Methods: Prais–Winsten Regression with AR(k) Errors

Implementation Notes for `praisk`

Ariel Linden
Linden Consulting Group, LLC

1. Overview

The `praisk` command implements the iterated Prais–Winsten generalized least squares (GLS) estimator for a linear regression model with AR(k) errors. The implementation is based on the exact nonlinear least squares (NLS) Prais–Winsten estimator described in (Vougas, 2021), extended to support higher-order AR processes, panel data, factor variables, and robust variance estimation. Where `praisk` follows the (Vougas, 2021) methodology exactly, this is explicitly noted, and any departure is described and justified (for example, to match the behavior of Stata's official `prais` command or to accommodate panel data).

2. The Statistical Model

The model consists of a linear regression equation with autoregressive errors. The following notation is used throughout. k denotes the AR lag order, i.e. the number of autoregressive terms. $\mathbf{p}$ denotes the $k \times 1$ vector of AR parameters $(p_1, \ldots, p_k)'$; when $k = 1$ the single element is written as the scalar p. q denotes the number of regression coefficients including the constant. $\mathbf{L}$ denotes the $n \times n$ Prais–Winsten transformation matrix, and $\mathbf{L}_0$ its $k \times k$ initialization block.

Let y_t denote the dependent variable at time t and $\mathbf{x}_t$ a $(q \times 1)$ vector of regressors. The model is:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{1}$$

$$u_t = p_1 u_{t-1} + p_2 u_{t-2} + \cdots + p_k u_{t-k} + \varepsilon_t \tag{2}$$

where $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$. The regression coefficients $\boldsymbol{\beta}$ and the AR parameter vector $\mathbf{p}$ are estimated jointly by the iterated GLS algorithm described below (Judge et al., 1985; Hamilton, 1994).

3. Model Specification by AR Order

This section presents the statistical model, transformation, and parameter estimation for each AR order in turn. The AR(1) case follows Stata's `prais` command exactly; AR(2) and AR($k > 2$) extend the framework following (Vougas, 2021).

3.1 AR(1)

Model. With $k = 1$, the model is:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{3}$$

$$u_t = p\, u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2) \tag{4}$$

Stationarity requires $|p| < 1$. Under this condition $\sigma_u^2 = \sigma_\varepsilon^2/(1-p^2)$ and the lag-ℓ autocovariance is $\gamma(\ell) = p^\ell \sigma_u^2$.

Prais–Winsten Transformation. For $t = 2, \ldots, n$ the quasi-differenced equation is:

$$y_t - p\, y_{t-1} = (\mathbf{x}_t - p\, \mathbf{x}_{t-1})'\boldsymbol{\beta} + \varepsilon_t \tag{5}$$

The first observation is retained by premultiplying by $\sqrt{1-p^2}$:

$$\tilde{y}_1 = \sqrt{1-p^2}\, y_1, \qquad \tilde{\mathbf{x}}_1 = \sqrt{1-p^2}\, \mathbf{x}_1 \tag{6}$$

This is the exact GLS transformation: $\sqrt{1-p^2}$ is the $(1,1)$ element of the Cholesky factor of Ω^{-1}, where $\Omega_{st} = p^{|s-t|}\sigma_u^2$. Stacking all n transformed observations, OLS on the system $\mathbf{y}_r = \mathbf{X}_r\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ yields the GLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}_r'\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{y}_r$. This is identical to Stata's `prais` command.

Yule–Walker Estimation of p. At each iteration p is estimated from the current residuals $\hat{u}_t$ by:

$$\hat{p} = \frac{\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{n} \hat{u}_{t-1}^2} \tag{7}$$

This is the moment estimator from the lag-1 Yule–Walker equation $\gamma(1) = p\gamma(0)$, solved for p. It is equation (5) of Vougas (2021) and matches the formula used by Stata's `prais`.

Error Variance and Standard Errors. After convergence the innovation variance is:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-q} \tag{8}$$

where $\hat{\boldsymbol{\varepsilon}} = \mathbf{y}_r - \mathbf{X}_r\hat{\boldsymbol{\beta}}$ and q is the number of estimated coefficients including the constant. The covariance matrix of $\hat{\boldsymbol{\beta}}$ is $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}_\varepsilon^2(\mathbf{X}_r'\mathbf{X}_r)^{-1}$, identical to Stata's `prais`. Standard errors for $\hat{p}$ are described in Section 9.

3.2 AR(2)

The AR(2) model and estimation procedure follow (Vougas, 2021). This is the first case not covered by Stata's `prais` command.

Model. With $k = 2$:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{9}$$

$$u_t = p_1 u_{t-1} + p_2 u_{t-2} + \varepsilon_t \tag{10}$$

Stationarity requires all eigenvalues of the companion matrix (equation (24)) to have modulus less than 1, which is equivalent to the triangle conditions $p_1 + p_2 < 1$, $p_2 - p_1 < 1$, and $|p_2| < 1$ (Hamilton, 1994).

Prais–Winsten Transformation. For $t = 3, \ldots, n$ the AR(2) filter is:

$$\tilde{u}_t = u_t - p_1 u_{t-1} - p_2 u_{t-2} \tag{11}$$

The first two observations require special treatment. The 2×2 autocovariance matrix of $(u_1, u_2)'$ is $\sigma_\varepsilon^2 V_2$, where V_2 is determined by the AR(2) parameters via the Yule–Walker equations. The two initial transformed observations are obtained by applying $\mathbf{L}_0$, the row-reversed Cholesky factor of

V_2^{-1} (equation (22)), to $(y_1, y_2)'$ and $([x_1, x_2])'$. The inverse autocovariance matrix V_2^{-1} is calculated using the formula Galbraith and Galbraith (1974). This is the approach of Vougas (2021) and differs from Stata's `prais`, which handles only the AR(1) case.

Yule–Walker Estimation of (p_1, p_2). The 2×2 Yule–Walker normal equations are (equation (11) of Vougas (2021)):

$$\begin{bmatrix} \sum \hat{u}_{t-1}^2 & \sum \hat{u}_{t-1}\hat{u}_{t-2} \\ \sum \hat{u}_{t-1}\hat{u}_{t-2} & \sum \hat{u}_{t-2}^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \sum \hat{u}_t \hat{u}_{t-1} \\ \sum \hat{u}_t \hat{u}_{t-2} \end{bmatrix} \tag{12}$$

Cross-products are accumulated over all contiguous segments before solving. The system is solved via LU decomposition (Golub and Van Loan, 1996).

3.3 AR($k > 2$)

For AR orders beyond 2, `praisk` follows the general algorithm of (Vougas, 2021). No analytic closed-form is presented in that paper beyond AR(2) so the implementation follows the MATLAB source code provided directly by the author.

Model. The general model is given by equations (1) and (2). Stationarity requires all eigenvalues of the $k \times k$ companion matrix C (equation (24)) to have modulus strictly less than 1. Individual coefficients p_j may exceed 1 in absolute value while the process remains stationary.

Prais–Winsten Transformation. For $t = k+1, \ldots, n$ the AR(k) filter is:

$$\tilde{u}_t = u_t - p_1 u_{t-1} - \cdots - p_k u_{t-k} \tag{13}$$

The first k observations are transformed using $\mathbf{L}_0$, the row-reversed Cholesky factor of V_k^{-1} (equation (22)), where V_k^{-1} is computed via (Galbraith and Galbraith, 1974). The general procedure is identical to that for AR(2) but applied to a $k \times k$ system.

Yule–Walker Estimation of $\mathbf{p}$. The $k \times k$ Yule–Walker system $A\mathbf{p} = \mathbf{b}$ is:

$$b[j] = \sum_{t=j+1}^{n} \hat{u}_t \hat{u}_{t-j}, \quad j = 1, \ldots, k \tag{14}$$

$$A[i,j] = \sum_{t=\max(i,j)+1}^{n} \hat{u}_{t-i}\hat{u}_{t-j}, \quad i,j = 1, \ldots, k \tag{15}$$

Cross-products are pooled across all segments and solved via LU decomposition. This generalizes equations (7) and (12) uniformly for any k.

4. The Iterative GLS Algorithm

The algorithm alternates ("zig-zags") between two steps until convergence (Prais and Winsten, 1954; Judge et al., 1985; Vougas, 2021). Starting from OLS residuals $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$, the steps are:

Step 1. Estimate the AR(k) parameters $\mathbf{p}$ from the current residuals by solving the pooled Yule–Walker normal equations (Section 5).

Step 2. Apply the exact Prais–Winsten GLS transformation using the current $\mathbf{p}$ (Section 6) and compute GLS estimates:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_r'\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{y}_r \tag{16}$$

where $\mathbf{y}_r$ and $\mathbf{X}_r$ are the Prais–Winsten transformed dependent variable and regressor matrix, respectively.

Convergence is declared when $\max_j |p_j^{(\text{new})} - p_j^{(\text{old})}| < \tau$, where τ is a user-specified tolerance (default 10^{-6}). This criterion is applied across all k AR coefficients simultaneously, and matches the convergence criterion used in the Vougas (2021) MATLAB code and in Stata's `prais` command.

Iteration 0 in the displayed output always shows $p = 0$ (or $\max|\lambda| = 0$ for AR($k > 1$)) as a reference point; the first Yule–Walker estimate appears at Iteration 1, matching the display convention of Stata's `prais` command.

5. Yule–Walker Estimation of AR(k) Parameters

At each iteration, the AR(k) parameter vector $\mathbf{p}$ is estimated from the current residuals by solving the $k \times k$ Yule–Walker normal equations $A\mathbf{p} = \mathbf{b}$ (Brockwell and Davis, 1991; Walker, 1931), where for each contiguous data segment of length n_s:

$$b[j] = \sum_{t=j+1}^{n_s} \hat{u}_t \hat{u}_{t-j}, \quad j = 1, \ldots, k \tag{17}$$

$$A[i,j] = \sum_{t=\max(i,j)+1}^{n_s} \hat{u}_{t-i} \hat{u}_{t-j}, \quad i,j = 1, \ldots, k \tag{18}$$

The matrix A is symmetric. Cross-products from all segments are accumulated before solving, yielding a single pooled estimate of $\mathbf{p}$ (see Section 8 for the panel data treatment). The system is solved via LU decomposition (Golub and Van Loan, 1996).

5.1 AR(1)

For $k = 1$, the solution is:

$$\hat{p} = \frac{\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{n} \hat{u}_{t-1}^2} \tag{19}$$

This is equation (5) of Vougas (2021) exactly.

5.2 AR(2)

For $k = 2$, the 2×2 normal equations are:

$$\begin{bmatrix} \sum \hat{u}_{t-1}^2 & \sum \hat{u}_{t-1}\hat{u}_{t-2} \\ \sum \hat{u}_{t-1}\hat{u}_{t-2} & \sum \hat{u}_{t-2}^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \sum \hat{u}_t \hat{u}_{t-1} \\ \sum \hat{u}_t \hat{u}_{t-2} \end{bmatrix} \tag{20}$$

This is equation (11) of Vougas (2021) exactly.

5.3 AR($k > 2$)

For $k > 2$, the general formula in equations (14) and (15) applies. This matches the MATLAB source code provided by the paper's author. The Vougas (2021) paper itself does not present a general closed-form expression beyond $k = 2$.

6. The Exact Prais–Winsten GLS Transformation

For a data segment of length n with AR(k) errors, the Prais–Winsten transformation matrix $\mathbf{L}$ is an $n \times n$ lower-triangular matrix that maps the regression equation to one with approximately iid errors (Prais and Winsten, 1954).

The lower $n - k$ rows of $\mathbf{L}$ apply the standard AR filter: row t (for $t > k$) produces:

$$\tilde{u}_t = u_t - p_1 u_{t-1} - p_2 u_{t-2} - \cdots - p_k u_{t-k} \tag{21}$$

The Cochrane–Orcutt estimator (Cochrane and Orcutt, 1949) discards the first k observations. The Prais–Winsten estimator retains them by using the exact covariance structure of the initial observations under the AR(k) process.

6.1 Exact Initialization of the First k Observations

The covariance matrix of $(u_1, \ldots, u_k)'$ is $\sigma_\varepsilon^2 V_k$, where V_k is the $k \times k$ autocovariance matrix of the AR(k) process. The top-left $k \times k$ block of $\mathbf{L}$ is set to the reversed Cholesky factor of V_k^{-1}:

$$\mathbf{L}_0 = \mathrm{chol}(V_k^{-1})[k{:}1,\ k{:}1] \tag{22}$$

The inverse covariance matrix V_k^{-1} is computed analytically using the closed-form expression of (Galbraith and Galbraith, 1974), which avoids matrix inversion and is numerically stable. See Hamilton (1994), p. 125 for a textbook derivation. This approach is identical to that used in the Vougas (2021) MATLAB code.

6.2 AR(1) Transformation

For $k = 1$, $V_k^{-1} = (1 - p^2)$ is a scalar. The first transformed observation is:

$$\tilde{y}_1 = \sqrt{1 - p^2}\, y_1 \tag{23}$$

and subsequent observations follow $\tilde{y}_t = y_t - p\, y_{t-1}$. This is the classical transformation of (Prais and Winsten, 1954).

6.3 AR(2) Transformation

For $k = 2$, the 2×2 Cholesky factor of V_2^{-1} produces two transformed initial observations that jointly capture the autocovariance at lags 1 and 2. Rows 3 through n use the standard two-lag filter.

6.4 AR($k > 2$) Transformation

For $k > 2$, the same procedure applies: compute the $k \times k$ Cholesky factor of V_k^{-1}, reverse its rows, and use these for the first k transformed observations. The Galbraith and Galbraith (1974) formula for V_k^{-1} applies uniformly for all k; no special-case treatment is required beyond AR(1).

7. AR Parameter Bounds, Stationarity, and Individual Coefficients Exceeding Unity

A common misconception is that AR coefficients must lie within $[-1, 1]$ for the process to be stationary. This holds for AR(1) but not for higher-order models (Hamilton, 1994; Brockwell and Davis, 1991).

7.1 AR(1)

For AR(1), $|p| < 1$ is both necessary and sufficient for covariance stationarity (Hamilton, 1994). In large samples, the Yule–Walker estimator applied to a stationary series will satisfy this. In finite samples, or with near-unit-root data, $|\hat{p}| \geq 1$ is possible though unusual.

7.2 AR(2)

The stationarity region for AR(2) is the triangle defined by three conditions: $p_2 + p_1 < 1$; $p_2 - p_1 < 1$; and $-1 < p_2 < 1$ (Hamilton, 1994). Within this region, individual coefficients can legitimately satisfy $|p_1| > 1$. For example, $p_1 = 1.2$ and $p_2 = -0.5$ satisfies all three conditions and defines a stationary process. The Yule–Walker estimator routinely produces such estimates when the true process has strong positive autocorrelation at lag 1 and negative autocorrelation at lag 2.

7.3 AR($k > 2$)

In general, the stationarity condition for AR(k) is that all eigenvalues of the companion matrix C have modulus strictly less than 1 (Hamilton, 1994; Brockwell and Davis, 1991), where:

$$C = \begin{bmatrix} p_1 & p_2 & \cdots & p_{k-1} & p_k \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{24}$$

There is no constraint on the magnitude of any individual p_j. The higher the AR order, the more freedom individual coefficients have to take large values while the overall process remains stationary.

7.4 No Stationarity Bounding in `praisk`

The Vougas (2021) MATLAB code does not enforce stationarity bounds on the Yule–Walker solution, nor does Stata's `prais` command. `praisk` follows the same convention: no bounding is applied. If the final converged estimates are non-stationary ($\max |\lambda| \geq 1$), a warning is printed but estimation completes. Non-stationary estimates indicate a misspecified model or insufficient data.

An earlier version of `praisk` enforced stationarity by bisecting $\mathbf{p}$ toward zero until $\max |\lambda| < 1$. This introduced bias relative to `prais` on benchmark data and was removed.

7.5 Iteration Display: p vs. $\max |\lambda|$

For AR(1), `praisk` prints p at each iteration, since $|p| < 1$ is the exact stationarity condition. For AR($k > 1$), printing individual AR coefficients would be misleading because values outside $[-1, 1]$ are routinely legitimate (Hamilton, 1994). Instead, `praisk` displays the maximum eigenvalue modulus of the companion matrix, $\max_j |\lambda_j|$, at each iteration. This scalar:

- correctly summarizes stationarity of the AR(k) process for all k;
- is directly interpretable: values below 1 indicate a stationary process; and
- collapses to $|p|$ for AR(1), making it a natural generalization of the AR(1) display.

This display choice is not present in the Vougas (2021) MATLAB code.

8. Panel Data and Multiple Segments

The Vougas (2021) paper and accompanying MATLAB code consider only a single time series. There is no discussion of panel data, multiple cross-sectional units, or gaps in the time index. The panel support in `praisk` is an independent extension designed to match the behavior of Stata's official `prais` command.

8.1 Segment Identification

A contiguous segment is a run of observations belonging to the same panel unit with no gaps in the time index. Boundaries occur wherever the panel identifier changes or the time variable is non-consecutive. The number of such boundaries is stored in `e(ngaps)`.

8.2 Pooled Yule–Walker

The cross-product matrices A and $\mathbf{b}$ from Section 5 are accumulated within each segment and summed across all segments before solving. This treats the problem as if the full stacked dataset were a single series for Yule–Walker purposes (Brockwell and Davis, 1991; Walker, 1931), ignoring cross-segment lags. The result is a single pooled $\mathbf{p}$ estimate common to all panel units, matching the behavior of Stata's `prais` in a panel setting (Park and Mitchell, 1980).

8.3 Independent GLS Transformation

The Prais–Winsten transformation is applied independently to each segment. The exact initialization (Section 6) is restarted at the beginning of every segment, so the first k observations of every panel unit or time gap receive the full Prais–Winsten treatment rather than being filtered as continuations of the previous segment (Prais and Winsten, 1954; Galbraith and Galbraith, 1974). Cross-segment autocorrelation is assumed to be zero.

8.4 Short Segments

A segment of length $n_s \leq k$ cannot contribute to the Yule–Walker accumulation and is skipped silently for that purpose. Its observations are still included in the GLS regression.

9. Standard Errors for AR Parameters

The asymptotic covariance of the Yule–Walker estimator is (Brockwell and Davis, 1991):

$$\mathrm{Var}(\hat{\mathbf{p}}) = \hat{\sigma}_\varepsilon^2 A^{-1} \tag{25}$$

where A is the pooled cross-product matrix from the final Yule–Walker solve and $\hat{\sigma}_\varepsilon^2$ is estimated as:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n_{\mathrm{ar}}} \sum_{\mathrm{segs}} \sum_{t=k+1}^{n_s} \left(\hat{u}_t - \hat{p}_1 \hat{u}_{t-1} - \cdots - \hat{p}_k \hat{u}_{t-k}\right)^2 \tag{26}$$

where n_{ar} is the total number of AR residuals across all segments. Standard errors are the square roots of the diagonal of $\hat{\sigma}_\varepsilon^2 A^{-1}$.

9.1 Departure from Vougas (2021)

Section 3 of Vougas (2021) suggests using V_k^{-1}/n as the asymptotic covariance, where V_k^{-1} is the inverse AR(k) autocovariance matrix evaluated at convergence. For AR(2), the diagonal elements of V_k^{-1} are mathematically equal (a consequence of the Yule–Walker structure) so this formula would produce identical standard errors for p_1 and p_2. This is incorrect in finite samples, where the diagonal elements of A differ because different lags have different numbers of cross-products. The implementation uses A^{-1} directly (Brockwell and Davis, 1991), producing correctly distinct standard errors for each AR coefficient.

10. GLS Coefficient Covariance

At convergence, the default covariance matrix of $\hat{\boldsymbol{\beta}}$ is (Judge et al., 1985):

$$\widehat{V}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}_r'\mathbf{X}_r)^{-1}, \qquad s^2 = \frac{\mathbf{e}'\mathbf{e}}{N-q} \tag{27}$$

where $\mathbf{e} = \mathbf{y}_r - \mathbf{X}_r\hat{\boldsymbol{\beta}}$ are the transformed residuals and q is the number of active regressors including the constant. Note that s^2 uses the degrees-of-freedom correction $(N-q)$, not the ML divisor N.

10.1 Robust Standard Errors

When `vce(robust)` is specified, the HC1 heteroskedasticity-consistent sandwich covariance matrix (White, 1980; MacKinnon and White, 1985) is computed on the GLS-transformed data:

$$\widehat{V}^{\text{HC1}}(\hat{\boldsymbol{\beta}}) = \frac{N}{N-q}(\mathbf{X}_r'\mathbf{X}_r)^{-1}\big[\mathbf{X}_r' \operatorname{diag}(\mathbf{e}^2)\,\mathbf{X}_r\big](\mathbf{X}_r'\mathbf{X}_r)^{-1} \tag{28}$$

Standard errors are labelled *Semirobust* following the convention of Stata's `prais` command, noting that these SEs are robust to heteroskedasticity in the innovations but not to misspecification of the AR(k) error structure.

10.2 Cluster-Robust Standard Errors

When `vce(cluster clustvar)` is specified, the cluster-robust sandwich estimator (Liang and Zeger, 1986) is used:

$$\widehat{V}^{\text{CR}}(\hat{\boldsymbol{\beta}}) = \frac{G}{G-1}\cdot\frac{N-1}{N-q}(\mathbf{X}_r'\mathbf{X}_r)^{-1}M(\mathbf{X}_r'\mathbf{X}_r)^{-1} \tag{29}$$

where $M = \sum_{g=1}^{G} S_g' S_g$, with S_g the sum of score vectors within cluster g, and G is the number of clusters. The small-sample correction $G/(G-1)\cdot(N-1)/(N-q)$ matches the standard Stata convention.

11. Exact Log Likelihood and Comparison with `prais`

The exact Gaussian log likelihood for the AR(k) model, accounting for the Prais–Winsten initialization, is (Hamilton, 1994; Vougas, 2021):

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \frac{1}{2}\ln|V_k^{-1}| - \frac{\mathbf{e}'\mathbf{e}}{2\sigma^2} \tag{30}$$

where $\sigma^2 = \mathbf{e}'\mathbf{e}/N$ (the ML estimator) and $|V_k^{-1}|$ is the determinant of the inverse AR(k) autocovariance matrix evaluated at the converged estimates.

11.1 The Determinant Correction Term

The term $\frac{1}{2}\ln|V_k^{-1}|$ arises from the exact distribution of the first k observations under the Prais–Winsten initialization (Galbraith and Galbraith, 1974; Hamilton, 1994). For AR(1) this equals $\frac{1}{2}\ln(1-p^2)$, which is negative for $|p|>0$ and therefore reduces the log likelihood relative to a formula that ignores initialization. This term is the correct adjustment for the non-iid contribution of the initialized observations and is the reason the Prais–Winsten likelihood differs from the Cochrane–Orcutt likelihood (Cochrane and Orcutt, 1949).

11.2 Difference from Stata's `prais`

Stata's `prais` command computes the log likelihood without the determinant term:

$$\ln L_{\texttt{prais}} = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{\mathbf{e}'\mathbf{e}}{2\sigma^2} \tag{31}$$

This treats the Prais–Winsten transformed observations as if all N were standard iid innovations. The difference between equations (30) and the above is exactly $\frac{1}{2}\ln|V_k^{-1}|$, which for AR(1) at $p=0.469$ equals approximately -0.124, matching the numerically observed discrepancy between the two commands.

Equation (30) is theoretically correct. However, because `prais` uses the alternative formula, AIC and BIC values from `praisk` (via `estat ic`) are not directly comparable to those from `prais`. They are comparable across different AR orders within `praisk`, since the determinant correction is applied consistently.

12. Information Criteria (AIC and BIC)

AIC (Akaike, 1974) and BIC (Schwarz, 1978) are computed by Stata's `estat ic` post-estimation command using the stored log likelihood from equation (30) and the model degrees of freedom $\mathrm{df} = \mathrm{rank}(\hat{V})$, which counts all non-zero diagonal entries of the posted variance-covariance matrix (i.e., active regressors including the constant, excluding omitted terms):

$$\mathrm{AIC} = -2\ln\hat{L} + 2\cdot\mathrm{df} \tag{32}$$

$$\mathrm{BIC} = -2\ln\hat{L} + \ln(N)\cdot\mathrm{df} \tag{33}$$

Note that the AR lag-order count k is not added to df in these expressions, matching the convention used by Stata's `prais` command.

13. Residual Autocorrelations as a Diagnostic

13.1 Motivation

After fitting an AR(k) model it is natural to ask whether the Prais–Winsten transformation has successfully whitened the errors. The standard approach is a formal portmanteau test applied to the AR innovation residuals $\hat{\mathbf{e}} = \mathbf{y}_r - \mathbf{X}_r\hat{\boldsymbol{\beta}}$. Under correct specification $\hat{\mathbf{e}}$ is approximately iid, and the test statistic follows an approximate χ^2 distribution for large n in a single time series.

`praisk` does not report the Q statistic. There are two reasons. First, and most importantly, the χ^2 approximation requires a single time series of length n; the degrees of freedom and the scaling factor $n(n+2)$ in the Ljung–Box formula both depend on this assumption. For panel data with

N units and T time periods, total observations are $n = NT$, but consecutive pairs of residuals are only meaningful within panels. Using $n = NT$ inflates Q dramatically (by a factor of approximately N) and produces a statistic with no valid null distribution. Second, even for pure time series, Q provides a single binary signal (reject or not reject) that obscures which lags are responsible for any remaining autocorrelation.

13.2 What `praisk` Displays

`praisk` instead reports the sample autocorrelations of the residuals at lags $1, \ldots, k$, for two residual series:

1. *Untransformed residuals*: the OLS residuals $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$, computed before the AR transformation. These establish the baseline level of serial correlation in the raw data.
2. *Transformed (innovation) residuals*: the GLS innovation residuals $\hat{\mathbf{e}} = \mathbf{y}_r - \mathbf{X}_r\hat{\boldsymbol{\beta}}$, computed after the Prais–Winsten transformation. Under correct AR(k) specification these should be approximately iid, so their autocorrelations should be near zero at all lags.

The lag-j sample autocorrelation of a residual series r_t is

$$\hat{\rho}(j) = \frac{\sum_{t=j+1}^{n}(r_t - \bar{r})(r_{t-j} - \bar{r})}{\sum_{t=1}^{n}(r_t - \bar{r})^2} \tag{34}$$

where $\bar{r}$ is the sample mean. The output table shows $\hat{\rho}(1), \ldots, \hat{\rho}(k)$ for both series side by side, enabling an immediate before-and-after comparison.

13.3 Interpretation

A successful AR(k) transformation is indicated by transformed autocorrelations close to zero at all displayed lags. A large residual autocorrelation at lag j in the transformed series suggests that the AR order may need to be increased to j or beyond. Under a correctly specified AR(k) model, autocorrelations beyond lag k are also zero in expectation; users who wish to inspect higher lags can obtain the full autocorrelation function with `ac` applied to the saved innovation residuals (`predict ue, ue`).

As a rough informal benchmark, for a single time series of length n the standard error of each $\hat{\rho}(j)$ under the white noise null is approximately $n^{-1/2}$ (Brockwell and Davis, 1991). For panel data this benchmark does not apply directly, but transformed autocorrelations substantially smaller in magnitude than the untransformed autocorrelations still constitute clear evidence that the transformation has been effective.

13.4 Panel Data: Validity of the Pooled Estimator

For panel data, `praisk` computes autocorrelations using Stata's lag operator, which returns a missing value for the first observation in each panel. Consequently, the estimator is evaluated only over within-panel consecutive pairs $(r_{it}, r_{i,t-j})$ for $t > j$, unit i, and all i. Cross-panel pairs never contribute.

The resulting estimate is a pooled within-panel autocorrelation,

$$\hat{\rho}_{\text{pooled}}(j) = \frac{\sum_{i=1}^{N}\sum_{t=j+1}^{T_i}(r_{it} - \bar{r})(r_{i,t-j} - \bar{r})}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(r_{it} - \bar{r})^2} \tag{35}$$

where T_i is the number of observations in panel i. This estimator is consistent for the common within-panel lag-j autocorrelation under the assumption that the autocorrelation structure is homogeneous across panels. It does not require a distributional assumption and no degrees-of-freedom argument is needed, which is precisely why `praisk` reports it in place of the Q statistic.

If heterogeneity across panels is a concern (for example, if one panel unit has a materially different residual autocorrelation structure from the others – such as a treated unit in a controlled interrupted time series study) the pooled figure should be interpreted with caution. In such cases it is advisable to inspect the residuals of individual panels separately using `predict` followed by panel-specific `ac` plots.

References

[1] Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723.

[2] Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*, 2nd ed. Springer.

[3] Cochrane, D., and G. H. Orcutt. 1949. Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association* 44(245): 32–61.

[4] Galbraith, R. F., and J. I. Galbraith. 1974. On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability* 11(1): 63–71.

[5] Golub, G. H., and C. F. Van Loan. 1996. *Matrix Computations*, 3rd ed. Johns Hopkins University Press.

[6] Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press.

[7] Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics*, 2nd ed. Wiley.

[8] Liang, K.-Y., and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73(1): 13–22.

[9] MacKinnon, J. G., and H. White. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29(3): 305–325.

[10] Park, R. E., and B. M. Mitchell. 1980. Estimating the autocorrelated error model with trended data. *Journal of Econometrics* 13(2): 185–201.

[11] Prais, S. J., and C. B. Winsten. 1954. Trend estimators and serial correlation. Cowles Commission Discussion Paper 383, Chicago.

[12] Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2): 461–464.

[13] Vougas, D. V. 2021. Prais–Winsten algorithm for regression with second or higher order autoregressive errors. *Econometrics* 9(3): 32.

[14] Walker, G. 1931. On periodicity in series of related terms. *Proceedings of the Royal Society of London, Series A* 131(818): 518–532.

[15] White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4): 817–838.