\eta = (\eta_1, \dots, \eta_T)^\top$ is a vector of independent $N(0, 1)$ -variables, c_α is the critical value corresponding to the significance level of the test, and the λ_t are the eigenvalues of $\Lambda = R^{\frac{1}{2}}(Q_1 - c_\alpha Q_2)R^{\frac{1}{2}}$.

Given the correlation matrix R , the probability (2) can be evaluated numerically. We give some examples in Section 3 below. Our main concern is to identify situations where the power of the Durbin-Watson test is exceptionally high or low. Both extremes are attained, depending on the regression matrix X , when the long memory parameter d from (1) approaches the boundary of the stationarity region, i.e. as $d \rightarrow \frac{1}{2}$. To see this, note that the λ_t are also the eigenvalues of

$$R^{\frac{1}{2}} \Lambda R^{-\frac{1}{2}} = R(Q_1 - c_\alpha Q_2),$$

and that, independently of $\phi(B)$ and $\theta(B)$,

$$\bar{R} := \lim_{d \rightarrow \frac{1}{2}} R(d) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} = ee^\top, \quad (3)$$

where $e = (1, 1, \dots, 1)^\top$. For the special case of fractionally integrated white noise, i.e. $\phi(B) \equiv \theta(B) \equiv 1$, this is a direct consequence of the representation

$$r_{t,t+h} = \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)} = \frac{(d)_h}{(1-d)_h}, \quad (4)$$

where $\Gamma(\cdot)$ is the gamma function and $(d)_h = d(d+1)\dots(d+h-1)$ is Pochhammer's symbol for the forward factorial function. For a general ARFIMA(p, d, q) process the same limiting result is valid, although the derivation is somewhat more cumbersome (Kleiber, 2001).

From (3) it follows that at most one of the eigenvalues of $\bar{R}(Q_1 - c_\alpha Q_2)$ is different from zero. Consider first the case where there is exactly one non-zero eigenvalue $\bar{\lambda}$. This is equivalent to

$$\bar{R}(Q_1 - c_\alpha Q_2) \neq 0,$$

which in turn requires that e is not an element of the column space of the regressor matrix X . Hence, $\bar{\lambda} \neq 0$ will not occur when the regression has an intercept. Given $\bar{\lambda} \neq 0$, the analysis of the limiting power (as $d \rightarrow \frac{1}{2}$) of the Durbin-Watson test is straightforward. If $\bar{\lambda} > 0$, the power drops to zero; if $\bar{\lambda} < 0$, the power tends to one. Note that this effect does not depend on the sample size, i.e. for any sample size T there are design matrices X such that the power of the

Durbin-Watson test is arbitrarily low for d in the vicinity of 0.5. We demonstrate this analytically at the end of this section, section 3 provides numerical examples.

The analysis is slightly more involved when $\bar{R}(Q_1 - c_\alpha Q_2) = 0$. We can then without loss of generality replace the λ_t in (2) by

$$\tilde{\lambda}_t = \frac{\lambda_t}{\frac{1}{2} - d},$$

since this does not affect the rejection probability. The $\tilde{\lambda}_t$ are the eigenvalues of $(\frac{1}{2} - d)^{-1}R(Q_1 - c_\alpha Q_2)$. Now, by l'Hopital's rule,

$$\begin{aligned} \lim_{d \rightarrow \frac{1}{2}} R(Q_1 - c_\alpha Q_2) &= \lim_{d \rightarrow \frac{1}{2}} \frac{1}{\frac{1}{2} - d} R(Q_1 - c_\alpha Q_2) \\ &= \lim_{d \rightarrow \frac{1}{2}} \frac{1}{\frac{1}{2} - d} (R - \bar{R})(Q_1 - c_\alpha Q_2) \\ &=: \lim_{d \rightarrow \frac{1}{2}} W(Q_1 - c_\alpha Q_2), \end{aligned}$$

where

$$W := \frac{1}{\frac{1}{2} - d} (R - \bar{R}).$$

For the remainder of this section we confine ourselves to the case $\phi(B) \equiv \theta(B) \equiv 1$, i.e. fractionally integrated white noise. We therefore need the W matrix for an ARFIMA $(0, d, 0)$ process, with correlations given by (4). In view of

$$(d)'_h = (d)_h \left(\sum_{j=1}^h \frac{1}{d+j-1} \right),$$

the typical entry of $W = (w_{s,t}(d))$ equals $w_h(d)$, $h := |t - s|$, where

$$\begin{aligned} w_h(d) &= - \left(\frac{(d)_h}{(1-d)_h} \right)' \\ &= - \frac{(d)_h \left(\sum_{j=1}^h \frac{1}{d+j-1} \right) + (d)_h \left(\sum_{j=1}^h \frac{1}{-d+j} \right)}{(1-d)_h} \\ &\xrightarrow{d \rightarrow \frac{1}{2}} - \frac{2 \left(\frac{1}{2} \right)_h \sum_{j=1}^h \frac{1}{j - \frac{1}{2}}}{\left(\frac{1}{2} \right)_h} \\ &= -2 \sum_{j=1}^h \frac{1}{j - \frac{1}{2}}. \end{aligned}$$

For instance, $w_1(\frac{1}{2}) = -4$, $w_2(\frac{1}{2}) = -16/3$, $w_3(\frac{1}{2}) = -92/15$, etc. Given $\lim_{d \rightarrow \frac{1}{2}} w_h(d)$, it is now possible to numerically compute the limiting rejection probability of the Durbin-Watson test for any regressor matrix X , section 3 provides some examples.

We stress that our results hold true for arbitrary weak white noise sequences $\{\varepsilon_t\}$, allowing, in particular, for nonlinearities of e.g. the GARCH type. We consider this to be an advantage over the more common large-sample approach in the long-memory area, for which asymptotics with nonlinear innovation sequences $\{\varepsilon_t\}$ are still in the development stage.

To conclude this section, we show that there exist regressors for which the power drops to zero irrespective of the sample size. As noted above, this will only occur in regressions without an intercept, for which the limiting eigenvalues of $R(Q_1 - c_\alpha Q_2)$ are zero except for one. The only non-zero eigenvalue may be computed via

$$\text{tr}(ee^\top(Q_1 - c_\alpha Q_2)) = e^\top(Q_1 - c_\alpha Q_2)e.$$

Let $k = 1$ and consider the regressor $x = (x_1, \dots, x_T)^\top$, where

$$x_t = \begin{cases} 0, & \text{for some } s \in \{2, \dots, T-1\}, \\ 1, & t \neq s. \end{cases}$$

Then $M_X e = (0, \dots, 0, 1, 0, \dots, 0)^\top$, where “1” occurs in the s th position, and $e^\top M_X e = 1$. Also, in view of $a_{ss} = 2$, for $s \in \{2, \dots, T-1\}$, $e^\top M_X A M_X e = 2$ and therefore

$$e^\top(Q_1 - c_\alpha Q_2)e = 2 - c_\alpha.$$

The latter expression is clearly positive, implying that the limiting power of the Durbin-Watson test equals zero for this regressor.

This example also shows how to avoid the zero-power trap: just compute $e^\top(Q_1 - c_\alpha Q_2)e$ and do not apply the test if this is positive.

3 Some numerical examples

In order to illustrate our results we consider two settings:

Model 1: (linear time trend)

$$y_t = \beta_1 + \beta_2 \cdot t + u_t, \quad t = 1, \dots, T.$$

Model 2: (alternating regressor)

$$y_t = \beta_1 + \beta_2 \cdot \{1 + (-1)^t\} + u_t, \quad t = 1, \dots, T.$$

In all examples, the significance level is 5%, and the power of the Durbin-Watson test is computed for an ARFIMA $(0, d, 0)$ process (4) with innovations $\varepsilon_t \sim \text{nid}(0, 1)$ for $d \in [0, 0.5]$, evaluated in steps of 0.01 using Imhof's (1961) method.

Figure 1 depicts the power in Model 1 for both a regression with and without a constant term for samples of sizes $T \in \{20, 40, 60, 80, 100\}$. In the regression without a constant (left panel), the power approaches one as $d \rightarrow \frac{1}{2}$. For the regression with a constant (right panel) the limiting power (as $d \rightarrow \frac{1}{2}$) is given in Table 1. It lies strictly between zero and one, approaching one as the sample size increases.

Although the Durbin-Watson test can have maximal power as $d \rightarrow \frac{1}{2}$, it is not (approximately) locally best invariant in the long-range dependent setting. From King and Evans (1988, p. 511), a necessary condition for this is that the derivative of the correlation matrix R is zero at lag 1 for $d = 0$, which is easily seen to be not the case for ARFIMA $(0, d, 0)$ disturbances.

Quite a different picture emerges for Model 2. Figure 2 shows that, in the regression without an intercept (left panel), the power approaches zero for $d \rightarrow \frac{1}{2}$, showing that the Durbin-Watson test is, in general, biased when the disturbances follow a strongly dependent process of the fractionally integrated type. For the regression with a constant term the power is considerably improved, approaching one as $d \rightarrow \frac{1}{2}$ with increasing sample size (see Table 1).

The left panel of figure 2 illustrates that a zero limiting power of the Durbin-Watson test is not an artifact of small samples. Although power increases monotonically for any $d \in (0, \frac{1}{2})$ as T increases, it still drops to zero as $d \rightarrow \frac{1}{2}$ irrespective of sample size.

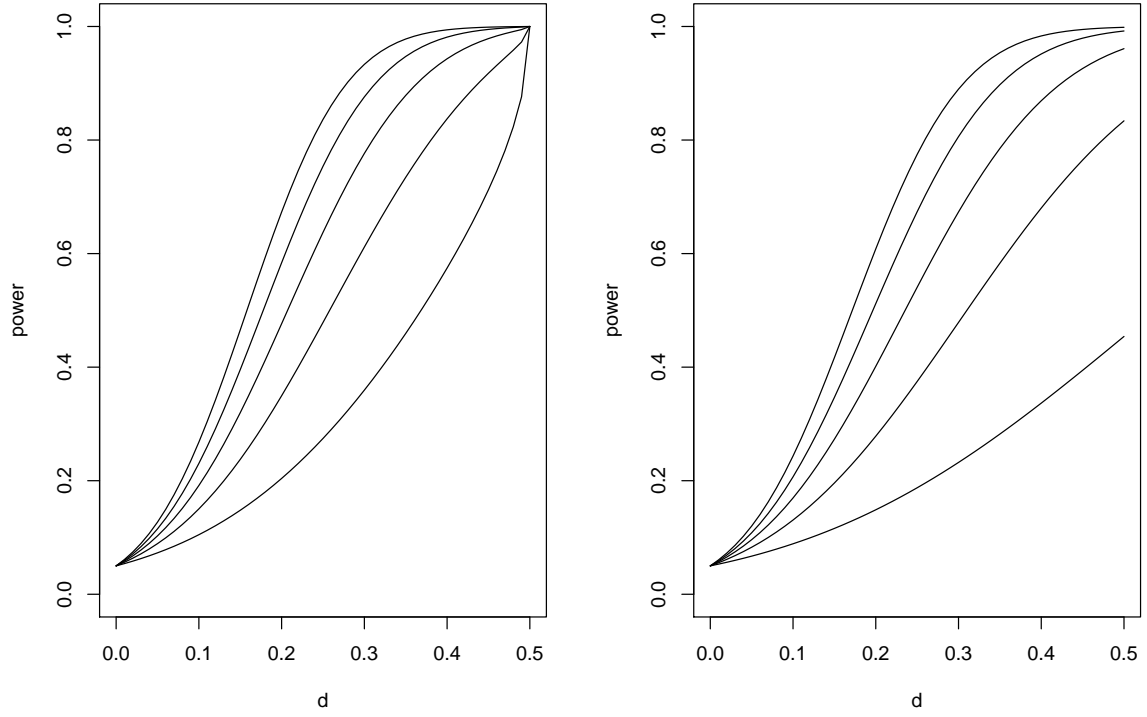


Figure 1: Power of the Durbin-Watson test, $\alpha = 0.05$, regressor $x_t = t$. Left panel: regression without intercept, right panel: regression with intercept. Sample sizes are 20, 40, 60, 80, and 100 (bottom to top).

Model	$T = 20$	$T = 40$	$T = 60$	$T = 80$	$T = 100$
$x_t = t$	0.4541	0.8338	0.9610	0.9920	0.9985
$x_t = 1 + (-1)^t$	0.5841	0.9020	0.9808	0.9966	0.9994

Table 1: Limiting power (as $d \rightarrow \frac{1}{2}$) of the Durbin-Watson test, regression with an intercept

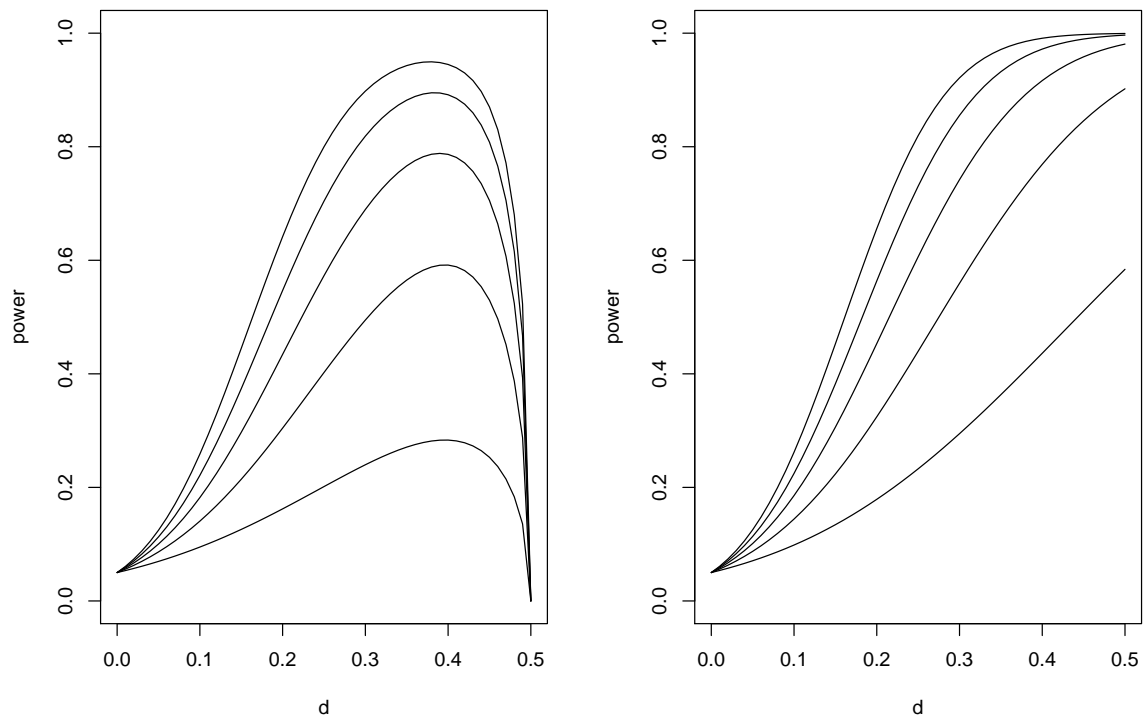


Figure 2: Power of the Durbin-Watson test, $\alpha = 0.05$, regressor $x_t = 1 + (-1)^t$. Left panel: regression without intercept, right panel: regression with intercept. Sample sizes are 20, 40, 60, 80, and 100 (bottom to top).

4 Extensions

The argument above generalizes in a straightforward fashion to all tests that are ratios of quadratic forms in u . These include (i) King's (1981) alternative Durbin-Watson test, where $Q_1 = M_X A_0 M_X$, and where A_0 equals A with top left and bottom right elements being 2 instead of 1, (ii) the Berenblut-Webb test (1973), where $Q_1 = B - BX(X^\top BX)^{-1}X^\top B$, $Q_2 = M_X$, and B equals A with top-left element being 2 instead of 1, (iii) King's (1985) point-optimal test, where Q_1 depends on a particular alternative, and (iv) various tests based on LUS residuals, as described in Krämer and Zeisel (1990).

Our results also extend to models of stationary long-range dependence that are not members of the ARFIMA class. These include fractional Gaussian noise, defined via first differences of fractional Brownian motion, with autocorrelation function

$$r_{t,t+h} = \frac{1}{2} \left[(h+1)^{2H} - 2h^{2H} + (h-1)^{2H} \right], \quad \frac{1}{2} < H < 1, \quad (5)$$

where H is the Hurst coefficient which is related to the memory parameter d of the ARFIMA model via $d = H - \frac{1}{2}$. As pointed out in Kleiber (2001), the correlation matrix for fractional Gaussian noise also tends to $\bar{R} = ee^\top$ as $H \rightarrow 1$, i.e., as $d \rightarrow \frac{1}{2}$.

Another example is the ‘‘Cauchy family’’ of long-memory models, proposed by Gneiting (2000). It is defined directly in terms of the autocorrelation function

$$r_{t,t+h} = \left(1 + \left| \frac{h}{c} \right|^\alpha \right)^{-(\beta/\alpha)}, \quad c > 0, \quad 0 < \alpha \leq 2, \quad \beta > 0. \quad (6)$$

If $\beta < 1$, this model exhibits long-range dependence with a Hurst coefficient $H = 1 - \beta/2$. It follows that $\bar{R} = ee^\top$ is the limiting correlation matrix as $\beta \rightarrow 0$.

Acknowledgements

Research supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 475. All computations were carried out in the R statistical software package (<http://www.r-project.org/>), Version 1.8.1.

References

- Bartels, R. (1992). On the power function of the Durbin-Watson test. *Journal of Econometrics*, 51, 101-12.
- Berenblut, I.I., and Webb, G.I. (1973). A new test for autocorrelated errors in the linear regression model. *Journal of the Royal Statistical Society*, B 35, 33-50.
- Gneiting, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *Journal of Applied Probability*, 37, 1104-1109.
- Hisamatsu, H., and Maekawa, K. (1994). The distribution of the Durbin-Watson statistic in integrated and near-integrated models. *Journal of Econometrics*, 61, 367-382.
- Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419-426.
- Kariya, T. (1988). The class of models for which the Durbin-Watson test is locally optimal. *International Economic Review*, 29, 167-175.
- King, M.L. (1981). The alternative Durbin-Watson test. *Journal of Econometrics*, 17, 51-66.
- King, M.L. (1985). A point optimal test against autoregressive disturbances. *Journal of Econometrics*, 27, 21-37.
- King, M.L., and Evans, M.A. (1988). Locally optimal properties of the Durbin-Watson test. *Econometric Theory*, 4, 509-516.
- Kleiber, C. (2001). Finite sample efficiency of OLS in linear regression models with long-memory disturbances. *Economics Letters*, 72, 131-136.
- Krämer, W. (1985). The power of the Durbin-Watson test for regressions without an intercept. *Journal of Econometrics*, 28, 363-370.
- Krämer, W., and Zeisel, H. (1990). Finite sample power of linear regression autocorrelation tests. *Journal of Econometrics*, 43, 363-372.
- Löbus, J.-U., and Ritter, L. (2000). The limiting power of the Durbin-Watson test. *Communications in Statistics—Theory and Methods*, 29, 2665-2676.
- Nabeya, S., and Tanaka, K. (1990). Limiting power of unit-root tests in time series regression. *Journal of Econometrics*, 46, 247-271.
- Nakamura, S., and Taniguchi, M. (1999). Asymptotic theory for the Durbin-Watson statistic under long-memory dependence. *Econometric Theory*, 15, 847-866.
- Tsay, W.-J. (1998). On the power of the Durbin-Watson statistic against fractionally integrated processes. *Econometric Reviews*, 17, 361-386.

Zeisel, H. (1989). On the power of the Durbin-Watson test under high autocorrelation. *Communications in Statistics—Theory and Methods*, 18, 3907-3916.