ESTEMPMM PACKAGE FOR ESTIMATING PARAMETERS OF TIME SERIES AND REGRESSION MODELS WITH ASYMMETRIC NON-GAUSSIAN ERRORS

Zabolotnii S.
Cherkasy, Ukraine

Accurate estimation of model parameters under non-Gaussian error distributions remains a fundamental challenge in statistical analysis and the assessment of measurement uncertainty. Classical least squares and maximum likelihood methods are optimal under the Gaussian assumption but lose efficiency when error distributions exhibit skewness or heavy tails. Such conditions frequently arise in financial time series, hydrological measurements, industrial process monitoring, and many other applied domains.

The Polynomial Maximization Method (PMM) [1] explicitly leverages information contained in higher-order moments of the error distribution, in particular skewness, which classical methods inherently ignore. Even the quadratic, second-order variant attains substantially lower variance of parameter estimates than classical approaches when innovations are non-Gaussian and asymmetric. The theoretical variance-reduction factor is expressed through the square of the skewness and the kurtosis of the error distribution, demonstrating a direct link between efficiency gains and asymmetry.

Despite the theoretical advantages established in prior studies, practical adoption has been limited by the lack of accessible and reliable software implementations. This work presents a comprehensive package for the R statistical environment, providing the first implementation of the second-degree PMM for both regression models and time-series analysis. The package was released on the CRAN repository in November 2025 under the name EstemPMM [2].

The package implements a three-step adaptive procedure that removes the need for a priori knowledge of the error distribution. First, preliminary estimates are obtained using classical methods; second, empirical moments up to the fourth order are computed from residuals; third, refined estimation is performed using the inferred moments. This adaptive strategy makes the method practically applicable without distributional assumptions.

For linear regression, the package provides PMM-based estimation with an interface aligned with standard R regression functions. Tools for direct comparison with ordinary least squares are included to quantify efficiency gains.

Another application area is time-series analysis. Full support is provided for the Box–Jenkins family, including autoregressive (AR), moving-average (MA), autoregressive moving-average (ARMA), and integrated variants (ARIMA). All functions share a unified interface with a method switch that allows users to choose between PMM and the classical Conditional Sum of Squares (CSS) approach. This flexibility enables head-to-head comparisons within a single functional framework. Forecasting is available via R’s standard mechanisms.

An important component of the package is inference. For linear models, bootstrap procedures are implemented to construct confidence intervals. For time series, a methodologically sound block bootstrap that preserves serial dependence is provided. This is critical because naively applying the ordinary bootstrap to time-series data breaks the autocorrelation structure. Facilities for visualizing bootstrap distributions of parameter estimates are included.

The package also includes a comprehensive suite for validation and method comparison. Functionality is provided for Monte Carlo simulations to enable thorough, simulation-based benchmarking. Comparison utilities are available for all model classes supported by the package.

The package architecture relies on base R components, ensuring stability, minimal external dependencies, and straightforward integration into existing analytical workflows.

Monte Carlo studies show substantial accuracy improvements under asymmetric innovations. For linear regression with moderately skewed data-typical of measurement errors with systematic bias – the polynomial maximization approach reduces standard errors of parameter estimates by 15–30% relative to classical methods. For time-series models, simulations for ARMA processes with sample size $n = 200$ and pronounced innovation asymmetry show mean-squared-error improvements of 25–35% for AR parameters and 12–20% for MA parameters compared with classical estimators. Information criteria consistently favor models estimated via PMM.

The proposed approach is particularly effective for asset-return series with asymmetric volatility, hydrological series exhibiting natural asymmetry, industrial process measurements with systematic asymmetric errors, and macroeconomic indicators with business-cycle asymmetry. Its key advantages include the explicit use of distributional asymmetry as an additional source of information, automatic adaptation to empirical data characteristics without specifying the distribution, robust bootstrap-based uncertainty quantification, and an open-source implementation available for R via CRAN.

The current roadmap includes a third-order variant PMM optimized for non-Gaussian symmetric distributions, the development of method-specific information criteria, and extensions to generalized linear and seasonal models.

References

1. Kunchenko, Y. (2002). Polynomial Parameter Estimations of Close to Gaussian Random Variables. Shaker Verlag, Aachen, Germany.

2. EstemPMM: Polynomial Maximization Method for Non-Gaussian Regression. R package version 0.1.1 (2025). The Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=EstemPMM