Documentation for python package 'qrbma'

Muhammad Fawdy Renardi Wahyu*

October 29, 2024

Mathematical Explanation of Operations in the Script

1. Data Preparation

a. Loading Data

Data is loaded from the Excel file `ALL_DATA.xlsx` for each sheet listed in `names`. Each sheet represents a different economic variable. The data in each sheet is read and stored in `cell_array`, containing:

- First column: Data from the sheet (excluding the first column and first row).
- Second column: Sheet name (variable name).
- Third column: Transformation code (`tcode_numbers`).

b. Data Transformation

To make the data stationary, each time series is transformed according to its specific `tcode` using the function `transx`.

Let x_i be the raw data for the i-th variable and $tcode_i$ the transformation code for that variable. The transformation is done as follows:

- **If** $tcode_i = 5$**:** (Annualized growth rate)

$$y_i = 400 \times \text{transx}(x_i, tcode_i)$$

- **If** $tcode_i = 4$ **or** $tcode_i = 2$**:** (Log or difference)

$$y_i = 4 \times \text{transx}(x_i, tcode_i)$$

*Corresponding author. Muhammad Fawdy Renardi Wahyu is affiliated with the Universitas Gadjah Mada, Indonesia. This study does not reflect the view of my institution. All errors are my own. Address: Department of Economics, Universitas Gadjah Mada, Sleman, Yogyakarta, Indonesia. Email: muhammad.fawdy.r@mail.ugm.ac.id

- **Otherwise:**

$$y_i = \text{transx}(x_i, tcode_i)$$

The function `transx` applies the transformation based on the given transformation code (differencing, logarithm, etc.).

c. Stacking Time Series

All transformed time series are stacked into the matrix `all_series` of size $T \times K$, where T is the number of time observations and K is the number of variables.

2. Defining Dependent and Independent Variables

a. Dependent Variable

The main dependent variable is inflation ($Y_{\text{raw\_full}}$), which is computed using the function `yfcsta`:

$$Y_{\text{raw\_full}} = \text{yfcsta}(y_1, tcode_1, n_{\text{fore}})$$

Where:

- y_1: Transformed time series for the inflation variable.
- $tcode_1$: Transformation code for the inflation variable.
- n_{fore}: Forecast horizon (in this case, $n_{\text{fore}} = 1$).

The lagged dependent variable ($Y_{\text{raw\_full\_lags}}$) is also set up as:

$$Y_{\text{raw\_full\_lags}} = y_1$$

b. Independent Variables

The independent variables (Z_{raw}) are all other variables except inflation:

$$Z_{\text{raw}} = \text{all\_series}[:, 2:]$$

c. Adjusting Data Size

Due to the use of lags and forecast horizon, the data is adjusted by removing the last observations:

$$\begin{aligned}
Y_{\text{raw_full}} &= Y_{\text{raw_full}}[: -n_{\text{fore}}] \\
Y_{\text{raw_full_lags}} &= Y_{\text{raw_full_lags}}[: -n_{\text{fore}}] \\
Z_{\text{raw}} &= Z_{\text{raw}}[: -n_{\text{fore}}, :]
\end{aligned}$$

d. Creating Lag of the Dependent Variable

Using the function `mlag2`, a lagged version of the dependent variable is created:

$$\text{ylag} = \text{mlag2}(Y_{\text{raw\_full\_lags}}, lags - 1)$$

With $lags = 2$, the lag used is 1.

e. Constructing the Matrix of Independent Variables X and Dependent Variable Y

If $lags > 0$:

$$X = \left[\mathbf{1}, Y_{\text{raw\_full\_lags}}[lags-1:], \text{ylag}[lags-1:], Z_{\text{raw}}[lags-1:]\right]$$

$$Y = Y_{\text{raw\_full}}[lags-1:]$$

Here, $\mathbf{1}$ is a vector of ones representing the intercept in the regression.

3. Calculating OLS Coefficients

The OLS coefficients β_{OLS} are calculated as:

$$\beta_{\text{OLS}} = (X^\top X)^{-1} X^\top Y$$

The OLS error variance σ^2_{OLS} is computed with:

$$\sigma^2_{\text{OLS}} = \frac{(Y - X\beta_{\text{OLS}})^\top (Y - X\beta_{\text{OLS}})}{T - p - 1}$$

Where:

- T: Number of observations.
- p: Number of parameters (number of columns in X).

4. Defining Bayesian Priors

a. Prior for Regression Coefficients β

The prior for β is assumed to be a multivariate normal distribution:

$$\beta \sim \mathcal{N}(0, V)$$

With the prior covariance matrix $V = 9I_p$, where I_p is the identity matrix of size $p \times p$.

b. Prior for Error Variance σ^2

The prior for σ^2 is assumed to be an Inverse-Gamma distribution:

$$\sigma^2 \sim \mathrm{IG}(a_0, b_0)$$

With $a_0 = 0.1$ and $b_0 = 0.1$.

5. Gibbs Sampling Procedure

For each iteration in Gibbs sampling, and for each quantile q, the following steps are performed:

a. Calculating Parameters τ_q^2 and θ_q

$$\tau_q^2 = \frac{2}{\alpha_q(1-\alpha_q)}$$

$$\theta_q = \frac{1-2\alpha_q}{\alpha_q(1-\alpha_q)}$$

Where α_q is the quantile level (e.g., 0.05, 0.10, ..., 0.95).

b. Updating Error Variance σ_q^2

Compute the posterior parameters for σ_q^2:

$$a_1 = a_0 + \frac{3T}{2}$$

Calculate the sum of squared errors (sse):

$$\mathrm{sse} = \left(Y - X\beta_q - \theta_q z_q\right)^2$$

Calculate a_2:

$$a_2 = b_0 + \sum_{t=1}^{T} \left(\frac{\mathrm{sse}_t}{2 z_{qt} \tau_q^2} \right) + \sum_{t=1}^{T} z_{qt}$$

Update σ_q^2 by drawing from an Inverse-Gamma distribution:

$$\sigma_q^2 \sim \mathrm{IG}\left(a_1, a_2\right)$$

c. Updating Regression Coefficients β_q

Define the diagonal matrix U:

$$U = \operatorname{diag}\left(\frac{1}{\sqrt{\sigma_q^2}\tau_q^2 z_q}\right)$$

Compute the centered vector Y_{tilde}:

$$Y_{\text{tilde}} = Y - \theta_q z_q$$

Calculate the posterior covariance V_{β_q} and posterior mean μ_{β_q}:

$$V_{\beta_q} = \left(X^\top U X + V^{-1}\right)^{-1}$$

$$\mu_{\beta_q} = V_{\beta_q}\left(X^\top U Y_{\text{tilde}}\right)$$

Draw β_q from a multivariate normal distribution:

$$\beta_q \sim \mathcal{N}\left(\mu_{\beta_q}, V_{\beta_q}\right)$$

d. Updating Latent Variable z_{qt} for $t = 1, \ldots, T$

For each t, calculate the parameters:

$$k1_t = \frac{\sqrt{\theta_q^2 + 2\tau_q^2}}{\left|Y_t - X_t\beta_q\right|}$$

$$k2_t = \frac{\theta_q^2 + 2\tau_q^2}{\tau_q^2}$$

Draw z_{qt} from an Inverse-Gaussian distribution (using the function `draw_ig`):

$$z_{qt} = \max\left(\frac{1}{\text{draw\_ig}(k1_t, k2_t)}, 1 \times 10^{-4}\right)$$

e. Storing Sample Results

After discarding the burn-in period, samples of β_q are stored:

$$\text{beta\_draws}[:, q, \text{irep} - \text{nburn}] = \beta_q$$

6. Additional Functions

a. Transformation Function `transx`

Transforms the data based on the transformation code *tcode*.

b. Function `mlag2`

Creates a lagged version of vector or matrix X with lag p:

$$X_{\text{lag}}[p:] = X[:-p]$$

c. Function `draw_ig`

Generates samples from the Inverse-Gaussian distribution with parameters μ and λ.

7. Final Results

After running the Gibbs sampling iterations, the final output is the posterior distribution of the regression coefficients β for each quantile q. This distribution can be used for prediction and further analysis of the effects of independent variables on various quantiles of inflation.

Additional Notes

- **Prior and Posterior Distributions:** In the Bayesian context, we update the prior distribution to the posterior distribution based on the observed data and specified model.
- **Using Gibbs Sampling:** Gibbs sampling is used due to the complex model structure and latent variables z_{qt}, which make the posterior distribution analytically intractable.
- **Latent Variable z_{qt}:** The latent variable z_{qt} handles heteroskedasticity in the quantile regression model, allowing different error variances across quantiles.