

Vougas, Dimitrios V.

**Article**

## Prais-Winsten algorithm for regression with second or higher order autoregressive errors

Econometrics

**Provided in Cooperation with:**

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Vougas, Dimitrios V. (2021) : Prais-Winsten algorithm for regression with second or higher order autoregressive errors, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 9, Iss. 3, pp. 1-6,  
<https://doi.org/10.3390/econometrics9030032>

This Version is available at:

<https://hdl.handle.net/10419/247622>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

Communication

# Prais–Winsten Algorithm for Regression with Second or Higher Order Autoregressive Errors

Dimitrios V. Vougas <sup>†</sup>

Swansea University, Swansea SA2 8PP, UK; d.v.vougas@swan.ac.uk or d.vougas@outlook.com

<sup>†</sup> Retired.

**Abstract:** There is no available Prais–Winsten algorithm for regression with AR(2) or higher order errors, and the one with AR(1) errors is not fully justified or is implemented incorrectly (thus being inefficient). This paper addresses both issues, providing an accurate, computationally fast, and inexpensive generic zig-zag algorithm.

**Keywords:** Prais–Winsten algorithm; fast zig-zag algorithm; linear regression; autoregressive errors; exact NLS; unconditional ML

## 1. Introduction

Both simulation and theoretical evidence show that the estimation of a regression with an autoregressive of order one (AR(1)) errors via exact (conditional) maximum likelihood (ML) is inferior to an exact nonlinear least squares (NLS) estimation (the correctly implemented Prais and Winsten (1954) (PW) method), at least for trending data. Park and Mitchell (1980) show via simulation that, for trending regressors, the PW method is more efficient than the exact/conditional ML (the Beach and MacKinnon 1978a) (BM) method). This is theoretically confirmed by Magge 1985; Magee 1989 who approximates biases of various two-step and iterative estimators. Correctly implemented, the PW method delivers exact NLS (equivalent to unconditional ML under normality) estimators. No normality requirement is needed for the PW method. Gurland (1954) formalises and resolves criticism against the estimation method of Cochrane and Orcutt (1949) (CO, hereafter) for losing the first observation,<sup>1</sup> which is employed by PW or re-proposed by Kadiyala (1968). Koopmans (1942) was the first to examine the stationary AR(1) correctly. The PW algorithm is a fast zig-zag (computationally inexpensive) algorithm, which retains the first observation and requires no numerical optimiser. However, in order to obtain an (econometrically efficient) exact NLS estimator, the correct (exact) closed form formula must be employed in the iterations for the update of the AR parameter. This is not the case for all available implementations of the PW method (see Magee (1989)). For example, (Judge et al. 1985, chp. 8) propose using the autocorrelation coefficient formula for the update, but this is not correct. Exact/conditional ML fast zig-zag algorithms are provided by (Beach and MacKinnon 1978a, 1978b) for AR(1)<sup>2</sup> and AR(2)<sup>3</sup> errors, respectively. None is available for higher than two AR orders. The PW algorithm for a regression with AR(1) errors is relatively well known, although usually incorrectly implemented. This is because an incorrect closed form estimator is used for the update. Park and Mitchell (1980) implement the PW algorithm correctly, although they do not give all the details. Furthermore, there is no zig-zag PW algorithm for a regression with AR(2) or higher order errors, and this literature gap is filled in this paper. Reliable standard errors are also calculated that are not affected by the presence of the lagged dependent variable as a regressor. In fact, our proposed method corrects the inefficient CO method and poor estimation via bad implementation of the Gauss–Newton algorithm.



**Citation:** Vougas, Dimitrios V. 2021. Prais–Winsten Algorithm for Regression with Second or Higher Order Autoregressive Errors. *Econometrics* 9: 32. <https://doi.org/10.3390/econometrics9030032>

Academic Editor: Marc S. Paoletta

Received: 9 June 2021

Accepted: 23 August 2021

Published: 27 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The paper is organised as follows: Section 2 fully discusses the closed form exact NLS estimation of pure AR(1), AR(2), and AR( $p$ ) coefficients, while Section 3 examines the iterative exact joint estimation of regression and autoregressive coefficients. Finally, Section 4 concludes.

## 2. Closed Form Exact NLS: AR(1,2,...,p)

The main ingredient for the correct PW algorithm is the correct<sup>4</sup> closed form update(s) for the autoregressive parameter(s), along with exact generalised least squares (GLS) iterations. Let (observed)  $y_t$  be generated via the stationary AR(1)

$$y_t = \theta y_{t-1} + v_t, \quad t = 2, \dots, n, \quad (1)$$

with

$$y_1 = \frac{v_1}{\sqrt{1 - \theta^2}}, \quad (2)$$

$v_t \sim iid(0, \sigma_v^2)$ ,  $t = 1, \dots, n$ . This assumption can be relaxed, and  $v_t$  can be a martingale difference sequence with finite fourth moment (see (Stock 1991), for example). Assuming normal  $v_t$ , the exact/conditional ML estimator of (1), conditional on (2), in closed form, is the solution to their cubic equation that BM provides. PW/exact NLS (eventually in closed form) requires the exact sum of squares to be minimised, that is,

$$\min_{\theta} \{y_1^2(1 - \theta^2) + \sum_2^n (y_t - \theta y_{t-1})^2\}, \quad (3)$$

see (Kmenta 1986, p. 319 (full proof in the present paper)). Here, we make use of  $\sigma_y/\sigma_v = \sqrt{1 - \theta^2}$ . Phillips (1977,1978) and Sawa (1978) (among others) employ an inexact version of (3), ignoring  $y_1^2(1 - \theta^2)$ . This results in the so-called LS (also incorrectly called the ML) estimator ( $\hat{\theta}_{LS}$ ), which is more biased and less efficient than the exact NLS estimator of (3) (also called the (original) PW estimator by Park and Mitchell (1980); PW2 in (Magee 1989, p. 662)). Minimisation is via

$$-2\theta y_1^2 - 2 \sum_2^n (y_t - \theta y_{t-1})y_{t-1} = 0, \quad (4)$$

which results in the closed form of the (genuine) PW estimator<sup>5</sup>

$$\hat{\theta}_{PW} = \frac{\sum_2^n y_t y_{t-1}}{\sum_3^n y_{t-1}^2}, \quad (5)$$

which is better than the OLS estimator  $\hat{\theta}_{LS} = \sum_2^n y_t y_{t-1} / \sum_2^n y_{t-1}^2$ .<sup>6</sup> An even worse estimator is the autocorrelation coefficient  $\sum_2^n y_t y_{t-1} / \sum_1^n y_t^2$ .<sup>7</sup>

For the stationary AR(2) model

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + v_t, \quad t = 3, \dots, n, \quad (6)$$

$$y_1 = v_1 / (\sqrt{1 - \theta_2^2} \sqrt{1 - \rho_1^2}), \quad y_2 = \rho_1 y_1 + v_2 / \sqrt{1 - \theta_2^2}, \quad \rho_1 = \theta_1 / (1 - \theta_2), \quad (7)$$

also  $\rho_2 = \theta_1 \rho_1 + \theta_2$ ,<sup>8</sup> no closed-form exact NLS/PW estimators of the autoregressive parameters are available in the literature. To derive them, the exact sum of squares to be minimised is

$$\min_{\theta_1, \theta_2} \{(1 - \theta_2^2)(1 - \rho_1^2)y_1^2 + (1 - \theta_2^2)(y_2 - \rho_1 y_1)^2 + \sum_3^n (y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2})^2\}. \quad (8)$$

Use has been made of the fact that in this case,  $\sigma_y/\sigma_v = \sqrt{1 - \theta_2^2} \sqrt{1 - \rho_1^2}$ . The required canonical equations (with the help of MATHEMATICA™ and imposing  $\rho_1 = \theta_1/(1 - \theta_2)$ ) are

$$\begin{aligned} (1 + \theta_2)y_1y_2 + \sum_3^n y_t y_{t-1} - \theta_1 \sum_3^n y_{t-1}^2 - \theta_2 \sum_3^n y_{t-1}y_{t-2} &= 0 \\ \theta_1 y_1y_2 + \theta_2(y_1^2 + y_2^2) + \sum_3^n y_t y_{t-2} - \theta_1 \sum_3^n y_{t-1}y_{t-2} - \theta_2 \sum_3^n y_{t-2}^2 &= 0 \end{aligned} \tag{9}$$

that become (after manipulations)

$$\begin{aligned} (y_1y_2 + \sum_3^n y_t y_{t-1}) - \theta_1 \sum_3^n y_{t-1}^2 - \theta_2(-y_1y_2 + \sum_3^n y_{t-1}y_{t-2}) &= 0 \\ \sum_3^n y_t y_{t-2} - \theta_1(-y_1y_2 + \sum_3^n y_{t-1}y_{t-2}) - \theta_2(-y_1^2 - y_2^2 + \sum_3^n y_{t-2}^2) &= 0. \end{aligned} \tag{10}$$

Re-arranging (10), the system of equations can be solved for the brand new (efficient/genuine) PW estimators of the AR(2) parameters. That is,

$$\begin{bmatrix} \hat{\theta}_{1PW} \\ \hat{\theta}_{2PW} \end{bmatrix} = \begin{bmatrix} \sum_3^n y_{t-1}^2 & \sum_4^n y_{t-1}y_{t-2} \\ \sum_4^n y_{t-1}y_{t-2} & \sum_5^n y_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_2^n y_t y_{t-1} \\ \sum_3^n y_t y_{t-2} \end{bmatrix}. \tag{11}$$

In view of (11), we can guess the brand new closed-form PW autoregressive estimators vector of any AR(p)

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + v_t, t = p + 1, \dots, n,$$

along with the required equations for observations 1, ..., p (not stated as they are difficult). That is,

$$\begin{bmatrix} \hat{\theta}_{1PW} \\ \hat{\theta}_{2PW} \\ \vdots \\ \hat{\theta}_{p-1PW} \\ \hat{\theta}_{pPW} \end{bmatrix} = \begin{bmatrix} \sum_3^n y_{t-1}^2 & \sum_4^n y_{t-1}y_{t-2} & \dots & \sum_{1+p}^n y_{t-1}y_{t-p+1} & \sum_{2+p}^n y_{t-1}y_{t-p} \\ \sum_4^n y_{t-1}y_{t-2} & \sum_5^n y_{t-2}^2 & \dots & \sum_{2+p}^n y_{t-2}y_{t-p+1} & \sum_{3+p}^n y_{t-2}y_{t-p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{1+p}^n y_{t-1}y_{t-p+1} & \sum_{2+p}^n y_{t-2}y_{t-p+1} & \dots & \sum_{2p-1}^n y_{t-p+1}^2 & \sum_{2p}^n y_{t-p+1}y_{t-p} \\ \sum_{2+p}^n y_{t-1}y_{t-p} & \sum_{3+p}^n y_{t-2}y_{t-p} & \dots & \sum_{2p}^n y_{t-p+1}y_{t-p} & \sum_{2p+1}^n y_{t-p}^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_2^n y_t y_{t-1} \\ \sum_3^n y_t y_{t-2} \\ \vdots \\ \sum_p^n y_t y_{t-p+1} \\ \sum_{p+1}^n y_t y_{t-p} \end{bmatrix}.$$

This expression is easily programmable in a matrix programming language.

### 3. Iteration

The closed-form estimators of the previous section are to be used in iterations for updating. Let observed  $Y_t$  be generated via

$$Y_t = x_t' \beta + y_t, \tag{12}$$

with (unobserved)  $y_t$  following (1) and (2);  $x_t$  is a  $k \times 1$  vector of regressors (the  $t$ -th row of regressor matrix  $X$ ), and  $\beta$  is the corresponding coefficient vector. The PW iterative algorithm has the following steps: (I) Apply OLS to (12) and derive residual  $\hat{y}_t$  and  $\hat{\theta}_{PW}$  from (5), replacing  $y_t$  with  $\hat{y}_t$ . (II) Re-estimate  $\beta$  from

$$Y_1 \sqrt{1 - \hat{\theta}_{PW}^2} = x_1' \beta \sqrt{1 - \hat{\theta}_{PW}^2} + v_1 \tag{13}$$

$$Y_t - \hat{\theta}_{PW}^2 Y_{t-1} = (x_t - \hat{\theta}_{PW}^2 x_{t-1})' \beta + v_t, t = 2, \dots, n. \tag{14}$$

For a PW algorithm that delivers exact NLS estimators, the first observation must be employed (see (13)), and the  $\theta$  update must be via (5). According to (Magee 1989, p. 663), (5) guarantees convergence. (III) Use the new estimate of  $\beta$  and obtain new residual  $\hat{y}_t$  from (12) and new estimate of  $\theta$ , and proceed to step (II) if convergence criterion is not met. In addition, estimates of  $\theta$  must be forced to be in  $(-1, 1)$ .  $P_1(\theta)$  is the exact GLS matrix of first order with the main diagonal  $\sqrt{1 - \theta^2}, 1, \dots, 1$ , first sub-diagonal  $-\theta, \dots, -\theta$ , and zero elsewhere. Let  $\tilde{v}$  be the converged innovation residual vector from (13) and (14), corresponding to converged  $\hat{\theta}_{PW}$ ; we define the innovation variance estimate  $\tilde{\sigma}_v^2 = \tilde{v}'\tilde{v}/(n - k)$ , and  $cov_{\tilde{\beta}} = \tilde{\sigma}_v^2 (X' P_1(\hat{\theta}_{PW})' P_1(\hat{\theta}_{PW}) X)^{-1}$  for the converged estimate of  $\beta$ ,  $\tilde{\beta}$ . For  $\hat{\theta}_{PW}$ , we rely on the fact that exact NLS estimation is identical to unconditional ML under normality and calculate the quasi-ML covariance  $cov_{\hat{\theta}_{PW}}$  as the inverse of minus the Hessian of the concentrated unconditional loglikelihood  $L(\theta) = -(n/2) \ln(Y^* - X^* \hat{\beta}_{GLS,\theta})' (Y^* - X^* \hat{\beta}_{GLS,\theta})$  evaluated at  $\hat{\theta}_{PW}$ , with  $\hat{\beta}_{GLS,\theta} = (X^* X^*)^{-1} (X^* Y^*)$ ,  $Y^* = P_1(\theta) Y$  and  $X^* = P_1(\theta) X$ , and the MLE innovation variance estimate  $\hat{\sigma}_{vMLE,\theta}^2 = \hat{v}(\theta)' \hat{v}(\theta) / n$ , with “residual”  $\hat{v}(\theta)$  for a given  $\theta$ ,  $\hat{v}(\theta) = Y - X \hat{\beta}_{GLS,\theta}$ . Alternatively, we may use the asymptotic covariance for  $\tilde{\theta}_{PW}$ , requiring no normality (see below). A third option could be to calculate the sandwich covariance. Note that all these covariances are not affected by the presence of the lagged dependent variable as a regressor.

Similarly, when  $y_t$  in (12) follows the AR(2) model, the genuine PW iterative algorithm has the following steps: (I) Apply OLS to (12) and derive residual  $\hat{y}_t$  and  $\hat{\theta}_{1PW}$  and  $\hat{\theta}_{2PW}$  from (11), replacing  $y_t$  with  $\hat{y}_t$ . (II) Re-estimate  $\beta$  from

$$Y_1 \sqrt{1 - \hat{\rho}_1^2} \sqrt{1 - \hat{\theta}_{2PW}^2} = x_1' \beta \sqrt{1 - \hat{\rho}_1^2} \sqrt{1 - \hat{\theta}_{2PW}^2} + v_1 \tag{15}$$

$$(Y_2 - \hat{\rho}_1 Y_1) \sqrt{1 - \hat{\theta}_{2PW}^2} = (x_2 - \hat{\rho}_1 x_1)' \beta \sqrt{1 - \hat{\theta}_{2PW}^2} + v_2 \tag{16}$$

$$Y_t - \hat{\theta}_{1PW} Y_{t-1} - \hat{\theta}_{2PW} Y_{t-2} = (x_t - \hat{\theta}_{1PW} x_{t-1} - \hat{\theta}_{2PW} x_{t-2})' \beta + v_t, t = 3, \dots, n. \tag{17}$$

Note that  $\hat{\rho}_1 = \hat{\theta}_{1PW} / (1 - \hat{\theta}_{2PW})$ . (In addition, estimates of  $\theta_1$  and  $\theta_2$  must be forced to be inside the stationarity triangle.) (III) Use the new estimator of  $\beta$  and obtain new residual  $\hat{y}_t$  from (12) and new estimators of  $\theta_1$  and  $\theta_2$  from (11), and proceed to step (II) if convergence criterion is not met.  $P_2(\theta_1, \theta_2)$  is the exact GLS matrix of second order with the main diagonal  $\sqrt{1 - \theta_2^2} \sqrt{1 - \rho_1^2}, \sqrt{1 - \theta_2^2}, 1, \dots, 1$ , first sub-diagonal  $-\rho_1 \sqrt{1 - \theta_2^2}, -\theta_1, \dots, -\theta_1$ , second subdiagonal  $-\theta_2, \dots, -\theta_2$ , and zero elsewhere. Let  $\tilde{v}$  be the converged innovation residual from (15)–(17), and define  $\tilde{\sigma}_v^2 = \tilde{v}'\tilde{v}/(n - k)$ . For the converged estimate of  $\beta$ ,  $\tilde{\beta}$ , we define  $cov_{\tilde{\beta}} = \tilde{\sigma}_v^2 (X' P_2(\hat{\theta}_{1PW}, \hat{\theta}_{2PW})' P_2(\hat{\theta}_{1PW}, \hat{\theta}_{2PW}) X)^{-1}$ , and  $\hat{\theta}_{1PW}$  and  $\hat{\theta}_{2PW}$  are the converged estimates of  $\theta_1$  and  $\theta_2$ , respectively. For  $\theta = \{\theta_1, \theta_2\}$ , under normality, we calculate the quasi-ML covariance  $cov_{\hat{\theta}_{PW}}$  from the inverse of minus the Hessian of the concentrated unconditional loglikelihood  $L(\theta_1, \theta_2) = -(n/2) \ln(Y^* - X^* \hat{\beta}_{GLS,\theta})' (Y^* - X^* \hat{\beta}_{GLS,\theta})$  evaluated at  $\hat{\theta}_{1PW}$  and  $\hat{\theta}_{2PW}$ , with  $\hat{\beta}_{GLS,\theta} = (X^* X^*)^{-1} (X^* Y^*)$ ,  $Y^* = P_2(\theta_1, \theta_2) Y$  and  $X^* = P_2(\theta_1, \theta_2) X$ , and the MLE innovation variance estimate  $\hat{\sigma}_{vMLE,\theta}^2 = \hat{v}(\theta)' \hat{v}(\theta) / n$ , with “residual”  $\hat{v}(\theta)$  for a given vector  $\theta$ ,  $\hat{v}(\theta) = Y - X \hat{\beta}_{GLS,\theta}$ . Alternatively, we may use the asymptotic covariance for  $\tilde{\theta}_{PW} = \{\hat{\theta}_{1PW}, \hat{\theta}_{2PW}\}$ , requiring no

normality (see below). A third option could be to calculate the sandwich covariance. Note that all these covariances are not affected by the presence of the lagged dependent variable as a regressor.

Finally, for the model  $Y = X\beta + y$  when  $y$  follows  $AR(p)$  with  $\theta = \{\theta_1, \dots, \theta_p\}'$ , OLS provides  $\hat{\beta}$  and  $\hat{y}$ , and the first step  $\hat{\theta}_{PW} = \{\hat{\theta}_{1PW}, \dots, \hat{\theta}_{pPW}\}'$ , using the generic formula above. Then the OLS (in fact GLS) regression of  $P(\hat{\theta}_{PW})Y$  on  $P(\hat{\theta}_{PW})X$  results in a new  $\hat{\beta}$ , new  $\hat{y} = Y - X\hat{\beta}$ , and new  $\hat{\theta}_{PW}$ , repeating until convergence and restricting the elements of  $\theta$  accordingly. The exact GLS matrix  $P$  has the correct Cholesky decomposition<sup>9</sup> of  $V_p^{-1}$  in positions  $1 : p$  and  $1 : p$ ,  $P[j, j] = 1$  for  $j = p + 1, \dots, n$ ,  $P[j + i, j] = -\theta_i$  for  $j = p + 1, \dots, n$  and  $i = 1, \dots, p$  with the restriction  $j + i \leq n$ , and zeros elsewhere. The fast calculation of  $P(\theta)Y$  and  $P(\theta)X$  may rely on the convolution procedure, except for the first  $p$  rows where it has to be implemented manually. The typical element of  $V_p^{-1}$ ,  $v_{ij}$ , is given in (Hamilton 1994, p. 125) or (Galbraith and Galbraith 1974, p. 70). For the converged  $\hat{\beta}$ ,  $\tilde{\beta}$ ,  $cov_{\tilde{\beta}} = \tilde{\sigma}_v^2 (X'P(\tilde{\theta}_{PW})'P(\tilde{\theta}_{PW})X)^{-1}$  is used, where  $\tilde{v}$  is the converged innovation residual and  $\tilde{\sigma}_v^2 = \tilde{v}'\tilde{v}/(n - k)$ . Assuming normality, we can calculate the quasi-ML covariance  $cov_{\hat{\theta}_{PW}}$  from the inverse of minus the Hessian of the concentrated unconditional loglikelihood  $L(\theta) = -(n/2) \ln(Y^* - X^*\hat{\beta}_{GLS,\theta})'(Y^* - X^*\hat{\beta}_{GLS,\theta})$  evaluated at converged  $\hat{\theta}_{PW}$  with  $\hat{\beta}_{GLS,\theta} = (X^*X^*)^{-1}(X^*Y^*)$ ,  $Y^* = P(\theta)Y$  and  $X^* = P(\theta)X$ , and the MLE innovation variance estimate  $\hat{\sigma}_{vMLE,\theta}^2 = \hat{v}(\theta)'\hat{v}(\theta)/n$ , with "residual"  $\hat{v}(\theta)$  for a given vector  $\theta$ ,  $\hat{v}(\theta) = Y - X\hat{\beta}_{GLS,\theta}$ . An alternative covariance for  $\hat{\theta}_{PW}$  is the asymptotic covariance  $\tilde{V}_p^{-1}/n$  ( $\tilde{V}_p^{-1}$  is  $V_p^{-1}$  evaluated at  $\tilde{\theta}_{PW} = \{\tilde{\theta}_{1PW}, \dots, \tilde{\theta}_{pPW}\}'$ ). A third covariance option is the sandwich covariance. Again, all these covariances are not affected by the presence of the lagged dependent variable as a regressor.

GAUSS<sup>TM</sup>/MATLAB<sup>TM</sup>/GNU Octave programmes for the method are available upon request.

#### 4. Conclusions

We have proposed the correct, fast, and generic PW algorithm for exact NLS estimation of a regression with AR errors of any order (requiring no normality assumption). However, if the innovations are normal, then our proposed estimators are identical to the unconditional ML estimators. No numerical optimiser is required either, although we have checked the results of our algorithm with the results of numerical optimisation. Quasi-normality may be assumed in constructing a covariance (quasi-ML or sandwich) for the AR parameters. Alternatively, the asymptotic covariance can also be calculated (with no normality requirement). The covariance of the regressor parameters is obtained from the GLS formulae. An information criterion may be employed to select the unknown AR order in practise, approximating any underlying ARMA process. For a recent paper on AR lag order selection, see Butler and Paoletta (2017). The estimation method is suitable and efficient for trending data. It is also applicable to dynamic regression with a lagged dependent variable included as a regressor. It provides reliable standard errors in the presence of the lagged dependent variable as a regressor. In addition, it can be used to derive efficient Wald-type unit root tests.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

#### Notes

- <sup>1</sup> The CO method of Oberhofer and Kmenta (1974) and Magnus (1978) is different. It is an ML estimation method, imposing zero initial conditions.
- <sup>2</sup> Magnus (1978) provides an alternative cumbersome algorithm which requires an equation solver.
- <sup>3</sup> There may be convergence issues with the BM algorithm for regression with AR(2) errors, but they are beyond the scope of this paper.



- <sup>4</sup> The one that obeys the corresponding first order condition for minimisation of the exact NLS sum of squares of the innovations.
- <sup>5</sup>  $\theta y_1^2 + \sum_2^n y_t y_{t-1} - \theta \sum_2^n y_{t-1}^2 = 0 \Rightarrow \sum_2^n y_t y_{t-1} = \theta(-y_1^2 + \sum_2^n y_{t-1}^2)$ .
- <sup>6</sup> Since  $\hat{\theta}_{PW} = \hat{\theta}_{LS}(1 + y_1^2 / \sum_3^n y_{t-1}^2) > \hat{\theta}_{LS}$ ,  $\hat{\theta}_{PW}$  is less biased than  $\hat{\theta}_{LS}$  as the bias of the LS estimator is negative. Magee (1989) gives approximate formulae for biases. We have verified superiority of the PW estimator.
- <sup>7</sup> Judge et al. (1985) recommend this estimator. However, according to our calculation (based on (Magee 1989)), this is the worst possible estimator amongst the ones examined.
- <sup>8</sup>  $\rho_1$  and  $\rho_2$  are the theoretical autocorrelation coefficients of order 1 and 2, respectively.
- <sup>9</sup> For this, one must use the command `rev(rev(chol(')))` instead of the `build` in `chol()` in matrix language programmes.

## References

- Beach, Charles M., and James G. MacKinnon. 1978a. A Maximum Likelihood Procedure for Regression with Autocorrelated Errors. *Econometrica* 46: 51–58. [CrossRef]
- Beach, Charles M., and James G. MacKinnon. 1978b. Full Maximum Likelihood Estimation of Second-Order Autoregressive Error Models. *Journal of Econometrics* 7: 187–98. [CrossRef]
- Butler, Ronald W., and Marc S. Paoletta. 2017. Autoregressive Lag-Order Selection Using Conditional Saddlepoint Approximations. *Econometrics* 5: 43. [CrossRef]
- Cochrane, Donald, and Guy H. Orcutt. 1949. Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. *Journal of the American Statistical Association* 44: 3–61.
- Galbraith, R. F., and J. I. Galbraith. 1974. Inverses of Some Patterned Matrices Arising in Theory of Stationary Time Series. *Journal of Applied Probability* 11: 63–71. [CrossRef]
- Gurland, John. 1954. An Example of Autocorrelated Disturbances in Linear Regression. *Econometrica* 22: 218–27. [CrossRef]
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.
- Kadiyala, Koteswara Rao. 1968. A Transformation Used to Circumvent the Problem of Autocorrelation. *Econometrica* 36: 93–96. [CrossRef]
- Kmenta, Jan. 1986. *Elements of Econometrics*. 2nd ed. New York: MacMillan.
- Koopmans, Tjalling. 1942. Serial Correlation and Quadratic Forms in Normal Variables. *Annals of Mathematical Statistics* 13: 14–33. [CrossRef]
- Magge, Lonnie. 1985. Efficiency of Iterative Estimators in the Regression Model with AR(1) Disturbances. *Journal of Econometrics* 29: 275–87. [CrossRef]
- Magee, Lonnie. 1989. An Edgeworth Test Size Correction for the Linear Model with AR(1) Errors. *Econometrica* 57: 661–74. [CrossRef]
- Magnus, Jan R. 1978. Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix. *Journal of Econometrics* 7: 281–12. [CrossRef]
- Oberhofer, Walter, and Jan Kmenta. 1974. A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models. *Econometrica* 42: 579–90. [CrossRef]
- Park, Rolla Edward, and Bridger M. Mitchell. 1980. Estimating the Autocorrelated Error Model with Trended Data. *Journal of Econometric* 13: 185–201. [CrossRef]
- Phillips, Peter C. B. 1977. Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation. *Econometrica* 45: 463–85. [CrossRef]
- Phillips, Peter C. B. 1978. Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autoregression. *Biometrika* 65: 91–98. [CrossRef]
- Sawa, Takamitsu. 1978. The Exact Moments of the Least Squares Estimator for the Autoregressive Model. *Journal of Econometrics* 8: 159–72. [CrossRef]
- Stock, James H. 1991. Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series. *Journal of Monetary Economics* 28: 435–59. [CrossRef]
- Prais, Sigbert J., and Christofer B. Winsten. 1954. *Trend Estimators and Serial Correlation*. Cowles Commission Discussion Paper No. 383, Chicago. Unpublished Work.