paneltime: Mathematical documentation

Espen Sirnes*

August 2024

Contents

1 Fundamentals **3**

1.1 Observed Residuals 4

1.2 ARIMA Correction 4

1.3 ARMA exact log likelihood and RE/FE correction 5

1.4 Random/Fixed Effects 9

1.5 GARCH Correction 11

2 The ARIMA Process **11**

Gradients 12

2.1 The Hessians 13

3 The GARCH process **13**

3.1 A note on terminology 15

3.2 Gradients 16

3.3 Hessians 17

3.4 Derivatives with respect to the ARIMA process 18

*Tromsø University Business School, University of Tromsø, The Arctic University, Tromsø, Norway.

4 Implemented group and time effects **20**
4.1 First and second derivatives of the implemented two-way effects 21

5 The LL gradient **23**
5.1 Adjusted Gaussian likelihood 24

6 The Hessian **25**

7 Simulation procedure **26**
7.1 Data generation 27
7.2 Estimation and simulation summaries 29

8 Expected-Information Standard Errors **30**
8.1 Information identity 31
8.2 Expected variance and the ratio approximation 32
8.3 Fourth moments 32
8.4 Time-series calculation for all designs 33
8.5 Static special case 35

9 Starting values **35**
9.1 Preliminary regression and full-order ARMA fit 35
9.2 Variance starts and relation to rugarch 38

10 Extensions **39**
10.1 Tobit model 39
10.2 Instrumental variables 40

In the following, we present the mathematical foundations underlying the code in the `paneltime` package. `paneltime` estimates panel data models with a dynamic ARIMA-type mean equation and a GARCH-type conditional variance equation, jointly with random or fixed effects, by maximum likelihood. We begin with fundamental definitions, including the normal distribution. While the reader is likely familiar with this, building from first principles clarifies all relevant variables and assumptions.

1 Fundamentals

We assume that the standardized residuals $\varepsilon_{i,t}/\sigma_{i,t}$ are normally distributed, so that the conditional density of $\varepsilon_{i,t}$ is

$$f\big(\varepsilon_{i,t}, \sigma_{i,t}\big) = \frac{1}{\sqrt{2\pi\,\sigma_{i,t}^2}} \exp\Big(-\tfrac{1}{2}\frac{\varepsilon_{i,t}^2}{\sigma_{i,t}^2}\Big), \tag{1}$$

The term $\varepsilon_{i,t}$ denotes the innovation after differencing (if requested), the ARMA filter, and the group/time effect transformation. These residuals correspond to group i, out of N total groups, at sequence position t, where t ranges up to a maximum of T. The dataset may be unbalanced, meaning some groups can have fewer than T observations. The function $\exp(\cdot)$ denotes the exponential function, and $\sigma_{i,t}^2$ is the conditional variance from a GARCH process (Engle, 1982; Bollerslev, 1986), specific to group i at time t.

In the remainder of the documentation, we will generally refer to the matrix representation of variables. That is, instead of working with scalar forms $\varepsilon_{i,t}$ and $\sigma_{i,t}^2$, we use their matrix counterparts ε and $\boldsymbol{\sigma}^2$. For instance, ε is an $N \times T$ matrix, where the entry in row i and column t corresponds to $\varepsilon_{i,t}$.

1.1 Observed Residuals

The regression residuals are defined as:

$$\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}, \tag{2}$$

where $\mathbf{y}$ is the dependent variable, $\mathbf{X}$ contains the regressors (including an intercept when present), and $\boldsymbol{\beta}$ is the vector of regression coefficients. For $d > 0$, the residual is formed after integer differencing. Regression, ARMA, effects and variance parameters are estimated jointly by maximum likelihood.

Implementation. The regressor array has shape $N \times T \times K$. Observations are sorted within each group and placed at the start of its row; trailing padding is excluded from likelihood calculations. Array columns therefore represent within-group sequence positions, not necessarily common calendar dates. When $d > 0$, the dependent variable and non-intercept regressors are differenced before forming $\mathbf{u}$; the intercept column is restored and the first d positions are zeroed. These boundary positions remain included in the conditional likelihood but are excluded by the exact likelihood correction. The processing order is residual formation, ARMA filtering, effects transformation, GARCH recursion, then summation of likelihood contributions.

1.2 ARIMA Correction

Serial dependence is removed via a Box–Jenkins ARIMA(p, d, q) filter (Box and Jenkins, 1970):

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{MA}^{-1} \cdot \mathbf{AR} \cdot \mathbf{u}, \tag{3}$$

where $\tilde{\boldsymbol{\varepsilon}}$ contains the innovations before the group/time effect corrections. The matrices $\mathbf{MA}$ and $\mathbf{AR}$ are the MA and AR lag operators, defined mathematically in Section 1.3. Differencing is a preprocessing step, not part of this ARMA equation.

Implementation. The lower-triangular filters act separately within each entity, following its observation order; histories never cross entity boundaries. Conditional filtering uses zero presample residuals and innovations. The following subsection explains how the covariance correction accounts for unknown stationary history instead.

1.3 ARMA exact log likelihood and RE/FE correction

A time series did not necessarily start when we began observing it. Its first recorded values can still depend on earlier values and shocks that we cannot see. Conditional ARMA calculations commonly initialize missing history at zero or another fixed value; `rugarch`, for example, uses initialized residuals rather than integrating over a stationary initial state. Zero initialization is a useful approximation, but it does not describe the actual unknown history of a process that was already running. Exact stationary Gaussian likelihood instead averages over the possible histories, with their stationary probabilities. The difference matters most near the start of each series.

The correction reuses the AR and MA filters already calculated by the package. For one entity, write $u_t = y_t - X_t\beta$ after any integer differencing. The ARMA model and its lag polynomials are

$$\begin{aligned} a(L)u_t &= b(L)\varepsilon_t, \qquad Lu_t = u_{t-1}, \\ a(L) &= 1 - \sum_{j=1}^{p} \rho_j L^j, \qquad b(L) = 1 + \sum_{j=1}^{q} \lambda_j L^j. \end{aligned} \tag{4}$$

Here ε_t is an independent Gaussian shock with variance σ^2. For a history of length T_i, let $\mathcal{L}$ be the T_i-by-T_i matrix with ones immediately below the

diagonal and zeros elsewhere. Define

$$\begin{aligned} \mathbf{AR} &= I - \sum_{j=1}^{p} \rho_j \mathcal{L}^j, \qquad \mathbf{MA} = I + \sum_{j=1}^{q} \lambda_j \mathcal{L}^j, \\ B &= \mathbf{MA}^{-1}\mathbf{AR}, \qquad e = Bu, \qquad B^{-1} = \mathbf{AR}^{-1}\mathbf{MA}. \end{aligned} \tag{5}$$

Thus B maps observed residuals to zero-history innovations, whereas B^{-1} maps innovations to residuals. Both matrices are lower triangular with unit diagonal, so $\det B = 1$. The inverse AR and MA operators are filters, not covariance inverses. With zero presample history, e equals the shocks; with unknown stationary history, it also contains a contribution from that history. The correction accounts for the covariance of this contribution.

Exact stationary ARMA construction. With constant innovation variance and no effects, the following construction includes the infinite presample history without forming a T_i-by-T_i covariance.

Let $r = \max(p, q)$ and let A_s be the r-dimensional AR companion matrix, with first row $(\rho_1, \ldots, \rho_p, 0, \ldots)$ and unit subdiagonal. Let $R = (1, 0, \ldots, 0)^\top$ and let Z be the padded row of AR coefficients plus the padded row of MA coefficients. The unit-innovation stationary state covariance solves the small Lyapunov equation

$$P = A_s P A_s^\top + RR^\top. \tag{6}$$

For $O_t = ZA_s^{t-1}$, the boundary forcing from applying the AR polynomial to the initial-state contribution is

$$D_t = O_t - \sum_{j=1}^{\min(p,t-1)} \rho_j O_{t-j}, \qquad 1 \le t \le r. \tag{7}$$

All subsequent forcing rows vanish by the companion recurrence. Stack these first r rows in D, and take F_r as the first r columns of the finite inverse MA operator, padded or clipped to the observation length. These columns are

shifted copies of the first column of $\mathbf{MA}^{-1}$. Then the initial-state innovation loading is $H = F_r D$, and

$$\mathrm{Cov}(e) = \sigma^2 \Omega, \qquad \Omega = I + HPH^\top = I + F_r V F_r^\top, \qquad V = DPD^\top. \quad (8)$$

The determinant lemma and Woodbury identity reduce the calculation to the positive-definite small system

$$\begin{aligned} P = L_P L_P^\top, \quad G = H^\top H, \quad S = I + L_P^\top G L_P, \\ R_P = L_P S^{-1} L_P^\top, \qquad x = R_P H^\top e. \end{aligned} \quad (9)$$

Both Cholesky factors are r-by-r; neither Ω nor the possibly singular V is inverted. A positive sum avoids cancellation in the Woodbury quadratic:

$$\begin{aligned} Q &= \|e - Hx\|^2 + \|L_P^\top (H^\top e - Gx)\|^2, \\ \ell_i &= -\frac{1}{2}\left[T_i \log(2\pi\sigma^2) + \log|S| + \frac{Q}{\sigma^2}\right]. \end{aligned} \quad (10)$$

Variance is fitted normally, not silently profiled. For $r = 0$ this reduces to independent Gaussian observations. Time and storage are linear in T_i for fixed orders and parameter count; second derivatives carry parameter-pair factors.

Scope and limitations with respect to GARCH and effects. The result above is exact for Gaussian ARMA with constant variance and no effects. The same missing-history argument also applies within each series of a panel with fixed or random effects, so the correction remains useful there and is enabled automatically without GARCH (Section 1.5). A small paired pilot with random effects and no GARCH suggested improved ARMA parameter RMSE, although the evidence is preliminary. The existing effects transformation is retained: the correction accounts for presample history, not exact joint integration over random effects. With effects or variance

regressors, its covariance calculation is approximate.

Exact stationary likelihood with GARCH is substantially harder: unobserved past shocks also determine the initial volatility, whose stationary distribution is generally not available in closed form. Integrating over these quantities requires specialized numerical methods, rather than the Gaussian covariance correction above. This is a research-level extension, not implemented here; conditional GARCH likelihood remains the standard practical approach. Applying the ARMA correction alone with GARCH does not perform this integration and should not be expected to improve estimates reliably. A small paired simulation pilot found no consistent RMSE gain with GARCH, with or without random effects; this is preliminary evidence, not a general result. Automatic mode therefore leaves GARCH models conditional.

Restrictions. The option requires regular observations and the built-in Gaussian model; fractional differencing, custom/non-Gaussian likelihoods, EGARCH, censoring and instrumental variables are rejected explicitly. Nonstationary AR roots and nonpositive or clipped variances invalidate the trial likelihood.

Implementation. By default, `exact_ll=None` in `Model.fit` enables the correction for supported ARIMA models without GARCH, including panels with fixed or random effects. `True` forces it on, including with GARCH; `False` forces conditional likelihood. Automatic mode keeps conditional likelihood for static models, fractional differencing, EGARCH, custom or non-Gaussian likelihoods, and censoring. Explicit `True` still rejects unsupported models. `results.exact_ll` records the resolved boolean.

The native columns `AMA_1` and `AMA_1AR` are the first columns of $\mathbf{M}\mathbf{A}^{-1}$ and B, respectively. `low_rank.py` implements the stationary construction above. Effects, variance regressors and explicitly enabled GARCH use a separate finite-history approximation with a dense covariance matrix; long histories

can therefore be expensive. The correction adds $\ell_{\text{corrected}} - \ell_{\text{conditional}}$ once per entity. Padding and integer-differencing boundaries are excluded, including for ragged panels. Differenced models condition on the original initial levels; this is not a diffuse likelihood for undifferenced levels.

Analytic derivatives of P and H through second order supplement the existing residual/variance chain rule in `calculus.py`, retaining all regressors and cross derivatives. The Hessian batches the two derivative orders into two small-system solves. Observation scores, whitening and forecasts use the same covariance. No new native kernel or airline-specific likelihood is used.

For repeated evaluations, `prepare_likelihood` and `evaluate` return the log likelihood, score and Hessian without fitting; only the last parameter vector is cached. For optimizer coordinates b mapped to full coefficients by $a = f(b)$,

$$g_b = J^\top g_a, \qquad H_b = J^\top H_a J + \sum_k g_{a,k} \nabla_b^2 f_k, \qquad J = \frac{\partial f}{\partial b^\top}. \tag{11}$$

The airline benchmark defines this map outside the engine and jointly fits both MA factors and variance. A final full-vector fixed fit creates results without further optimization; inference for tied coordinates requires the transformed curvature.

1.4 Random/Fixed Effects

The group and/or time effects are removed from the ARMA innovations by centered quasi-demeaning corrections:

$$\varepsilon_{i,t} = \tilde{\varepsilon}_{i,t} + \theta_{g,i} \left(\bar{\tilde{\varepsilon}} - \bar{\tilde{\varepsilon}}_{g,i} \right) + \theta_{t,t} \left(\bar{\tilde{\varepsilon}} - \bar{\tilde{\varepsilon}}_t \right), \tag{12}$$

where $\bar{\tilde{\varepsilon}}$ is the grand mean over included observations, $\bar{\tilde{\varepsilon}}_{g,i}$ is the mean for group i, and $\bar{\tilde{\varepsilon}}_t$ is the mean for actual calendar date t. Fixed effects have weight one; random effects use a variance-dependent weight. Thus $\boldsymbol{\varepsilon}$ is a

transformed innovation, not an innovation plus separately estimated v_i and w_t coefficients. The RE transformation approaches FE as the effect variance grows relative to the idiosyncratic variance.

The effects are removed after the ARMA filter for two reasons. First, since all groups share the same filter, the effects keep their structure in $\tilde{\boldsymbol{\varepsilon}}$: if $u_{i,t} = v_i + w_t + \eta_{i,t}$, the filtered group effect is, after the first $\max(p, q)$ positions, the constant $v_i \left(1 - \sum_j \rho_j\right) / \left(1 + \sum_j \lambda_j\right)$ within each group, and in a balanced panel the filtered time effect is common to all groups at each date. Second, the GARCH recursion then operates on innovations from which the effects have been removed, so the effects are not interpreted as volatility. The correction is not exact during the first $\max(p, q)$ positions of each group, and the filtered time effect is serially correlated (for AR(1), $w_t - \rho w_{t-1}$), which scalar quasi-demeaning does not remove.

The transform is thus exact when the effects enter the innovation equation, $\mathbf{u} = \mathbf{AR}^{-1}\mathbf{MA}\,(v_i + w_t + \boldsymbol{\eta})$, since the filter then returns $v_i + w_t + \eta_{i,t}$ directly. When the effects are added to $\mathbf{u}$ itself, as above, it is an approximation. In simulations with ARMA(2,2)–GARCH(2,2) residuals, unit-variance group and time effects, $N = 20$ and $T = 100$, the ARMA estimates were essentially the same whether the effects were added to $\mathbf{u}$ or entered the innovation equation (bias about 0.04 in both cases), so the approximation error is small relative to the ordinary finite-sample error.

Implementation. A term is omitted when its effect is `none`. Random-effects weights are estimated from variance components as described in Section 4. In the two-way case, both corrections are computed from the same $\tilde{\varepsilon}$ and added, not applied sequentially. Means use actual group membership and calendar dates, not aligned array columns, so unbalanced panels retain the correct date structure.

1.5 GARCH Correction

Heteroscedasticity is then corrected by a GARCH(k,m) transformation (Bollerslev, 1986):

$$\sigma^2 = \mathbf{G}\mathbf{A}_\sigma^{-1}\,\mathbf{G} \;+\; \mathbf{G}\mathbf{A}_\sigma^{-1}\,\mathbf{A}\mathbf{R}_\sigma\,h(\varepsilon, z) \tag{13}$$

where σ^2 and $h(\varepsilon, z)$ are $N \times T$ arrays, and $\mathbf{G}\mathbf{A}_\sigma$ and $\mathbf{A}\mathbf{R}_\sigma$ are $T \times T$ GARCH and ARCH lag operators. The variance-regressor design $\mathbf{W}$ contains a constant and any supplied variance predictors, with $\mathbf{G} = \mathbf{W}\boldsymbol{\omega}$. In the ordinary GARCH model, $h(\varepsilon) = \varepsilon^2$; the resulting variances standardize the effect-corrected innovations for likelihood evaluation.

Implementation. At the first observation of each group, the initial variance replaces $\mathbf{G}$'s corresponding entry. It is fixed from the starting-value calculation unless `include_initvar=True`, which makes it an estimated parameter. Presample innovations and variances are initialized at zero. The default Gaussian calculation uses $h(\varepsilon) = \varepsilon^2 + 10^{-8}$ and the same adjusted square in the quadratic likelihood term; conditional variances are clipped to $[10^{-30}, 10^{30}]$ for numerical stability. These safeguards do not change the statistical model in Equation 1. Custom h-functions and EGARCH use their configured transformations; details and derivatives follow in Section 3.

2 The ARIMA Process

As described above, $\mathbf{u}$ contains regression residuals after differencing (if $d > 0$). These residuals are linked to the ARMA innovations $\tilde{\boldsymbol{\varepsilon}}$ through:

$$\left(\mathbf{I} - \sum_{j=1}^{p} \rho_j \cdot \mathbf{L}^j\right) \cdot \mathbf{u} = \left(\mathbf{I} + \sum_{j=1}^{q} \lambda_j \cdot \mathbf{L}^j\right) \cdot \tilde{\boldsymbol{\varepsilon}}, \tag{14}$$

where $\mathbf{L}$ is the lag operator (or lag matrix). The moving average component is defined as:

$$\mathbf{MA} = \mathbf{I} + \sum_{j=1}^{q} \lambda_j \cdot \mathbf{L}^j, \tag{15}$$

and the autoregressive component is defined as:

$$\mathbf{AR} = \mathbf{I} - \sum_{j=1}^{p} \rho_j \cdot \mathbf{L}^j. \tag{16}$$

We can then rewrite Equation 14 as:

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{MA}^{-1} \cdot \mathbf{AR} \cdot \mathbf{u}, \tag{17}$$

which expresses the innovations $\tilde{\boldsymbol{\varepsilon}}$ as a transformation of the already-differenced residuals $\mathbf{u}$.

Gradients

The derivatives in this section are those of the ARMA innovations $\tilde{\boldsymbol{\varepsilon}}$, before the group/time effect corrections. The derivatives of the effect-corrected innovations $\boldsymbol{\varepsilon}$ used in the likelihood follow from these through the effect transform, as given in Equations 66 and 72. With fixed or no effects, the transform is linear with constant weights, so the derivatives of $\boldsymbol{\varepsilon}$ are the same transform applied to the derivatives of $\tilde{\boldsymbol{\varepsilon}}$.

When writing down the gradients, it is convenient to define the following function

$$\mathbf{D}(n) = -\mathbf{MA}^{-1} \cdot [\mathbf{L}^1, \cdots, \mathbf{L}^n] \tag{18}$$

The partial derivatives of $\tilde{\boldsymbol{\varepsilon}}$ with respect to the model parameters are:

$$\frac{\partial \tilde{\boldsymbol{\varepsilon}}}{\partial \boldsymbol{\rho}} = \mathbf{D}(p)\,\mathbf{u} \tag{19}$$

$$\frac{\partial \tilde{\boldsymbol{\varepsilon}}}{\partial \boldsymbol{\lambda}} = \mathbf{D}(q)\,\tilde{\boldsymbol{\varepsilon}} \tag{20}$$

Note that the derivative with respect to $\boldsymbol{\lambda}$ is taken with respect to $\tilde{\boldsymbol{\varepsilon}}$ itself rather than $\mathbf{u}$: since $\mathbf{MA}$ is inverted in Equation 3, differentiating $\mathbf{MA}^{-1}$ with respect to λ_j reintroduces a factor of $\mathbf{MA}^{-1}$, i.e. $\partial\mathbf{MA}^{-1}/\partial\lambda_j = -\mathbf{MA}^{-1}\mathbf{L}^j\mathbf{MA}^{-1}$, so that $\partial\tilde{\boldsymbol{\varepsilon}}/\partial\lambda_j = -\mathbf{MA}^{-1}\mathbf{L}^j\tilde{\boldsymbol{\varepsilon}}$.

$$\frac{\partial\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\beta}} = -\mathbf{MA}^{-1}\cdot\mathbf{AR}\cdot\mathbf{X} \tag{21}$$

2.1 The Hessians

The Hessians can now be written as

$$\frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\rho}} = \frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} = 0 \tag{22}$$

The mixed $\boldsymbol{\rho}$–$\boldsymbol{\beta}$ derivative is generally nonzero; it is given by Equation 26.

$$\frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\lambda}'\partial\boldsymbol{\lambda}} = 2\mathbf{D}(q)'\cdot\frac{\partial\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\lambda}} \tag{23}$$

$$\frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\lambda}'\partial\boldsymbol{\rho}} = \mathbf{D}(q)'\cdot\frac{\partial\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\rho}} \tag{24}$$

$$\frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\lambda}'\partial\boldsymbol{\beta}} = \mathbf{D}(q)'\cdot\frac{\partial\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\beta}} \tag{25}$$

$$\frac{\partial^2\tilde{\boldsymbol{\varepsilon}}}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\beta}} = -\mathbf{D}(p)'\cdot\mathbf{X}, \tag{26}$$

3 The GARCH process

The underlying objective of the GARCH transformation is to model serially dependent conditional heteroskedasticity (Engle, 1982; Bollerslev, 1986). What is required is therefore a kernel that captures this serial dependence. Hence, the objective of GARCH is to obtain the serially correlated conditional standard deviations σ that deflate the stationarized ARIMA residuals

ε into white noise ε/σ.

This is, in a sense, the opposite of the objective of the ARIMA transformation, which tries to *remove* serial correlation in the conditional mean.

We shall assume a general GARCH(k, m) process of the form

$$\sigma^2 = \mathbf{G} + \sum_{j=1}^{m} \psi_j \mathbf{L}^j h\left(\varepsilon, z\right) + \sum_{j=1}^{k} \gamma_j \cdot \mathbf{L}^j \sigma^2 \tag{27}$$

For the ordinary GARCH model $h(\varepsilon) = \varepsilon^2$. The package's `vol=EGARCH` option instead uses a log-variance recursion driven by log-squared innovations, as specified in Equation 34. This is not the standard EGARCH specification of Nelson (1991), whose log-variance equation uses signed and centered absolute standardized innovations. The citation provides the exponential-volatility background, not the package's exact innovation transformation. In the package, the user can specify the function $h(\cdot)$, which permits substantial freedom in the specification of the variance model. The user also has the option to let $h(\cdot)$ depend on a shape parameter z, which is then estimated jointly with the other parameters by maximizing the log likelihood.

The default h-function is the usual GARCH model:

$$h(\varepsilon) = \varepsilon^2 \tag{28}$$

If a different $h(\cdot)$ is supplied, its first and second derivative with respect to ε and, if used, z, must be defined also.

$\mathbf{W}$ contains a variance constant and any supplied variance predictors. The initial variance replaces, rather than adds to, the variance-regressor contribution at the first included observation of each group. Writing that initial position as $t = 1$ within each row,

$$G_{i,t} = \begin{cases} \sigma_0^2, & t = 1, \\ W_{i,t}\boldsymbol{\omega}, & t > 1. \end{cases} \tag{29}$$

The initial variance is fixed at its starting value by default; `include_initvar=True` adds it as an estimated parameter. In the recursions, presample innovations and variances are initialized at zero.

The equation above can be reformulated as

$$\left(1 - \sum_{j=1}^{k} \gamma_j \cdot \mathbf{L}^j\right) \sigma^2 = \mathbf{G} + \sum_{j=1}^{m} \psi_j \mathbf{L}^j h\left(\varepsilon, z\right) \tag{30}$$

Let

$$\mathbf{GA}_\sigma = I - \sum_{j=1}^{k} \gamma_j \cdot \mathbf{L}^j \tag{31}$$

$$\mathbf{AR}_\sigma = \sum_{j=1}^{m} \psi_j \cdot \mathbf{L}^j \tag{32}$$

Then the closed form solution of the GARCH process in matrix notation is, similar to the ARIMA model

$$\sigma^2 = \mathbf{GA}_\sigma^{-1} \cdot (\mathbf{G} + \mathbf{AR}_\sigma \cdot h(\varepsilon, z)) \tag{33}$$

Equations 27–33 use the variance scale and describe ordinary GARCH. For `vol=EGARCH`, the implementation instead recurses on the log-variance state $\boldsymbol{\ell} = \log \boldsymbol{\sigma}^2$:

$$\boldsymbol{\ell} = \mathbf{G} + \sum_{j=1}^{m} \psi_j \mathbf{L}^j \log(\boldsymbol{\varepsilon}^2 + 10^{-8}) + \sum_{j=1}^{k} \gamma_j \mathbf{L}^j \boldsymbol{\ell}, \qquad \boldsymbol{\sigma}^2 = \exp(\boldsymbol{\ell}). \tag{34}$$

The Gaussian likelihood is evaluated using $\boldsymbol{\sigma}^2 = \exp(\boldsymbol{\ell})$.

3.1 A note on terminology

The labels ARCH and GARCH follow the established terminology of Engle (1982) and Bollerslev (1986). However, the word "autoregressive" in the

name ARCH should not be read as saying that the $h(\boldsymbol{\varepsilon}, z)$ term in Equation 27 is itself an autoregressive feedback process. The $\boldsymbol{\psi}$ lag polynomial applies transformed past innovations to the variance equation. Since h is a function of innovations rather than lagged conditional variances, this term describes the short-run response to shocks and is analogous in role to the moving-average side of an ARMA model; it is not literally an MA process, because h may be nonlinear.

By contrast, the $\boldsymbol{\gamma}$ lag polynomial acts on past conditional variances and feeds them back into the current variance. This is the autoregressive, persistence-producing part of the recursion and is analogous in role to the AR side of an ARMA model. The $\boldsymbol{\psi}$ term is analogous to the MA-side shock response, while remaining a distributed lag of the generally nonlinear transform $h(\boldsymbol{\varepsilon}, z)$ rather than a literal MA process. We retain the conventional ARCH and GARCH names, while using $\mathbf{AR}_\sigma$ for the innovation-response operator and $\mathbf{GA}_\sigma$ for the lagged-variance feedback operator in Equation 13.

3.2 Gradients

When writing down the gradients, it is convenient to define the following function, analogous to $\mathbf{D}(n)$ in Equation 18:

$$\mathbf{B}(n) = \mathbf{GA}_\sigma^{-1} \cdot [\mathbf{L}^1, \cdots, \mathbf{L}^n] \tag{35}$$

The first derivatives are then:

$$\frac{\partial \sigma^2}{\partial \psi} = \mathbf{B}(m) \cdot h(\varepsilon, z) \tag{36}$$

$$\frac{\partial \sigma^2}{\partial \boldsymbol{\omega}} = \mathbf{GA}_\sigma^{-1} \cdot \mathbf{W}^* \tag{37}$$

where $\mathbf{W}^*$ is $\mathbf{W}$ with its first included entry in each group row set to zero, since that entry is replaced by the initial variance.

$$\frac{\partial \sigma^2}{\partial \sigma_0^2} = \mathbf{G}\mathbf{A}_\sigma^{-1} \cdot \mathbf{e} \tag{38}$$

Here $\mathbf{e}$ has ones at the first included position in each group row and zeros elsewhere; this derivative is used only when the initial variance is estimated with `include_initvar=True`.

$$\frac{\partial \sigma^2}{\partial z} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma \cdot \frac{\partial h}{\partial z} \tag{39}$$

$$\frac{\partial \sigma^2}{\partial \varepsilon} = \mathbf{G}\mathbf{A}_\sigma^{-1} \cdot \mathbf{A}\mathbf{R}_\sigma \cdot \frac{\partial h}{\partial \varepsilon} \tag{40}$$

For $\boldsymbol{\gamma}$, which enters through the inverted operator $\mathbf{G}\mathbf{A}_\sigma$ itself (the GARCH counterpart of $\boldsymbol{\lambda}$ in Equation 20), differentiating $\mathbf{G}\mathbf{A}_\sigma^{-1}$ reintroduces a factor of $\mathbf{G}\mathbf{A}_\sigma^{-1}$, giving

$$\frac{\partial \sigma^2}{\partial \gamma} = \mathbf{B}(k) \cdot \boldsymbol{\sigma}^2 \tag{41}$$

(not $\mathbf{B}(k)\cdot(\mathbf{G}+\mathbf{A}\mathbf{R}_\sigma\, h(\boldsymbol{\varepsilon}, z))$: since $\boldsymbol{\sigma}^2 = \mathbf{G}\mathbf{A}_\sigma^{-1}(\mathbf{G}+\mathbf{A}\mathbf{R}_\sigma h)$ already contains one factor of $\mathbf{G}\mathbf{A}_\sigma^{-1}$, and this does not cancel against the $\mathbf{G}\mathbf{A}_\sigma^{-1}$ inside $\mathbf{B}(k)$).

3.3 Hessians

The corresponding second derivatives, used in the Hessian, are

$$\frac{\partial^2 \sigma^2}{\partial \psi' \partial \psi} = 0 \tag{42}$$

$$\frac{\partial^2 \sigma^2}{\partial \psi'\, \partial z} = \mathbf{B}(m)\, h_z\, (\varepsilon, z) \tag{43}$$

$$\frac{\partial^2 \sigma^2}{\partial \psi' \partial \varepsilon} = \mathbf{B}(m)\, h_\varepsilon\, (\varepsilon, z) \tag{44}$$

$$\frac{\partial^2\sigma^2}{\partial z\,\partial z'} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma\, h_{zz}\,(\varepsilon, z) \tag{45}$$

$$\frac{\partial^2\sigma^2}{\partial z\,\partial \varepsilon'} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma\, h_{z\varepsilon}\,(\varepsilon, z) \tag{46}$$

$$\frac{\partial^2\sigma^2}{\partial \varepsilon\,\partial \varepsilon'} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma\, h_{2\varepsilon}\,(\varepsilon, z) \tag{47}$$

All other second derivatives among ψ, ω, σ_0^2 are zero (since $\mathbf{G} + \mathbf{A}\mathbf{R}_\sigma h$ is linear in each of these).

For $\boldsymbol{\gamma}$ paired with any of $\psi, \omega, \sigma_0^2, z, \varepsilon$, only $\mathbf{G}\mathbf{A}_\sigma^{-1}$ carries $\boldsymbol{\gamma}$-dependence, and the generic notation is genuinely simpler here: writing ζ for the second parameter,

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\gamma}\,\partial\zeta'} = \mathbf{B}(k)\cdot\frac{\partial\sigma^2}{\partial\zeta'}, \tag{48}$$

and for $\boldsymbol{\gamma}$ against itself, differentiating both factors of Equation 41,

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\gamma}'} = 2\,\mathbf{B}(k)'\cdot\frac{\partial\sigma^2}{\partial\boldsymbol{\gamma}}. \tag{49}$$

3.4 Derivatives with respect to the ARIMA process

For $\boldsymbol{\xi}, \boldsymbol{\zeta} \in \{\boldsymbol{\rho}, \boldsymbol{\lambda}, \boldsymbol{\beta}\}$, σ^2 depends on these parameters only through $\boldsymbol{\varepsilon}(\boldsymbol{\xi})$, so the chain rule applied to Equation 40 gives

$$\frac{\partial\sigma^2}{\partial\boldsymbol{\xi}} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma\, h_\varepsilon\,(\varepsilon, z)\cdot\frac{\partial\varepsilon}{\partial\boldsymbol{\xi}}, \tag{50}$$

and, differentiating a second time—an ordinary product rule, since h_ε is itself a function of $\boldsymbol{\varepsilon}(\boldsymbol{\zeta})$—this genuinely benefits from the generic $\boldsymbol{\xi}, \boldsymbol{\zeta}$ notation, since it unifies all six $\{\rho, \lambda, \beta\}$-pairs into one equation:

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\xi}\,\partial\boldsymbol{\zeta}'} = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{A}\mathbf{R}_\sigma\left[\, h_{2\varepsilon}\,(\varepsilon, z)\cdot\frac{\partial\varepsilon}{\partial\boldsymbol{\xi}}\frac{\partial\varepsilon}{\partial\boldsymbol{\zeta}'} \;+\; h_\varepsilon\,(\varepsilon, z)\cdot\frac{\partial^2\varepsilon}{\partial\boldsymbol{\xi}\,\partial\boldsymbol{\zeta}'}\,\right], \tag{51}$$

valid for both the diagonal ($\boldsymbol{\xi} = \boldsymbol{\zeta}$) and off-diagonal case, with $\partial^2\boldsymbol{\varepsilon}/\partial\boldsymbol{\xi}\,\partial\boldsymbol{\zeta}'$ given by Equation 72. With no or fixed effects this equals the effect transform of Equations 22–26, which is zero except for pairs involving $\boldsymbol{\lambda}$ and the $\boldsymbol{\rho}$–$\boldsymbol{\beta}$ pair; with random effects all pairs are generally nonzero, because the weights θ_j depend on all three parameter groups.

For an ARIMA parameter $\boldsymbol{\xi} \in \{\boldsymbol{\rho}, \boldsymbol{\lambda}, \boldsymbol{\beta}\}$ crossed with a GARCH-native parameter, only $\boldsymbol{\varepsilon}(\boldsymbol{\xi})$ carries the $\boldsymbol{\xi}$-dependence, so each cross term follows directly from the ε-derivatives already given above:

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\xi}\,\partial\boldsymbol{\gamma}'} = \mathbf{B}(k)\cdot\frac{\partial\sigma^2}{\partial\boldsymbol{\xi}}, \tag{47a}$$

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\xi}\,\partial\psi'} = \mathbf{B}(m)\,h_{\varepsilon}\,(\varepsilon, z)\cdot\frac{\partial\varepsilon}{\partial\boldsymbol{\xi}}, \tag{47b}$$

$$\frac{\partial^2\sigma^2}{\partial\boldsymbol{\xi}\,\partial z'} = \mathbf{G}\mathbf{A}_{\sigma}^{-1}\mathbf{A}\mathbf{R}_{\sigma}\,h_{z\varepsilon}\,(\varepsilon, z)\cdot\frac{\partial\varepsilon}{\partial\boldsymbol{\xi}}, \tag{47c}$$

(note, in particular, no leading factor of 2 in the $\boldsymbol{\psi}$ case), while $\partial^2\sigma^2/\partial\boldsymbol{\xi}\,\partial\omega' = \partial^2\sigma^2/\partial\boldsymbol{\xi}\,\partial\sigma_0^{2\prime} = 0$.

Finally, the log-variance derivatives needed for the likelihood (in the section "The LL gradient" below) are worth keeping in generic form, since the same two-line identity applies uniformly to every parameter pair rather than needing to be listed separately for each:

$$\frac{\partial\ln(\sigma^2)}{\partial\xi} = \frac{1}{\sigma^2}\frac{\partial\sigma^2}{\partial\xi}, \tag{52}$$

$$\frac{\partial^2\ln(\boldsymbol{\sigma}^2)}{\partial\xi\,\partial\zeta'} = \frac{1}{\boldsymbol{\sigma}^2}\frac{\partial^2\boldsymbol{\sigma}^2}{\partial\xi\,\partial\zeta'} - \frac{\partial\ln(\boldsymbol{\sigma}^2)}{\partial\xi}\frac{\partial\ln(\boldsymbol{\sigma}^2)}{\partial\zeta'}. \tag{53}$$

4 Implemented group and time effects

The likelihood applies the effects transform to the ARMA innovations $\tilde{\varepsilon}$ of Equation 3. For an included observation, define the grand mean $\bar{\tilde{\varepsilon}}$, the group mean $\bar{\tilde{\varepsilon}}_{g,i}$, and the mean for its actual calendar date $\bar{\tilde{\varepsilon}}_t$. The correction in Equation 12 uses a separate weight for each enabled dimension. A 'none' effect has weight zero, a 'fixed' effect has weight one, and a 'random' effect uses the estimated variance components below. The time means are formed from the date map, not from a common matrix-column position.

For the two-way random-effects calculation, let $n_{g,i} = T_i$ be the number of observations in group i, and let n_t be the number of groups observed at date t. Define

$$a_g = \frac{1}{\text{mean}_i(n_{g,i})}, \qquad a_t = \frac{1}{\text{mean}_t(n_t)}, \qquad \mu = \bar{\bar{\tilde{\varepsilon}}}. \tag{54}$$

Let K_s be the number of slope regressors, excluding the intercept, and NT the number of included observations. The implementation estimates the idiosyncratic variance from the two-way demeaned innovation:

$$\tilde{\varepsilon}_{w,i,t} = \tilde{\varepsilon}_{i,t} - \bar{\bar{\tilde{\varepsilon}}}_{g,i} - \bar{\bar{\tilde{\varepsilon}}}_t + \mu, \qquad \widehat{\sigma}^2_\epsilon = \frac{\text{mean}_{\mathcal{O}}(\tilde{\varepsilon}^2_{w,i,t})}{1 - a_g - a_t + (1 - K_s)/NT}. \tag{55}$$

For dimension $j \in \{g, t\}$, first remove the other dimension's mean and add the grand mean; let $\bar{\bar{\tilde{\varepsilon}}}_{j,\text{adj}}$ denote the resulting mean within the own dimension. Set $n_j^{\text{eff}} = NTa_j$ and $d_j = \max(n_j^{\text{eff}} - 1 - K_s, 1)$. The own-dimension variance estimate is

$$\widehat{\sigma}^2_{v_j} = \max\left\{0, \frac{n_j^{\text{eff}}}{d_j}\,\text{mean}_{\mathcal{O}}\left[(\bar{\bar{\tilde{\varepsilon}}}_{j,\text{adj}} - \mu)^2\right] - a_j\widehat{\sigma}^2_\epsilon\right\}. \tag{56}$$

The quasi-demeaning weight at an observation has the standard variance-

components form (Swamy and Arora, 1972; Greene, 2012):

$$\theta_j = 1 - \sqrt{\frac{\widehat{\sigma}_\epsilon^2}{\widehat{\sigma}_\epsilon^2 + n_j \widehat{\sigma}_{v_j}^2}}, \tag{57}$$

where n_j is that group's observation count for $j = g$, or that date's observed-group count for $j = t$. Negative estimated effect variances are truncated to zero, giving $\theta_j = 0$. The variance components are recomputed from the current ARMA innovations during likelihood evaluation; they are not separate entries in the optimizer parameter vector. Since $\tilde{\varepsilon}$ depends on $\boldsymbol{\beta}$, $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$, so do the weights. The likelihood applies group and time effects to the innovations of the mean equation; the 'effects.variance' setting does not add a variance-effects transform.

The references give the variance-components and quasi-demeaning background, not the package's complete two-way algorithm. In particular, Equations 55–56 specify the implemented variance estimates, and Equation 12 applies the two centered corrections additively to the ARMA innovations, rather than sequentially or to variance-standardized residuals.

4.1 First and second derivatives of the implemented two-way effects

The following derivatives describe the effect transform of the ARMA innovations $\tilde{\varepsilon}$. Let $\langle x \rangle = (NT)^{-1} \sum_{\mathcal{O}} x$ be the mean over included observations, let $\bar{x}_j$ be the mean within the observation's own group ($j = g$) or actual calendar date ($j = t$), and define the centered operator

$$\mathcal{C}_j(x) = \langle x \rangle - \bar{x}_j, \qquad \varepsilon = \tilde{\varepsilon} + \sum_{j \in \{g,t\}} \theta_j \mathcal{C}_j(\tilde{\varepsilon}). \tag{58}$$

For parameters $h, g \in \{\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\lambda}\}$, write $\tilde{\varepsilon}_h = \partial\tilde{\varepsilon}/\partial h$ and $\tilde{\varepsilon}_{hg} = \partial^2\tilde{\varepsilon}/\partial h\,\partial g$, given by Equations 19–26. For any such derivative $x \in \{\tilde{\varepsilon}_h, \tilde{\varepsilon}_{hg}\}$, define the

linear maps

$$w(x) = x - \bar{x}_g - \bar{x}_t + \langle x \rangle, \tag{59}$$

$$b_j(x) = \overline{(x - \bar{x}_{-j} + \langle x \rangle)}_j - \langle x \rangle, \tag{60}$$

where $\bar{x}_{-j}$ is the mean over the other panel dimension, and the overbar outside parentheses denotes the mean over dimension j. Write $\tilde{\varepsilon}_{w,h} = w(\tilde{\varepsilon}_h)$, $b_{j,h} = b_j(\tilde{\varepsilon}_h)$, $\tilde{\varepsilon}_{w,hg} = w(\tilde{\varepsilon}_{hg})$ and $b_{j,hg} = b_j(\tilde{\varepsilon}_{hg})$. Let $D = 1 - a_g - a_t + (1 - K_s)/NT$ be the denominator in Equation 55, $n_j^{\text{eff}} = NTa_j$, $d_j = \max(n_j^{\text{eff}} - 1 - K_s, 1)$, and $c_j = n_j^{\text{eff}}/(d_j NT)$. On the region where the variance components are positive, their first derivatives are

$$\tau_h \equiv \frac{\partial \widehat{\sigma}_\epsilon^2}{\partial h} = \frac{2}{NTD} \sum_{\mathcal{O}} \tilde{\varepsilon}_w \tilde{\varepsilon}_{w,h}, \tag{61}$$

$$\nu_{j,h} \equiv \frac{\partial \widehat{\sigma}_{v_j}^2}{\partial h} = 2c_j \sum_{\mathcal{O}} (\bar{\tilde{\varepsilon}}_{j,\text{adj}} - \mu) b_{j,h} - a_j \tau_h. \tag{62}$$

Here $\tau = \widehat{\sigma}_\epsilon^2$ and $\nu_j = \widehat{\sigma}_{v_j}^2$. For an observation with own-dimension count n_j, the weight derivatives are

$$\theta_{j,\tau} = -\frac{1}{2\tau}(1 - \theta_j)\theta_j(2 - \theta_j), \tag{63}$$

$$\theta_{j,\nu} = \frac{n_j}{2\tau}(1 - \theta_j)^3, \tag{64}$$

$$\theta_{j,h} = \theta_{j,\tau}\tau_h + \theta_{j,\nu}\nu_{j,h}. \tag{65}$$

Thus the first derivative of the transformed innovation is

$$\varepsilon_h = \tilde{\varepsilon}_h + \sum_{j \in \{g,t\}} \left[\theta_j \mathcal{C}_j(\tilde{\varepsilon}_h) + \theta_{j,h} \mathcal{C}_j(\tilde{\varepsilon}) \right]. \tag{66}$$

Unlike the regression residual u, the innovation $\tilde{\varepsilon}$ is not linear in the parameters, so $\tilde{\varepsilon}_{hg}$ is generally nonzero (Equations 22–26). The second deriva-

tives of the variance components are therefore

$$\tau_{hg} = \frac{2}{NTD} \sum_{\mathcal{O}} \left(\tilde{\varepsilon}_{w,h} \tilde{\varepsilon}_{w,g} + \tilde{\varepsilon}_{w} \tilde{\varepsilon}_{w,hg} \right), \qquad \nu_{j,hg} = 2c_j \sum_{\mathcal{O}} \left[b_{j,h} b_{j,g} + (\bar{\tilde{\varepsilon}}_{j,\text{adj}} - \mu) b_{j,hg} \right] - a_j \tau_{hg}. \tag{67}$$

The second derivatives of each weight follow from $\theta_j = \theta_j(\tau, \nu_j)$:

$$\theta_{j,\tau\tau} = -\frac{\theta_{j,\tau}}{2\tau} \left[4 - 3(2 - \theta_j)\theta_j \right], \tag{68}$$

$$\theta_{j,\nu\nu} = -\frac{3}{4} \left(\frac{n_j}{\tau} \right)^2 (1 - \theta_j)^5, \tag{69}$$

$$\theta_{j,\tau\nu} = -\frac{\theta_{j,\nu}}{2\tau} \left[3(\theta_j - 2)\theta_j + 2 \right], \tag{70}$$

$$\theta_{j,hg} = \theta_{j,\tau\tau} \tau_h \tau_g + \theta_{j,\tau\nu} (\tau_h \nu_{j,g} + \nu_{j,h} \tau_g) + \theta_{j,\nu\nu} \nu_{j,h} \nu_{j,g} + \theta_{j,\tau} \tau_{hg} + \theta_{j,\nu} \nu_{j,hg}. \tag{71}$$

The corresponding second derivative of the transformed innovation is

$$\varepsilon_{hg} = \tilde{\varepsilon}_{hg} + \sum_{j \in \{g,t\}} \left[\theta_j \mathcal{C}_j(\tilde{\varepsilon}_{hg}) + \theta_{j,h} \mathcal{C}_j(\tilde{\varepsilon}_g) + \theta_{j,g} \mathcal{C}_j(\tilde{\varepsilon}_h) + \theta_{j,hg} \mathcal{C}_j(\tilde{\varepsilon}) \right]. \tag{72}$$

These derivatives apply to a random-effects dimension while $\tau > 0$, $\nu_j > 0$, and $n_j > 1$. For a fixed effect, $\theta_j = 1$ is constant; for a disabled effect, $\theta_j = 0$, so its weight derivatives vanish and Equations 66 and 72 reduce to the linear transform of $\tilde{\varepsilon}_h$ and $\tilde{\varepsilon}_{hg}$. The implementation also sets the random-effects weight and its derivatives to zero when the estimated own-dimension variance is clipped at zero or the group/date has at most one observation.

5 The LL gradient

Given the normal distribution in Equation 1, it follows that the log-likelihood contribution of group i, summed over its T_i time periods, is

$$LL_i = -\frac{1}{2}\ln(2\pi)\cdot T_i - \frac{1}{2}\mathbf{1}'\left(\ln\left(\sigma_{i,t}^2\right) + \frac{\epsilon_{i,t}^2}{\sigma_{i,t}^2}\right) \tag{73}$$

so that, pointwise in t,

$$\frac{\partial LL}{\partial \epsilon_{i,t}} = -\frac{\epsilon_{i,t}}{\sigma_{i,t}^2} \tag{74}$$

and

$$\frac{\partial LL}{\partial \ln\left(\sigma_{i,t}^2\right)} = -\frac{1}{2}\left(1 - \frac{\epsilon_{i,t}^2}{\sigma_{i,t}^2}\right) \tag{75}$$

The likelihood code exposes derivatives with respect to the variance state used by each model. For ordinary GARCH, this state is $v = \sigma^2$; for EGARCH, it is the log-variance $r = \ln v$. The log-variance derivatives above convert to variance derivatives as

$$\frac{\partial LL}{\partial v} = \frac{1}{v}\frac{\partial LL}{\partial r}, \qquad \frac{\partial^2 LL}{\partial \epsilon\, \partial v} = \frac{1}{v}\frac{\partial^2 LL}{\partial \epsilon\, \partial r}, \qquad \frac{\partial^2 LL}{\partial v^2} = \frac{1}{v^2}\left(\frac{\partial^2 LL}{\partial r^2} - \frac{\partial LL}{\partial r}\right). \tag{76}$$

5.1 Adjusted Gaussian likelihood

When `GARCH_assist` or `kurtosis_adj` is nonzero, the implementation selects its `Hyperbolic` likelihood instead of the default Gaussian likelihood. Let $q = \epsilon^2 + 10^{-8}$, $s = \operatorname{sign}(q - v)$, $a =$ `GARCH_assist`, $k =$ `kurtosis_adj`, and let m_v denote the variance-boundary mask (zero at an active bound and one in the interior). The implemented pointwise likelihood is

$$L_H = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\left[\ln v + (1-k)\frac{q}{v} + a\frac{|q-v|}{v} + \frac{k}{3}\frac{q^2}{v^2}\right]. \tag{77}$$

Away from the kink $q = v$, its score and Hessian entries are

$$\frac{\partial L_H}{\partial \epsilon} = -\frac{1}{2}\left[2(1-k)\frac{\epsilon}{v} + 2as\frac{\epsilon}{v} + \frac{4k}{3}\frac{q\epsilon}{v^2}\right], \tag{78}$$

$$\frac{\partial L_H}{\partial v} = -\frac{1}{2}m_v\left[\frac{1}{v} - (1-k)\frac{q}{v^2} - a\frac{s}{v} - a\frac{|q-v|}{v^2} - \frac{2k}{3}\frac{q^2}{v^3}\right], \tag{79}$$

$$\frac{\partial^2 L_H}{\partial \epsilon^2} = -\frac{1}{2}\left[\frac{2(1-k)}{v} + \frac{2as}{v} + \frac{4k}{3}\frac{q+2\epsilon^2}{v^2}\right], \tag{80}$$

$$\frac{\partial^2 L_H}{\partial v\,\partial \epsilon} = -\frac{1}{2}m_v\left[-2(1-k)\frac{\epsilon}{v^2} - 2as\frac{\epsilon}{v^2} - \frac{8k}{3}\frac{q\epsilon}{v^3}\right], \tag{81}$$

$$\frac{\partial^2 L_H}{\partial v^2} = -\frac{1}{2}m_v\left[-\frac{1}{v^2} + 2(1-k)\frac{q}{v^3} + 2a\frac{s}{v^2} + 2a\frac{|q-v|}{v^3} + 2k\frac{q^2}{v^4}\right]. \tag{82}$$

At $q = v$, the absolute-value term is not differentiable; the implementation uses $\text{sign}(0) = 0$ there.

$$\frac{\partial LL}{\partial \xi} = \mathbf{1}'\left(\frac{dLL}{d\epsilon_{i,t}}\frac{\partial \epsilon_{i,t}}{\partial \xi} + \frac{\partial LL}{\partial ln\left(\sigma^2_{i,t}\right)}\frac{\partial ln\left(\sigma^2_{i,t}\right)}{\partial \xi}\right) \tag{83}$$

For $\xi \in \{\beta,\rho,\lambda,\gamma,\psi,\omega,z\}$. The first term is zero for $\xi \in \{\psi,\omega,z\}$ since these parameters do not affect the residuals.

6 The Hessian

$$\frac{\partial^2 dLL}{\partial \varepsilon_i^2} = -\mathbf{1}'\frac{1}{\sigma_i^2} \tag{84}$$

$$\frac{\partial^2 dLL}{\partial ln\left(\sigma_i^2\right)\partial \varepsilon_i} = \mathbf{1}'\frac{\epsilon_i}{\sigma_i^2} \tag{85}$$

$$\frac{\partial^2 dLL}{\partial ln\left(\sigma_i^2\right)^2} = -\frac{1}{2}\frac{{\epsilon_i}^2}{\sigma_i^2} \tag{86}$$

For parameter ξ and ζ, the corresponding hessian elements are

$$
\begin{aligned}
\frac{\partial^2 dLL}{\partial \boldsymbol{\xi}\,\partial \boldsymbol{\zeta}'} = \mathbf{1}'\Bigg(& \frac{\partial^2 LL}{\partial \boldsymbol{\epsilon}_i^2} \cdot \frac{\partial \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\xi}} \cdot \frac{\partial \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\zeta}'} \\
& + \frac{\partial^2 dLL}{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)\,\partial \boldsymbol{\epsilon}_i} \cdot \left(\frac{\partial \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\xi}} \cdot \frac{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)}{\partial \boldsymbol{\zeta}'} + \frac{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)}{\partial \boldsymbol{\xi}} \cdot \frac{\partial \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\zeta}'} \right) \\
& + \frac{\partial^2 dLL}{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)^2} \cdot \frac{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)}{\partial \boldsymbol{\xi}} \cdot \frac{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)}{\partial \boldsymbol{\zeta}'} \\
& + \frac{\partial LL}{\partial \boldsymbol{\epsilon}_i} \cdot \frac{\partial^2 \boldsymbol{\epsilon}_i}{\partial \boldsymbol{\xi}\,\partial \boldsymbol{\zeta}'} + \frac{\partial LL}{\partial \ln\left(\boldsymbol{\sigma}_i^2\right)} \cdot \frac{\partial^2 \ln\left(\boldsymbol{\sigma}_i^2\right)}{\partial \boldsymbol{\xi}\,\partial \boldsymbol{\zeta}'} \Bigg)
\end{aligned} \tag{87}
$$

Where $\frac{\partial \varepsilon_i}{\partial \xi} = \frac{\partial^2 \varepsilon_i}{\partial \xi \partial \zeta'} = 0$ for γ, ψ and ω. The hessian is then

$$
\begin{array}{ccccccc}
\frac{\partial^2 LL}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}'} & & & & & & \\
\frac{\partial^2 LL}{\partial \boldsymbol{\rho}\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\rho}\,\partial \boldsymbol{\rho}'} & & & & & \\
\frac{\partial^2 LL}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\rho}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\lambda}\,\partial \boldsymbol{\lambda}'} & & & & \\
\frac{\partial^2 LL}{\partial \gamma\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial \gamma\,\partial \boldsymbol{\rho}'} & \frac{\partial^2 LL}{\partial \gamma\,\partial \boldsymbol{\lambda}'} & \frac{\partial^2 LL}{\partial \gamma\,\partial \gamma'} & & & \\
\frac{\partial^2 LL}{\partial \psi\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial \psi\,\partial \boldsymbol{\rho}'} & \frac{\partial^2 LL}{\partial \psi\,\partial \boldsymbol{\lambda}'} & \frac{\partial^2 LL}{\partial \psi\,\partial \gamma'} & \frac{\partial^2 LL}{\partial \psi\,\partial \psi'} & & \\
\frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \boldsymbol{\rho}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \boldsymbol{\lambda}'} & \frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \gamma'} & \frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \psi'} & \frac{\partial^2 LL}{\partial \boldsymbol{\omega}\,\partial \boldsymbol{\omega}'} & \\
\frac{\partial^2 LL}{\partial z\,\partial \boldsymbol{\beta}'} & \frac{\partial^2 LL}{\partial z\,\partial \boldsymbol{\rho}'} & \frac{\partial^2 LL}{\partial z\,\partial \boldsymbol{\lambda}'} & \frac{\partial^2 LL}{\partial z\,\partial \gamma'} & \frac{\partial^2 LL}{\partial z\,\partial \psi'} & \frac{\partial^2 LL}{\partial z\,\partial \boldsymbol{\omega}'} & \frac{\partial^2 LL}{\partial z \partial z'}
\end{array}
$$

7 Simulation procedure

The simulation code, parameter settings, benchmark runners, and analysis scripts are available in the testing repository at `https://github.com/paneltime/paneltime_test`.

We describe the general `panel_time` design only. The static-panel and ordinary time-series designs are restrictions of this model: the former removes the ARMA and GARCH dynamics, while the latter sets $N = 1$ and removes both effects.

7.1 Data generation

Each replication is a balanced panel generated as

$$y_{it} = \beta_0 + \beta_1 x_{0,it} + \beta_2 x_{1,it} + \alpha_i + \lambda_t + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{88}$$

The two regressors and the disturbance u_{it} are generated independently, each with the same ARMA(2,2)–GARCH(2,2) coefficients. For any one of these processes, denoted s_{it}, the recursion is

$$s_{it} = \rho_1 s_{i,t-1} + \rho_2 s_{i,t-2} + e_{it} + \lambda_1 e_{i,t-1} + \lambda_2 e_{i,t-2}, \tag{89}$$

$$e_{it} = \sqrt{h_{it}} z_{it}, \qquad z_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \tag{90}$$

$$h_{it} = \omega_s + \psi_1 e_{i,t-1}^2 + \psi_2 e_{i,t-2}^2 + \gamma_1 h_{i,t-1} + \gamma_2 h_{i,t-2}. \tag{91}$$

The MA coefficients λ_1, λ_2 are distinct from the time effect λ_t. The group effects $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ are constant over time, and the time effects $\lambda_t \sim \mathcal{N}(0, \sigma_\lambda^2)$ are common to all groups at a given date. All effects, regressor innovations and disturbance innovations are mutually independent.

Table 1: Assumed parameters of the `panel_time` simulation

Quantity	Symbol	Value
Replications	R	100
Groups per panel	N	20
Retained periods per group	T	100
Observations per replication	NT	2,000
Slope regressors	K_s	2
Discarded burn-in periods	B	100
Differencing order	d	0
Regression intercept	β_0	0
Regression slopes	(β_1, β_2)	$(1, 1)$
AR coefficients	(ρ_1, ρ_2)	$(0.5, -0.2)$
MA coefficients	(λ_1, λ_2)	$(0.3, 0.2)$
ARCH coefficients	(ψ_1, ψ_2)	$(0.05, 0.03)$
GARCH coefficients	(γ_1, γ_2)	$(0.70, 0.15)$
Regressor innovation SD	σ_x	1
Disturbance innovation SD	σ_ϵ	2
Group-effect SD	σ_α	1
Time-effect SD	σ_λ	1
Regressor variance constant	ω_x	0.07
Disturbance variance constant	ω_u	0.28

The variance constants are chosen separately for the regressors and disturbance using

$$\omega_s = \sigma_s^2 \left(1 - \psi_1 - \psi_2 - \gamma_1 - \gamma_2\right). \tag{92}$$

Thus the common variance persistence is 0.93, and the target unconditional innovation variances are 1 and 4, respectively. These are innovation variances, not the unconditional variances of the ARMA-filtered series.

The generator starts each series and its innovations at zero and initializes the two presample conditional variances to the target innovation variance. It generates $T + B + 2$ positions and retains the last T, discarding the two

initialization positions and $B = 100$ burn-in periods. The group and time effects are drawn for the retained panel and added after the dynamic processes have been generated, as in Equation 88; they are not passed through the ARMA recursion.

Replications are indexed by $r = 0, \ldots, 99$. For reproducibility, NumPy's `default_rng` is seeded with $10{,}000 + r$ for the regressors, $20{,}000 + r$ for the disturbance, and $30{,}000 + r$ for the group and time effects. Each generated dataset is saved together with its innovations, conditional standard deviations, effects and panel identifiers; the parameter settings and implied variance constants are saved alongside the datasets.

7.2 Estimation and simulation summaries

Each retained panel is fitted using `paneltime` with an intercept and both regressors, ARMA(2,2) with no differencing, GARCH(2,2), and random effects in both the group and time dimensions. The coefficients are estimated jointly by maximum likelihood using the transformations described above. In particular, the effects enter the simulated regression residual before the fitted ARMA filter, so the qualification in Section 1 concerning the effect transformation applies.

For each replication, the simulation runner records convergence, log likelihood, elapsed fitting time, parameter estimates, robust standard errors, and curvature-based standard errors evaluated at the true parameters. It also compares estimated group and time effects with their centered simulated values. The analysis separately computes the theoretical reference standard errors described in Section 8 from the simulation settings alone. Across replications, the sampling standard deviation of an estimate and the median reported robust standard error are distinct summaries; neither is used to generate the expected-information reference.

8 Expected-Information Standard Errors

For simulation studies it is useful to have a reference standard error computed from the data-generating parameters alone, without simulating or observing any data. We use the same no-effects time-series approximation for every design. For a panel, we compute information $\mathcal{I}_T$ for one length-T series and add N independent contributions, $\mathcal{I} = N\mathcal{I}_T$. Thus the reference SEs are the single-series SEs divided by $\sqrt{N}$. The time-only design has $N = 1$. The reference is

$$\mathrm{SE}_j = \sqrt{\left[\mathcal{I}^{-1}\right]_{jj}}, \qquad \mathcal{I} = \mathbb{E}_{\boldsymbol{\vartheta}_0}[-\mathbf{H}(\boldsymbol{\vartheta}_0)]\,, \tag{93}$$

where $\mathbf{H}$ is the Hessian of the hypothetical no-panel model, and the implementation approximates its expected information as described below. It is not the expected Hessian of the fitted panel likelihood, a robust covariance, or the sampling covariance of the estimates.

The information calculation ignores group and time effects, quasi-demeaning, and panel degrees-of-freedom corrections, but preserves the separate fixed initial variance and GARCH derivative paths of each length-T series. It does not concatenate the groups into one long series. One scalar correction is added to the intercept variance:

$$\mathrm{SE}_{\mathrm{th}}(\widehat{\beta}_0) = \sqrt{[\mathcal{I}^{-1}]_{00} + \frac{\sigma_\alpha^2}{N} + \frac{\sigma_\lambda^2}{T}}. \tag{94}$$

The correction accounts for the simulated effect means, which are indistinguishable from an intercept shift. It does not inflate the innovation variance or change any other reference SE, and $\mathcal{I}$ itself remains the no-effects information matrix. Fixed- and random-effects fits share this unconditional simulation reference because the effects are redrawn across replications; it is not the intercept SE conditional on fixed effects. This simplification affects only the benchmark, not estimation or the standard errors reported by the fitted models.

We assume stationary and invertible ARMA polynomials, independent zero-mean regressors with the same ARMA dynamics as the disturbance, positive innovation variances, nonnegative GARCH coefficients with $\sum_m \psi_m + \sum_k \gamma_k < 1$, and a finite fourth moment (Equation 98). GARCH coefficients require at least one nonzero ARCH coefficient for identification. The calculation uses no differencing. For the fractional time-series design, shared parameters reuse this ordinary ARMA reference; no theoretical reference for the fractional parameter d is supplied.

8.1 Information identity

For one observation write $e = \varepsilon_{it}$, $v = \sigma_{it}^2$, $\mathbf{d} = \partial e/\partial \boldsymbol{\vartheta}$ and $\mathbf{r} = \partial v/\partial \boldsymbol{\vartheta}$. If $e = \sqrt{v}\, z$ with z standard normal and independent of the information $\mathcal{F}$ that determines v, $\mathbf{d}$, $\mathbf{r}$ and the second derivatives, then the expectation of the Gaussian Hessian of Section 6 conditional on $\mathcal{F}$ is

$$\mathbb{E}[-\ell_{ab} \mid \mathcal{F}] = \frac{d_a d_b}{v} + \frac{r_a r_b}{2v^2}, \tag{95}$$

since all terms that are multiplied by e or by $e^2 - v$ have conditional mean zero. Two exact consequences follow from the symmetry of z. Changing the sign of a single standardized innovation z_s changes the sign of ε_s, but leaves every variance, every squared residual, every other residual and every derivative $d_{\cdot,s'}$ with $s' \le s$ unchanged, since these depend on squares or on earlier innovations only. Hence:

1. *The mean–variance block vanishes.* With $\mathbf{F} = \sum_m \psi_m \mathbf{A}_m$ the map from squared residuals to variances, $r_{a,t} = 2\sum_s F_{ts}\varepsilon_s d_{a,s}$ for a mean parameter a, while $r_{\zeta,t}$ is unchanged by the sign of z_s for a variance parameter ζ. Each term of $r_a r_\zeta / v^2$ is therefore odd in one z_s, and its expectation is zero.

2. *Only equal-time products remain in the mean block.* In

$$r_a r_b = 4 \sum_{s,s'} F_{ts} F_{ts'} \varepsilon_s \varepsilon_{s'} d_{a,s} d_{b,s'}$$

a term with $s < s'$ is odd in $z_{s'}$, and has expectation zero.

Here $\mathbf{A}_m = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{L}^m$ (Equation 33). These identities concern the hypothetical time-series model without any panel effect transform.

8.2 Expected variance and the ratio approximation

With $\mathbf{g} = (\sigma^2_{\text{init}}, \omega, \ldots, \omega)'$ and $\mathbf{h}_i = \boldsymbol{\varepsilon}_i \circ \boldsymbol{\varepsilon}_i$, the variance is $\mathbf{v}_i = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{g} + \mathbf{F}\mathbf{h}_i$ and

$$\bar{\mathbf{v}} = \mathbb{E}[\mathbf{v}_i] = \mathbf{G}\mathbf{A}_\sigma^{-1}\mathbf{g} + \mathbf{F}\boldsymbol{\mu}_h. \tag{96}$$

The moment approximation used in the calculation is

$$\mathbb{E}\left[\frac{x_t}{v_t}\right] \approx \frac{\mathbb{E}[x_t]}{\bar{v}_t}, \qquad \mathbb{E}\left[\frac{x_t}{v_t^2}\right] \approx \frac{\mathbb{E}[x_t]}{\bar{v}_t^2}, \qquad \mathbb{E}\left[\varepsilon_s^2 x_s\right] \approx \mu_{h,s}\,\mathbb{E}[x_s], \tag{97}$$

for the derivative products x. It is exact when the variance is deterministic ($\psi = 0$). Otherwise it errs in both directions: $\mathbb{E}[1/v] > 1/\mathbb{E}[v]$ by Jensen's inequality, but the numerators are positively correlated with v. The conditional-variance coefficient of variation $\sqrt{\mathrm{Var}(\sigma^2_{it})}/\sigma^2_\epsilon$ indicates its size. Evaluating the ratios exactly would require the stationary joint distribution of $(v_t, \mathbf{r}_t)$, which has no closed form.

8.3 Fourth moments

The squared-innovation ARMA representation of GARCH and its correlation structure are described by Bollerslev (1986, 1988): $\epsilon_t^2 - \sigma_\epsilon^2$ has autoregressive coefficients $\psi_j + \gamma_j$ and moving-average coefficients $-\gamma_j$, driven by $\eta_t = \epsilon_t^2 - \sigma_t^2$. Applying that representation to the Gaussian-innovation assumption used here gives the following moment formulas. With $S = \sum_{j \geq 0} \pi_j^2$ the sum of its

squared MA(∞) weights,

$$\begin{gathered}\operatorname{Var}(\eta_{it}) = \frac{2\sigma_\epsilon^4}{3-2S}, \qquad \operatorname{Var}(\sigma_{it}^2) = (S-1)\operatorname{Var}(\eta_{it}), \\ \mathbb{E}[\epsilon_{it}^4] < \infty \iff S < \tfrac{3}{2},\end{gathered} \tag{98}$$

which for GARCH(1,1) is $3\psi^2 + 2\psi\gamma + \gamma^2 < 1$. The squared-innovation autocovariance $H(k) = \operatorname{Cov}(\epsilon_t^2, \epsilon_{t-k}^2)$ is obtained from these MA weights and $\operatorname{Var}(\eta_t)$.

8.4 Time-series calculation for all designs

Each hypothetical series has no effects. The following equations calculate $\mathcal{I}_T$ for one series; the total information is $\mathcal{I} = N\mathcal{I}_T$. The regressors and disturbance have the same ARMA law but independent innovations, with variances σ_x^2 and σ_ϵ^2. Write

$$A(z) = 1 - \sum_j \rho_j z^j, \qquad B(z) = 1 + \sum_j \lambda_j z^j, \qquad G(z) = 1 - \sum_j \gamma_j z^j.$$

Stationary residual derivatives for the AR and MA parameters have filters $D_{\rho_j}(z) = -z^j/A(z)$ and $D_{\lambda_j}(z) = -z^j/B(z)$ applied to ϵ_t. Consequently, with $D_{a,h}$ denoting lag coefficients,

$$p_{ab} = \sigma_\epsilon^2 \sum_{h\geq 0} D_{a,h} D_{b,h}, \qquad p_{00} = \left(\frac{A(1)}{B(1)}\right)^2, \qquad p_{ll} = \sigma_x^2. \tag{99}$$

All remaining mean-parameter cross moments vanish for the independent, zero-mean regressor design.

It is important to retain the finite GARCH initialization: replacing it by a fully stationary variance-derivative process can substantially change the SEs in this design. Let b_h and f_h be the coefficients of $b(z) = 1/G(z)$ and $f(z) = b(z)\sum_j \psi_j z^j$, and index the observations by $t = 0, \ldots, T-1$. With

$g_0 = \sigma_\epsilon^2$ and $g_t = \omega$ for $t > 0$,

$$\bar{v}_t = \sum_{h=0}^{t} b_h g_{t-h} + \sigma_\epsilon^2 \sum_{h=0}^{t} f_h, \qquad (\mathcal{I}_T)_{ab} \approx p_{ab} \sum_{t=0}^{T-1} \left[\frac{1}{\bar{v}_t} + \frac{2\sigma_\epsilon^2}{\bar{v}_t^2} \sum_{h=0}^{t} f_h^2 \right]. \quad (100)$$

The first term measures information in the innovation itself; the second includes its effect on future conditional variances.

The variance derivatives are finite convolutions of $h_t = \epsilon_t^2$. Their random-part filters and means are

$$\begin{aligned} c_{\psi_j}(z) &= z^j b(z), & M_{\psi_j,t} &= \sigma_\epsilon^2 \sum_{h=0}^{t} c_{\psi_j,h}, \\ c_{\gamma_j}(z) &= z^j b(z) f(z), & M_{\gamma_j,t} &= \sum_{h=0}^{t} [z^j b(z)]_h \, \bar{v}_{t-h}, \\ c_\omega(z) &= 0, & M_{\omega,t} &= \sum_{h=0}^{t-1} b_h. \end{aligned} \quad (101)$$

The empty sum at $t = 0$ is zero, since the initial variance is fixed. If $H(k) = \mathrm{Cov}(\epsilon_t^2, \epsilon_{t-k}^2)$ is obtained from the squared-innovation ARMA representation above, then

$$\begin{aligned} V_{ab,t} &= \sum_{i=0}^{t} \sum_{j=0}^{t} c_{a,i} c_{b,j} H(|i-j|), \\ (\mathcal{I}_T)_{ab} &\approx \frac{1}{2} \sum_{t=0}^{T-1} \frac{M_{a,t} M_{b,t} + V_{ab,t}}{\bar{v}_t^2}. \end{aligned} \quad (102)$$

These sums can be evaluated by lag convolutions and cumulative sums, without panel matrices or simulated observations. The implementation uses stationary innovation and ARMA-derivative moments, omitting the transient in residuals caused by the zero-presample ARMA filter. It retains the fixed initial variance and finite GARCH derivative paths separately for every length-T series. This matters for identification of the individual GARCH coefficients:

replacing repeated short-series initialization by one long-series initialization can substantially inflate their reference SEs. The ARMA boundary simplification is additional to the GARCH ratio/product approximation of Equation 97.

8.5 Static special case

With no ARMA or GARCH terms, $\sigma^2 = \sigma_\epsilon^2 = \omega$ and $n = NT$, and the same calculation reduces to

$$\mathrm{SE}_{\mathrm{th}}(\widehat{\beta}_0) = \sqrt{\frac{\sigma^2}{n} + \frac{\sigma_\alpha^2}{N} + \frac{\sigma_\lambda^2}{T}}, \qquad \mathrm{SE}_{\mathrm{th}}(\widehat{\beta}_l) = \frac{\sigma}{\sqrt{n\sigma_x^2}}, \qquad \mathrm{SE}_{\mathrm{th}}(\widehat{\omega}) = \sigma^2\sqrt{\frac{2}{N(T-1)}}. \tag{103}$$

The $N(T-1)$ in the variance-constant reference reflects one fixed initial variance per series, following the time-only convention applied independently to each group. Only the intercept receives the effect-mean correction. For the static panel design (n = 2,000, σ = 2, σ_x = 1, unit effect variances), the slope reference is 0.04472 and the unconditional intercept reference is 0.24900.

9 Starting values

The joint likelihood is optimized from one deterministic starting vector by default. Initialization is not a separate estimate reported to the user, and does not alter the derivatives, Hessian, or parameter restrictions of the final likelihood. Explicit user-supplied starting coefficients take precedence.

9.1 Preliminary regression and full-order ARMA fit

First, ordinary least squares is applied to the differenced data, after the existing preliminary group/time effects transformations. Write these working variables as $\widetilde{y}_{it}$ and $\widetilde{\mathbf{x}}_{it}$ and put $u_{it} = \widetilde{y}_{it} - \widetilde{\mathbf{x}}_{it}'\boldsymbol{\beta}$. The effects transformation

is held fixed during this preliminary fit. In particular, this is an initialization approximation, not the final likelihood: the final estimator removes effects *after* ARMA filtering and recomputes random-effect weights as described in Section 4. With $d > 0$, the first d differencing-boundary observations are excluded from the preliminary fit.

For requested orders (p, q), a Hannan–Rissanen-style two-regression start Hannan and Rissanen (1982) supplies all AR and MA coefficients. An auxiliary AR(r) regression of the OLS residuals gives approximate innovations $\widehat{a}_{it}$. A second pooled regression estimates

$$u_{it} = \sum_{j=1}^{p} \rho_j u_{i,t-j} + \sum_{j=1}^{q} \lambda_j \widehat{a}_{i,t-j} + \text{error}_{it}.$$

Only complete lag windows within an individual are used; no lag crosses a group boundary or an excluded observation. The auxiliary order is $r = \min\{20, \max[2\max(p, q) + 1, \lceil \log T_{\max} \rceil]\}$, reduced when necessary to leave more usable rows than regression columns. If the auxiliary regressions cannot be identified, a warning is issued and zero ARMA coefficients are used to seed the subsequent conditional-sum-of-squares fit.

The regression and all requested ARMA coefficients are then jointly refined by minimizing

$$e_{it} = u_{it} - \sum_{j=1}^{p} \rho_j u_{i,t-j} - \sum_{j=1}^{q} \lambda_j e_{i,t-j}, \tag{104}$$

$$Q(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\lambda}) = \frac{1}{n_c} \sum_{(i,t) \in \mathcal{C}} e_{it}^2. \tag{105}$$

Here presample residuals and innovations are zero, the recursion restarts at each group or excluded-observation boundary, and $\mathcal{C}$ omits the first p observations of each contiguous run, as in conditional least squares. GARCH is deliberately ignored at this stage. The implementation evaluates the inverse-

MA filtering of innovations and their derivative channels through the existing native ARMA recursion, packing contiguous runs as independent series. This changes only the computation, not the criterion or its boundary conditions.

Stationarity and invertibility of the preliminary ARMA coefficients are enforced through reflection coefficients Monahan (1984). For an unconstrained vector $\boldsymbol{\eta}$, let $\kappa_j = \tanh \eta_j$ and construct

$$a_j^{(j)} = \kappa_j, \qquad a_\ell^{(j)} = a_\ell^{(j-1)} - \kappa_j a_{j-\ell}^{(j-1)}, \quad 1 \le \ell < j.$$

Use $\boldsymbol{\rho} = \mathbf{a}^{(p)}$ and $\boldsymbol{\lambda} = -\mathbf{a}^{(q)}$, reflecting the MA sign convention $1 + \sum_j \lambda_j L^j$. Auxiliary roots outside a radius of 0.98 are reflected inside that radius before conversion to reflection coordinates. These are safeguards on the *preliminary* fit only, not new restrictions on the final likelihood.

The CSS refinement uses damped Gauss–Newton steps, with at most 10 iterations, and analytic first derivatives. It deliberately stops well before a precise standalone ARMA estimate is needed: in internally rescaled coordinates it stops when $\|J'\mathbf{e}/n_c\|_\infty \le 10^{-4}$ or an accepted step reduces Q by at most $10^{-4}\max(Q, 10^{-12})$, where J is the innovation Jacobian. The joint likelihood optimizer subsequently refines these starting values. For example, before the reflection-coordinate chain rule,

$$\frac{\partial e_{it}}{\partial \boldsymbol{\beta}} = -\widetilde{\mathbf{x}}_{it} + \sum_{j=1}^{p} \rho_j \widetilde{\mathbf{x}}_{i,t-j} - \sum_{j=1}^{q} \lambda_j \frac{\partial e_{i,t-j}}{\partial \boldsymbol{\beta}},$$

and the direct AR and MA derivative inputs are respectively $-u_{i,t-j}$ and $-e_{i,t-j}$, followed by the same inverse-MA recursion. Variables are rescaled internally to improve conditioning and converted back afterwards. Only objective-decreasing steps are accepted. Damping increases when the actual objective reduction is less than one quarter of the Gauss–Newton prediction, decreases when it exceeds three quarters, and is otherwise retained. This avoids repeatedly accepting overly aggressive steps that make little progress.

An iteration-limit or numerical-stall warning distinguishes a bounded preliminary search from a converged fit; the best accepted preliminary coefficients remain usable as starting values. These initialization-specific warnings are reported without aborting the final likelihood fit; unrelated numerical warnings remain errors.

9.2 Variance starts and relation to rugarch

Let s_e^2 be the sample variance of the full-order preliminary innovations over included observations, including the AR conditioning boundary. The existing squared-residual moment rule supplies the first ARCH and GARCH starts, $\psi_1^{(0)}$ and $\gamma_1^{(0)}$; higher variance lags initially remain zero. For the ordinary GARCH specification,

$$v_{\text{initial}} = s_e^2, \qquad \omega^{(0)} = h(0, s_e^2, s_e^2)\big(1 - \psi_1^{(0)} - \gamma_1^{(0)}\big).$$

Thus the initial innovation variance no longer relies on an ARMA(1,1) variance formula when higher orders were requested. EGARCH retains its existing transformed-variance initialization. With no ARMA terms, the existing OLS and variance initialization are unchanged.

The broad strategy is shared with `rugarch` (Galanos, 2026): for the ordinary ARMA mean equation without an in-mean term, version 1.5-5 initializes the mean equation by calling `stats::arima(method="CSS")` at the full requested order R Core Team (2026). That routine starts ARMA coefficients at zero and uses BFGS, without its stationarity transformation on the CSS path. Paneltime instead uses the auxiliary-regression seed, constrained reflection coordinates, and a short damped least-squares refinement, extended to pooled panel runs. It does not copy rugarch's GARCH starts: for GARCH(2,2), rugarch defaults to ARCH coefficients $(0.025, 0.025)$, GARCH coefficients $(0.45, 0.45)$, and a variance constant equal to the variance of its input series divided by 1000 (after scaling, if requested). These version-

specific defaults are specified by `.meqstartpars()` and `.aparchstart()` in the package's `rugarch-startpars.R` source file. Neither preliminary procedure guarantees the global maximum of the joint ARMA–GARCH likelihood.

10 Extensions

10.1 Tobit model

For censored observations, i.e. where the underlying, unobserved dependent variable lies beyond a threshold (Tobin, 1958; Amemiya, 1973), the contribution to the likelihood is given by the normal cumulative distribution function evaluated at the threshold,

$$F\left(\gamma\frac{\epsilon_{i,t}}{\sigma_{i,t}}\right) = \int_{-\infty}^{\gamma\epsilon_{i,t}/\sigma_{i,t}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\varepsilon^2\right)d\varepsilon \tag{106}$$

where $\epsilon_{i,t}$ is the residual evaluated at the censoring threshold. Let γ indicate the type of threshold, with $\gamma = 1$ for a lower threshold and $\gamma = -1$ for an upper threshold. Define $x = \gamma\epsilon_{i,t}/\sigma_{i,t}$, $\lambda(x) = f(x)/F(x)$, and $s = \ln\sigma_{i,t}$. The derivatives with respect to residual and log standard deviation are

$$\frac{\partial \ln F(x)}{\partial \epsilon_{i,t}} = \frac{\gamma\lambda(x)}{\sigma_{i,t}} \tag{107}$$

$$\frac{\partial \ln F(x)}{\partial s} = -x\lambda(x) \tag{108}$$

$$\frac{\partial^2 \ln F(x)}{\partial \epsilon_{i,t}^2} = -\left[\gamma\lambda(x)\frac{\epsilon_{i,t}}{\sigma_{i,t}^3} + \frac{\lambda(x)^2}{\sigma_{i,t}^2}\right] \tag{109}$$

$$\frac{\partial^2 \ln F(x)}{\partial \epsilon_{i,t}\,\partial s} = \frac{\gamma\lambda(x)}{\sigma_{i,t}}\left[x^2 + x\lambda(x) - 1\right] \tag{110}$$

$$\frac{\partial^2 \ln F(x)}{\partial s^2} = x\lambda(x) - x^3\lambda(x) - x^2\lambda(x)^2 \tag{111}$$

The optimizer's log-variance coordinate is $r = \ln \sigma_{i,t}^2 = 2s$. Consequently,

$$\frac{\partial L}{\partial r} = \frac{1}{2}\frac{\partial L}{\partial s}, \qquad \frac{\partial^2 L}{\partial \epsilon\, \partial r} = \frac{1}{2}\frac{\partial^2 L}{\partial \epsilon\, \partial s}, \qquad \frac{\partial^2 L}{\partial r^2} = \frac{1}{4}\frac{\partial^2 L}{\partial s^2}. \tag{112}$$

When the variance state is $v = \sigma^2$ rather than r, the conversion in Equation 76 is also required. The current implementation does not consistently perform this conversion for Tobit observations in ordinary GARCH, and its coded $\partial^2 L/\partial r^2$ term has the wrong sign on the $x^2\lambda(x)^2$ component. Thus the formulas above are the analytic reference; they do not describe the current Tobit Hessian implementation exactly.

10.2 Instrumental variables

The current instrumental-variable code path does not implement the conventional two-stage least-squares projection described by $\mathbf{P}_Z\mathbf{X}$. When a non-constant instrument matrix is supplied, `Panel.set_instrumentals()` assigns that matrix directly as `XIV`, and the likelihood forms its regression residual using `XIV`. Thus the current implementation uses the supplied instrument columns in place of the regressor design for the likelihood transformation; it does not estimate first-stage fitted regressors. The earlier description as standard 2SLS was inaccurate. This behavior should not be interpreted as a general IV estimator.

References

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, 41(6):997–1016.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1988). On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis*, 9(2):121–131.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Galanos, A. (2026). *rugarch: Univariate GARCH Models*. The Comprehensive R Archive Network. R package version 1.5-5, released 15 March 2026. Versioned metadata and initialization source: `https://github.com/cran/rugarch/tree/1.5-5`.

Greene, W. H. (2012). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 7th edition.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1):81–94.

Monahan, J. F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71(2):403–404.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

R Core Team (2026). *ARIMA Modelling of Time Series: stats::arima*. R Foundation for Statistical Computing. Development manual for the stats package, version 4.6.0; accessed 5 November 2026.

Swamy, P. A. V. B. and Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica*, 40(2):261–275.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):24–36.