`qregark`: Quantile regression with AR(k) errors, by Prais–Winsten or Cochrane–Orcutt estimation

Ariel Linden
Linden Consulting Group, LLC
alinden@lindenconsulting.org

November 2026

Abstract

This document describes the methods implemented in the Stata command `qregark`, which fits linear quantile regression models with autoregressive errors of order k. Analogous to `prais` for least squares, `qregark` implements a Prais–Winsten-type estimator (the default), which retains the first k observations through a stationary transformation, and a conditional, Cochrane–Orcutt-type estimator (option `corc`), which conditions on them. The document gives the model, the estimators, the standard errors, the optional small-sample bias correction, and the stored diagnostics, offers practical advice on their use, and summarizes the results of the testing that led to particular implementation choices. It accompanies the help file, which describes the syntax and options.

1 Introduction

Interrupted time series analysis (ITSA) is widely used in many research disciplines to evaluate interventions when randomization is infeasible [1, 2, 3]. ITSA has been argued to have strong internal validity, primarily through its control over regression to the mean [1, 2, 4], and generally has good external validity, particularly when the unit of measurement is at the population level or when results can be generalized to other units, treatments, or settings [2, 5]. In the single-group design, observations collected at regular intervals before and after an intervention are used to estimate the pre-intervention trend and changes in the level and trend of the outcome following the intervention [3]. Valid inference depends not only on specifying these intervention-associated changes but also on using a model appropriate for the outcome and its temporal dependence [6].

ITSA models are usually estimated by ordinary least squares (OLS) or by generalized least squares methods that allow for autocorrelated errors, such as the Prais–Winsten estimator [7, 3]. These estimators target the conditional *mean* of the outcome. In much research, however, many outcomes analyzed with ITSA are skewed, heavy-tailed, or subject to occasional extreme values (for example, costs, wages, lengths of stay, or utilization rates), and an intervention may affect parts of the outcome distribution differently: a program aimed at high utilizers, for instance, may change the upper tail while leaving the median largely unchanged. In such settings the mean can be a poor summary of the typical level, a few extreme observations can dominate the least-squares fit, and the question of interest may concern a quantile rather than the mean.

Quantile regression [8, 9] addresses these concerns. It estimates the conditional τth quantile of the outcome for any $\tau \in (0, 1)$; median regression ($\tau = .5$) is a robust alternative to least squares for the typical level, and other quantiles describe how the intervention shifts the lower or upper part of

the distribution. The level and trend changes of an ITSA model then have the same interpretation as in the usual model, but for the τth quantile of the outcome rather than its mean.

Conventional quantile regression inference, however, treats the observations as independent, whereas the errors of a time series are typically autocorrelated. Independence is not needed to define the regression quantiles, and, provided the specified quantile restriction is correct, the point estimates generally remain consistent under serial dependence; but standard errors computed as if the observations were independent do not account for the dependence and are usually too small when the autocorrelation is positive.

`qregark` extends Stata's `qreg` command [10] to regression models with AR(k) errors, in the same way that `prais` [10] extends `regress`: it models the errors as an autoregressive process of order k, estimates the regression and autoregressive (AR) coefficients jointly, and offers both of the classical approaches to the first observations of a series.

- With option `corc`, `qregark` uses the conditional, Cochrane–Orcutt-type estimator of Tian et al. [11]: it minimizes their conditional objective with an EM-based algorithm, with the numerical modifications described in Section 2. The first k observations of each series are conditioned on and do not contribute to the objective. This is the quantile analogue of `prais, corc`.
- By default, `qregark` uses a Prais–Winsten-type estimator, the quantile analogue of the default `prais` estimator. Like the Prais–Winsten method for least squares [7, 12], it retains the first k observations through a transformation based on the stationary covariance of the errors. It follows the iterative Prais–Winsten-type least absolute value estimator of Dielman and Rose [13], extended to AR(k) errors and to quantiles other than the median.

With k set to zero, both would reduce to `qreg`; `qregark` requires $k \geq 1$. Throughout, the AR coefficients are written $\phi_1, \ldots, \phi_k$; the help file and output label them $\rho_1, \ldots, \rho_k$, as in `prais`.

One consequence of modeling the errors needs to be stated at the outset. In the model of Section 2, the quantile restriction is placed on the *innovations* of the AR process: their τth quantile is zero, so the coefficients describe the conditional τth quantile of the outcome given its own past. The τth quantile of the outcome itself, at given values of the regressors, additionally involves the τth quantile of the stationary AR errors, which away from the median is generally not zero. The slope coefficients, and hence the intervention effects, are unaffected, but the structural intercept is not the starting level of the τth quantile of the outcome. For this reason `qregark` reports as `_cons` the baseline level of the outcome quantile (Section 3).

Section 2 describes the model and the two estimators. Section 3 explains how the intercept is reported, Section 4 the standard errors, and Section 5 the optional small-sample bias correction. Section 6 describes the diagnostics and stored flags, Section 7 summarizes the simulation designs and results referred to throughout, Section 8 gives practical advice, and Section 9 summarizes limitations.

2 Model and estimation

2.1 The model

For a single series observed at $t = 1, \ldots, T$, `qregark` fits

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad \varepsilon_t = \phi_1\varepsilon_{t-1} + \cdots + \phi_k\varepsilon_{t-k} + \eta_t, \tag{1}$$

where $\mathbf{x}_t$ contains the regressors (in an ITSA model, the time trend, the post-intervention indicator, and their interaction, plus a constant), and the innovations η_t are independent and identically

distributed with τth quantile zero. The conditional τth quantile of y_t given its past is then

$$Q_\tau(y_t \mid y_{t-1}, \dots) = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{j=1}^{k} \phi_j \, (y_{t-j} - \mathbf{x}_{t-j}'\boldsymbol{\beta}), \tag{2}$$

which is the one-step quantile returned by `predict, quantile`.

Stationarity. Let $\mathbf{A}$ be the $k \times k$ companion matrix of the AR process (first row $\boldsymbol{\phi}'$, an identity matrix below it), and let $\lambda(\boldsymbol{\phi})$ denote the largest modulus of its eigenvalues. The process is stationary when $\lambda(\boldsymbol{\phi}) < 1$, equivalently when all roots of the AR polynomial $1 - \phi_1 z - \cdots - \phi_k z^k$ lie outside the unit circle. Throughout, the limits used by `qregark` are stated in terms of $\lambda(\boldsymbol{\phi})$; for AR(1), $\lambda(\phi) = |\phi|$.

Segments. With panel data or gaps in the time variable, the sample is split into contiguous segments (a new segment starts at each panel change and each gap), and the AR recursion restarts in each segment. The regression coefficients and the AR coefficients are common to all segments.

Objective and reported fit statistics. Both estimators minimize the quantile check function $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ of the residuals of a transformed model. Their statistical status differs:

- For the conditional estimator, the objective is the innovation check loss, and minimizing it is equivalent to maximizing a conditional working likelihood under which the innovations follow an asymmetric Laplace distribution [11].
- For the Prais–Winsten-type estimator, the objective is an *augmented* check-loss criterion: the innovation check loss plus the check loss of the transformed first observations. The transformed first observations are not, in general, independent asymmetric Laplace variables, and an exact likelihood for the original observations under a parameter-dependent transformation would also require its Jacobian. The criterion is therefore not an exact stationary-data likelihood.

For both, `qregark` reports $\hat{\sigma} = N^{-1} \sum \rho_\tau(\hat{r}_i)$, the mean check loss over the N rows of the estimating system, and the asymmetric-Laplace working criterion $N \ln\{\tau(1-\tau)/\hat{\sigma}\} - N$ as a likelihood-like fit statistic (`e(ll)`).

2.2 The conditional estimator (option `corc`)

For $t = k+1, \dots, T$ in each segment, the innovations are $\eta_t(\boldsymbol{\beta}, \boldsymbol{\phi}) = \varepsilon_t - \sum_j \phi_j \varepsilon_{t-j}$, with $\varepsilon_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$. The conditional estimator minimizes

$$S_{\mathrm{CO}}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{\text{segments}} \sum_{t=k+1}^{T} \rho_\tau\{\eta_t(\boldsymbol{\beta}, \boldsymbol{\phi})\}, \tag{3}$$

so the first k observations of each segment are conditioned on and do not contribute to the objective. This is the objective of Tian et al. [11]. It is minimized with their EM algorithm, which represents the asymmetric Laplace distribution as a normal–exponential mixture: given the current estimates, the E-step yields weights proportional to $1/|\eta_t|$, and the M-step updates $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ by weighted least squares. No restriction is placed on $\lambda(\boldsymbol{\phi})$.

Staged floors. As the algorithm approaches a solution, the weights of innovations that reach zero grow without bound and can hold those innovations at zero, so that the algorithm stops short of a minimum. `qregark` therefore bounds $|\eta_t|$ below by a floor that is lowered in nine stages from 10^{-2} to 10^{-10} times the mean absolute innovation, each stage started from the estimates of the previous one and run to convergence. This helps the algorithm move off such points; it does not guarantee a global minimum, because the objective is not convex in $(\boldsymbol{\beta}, \boldsymbol{\phi})$ jointly. The stages are the lines of the iteration log.

2.3 The Prais–Winsten-type estimator (default)

The conditional estimator discards the information in the first k observations of each segment. In ITSA models this information matters, because with persistent errors the constant and the trend enter the quasi-differenced rows only through $1 - \sum_j \phi_j$ (and a correspondingly flattened trend), so the level of the series is determined largely by the first observations [14]. The default estimator therefore also includes the first k observations of each segment, transformed so that, under the model, they have the covariance structure of innovations.

Let $\boldsymbol{\Gamma}_k(\boldsymbol{\phi})$ denote the covariance matrix of $(\varepsilon_1, \ldots, \varepsilon_k)'$ under the stationary AR(k) process with unit innovation variance. It is obtained from $(\mathbf{I} - \mathbf{A} \otimes \mathbf{A})\,\mathrm{vec}(\boldsymbol{\Gamma}) = \mathrm{vec}(\mathbf{e}_1\mathbf{e}_1')$, where $\mathbf{e}_1$ is the first unit vector. The companion-state covariance is ordered in reverse time; for a stationary scalar process its symmetric Toeplitz form is also the covariance of the chronological initial-error vector. With $\boldsymbol{\Gamma}_k = \mathbf{L}\mathbf{L}'$ (Cholesky factorization), the transformed first-block residuals are

$$\mathbf{r}_F = \mathbf{L}^{-1}(\varepsilon_1, \ldots, \varepsilon_k)' - \mathbf{d}, \tag{4}$$

which for AR(1) reduces to the Prais–Winsten weight, $r_1 = \sqrt{1 - \phi_1^2}\,\varepsilon_1 - d$ [7]; for AR(k) the transformation is that of Vougas [12]. The Prais–Winsten-type estimator minimizes the augmented criterion

$$S_{\mathrm{PW}}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{\text{segments}} \Big[\sum_{j=1}^{k} \rho_\tau(r_{F,j}) + \sum_{t=k+1}^{T} \rho_\tau\{\eta_t(\boldsymbol{\beta}, \boldsymbol{\phi})\} \Big] \tag{5}$$

subject to $\lambda(\boldsymbol{\phi}) < .995$.

First-block offsets. The elements of $\mathbf{L}^{-1}(\varepsilon_1, \ldots, \varepsilon_k)'$ are uncorrelated, with variance equal to the innovation variance, but they are weighted sums of current and past innovations, so their distribution is in general not that of the innovations and their τth quantile is in general not zero. The offsets $\mathbf{d}$ in (4) are their τth quantiles. `qregark` estimates them by simulation: it generates 2,000 stationary paths of the fitted AR process driven by innovations resampled from the fitted ones, applies $\mathbf{L}^{-1}$ to the first k values, and takes the τth quantile of each component. The simulation uses its own random-number stream with a fixed seed, so results are reproducible and the user's random-number state is left unchanged. The offsets are estimated at every τ, including the median: they are zero at the median only when the innovations are symmetric.

Computation. Given $\boldsymbol{\phi}$ (and $\mathbf{d}$), the criterion (5) is a linear quantile regression of the transformed outcome on the transformed regressors, which `qregark` solves by iteratively reweighted least squares with the same staged floors as the EM algorithm. The AR coefficients are found by minimizing the resulting profile criterion with the Nelder–Mead simplex method, started from the conditional (`corc`) estimates and, for AR(1), from the best point of a grid from $-.9$ to .9 in steps of .1, or, for AR(k), also from $\boldsymbol{\phi} = \mathbf{0}$ and $\boldsymbol{\phi} = (.5, 0, \ldots, 0)'$; the best solution is refined by a restart with a smaller

simplex. The offsets are then re-estimated from the fitted innovations and the AR coefficients re-optimized locally, twice. The iteration log reports the number of function evaluations and the minimized criterion. Unlike the classical Prais–Winsten procedure, which alternates between $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, `qregark` minimizes the profile criterion over $\boldsymbol{\phi}$ directly. The criterion is not convex in $(\boldsymbol{\beta}, \boldsymbol{\phi})$ jointly, so a converged result is a candidate local solution; options `start()` and `from()` can be used to check that the estimates do not depend on the starting values.

Implementation note: Prais–Winsten as the default. In simulations of eleven single-group ITSA designs (Section 7, Table 1, study A), the Prais–Winsten-type estimator estimated the level and pre-intervention trend of the series much more precisely than the conditional estimator. With the small-sample bias correction of Section 5 applied to both, the median root mean squared error across designs was 1.81 against 10.12 for the intercept and .168 against .635 for the pre-intervention trend. This agrees with the results of Park and Mitchell [14] for least squares with trending regressors and of Dielman and Rose [13] and Dielman [15] for least absolute value regression.

Implementation note: offsets at the median. With skewed (χ^2_2) innovations centered at their median, 30 observations, AR(1) errors with $\phi = .9$, and $\tau = .5$ (Table 1, study B), setting the offsets to zero biased the pre-intervention trend by $-.137$ (root mean squared error .452), against $-.036$ (.418) with simulated offsets. With $\phi = .5$ the differences were small. With symmetric innovations, estimating the offsets at the median changes the estimates only in the last digits.

3 The intercept

In model (1), the τth quantile of the innovations η_t is zero, but that of the stationary AR errors ε_t generally is not. The stationary errors are a weighted sum of current and past innovations, so their τth quantile depends on the innovation distribution and, strongly when the errors are persistent, on the AR coefficients. The structural intercept β_0 of (1) therefore is not the starting level of the τth quantile of the outcome.

Write $\mathbf{x}_t = (\mathbf{x}'_{1t}, 1)'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \beta_0)'$. `qregark` reports as `_cons` the *baseline level*

$$\hat{b}_0 = \hat{Q}_\tau\big(y_t - \mathbf{x}'_{1t}\hat{\boldsymbol{\beta}}_1\big), \tag{6}$$

the sample τth quantile over the estimation sample of the outcome minus the terms of the other regressors, which estimates $\beta_0 + Q_\tau(\varepsilon)$, the τth quantile of the outcome when all other regressors are zero. When time is coded zero at the intended baseline and the other regressors are at their reference values there, as in the usual ITSA parameterization [3, 16], this is the starting level of the τth quantile of the series. It plays the role that `_cons` plays in `prais`, where the errors have mean zero. It avoids the particular sensitivity of the structural intercept to the estimated persistence, although it still depends on the persistence indirectly, through the estimated slopes. Comparisons of baseline levels across groups also require that the groups share the error distribution assumed by the model (a common innovation distribution and common AR coefficients); the definition of the intercept alone does not establish comparability. At the median with symmetric innovations the two intercepts estimate the same quantity. The slope coefficients, including group contrasts and intervention effects, are the same under either definition. The structural intercept $\hat{\beta}_0$ is stored in `e(b0_struct)` and is used by `predict` for the one-step quantile (2) and the innovations.

Standard error. The variance of $\hat{b}_0$ is approximated by

$$\widehat{\mathrm{Var}}(\hat{b}_0) = v_q + \bar{\mathbf{x}}'_1\, \widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_1)\, \bar{\mathbf{x}}_1, \qquad \widehat{\mathrm{Cov}}(\hat{b}_0, \hat{\boldsymbol{\theta}}_j) = -\bar{\mathbf{x}}'_1\, \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\theta}}_j), \tag{7}$$

where $\bar{\mathbf{x}}_1$ is the sample mean of $\mathbf{x}_{1t}$, $\hat{\boldsymbol{\theta}}_j$ is any other coefficient, and v_q is the sampling variance of the sample τth quantile of the fitted error process. v_q is estimated by simulating 1,000 replicates of the AR errors with the estimation sample's segments (100-observation burn-in per segment), driven by innovations resampled from the fitted ones, and taking the variance of their sample τth quantiles; the simulation uses its own random-number stream with a fixed seed. (7) treats the sample quantile and the slope estimates as uncorrelated. With `vce(bootstrap)`, (6) is instead recomputed in each replication.

Small-sample adjustment (option `biascorrect`). With persistent errors, the sample quantile in (6) has a finite-sample bias, even though the process is stationary: a short window of a persistent series covers little of its stationary distribution, and the slope terms are removed as estimated, which absorbs part of the error. With `biascorrect`, `qregark` estimates this bias from the same simulated error replicates: in each replicate, the slope terms are removed by a Gaussian (least-squares) projection of the transformed simulated errors on the transformed regressors with $\boldsymbol{\phi}$ fixed at its estimate, and the sample τth quantile of the result is computed. The bias is the mean of these quantiles minus the τth quantile of all simulated errors pooled, and it is subtracted from $\hat{b}_0$. The unadjusted value is stored in `e(b0_unadjusted)`.

Both the adjustment for small-sample bias and the variance approximation (7) are specific to `qregark` and are supported by simulation only (Section 7).

4 Standard errors

4.1 Default variance (`vce(iid)`)

Let $r_i(\boldsymbol{\beta}, \boldsymbol{\phi})$, $i = 1, \ldots, N$, denote the residuals of the estimating system: the innovations η_t and, for the Prais–Winsten-type estimator, the transformed first-block residuals (4). Let $\mathbf{G}$ be the $N \times (p + k)$ matrix whose ith row is the derivative of r_i with respect to $(\boldsymbol{\beta}', \boldsymbol{\phi}')$, where p is the number of regression coefficients. For the innovations this row is $(\mathbf{x}_t - \sum_j \phi_j \mathbf{x}_{t-j},\ \varepsilon_{t-1}, \ldots, \varepsilon_{t-k})$ up to sign; for the first-block rows the derivative with respect to $\boldsymbol{\beta}$ is $-\mathbf{L}^{-1}$ applied to the corresponding rows of $\mathbf{X}$, and that with respect to $\boldsymbol{\phi}$ is computed numerically. `qregark` uses

$$\widehat{\mathbf{V}} = \tau(1 - \tau)\, \hat{s}^2\, (\mathbf{G}'\mathbf{G})^{-1}, \tag{8}$$

where $s = 1/f_\eta(0)$ is the sparsity, the reciprocal of the innovation density at zero.

Status of (8). For the conditional estimator, (8) is the usual asymptotic variance of a linear quantile regression [9] applied to the innovation equations, valid under i.i.d. innovations. For the Prais–Winsten-type estimator it is a working approximation: it applies one sparsity to all rows. Under the stationary AR model the transformed first observations are independent of the subsequent innovations $\eta_{k+1}, \ldots, \eta_T$, but the components within the transformed first block need not be mutually independent, and their marginal distributions, and densities at zero, need not match those of the innovations. The approximation is motivated by an asymptotic setting with a fixed number of long segments, in which the first-block rows are a vanishing fraction of the sample and (8) reduces to the conditional-estimator form. With short segments, or many of them, the first-block rows are a substantial fraction of the sample and that argument does not apply; the adequacy of (8) there rests on the simulation evidence of Section 7.

Sparsity. As in `qreg, vce(iid, residual)`, the sparsity is estimated by the difference quotient of the residual quantiles,

$$\hat{s} = \frac{\hat{Q}_{\tau+h}(r) - \hat{Q}_{\tau-h}(r)}{2h},$$

computed after removing the residuals that define the solution, with bandwidth h chosen by the Hall–Sheather (default), Bofinger, or Chamberlain rule (option `bwidth()`). The Prais–Winsten-type optimum over $\boldsymbol{\phi}$ is located by Nelder–Mead with a tolerance of about 10^{-6}, so its solution-defining residuals are zero only to about that precision; residuals smaller than 10^{-5} times the mean absolute residual are therefore removed for that estimator, against 10^{-7} for the conditional estimator. As in `qreg`, the sparsity cannot be estimated when $\tau \pm h$ falls outside $(0, 1)$, which can happen at extreme quantiles in short series.

With a constant, the row and column of the intercept are then replaced by (7). The Wald statistic `e(chi2)` tests the regression coefficients other than the constant.

4.2 Residual bootstrap (`vce(bootstrap)`)

The innovations are resampled with replacement, the AR errors are rebuilt recursively within each segment from the observed initial errors, and the model is refit to $y_t^* = \mathbf{x}_t'\hat{\boldsymbol{\beta}} + \varepsilon_t^*$. This is a conditional-initialization bootstrap: the first k errors of each segment are held at their observed values, so the bootstrap describes uncertainty conditional on them and does not reproduce their sampling variation over stationary initial blocks. For the Prais–Winsten-type estimator, which uses those observations, this is a material qualification, particularly with short or many segments. The conditional estimator is refit by the EM algorithm from the full-sample estimates. The Prais–Winsten-type estimator is refit by a local Nelder–Mead search started at the full-sample estimates; the first-block offsets are then re-estimated from the replicate's innovations and the search repeated. The offsets shift the transformed first observations and hence the intercept, so holding them at their full-sample values would understate the variability of the intercept. The baseline level (6) is recomputed in each replication. The reported variance is the sample variance of the replicates, and percentile confidence intervals are stored in `e(ci_percentile)`. The default of 500 replications follows Tian et al. [11].

5 Small-sample bias correction

5.1 Motivation

In short series, and particularly with the trend and intervention terms of an ITSA model, the estimated AR coefficients are biased toward zero; in our simulations by about $-.2$ to $-.4$ for $\phi = .8$ with 30 observations. Separately, in our simulations, even errors with a unit root typically gave estimates near .6 with 30 observations. The direction of the bias is well known for autoregressive estimates: for the least-squares estimator of a stationary AR(1) process with an estimated mean, the classical approximation is $-(1 + 3\phi)/n$ [17, 18], and the bias is larger when a time trend is also estimated [19]. These results are for least-squares estimators, not for the joint quantile estimator of `qregark` or for general AR(k) models; they explain the direction and rough size of the bias, which for `qregark` was established by simulation. The estimated persistence then understates the uncertainty in the level and trends, whose standard errors are too small and whose confidence intervals undercover. Option `biascorrect` corrects the AR coefficients and carries the correction into the standard errors.

5.2 Calibration

The correction is a simulation-based calibration in the spirit of indirect inference [20, 21]. Let $\hat{\boldsymbol{\phi}}$ be the estimate, and let $\mathbf{m}(\boldsymbol{\phi})$ be the mean of a surrogate estimator applied to series simulated with true AR coefficients $\boldsymbol{\phi}$. The corrected estimate $\tilde{\boldsymbol{\phi}}$ is an approximate, projected calibration targeting $\mathbf{m}(\tilde{\boldsymbol{\phi}}) = \hat{\boldsymbol{\phi}}$ within the region described below, computed by the iteration of MacKinnon and Smith [21, eqs. 15–16],

$$\boldsymbol{\phi}^{(i+1)} = \Pi\big\{\boldsymbol{\phi}^{(i)} + \hat{\boldsymbol{\phi}} - \mathbf{m}(\boldsymbol{\phi}^{(i)})\big\}, \qquad \boldsymbol{\phi}^{(0)} = \Pi(\hat{\boldsymbol{\phi}}), \tag{9}$$

run for eight iterations, where Π restricts the result to that region. Eight iterations do not guarantee that the equation is solved, and an active restriction may prevent it from being solved exactly; the remaining discrepancy $\mathbf{m}(\tilde{\boldsymbol{\phi}}) - \hat{\boldsymbol{\phi}}$ is not currently stored.

- *Surrogate estimator.* The quantile estimator itself would be much too costly to simulate. $\mathbf{m}(\cdot)$ is therefore the mean of a Gaussian least-squares estimator of the same form: iterated Prais–Winsten for the default estimator, iterated Cochrane–Orcutt with `corc`, with $\boldsymbol{\phi}$ updated by regressing the residuals on their lags, as `prais` does. This assumes that the least-squares and quantile estimators have similar small-sample bias, an assumption assessed by simulation: it is supported for symmetric innovations and, with skewed innovations, near the median, but not in the tails of a skewed distribution, where the correction over-corrects in the short tail and under-corrects in the long tail (Section 7).
- *How $\mathbf{m}(\boldsymbol{\phi})$ is computed.* To evaluate $\mathbf{m}$ at a candidate $\boldsymbol{\phi}$, `qregark` generates 500 artificial error series from an AR process with coefficients $\boldsymbol{\phi}$, fits the surrogate estimator to each one with the user's own regressors (the trend, intervention, and any other terms) and the same segments as the data, and averages the 500 estimated AR coefficients. This is the standard simulation estimate of the binding function in indirect inference [20, 21]: samples of the same form as the data, simulated with the observed regressors, with the estimator of interest applied to each. Using the model's own design matters because the bias depends on it: on the series lengths, the gaps and panels, and especially on the trend and intervention terms that are estimated alongside the AR coefficients. Because the surrogate is a least-squares estimator, the simulated innovations are standard normal; each segment is preceded by a 50-observation burn-in, so the simulated errors start approximately from the stationary distribution assumed by the model (less closely for coefficients very near the stationarity limit). The same 500 sets of random numbers are reused at every candidate $\boldsymbol{\phi}$ (common random numbers). Differences between $\mathbf{m}$ at two values of $\boldsymbol{\phi}$ then reflect the change in $\boldsymbol{\phi}$ rather than new simulation noise, which keeps the iteration (9) and the derivative $\mathbf{J}$ used for the variance below stable. The random numbers are generated from a fixed seed set inside the program, so the correction is the same every time the model is fit, and the user's random-number state is restored afterwards.
- *Stationarity.* Π scales $\boldsymbol{\phi}$ toward zero, if needed, so that $\lambda(\boldsymbol{\phi}) < .99$, and then applies the calibration bound below.

Calibration bound. In very short or very persistent series $\mathbf{m}(\boldsymbol{\phi})$ can be only weakly sensitive to the true persistence, so the calibration cannot distinguish moderate from strong persistence and (9) can move far from the estimate. `qregark` therefore limits the sum of the corrected coefficients. Along the direction of the estimate, $\boldsymbol{\phi} = c\,\hat{\boldsymbol{\phi}}/\sum_j \hat{\phi}_j$, it evaluates the sum of $\mathbf{m}$ at $c = .5, .6, .7, .8, .85, .9, .95, .99$ and takes as the bound the largest c up to which the slope of the sum of $\mathbf{m}$ with respect to c remains at least .5. The bound is never below $\sum_j \hat{\phi}_j$, and no bound

is applied when $\sum_j \hat{\phi}_j \leq 0$. When the bound or the stationarity restriction binds at the solution, `e(capped)` is 1, the log shows "(at bound)", and the bound is stored in `e(phi_bound)`. This is a safeguard on the sum, not a test of identification of each coefficient. It restrains the correction in a way similar in purpose to the stationarity adjustment of Kilian [22], but it is a rule specific to `qregark`.

No correction at the limit. If $\lambda(\hat{\boldsymbol{\phi}}) \geq .99$, no correction is made: the uncorrected coefficients are kept, `e(capped)` is 1, and the variance matrix is not adjusted. Because the stationarity restriction of the correction ($\lambda < .99$) is tighter than that of the Prais–Winsten-type estimator ($\lambda < .995$), a "correction" of an estimate between the two could only lower the persistence; in a controlled ITSA of the California tobacco-control data [23], for example, it would have moved λ from .995 to .990.

Re-estimation. With $\tilde{\boldsymbol{\phi}}$ held fixed, the regression coefficients are re-estimated: by the staged algorithm on the transformed system, with the offsets re-estimated, for the default estimator, and by the EM algorithm with $\boldsymbol{\phi}$ fixed with `corc`. The uncorrected coefficients are stored in `e(rho_uncorrected)`, and the baseline level is adjusted as described in Section 3.

5.3 Variance after the correction

The variance matrix is adjusted by a regularized, local delta-method approximation. Write $\mathbf{J} = \partial \mathbf{m}/\partial \boldsymbol{\phi}'$ at $\tilde{\boldsymbol{\phi}}$, estimated by forward differences (step .02) with the same random numbers. At an interior exact solution of the calibration equation, with a locally nonsingular Jacobian of the binding function, the derivative of the inverse map $\mathbf{m}^{-1}$ at $\hat{\boldsymbol{\phi}}$ is $\mathbf{J}^{-1}$ [21, eq. 29]; (10) applies this derivative as a local approximation to the implemented finite-iteration calibration. The regression coefficients, re-estimated given $\tilde{\boldsymbol{\phi}}$, move with it at the rate $\mathbf{B} = -\mathbf{H}_{\beta\beta}^{-1}\mathbf{H}_{\beta\phi}$, as for two-step estimators [24, Theorem 6.1], where $\mathbf{H} = \mathbf{G}'\mathbf{G}$ approximates the curvature of the fitted criterion. The adjusted variance is

$$\widetilde{\mathbf{V}} = \mathbf{A}\widehat{\mathbf{V}}\mathbf{A}', \qquad \mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{B}(\mathbf{J}_*^{-1} - \mathbf{I}) \\ \mathbf{0} & \mathbf{J}_*^{-1} \end{pmatrix}, \tag{10}$$

with $\widehat{\mathbf{V}}$ from (8). In the settings considered, the estimated calibration map can become weakly sensitive to $\boldsymbol{\phi}$ or ill-conditioned, so that $\mathbf{J}^{-1}$ would be very large. $\mathbf{J}_*$ is therefore $\mathbf{J}$ with its singular values floored at .5, the slope that defines the calibration bound, which bounds the operator norm of $\mathbf{J}_*^{-1}$ by 2. This limits the inflation of the AR-coefficient block of the variance, but not that of the regression coefficients: through the term $\mathbf{B}(\mathbf{J}_*^{-1} - \mathbf{I})$, their variances can increase by much more, as in the capped fits of Section 7.

Two further qualifications apply. When the calibration bound or the stationarity restriction is active, $\tilde{\boldsymbol{\phi}}$ is not $\mathbf{m}^{-1}(\hat{\boldsymbol{\phi}})$, and (10) does not differentiate the constrained procedure; it remains an unconstrained local approximation. And the delta-method approximation is not exact near a unit root [25]. Its accuracy was assessed by simulation. Option `biascorrect` may not be combined with `vce(bootstrap)`.

6 Diagnostics and stored flags

After estimation, `qregark` displays the autocorrelations at lags $1, \ldots, k$ of the residuals $y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$ and of the innovations $\hat{\eta}_t$, computed within segments, as `praisk` [26] and `betark` [27] do; they are stored in `e(ac_res#)` and `e(ac_inn#)`. The residual autocorrelations show the serial dependence that the AR model accounts for; the innovation autocorrelations should be near zero if the order

k is adequate. The innovations can be examined further with `predict, innovations` and Stata's `ac` command.

Implementation note: no portmanteau test. A Ljung–Box test of the innovations and a score (Lagrange multiplier) test of omitted AR terms were evaluated by simulation (Table 1, study E), with the fitted order equal to the true order. At nominal 5%, the Ljung–Box test rejected in 4% to 15% of samples with 30 to 100 observations. The score test had about the nominal size in single series, but rejected in 17% to 27% of samples in panels of ten 20-observation series, and with 30 observations it detected an omitted second lag ($\phi = (.5, .3)$, AR(1) fitted) in only 2% to 6% of samples. Neither test is reported, which also keeps `qregark` consistent with `praisk` and `betark`.

The following flags are stored; their interpretation is described here and in the help file rather than in warnings printed with the output, because the conditions call for interpretation rather than a specific action. The non-convergence warning and the note for a nonstationary estimate are printed.

`e(converged)`
1 if the EM algorithm and, for the default estimator, the Nelder–Mead search converged.

`e(stationary)`
1 if $\lambda(\hat{\boldsymbol{\phi}}) < 1$, that is, all eigenvalues of the companion matrix lie inside the unit circle. The conditional estimator does not restrict λ; the Prais–Winsten-type estimator restricts it to below .995.

`e(at_limit)`
1 if λ of the uncorrected AR coefficients is at the limit of .995. This is an optimization diagnostic, not a test for a unit root: the fitted persistence has reached the imposed numerical boundary, where ordinary interior-point inference is unreliable. The standard errors of the AR coefficients are then not valid, and the level terms and trends are imprecisely estimated. This happens mainly with long series or pooled panels; in short series with trend terms the downward bias described in Section 5 keeps the estimates well below the limit.

`e(capped)`
With `biascorrect`, 1 if the calibration bound or the stationarity restriction binds, or if no correction was made because $\lambda(\hat{\boldsymbol{\phi}}) \geq .99$; `e(phi_bound)` holds the bound.

`e(N_misreps)`
With `vce(bootstrap)`, the number of replications whose refit did not converge.

7 Simulation evidence

Table 1 lists the simulation studies referred to in this document, and Tables 2 and 3 summarize the main results for the bias correction. Except for the panel scenarios in study E, the data were single-group ITSA series

$$y_t = 10 + .5\,t - 3\,x_t - .4\,x_t(t - T_0) + \varepsilon_t, \qquad t = 0, \ldots, n-1, \quad x_t = I(t \geq T_0), \quad T_0 = n/2, \quad (11)$$

with AR errors ε_t started from a burn-in of 100 to 200 observations so that the series are stationary, and innovations shifted so that their τth quantile is zero. In the panel scenarios of study E, each of the 2 or 10 series followed (11) with the same intervention time, slopes, and AR coefficients, but its own intercept $(10, 11, \ldots)$ and its own independently generated AR errors, and the model was fit with panel-specific intercepts. In one scenario the series instead shared the intercept 10, and a

common intercept was fitted. The true baseline level is $10+Q_\tau(\varepsilon)$, computed from a long stationary simulation (200,000 observations in study C). Fits that did not converge were excluded, so coverage, bias, and root mean squared error are computed over the converged fits of each design. With 500 replications, the Monte Carlo standard error of a coverage rate near .95 is about .010, so departures of a few points are within simulation noise, but coverage of .83 is meaningful undercoverage.

Table 1: Simulation studies.

Study	Purpose	n	AR coefficients	Innovations, τ	Reps
A	PW vs. `corc`	30, 50	AR(1)–AR(3)[a]	normal, t_3, χ^2_2; .25, .5, .75	150
B	offsets at the median	30	.5, .9	χ^2_2; .5	300
C	`biascorrect`	30, 50, 100	.5, .8, .9, (.5, .3)	normal, t_3; .25, .5, .75	500
D	bootstrap-after-bootstrap	30, 50	.5, .9	normal; .25, .5	100–200[b]
E	portmanteau tests	20–100	.3, .6, .9, (.5, .2), (.7, .2)[c]	normal, t_3, χ^2_2; .25, .5	200–400
F	`biascorrect`, skewed innovations	30, 50	.5, .8, .9, (.5, .3)	χ^2_2; .25, .5, .75	500

[a] Eleven designs, written (n; $\boldsymbol{\phi}$; τ; innovations): (30; .5; .5; normal), (30; .8; .5; normal), (30; .8; .5; t_3), (30; .8; .75; normal), (30; .9; .5; normal), (30; (.5,.3); .5; normal), (50; .8; .25; χ^2_2), (50; .8; .5; normal), (50; .9; .5; normal), (50; (.5,.3); .5; normal), and (50; (.4,.2,.2); .5; normal). Study C is the full factorial of its n, $\boldsymbol{\phi}$, innovations, and τ (72 designs). [b] With 99 or 149 bootstrap replications per sample. [c] Power against an omitted lag with $\phi = (.5, .3)$ and AR(1) fitted. Study E also included panels of 2 and 10 series.

Table 2: Study C: coverage of nominal 95% intervals (range over the 24 designs at each n and the five coefficients: intercept, pre- and post-intervention trends, level and slope changes), share of `biascorrect` fits that were capped, and non-converged fits.

	$n = 30$	$n = 50$	$n = 100$
Coverage without `biascorrect`	.68–.97	.76–.96	.82–.97
Coverage with `biascorrect`	.83–.98	.91–.98	.92–.98
Share capped, all designs (mean)	.56	.41	.25
Share capped, $\phi = .9$	.61–.81	.50–.77	.43–.66
Fits not converged (of 12,000)[a]	2	2	1

Coverage uses normal critical values; with t critical values it is higher by at most .03. [a] Per estimator; each design had 500 replications, so coverage is computed over 499 or 500 fits.

Bias correction. Across the designs of study C, `biascorrect` raised the lowest coverage of nominal 95% intervals from .68 to .83 with 30 observations, from .76 to .91 with 50, and from .82 to .92 with 100 (Table 2). The cost in precision was modest: the median ratio of root mean squared errors with and without the correction, over designs and coefficients, was 1.01 for $\phi = .9$, 1.08 for $\phi = .8$, 1.06 for $\phi = (.5, .3)$, and 1.11 for $\phi = .5$ (range .85 to 1.47). The correction therefore improves coverage substantially but does not achieve nominal coverage in all designs: with 30 observations and $\phi = .9$ or $(.5, .3)$, coverage of .83 to .92 for the intercept and .87 to .95 for the pre- and post-intervention trends remains below .95 in most designs.

Capped and uncapped fits. Coverage also differs between subsets of samples. With 50 observations, $\phi = .9$, and $\tau = .25$, samples in which the calibration bound was reached had very conservative intervals (median standard error of the intercept 23.8, against a sampling standard

Table 3: Study F (χ^2_2 innovations, which are right-skewed, so that $\tau = .25$ lies in the short tail and $\tau = .75$ in the long tail): coverage of nominal 95% intervals (range over the four AR designs and five coefficients), mean bias of the sum of the AR coefficients, share of capped fits, and median ratio of root mean squared errors with and without `biascorrect`.

		$\tau = .25$	$\tau = .5$	$\tau = .75$
Coverage without `biascorrect`	$n = 30$	.77–.96	.75–.94	.69–.95
	$n = 50$	.78–.96	.81–.94	.76–.96
Coverage with `biascorrect`	$n = 30$	.82–.99	.87–.99	.80–.97
	$n = 50$	.84–1.00	.93–.98	.88–.97
Bias of $\sum_j \hat{\phi}_j$ without		$-.22$ to $-.03$	$-.46$ to $-.09$	$-.74$ to $-.15$
Bias of $\sum_j \hat{\phi}_j$ with		$-.08$ to $+.21$	$-.19$ to $+.12$	$-.42$ to .00
Share capped		.01–.95	.02–.81	.08–.43
RMSE ratio (median; maximum)		1.10; 1.75	1.04; 1.30	1.07; 1.28

Coverage uses normal critical values. One fit per estimator (of 12,000) did not converge.

deviation of 2.64; coverage 1), and the other samples had intervals that were too narrow (coverage .83 for the intercept and .87 for the pre-intervention trend). Across designs, the median standard errors of capped fits were 6 to 10 times the sampling standard deviation. The pooled coverage rates of Table 2 thus average over two groups of fits whose intervals err in opposite directions.

Skewed innovations. Study F repeated the design of study C with χ^2_2 innovations (Table 3). The quantile estimator's small-sample bias then depends on where τ lies in the innovation distribution: in the short (lower) tail, at $\tau = .25$, the uncorrected AR coefficients were only mildly biased, whereas in the long (upper) tail, at $\tau = .75$, they were biased more strongly than with symmetric innovations. Because the Gaussian surrogate does not reflect this, `biascorrect` over-corrected at $\tau = .25$, by up to $+.21$ in the sum of the AR coefficients when $\phi = .5$; the intervals for the intercept and trends were then conservative (coverage .91 to 1.00 for the intercept), and the root mean squared error increased by up to 75%. At $\tau = .75$ the correction removed only part of the bias; coverage rose from as little as .69 to .80 or more but remained below nominal in the most persistent designs. At the median the correction behaved much as with symmetric innovations, raising the lowest coverage from .75 to .87 with 30 observations and from .81 to .93 with 50. Away from the median, the lowest coverage with `biascorrect` (.82 at $\tau = .25$) was for the level change, consistent with item 3 of Section 9.

Alternatives evaluated. Two alternatives covered less well than the approach adopted. The bootstrap-after-bootstrap intervals of Kilian [22] (study D), with the bias-corrected coefficients and the stationarity adjustment in each replication, covered .77 to .93 for the trends and the level change with 30 to 50 observations, against .91 to 1.00 for the delta-method intervals in the same samples, and were narrower. A rule that applied the correction only when the uncorrected sum of AR coefficients was between .3 and .95 covered .76 to .88 for the intercept with 30 observations, normal innovations, and $\phi = .9$ or AR(2) errors, against .84 to .92 when the correction was always applied; in short series the uncorrected coefficients are themselves biased toward zero and are a poor guide to whether the correction is needed. With weakly correlated errors the correction costs little (coverage .95 to .97 with $\phi = .5$ and 30 observations).

8 Practical advice

Choosing the estimator. Use the default Prais–Winsten-type estimator for ITSA and other trending series: it estimates the level and trends much more precisely than the conditional estimator when the errors are persistent (Section 2). Use `corc` for the conditional objective of Tian et al. [11]. The conditional estimator does not impose stationarity, but the baseline-level interpretation and the simulation-based adjustments of this document assume a stationary error process (Section 9, item 6).

Choosing the AR order. Start with the order suggested by the data (for example, $k = 1$ for many annual series, or a seasonal order such as $k = 4$ or $k = 12$ for quarterly or monthly series when seasonality is not otherwise modeled), and check the innovation autocorrelations in the output, which should be near zero. For a fuller look, obtain the innovations with `predict, innovations` and examine them with `ac` (adding `if e(sample)`). No portmanteau test is reported, because none was accurate in short series or panels (Section 6).

Interpreting `_cons`. `_cons` is the baseline level of the τth quantile of the outcome, not the structural intercept of the AR model (Section 3). With time coded zero at the start of the series and the other regressors at their reference values, it is the estimated starting level of the τth quantile, as `_cons` is the starting level of the outcome in the usual ITSA regression model [3, 16]. It can be compared across quantiles, across estimators, and with `_cons` from `prais` or `praisk` [26]. In a multiple-group ITSA, it is the starting level of the reference group; the group contrasts and all intervention effects do not depend on how the intercept is defined, and comparisons of levels across groups assume a common error distribution. Post-intervention trends are obtained as usual, for example with `lincom _t + _x_t`. The structural intercept, if needed, is in `e(b0_struct)`; `predict, quantile` and `predict, innovations` use it automatically, whereas `predict, xb` and `predict, residuals` use `_cons`.

Using `biascorrect`.

- Within the settings evaluated (single-group ITSA series of 30 to 100 observations with AR(1) or AR(2) errors and symmetric innovations, and of 30 or 50 observations with skewed innovations), the correction improved coverage when the time segments were short and the errors autocorrelated, although coverage remained below nominal in some designs. It is therefore recommended in such settings. Do not decide from the size of the uncorrected AR coefficients: in short series they are biased toward zero, and a rule based on them covered less well than always correcting (Section 7).
- Expect the intervention effects to change little and the confidence intervals for the level and trends to widen: the correction mainly repairs the standard errors. The uncorrected AR coefficients are in `e(rho_uncorrected)`.
- Check `e(capped)`. If it is 1, the calibration bound was reached and the intervals for the level and trends are likely to be much wider than needed; if it is 0 in a very short, persistent series, they may still be too narrow.
- With skewed innovations, the correction over-corrects at quantiles in the short tail of the distribution (for example, a right-skewed outcome at $\tau = .25$), giving conservative intervals, and under-corrects in the long tail. Report results with and without `biascorrect` as a sensitivity analysis when the two differ materially.

- The correction makes little difference with long segments, and none when the AR coefficients are at the limit.

When `e(at_limit)` is 1. The fitted persistence has reached the imposed boundary, which happens mainly with long series or pooled panels. This is not a unit-root test, but ordinary inference is unreliable at the boundary: the standard errors of the AR coefficients are not valid, and the level terms and trends are imprecisely estimated. Reconsider the specification, for example whether the mean structure omits a trend or level shift that the errors are absorbing, or whether a model in first differences is more appropriate.

Quantiles away from the median. At extreme quantiles in short series the sparsity, and hence the default standard errors, may not be computable (for example, $\tau = .9$ with fewer than about 50 observations, or $\tau = .75$ with fewer than about 30). Use a quantile closer to the median or `vce(bootstrap)`, keeping in mind that the bootstrap tends to undercover at extreme quantiles. The estimated level change is also somewhat biased away from the median (Section 9).

Checking the solution. Because the criterion is not convex, confirm that `e(converged)` is 1, and refit with `start(median)` or `from()` when the estimates look implausible.

Reproducibility. The first-block offsets, the variance of the baseline level, and the bias correction use internal random-number seeds, so repeated runs give identical results and the user's random-number state is unchanged. With `vce(bootstrap)`, specify `seed()`.

9 Limitations

Each limitation is followed by its practical implications for current use. Where a check is suggested, it is a sensitivity analysis: disagreement between two fits indicates that the results are sensitive to the choice, not which fit is correct. Possible extensions of the program are collected in Section 9.1.

1. Inference for the Prais–Winsten-type estimator rests on a working variance (8) and, with `biascorrect`, on a regularized local approximation (10); neither is an exact asymptotic result for short segments, and their adequacy is supported by simulation only. *Practical implications:* treat the intervals as approximate, particularly in short series, and consider comparing the default standard errors with those from `vce(bootstrap)` (see item 8).

2. With about 30 observations and strong persistence (ϕ around .9, or AR(2) errors with a sum of coefficients around .8), the calibration bound binds in 53% to 81% of the samples, and the intervals for the intercept (coverage .83 to .92) and the trends (.87 to .95) remain too narrow in most designs even with `biascorrect`. Capped fits are likely to have intervals much wider than needed, and uncapped fits intervals that are too narrow. *Practical implications:* check `e(capped)`. In the simulations, the intervals of capped fits were conservative and those of uncapped fits in very short, persistent series were too narrow; longer pre- and post-intervention periods reduce both problems.

3. At $\tau \neq .5$ the estimated level change is biased, upward at low quantiles and downward at high quantiles; in the simulations, by a tenth to a third of its standard error. With persistent errors the first observation after the intervention carries much of the information about the step. `biascorrect` increases this bias somewhat; coverage remained above .9. *Practical implications:*

report the contrasts that answer the research question (the immediate level change, the effect at a specified post-intervention time, or a cumulative effect summed over several periods), preferably prespecified, with their uncertainty. These are different estimands; the simulations do not show that any of them is less biased than the level change, and a contrast that combines the level and slope changes can inherit the bias of both.

4. The bias correction uses a Gaussian least-squares surrogate. With skewed innovations, the small-sample bias of the quantile estimator depends on where τ lies in the distribution. In the short tail (for example, χ^2 innovations at $\tau = .25$) the quantile estimator is much less biased than the least-squares estimator, and `biascorrect` over-corrects, giving conservative intervals and a larger root mean squared error; in the long tail ($\tau = .75$) it under-corrects, and intervals in short, persistent series remain too narrow (Table 3). *Practical implications:* examine the distribution of the innovations (`predict, innovations`); when it is clearly skewed and τ lies in its short or long tail, report results with and without `biascorrect` as a sensitivity analysis.

5. The first-block transformation (4) gives the transformed initial errors the innovation variance and removes their pairwise covariance. It does not generally reproduce the innovation distribution: the offsets align their τth quantiles with zero but do not otherwise equate their distributions. Exact distributional matching holds in the zero-mean Gaussian case, including Gaussian median regression, but not generally under the innovation-quantile normalization at other quantiles. With long segments this affects a small fraction of the sample; with short segments the first observations are a substantial and, by design, influential fraction of it. *Practical implications:* with short or many segments, comparing the default estimates with those from `corc` shows how much the results depend on the first observations. A difference may reflect the information those observations are intended to contribute, not only the approximation, so it indicates sensitivity rather than error.

6. The Prais–Winsten-type estimator requires a stationary AR process. The conditional estimator can be computed without imposing stationarity, but a nonstationary fit requires separate consideration of the estimand and of inference: the stationary baseline-level interpretation (Section 3) and the simulation-based adjustments described here should not automatically be applied. *Practical implications:* if the errors may be nonstationary, or `e(at_limit)` is 1, reconsider the specification, for example whether the mean structure omits a trend or level shift, or whether a model in first differences is more appropriate.

7. The criterion is not convex jointly in $(\boldsymbol{\beta}, \boldsymbol{\phi})$, so a converged result is a candidate local solution, not necessarily the global minimum. *Practical implications:* confirm that `e(converged)` is 1 and refit with `start(median)` and with `from()` at other values, including the `corc` estimates; prefer the solution with the smallest criterion value (`e(sum_adev)`).

8. The bootstrap holds the initial errors of each segment fixed, so it does not reflect their sampling variation; it tends to undercover at extreme quantiles and is not available with `biascorrect`. *Practical implications:* the bootstrap intervals condition on the observed initial errors and should not be interpreted as providing validated unconditional coverage for baseline levels in short series.

9. Panels share the regression and AR coefficients, and no correlation across panels is modeled. *Practical implications:* when panels may differ in persistence, fitting them separately (or checking the innovation autocorrelations by panel) shows whether a common AR process is reasonable. Cross-panel dependence requires a design or covariance procedure that explicitly accommodates it; a contrast with a control series may remove common shocks when the groups have comparable

exposure to them, but this is an additional assumption.

9.1 Future development

The following extensions could address some of the limitations above. Each would require derivation and evaluation before it could be relied on.

- An unconditional bootstrap that draws the initial errors of each replication from the fitted stationary distribution (items 1 and 8).
- A variance for the Prais–Winsten-type estimator that treats the first-block rows separately, for example with their own sparsity; this alone would not account for the dependence within the block or for the estimation of its offsets (item 1).
- Intervals for the level and trends that treat the persistence as a nuisance parameter, inspired by the grid bootstrap of Hansen [28], which inverts bootstrap tests for an autoregressive parameter; extending this to intervention contrasts would require an additional construction (item 2).
- A simulation-based adjustment of the level change away from the median, analogous to that of the baseline level (item 3).
- A calibration that uses the quantile estimator itself, or a surrogate with resampled rather than Gaussian innovations, at a higher computational cost (item 4).
- Simulation of the full distribution of the transformed first observations rather than only their quantile offsets (item 5).
- Panel-specific AR coefficients (item 9).

References

[1] D. T. Campbell, J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago, 1966.

[2] W. R. Shadish, T. D. Cook, D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston, 2002.

[3] A. Linden, Conducting interrupted time-series analysis for single- and multiple-group comparisons, *Stata Journal* 15 (2) (2015) 480–500. doi:10.1177/1536867X1501500208.

[4] A. Linden, Estimating the effect of regression to the mean in health management programs, *Disease Management and Health Outcomes* 15 (1) (2007) 7–12.

[5] A. Linden, J. Adams, N. Roberts, The generalizability of disease management program results: getting from here to there, *Managed Care Interface* 17 (7) (2004) 38–45.

[6] J. Lopez Bernal, S. Soumerai, A. Gasparrini, A methodological framework for model selection in interrupted time series studies, *Journal of Clinical Epidemiology* 103 (2018) 82–91. doi:10.1016/j.jclinepi.2018.05.026.

[7] S. J. Prais, C. B. Winsten, Trend estimators and serial correlation, Cowles Commission Discussion Paper: Statistics No. 383, Cowles Commission, Chicago, 1954.

[8] R. Koenker, G. Bassett, Regression quantiles, *Econometrica* 46 (1) (1978) 33–50.

[9] R. Koenker, *Quantile Regression*, Cambridge University Press, New York, 2005.

[10] StataCorp, *Stata 19 Base Reference Manual*, Stata Press, College Station, TX, 2025.

[11] Y. Tian, M. Tang, Y. Zang, M. Tian, Quantile regression for linear models with autoregressive errors using EM algorithm, *Computational Statistics* 33 (2018) 1605–1625.

[12] D. V. Vougas, Prais–Winsten algorithm for regression with second or higher order autoregressive errors, *Econometrics* 9 (3) (2021) 32. doi:10.3390/econometrics9030032.

[13] T. E. Dielman, E. L. Rose, Estimation in least absolute value regression with autocorrelated errors, *Journal of Statistical Computation and Simulation* 50 (1994) 29–43. doi:10.1080/00949659408811597.

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

[15] T. E. Dielman, Estimation and hypothesis testing in LAV regression with autocorrelated errors: is correction for autocorrelation helpful?, *Journal of Modern Applied Statistical Methods* 10 (2011) 539–548.

[16] A. Linden, A comprehensive set of postestimation measures to enrich interrupted time-series analysis, *Stata Journal* 17 (1) (2017) 73–88. doi:10.1177/1536867X1701700105.

[17] M. G. Kendall, Note on bias in the estimation of autocorrelation, *Biometrika* 41 (3–4) (1954) 403–404. doi:10.1093/biomet/41.3-4.403.

[18] F. H. C. Marriott, J. A. Pope, Bias in the estimation of autocorrelations, *Biometrika* 41 (3–4) (1954) 390–402.

[19] D. W. K. Andrews, Exactly median-unbiased estimation of first order autoregressive/unit root models, *Econometrica* 61 (1) (1993) 139–165. doi:10.2307/2951781.

[20] C. Gouriéroux, A. Monfort, E. Renault, Indirect inference, *Journal of Applied Econometrics* 8 (1993) S85–S118. doi:10.1002/jae.3950080507.

[21] J. G. MacKinnon, A. A. Smith, Jr., Approximate bias correction in econometrics, *Journal of Econometrics* 85 (2) (1998) 205–230. doi:10.1016/S0304-4076(97)00099-7.

[22] L. Kilian, Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* 80 (2) (1998) 218–230. doi:10.1162/003465398557465.

[23] A. Abadie, A. Diamond, J. Hainmueller, Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program, *Journal of the American Statistical Association* 105 (490) (2010) 493–505.

[24] W. K. Newey, D. McFadden, Large sample estimation and hypothesis testing, in: R. F. Engle, D. L. McFadden (Eds.), *Handbook of Econometrics*, vol. 4, Elsevier, Amsterdam, 1994, pp. 2111–2245.

[25] P. C. B. Phillips, Folklore theorems, implicit maps, and indirect inference, *Econometrica* 80 (1) (2012) 425–454.

[26] A. Linden, PRAISK: Stata module for computing iterated Prais–Winsten regression with AR(k) errors, Statistical Software Components s459648, Boston College Department of Economics, 2026.

[27] A. Linden, BETARK: Stata module for computing beta regression with autoregressive-corrected errors for proportion outcomes, by joint conditional maximum likelihood, Statistical Software Components s459773, Boston College Department of Economics, 2026.

[28] B. E. Hansen, The grid bootstrap and the autoregressive model, *Review of Economics and Statistics* 81 (4) (1999) 594–607. doi:10.1162/003465399558463.