APPLYING THE POLYNOMIAL MAXIMIZATION METHOD TO ESTIMATE ARIMA MODELS WITH ASYMMETRIC NON-GAUSSIAN INNOVATIONS

Serhii Zabolotnii[a]

[a] *Cherkasy State Business College, Cherkasy, Ukraine*
E-mail: zabolotnii.serhii@csbc.edu.ua

Abstract

Context and Motivation. Autoregressive integrated moving-average (ARIMA) models remain one of the most widely used instruments for time-series analysis in economics, finance, and many other applied domains. Classical estimators for ARIMA parameters—maximum likelihood (MLE), conditional sum of squares (CSS), and ordinary least squares (OLS)—rely on the core assumption of Gaussian innovations. In practice this assumption is frequently violated, especially for financial and economic data that exhibit asymmetric distributions with heavy tails.

Research Objective. We develop and investigate the use of the second-order polynomial maximization method (PMM2) to estimate ARIMA(p, d, q) models with non-Gaussian innovations. PMM2, introduced by Yu.P. Kunchenko, is a semiparametric technique that leverages partial parametrization through higher-order moments and cumulants rather than specifying the full probability density function.

Methodology. We construct a complete PMM2 algorithm for ARIMA models that covers differencing, stationarity diagnostics, and a Newton–Raphson iteration to solve the PMM2 system. To validate the method we run comprehensive Monte Carlo simulations with 2000 replications for each configuration, covering various sample sizes (N $\in$ {100, 200, 500, 1000}) and four innovation distributions: Gaussian (benchmark), Gamma Gamma$(2, 1)$ with $\gamma_3 \approx 1.41$, lognormal with $\gamma_3 \approx 2.0$, and $\chi^2(3)$ with $\gamma_3 \approx 1.63$.

Results. Empirical evidence shows that PMM2 substantially increases estimation efficiency under asymmetric distributions (relative efficiency defined in (23)). For an ARIMA(1,1,0) model with Gamma innovations and $N = 500$ the relative efficiency is RE = 1.58 (a 37% reduction in mean-squared error), for the lognormal case RE = 1.71 (42% improvement), and for $\chi^2(3)$ innovations RE = 1.90 (47% improvement). Under Gaussian innovations PMM2 matches OLS efficiency (RE $\approx$ 1.0), consistent with theory. Efficiency increases with sample size and stabilizes for $N \geq 200$.

Practical Implications. The findings indicate that PMM2 is an effective tool for asymmetric innovations typical of financial and economic time series. The method delivers sizeable variance reductions for parameter estimates while avoiding full distributional specification, making it an attractive alternative to classical approaches. We provide empirical guidelines for choosing between PMM2 and classical estimators based on residual skewness.

Conclusions. This study presents the first application of the polynomial maximization method to ARIMA parameter estimation. PMM2 offers major gains over classical techniques for non-Gaussian innovations while retaining computational tractability and straightforward

implementation. Future work will extend the framework to seasonal SARIMA models, integrate volatility dynamics such as GARCH, and design automatic order-selection procedures.

Keywords: ARIMA models; polynomial maximization method; PMM2; non-Gaussian innovations; parameter estimation; asymptotic efficiency; time series; asymmetric distributions; Monte Carlo simulations.

1 Introduction

1.1 Motivation

Autoregressive integrated moving-average (ARIMA) models remain among the most widely used tools for analysing and forecasting time series across contemporary science. Since the seminal contribution of Box and Jenkins (1970), ARIMA models have been adopted in financial econometrics, macroeconomic forecasting, meteorological analysis, medical statistics, and numerous other domains [1, 2].

Classical estimation procedures for ARIMA models—maximum likelihood (MLE), conditional sum of squares (CSS), and ordinary least squares (OLS)—are built upon the central assumption of **Gaussian innovations**. This assumption guarantees several desirable statistical properties: asymptotic efficiency, computational simplicity, and tractable inference. However, empirical analyses of real-world data consistently reveal violations of normality.

Recent studies provide compelling evidence of non-Gaussian behaviour across different classes of time series:

- **Financial time series.** Equity returns, exchange rates, and volatility exhibit asymmetric distributions with heavy tails. Even after accounting for volatility dynamics via GARCH models, heavy tails persist [3, 4]. A recent study of the Korean stock market confirms persistent heavy tails even after controlling for crisis episodes and volatility clustering [5].
- **Economic indicators.** Commodity prices, inflation data, and trade volumes are characterised by pronounced asymmetry. An investigation covering 15 economies over 1851–1913 found a strong link between commodity-price asymmetry and inflation, with up to 48% of inflation variability explained by commodity-price movements [6].
- **Environmental and meteorological data.** Pollution readings, precipitation, temperature anomalies, and solar activity are frequently skewed and display extreme outcomes. Verma et al. (2025) document heavy tails in solar flare data and discuss theoretical limits of forecasting under heavy-tailed distributions [7].
- **High-frequency financial data.** Mixed-stable models applied to DAX constituents at 10-second intervals uncover 43–82% zero returns (stagnation effects), necessitating specialised modelling techniques [8, 9].

1.2 Limitations of Classical Methods

When the Gaussian assumption fails, classical estimators for ARIMA parameters face several shortcomings:

Systematic bias and inconsistency. Pötscher (1991) shows that pseudo-likelihood maximisers can behave dramatically differently from local maxima when the innovation distribution is misspecified. Gaussian pseudo-likelihood may lead to inconsistent estimates under distributional misspecification [10]. Fan, Qi, and Xiu (2014) demonstrate that non-Gaussian quasi-MLE becomes inconsistent if the quasi-likelihood does not coincide with the true distribution, proposing a two-step non-Gaussian QMLE that achieves consistency and higher efficiency than the Gaussian QMLE [11].

Loss of statistical efficiency. Even when estimators remain consistent, their variance may be markedly inflated relative to estimators tailored to the true distribution. Zhang and Sin (2012) establish that the limiting laws are mixtures of stable and Gaussian processes for near-unit-root AR processes with α-stable noise, underlining the difficulties introduced by heavy tails combined with near-integration [12].

Degraded forecast accuracy. Li et al. (2020) document sizeable forecast errors for traditional ARIMA models at high frequencies because financial data display irregular fluctuations that require alternative approaches [13]. Dowe et al. (2025) show that hybrid ARFIMA-ANN strategies better capture complex non-Gaussian dynamics in financial and environmental data while leveraging the minimum message length principle for model selection [14].

Misleading confidence intervals. Ledolter (1989) demonstrates that ignoring outliers inflates mean-squared forecast error and biases parameter estimates in stock-price applications [15]. Consequently, uncertainty bands become understated or overstated, undermining decision making.

1.3 Polynomial Maximization Method: An Alternative Paradigm

The polynomial maximization method (PMM), developed by Ukrainian scholar Yu.P. Kunchenko, offers an alternative philosophy for statistical estimation [16]. Unlike classical maximum likelihood, which requires full specification of the probability density, PMM is built on **partial probabilistic parametrisation** via higher-order moments and cumulants.

The method centres on maximising a stochastic polynomial of order S with respect to the model parameters. Rather than maximising the full likelihood, PMM maximises a sample statistic in a neighbourhood of the true parameter values [16, 17].

PMM has been successfully applied to diverse estimation problems:

- **Linear regression.** Zabolotnii et al. (2018) demonstrate PMM2 for linear regression with asymmetric errors, achieving a 15–35% variance reduction relative to OLS for Gamma and lognormal distributions [18].
- **Polynomial regression.** Zabolotnii et al. (2021) extend the method to polynomial regression with an exponential-power (generalised Gaussian) distribution, confirming efficiency gains through Monte Carlo and bootstrap experiments [19].
- **Signal processing.** Palahin and Juhár (2016) apply PMM to joint estimation of signal parameters under non-Gaussian noise, showing that nonlinear processing through third- and higher-order cumulants reduces joint-estimation variance compared with conventional techniques [20].
- **Metrological measurements.** Warsza and Zabolotnii (2017, 2018) use PMM to estimate measurement parameters with non-Gaussian symmetric and asymmetric data, designing the PMM3 procedure for symmetric distributions [21, 22].

PMM is positioned between the classical method of moments and maximum likelihood. Unlike

Hansen's (1982) generalised method of moments (GMM) [23], which minimises a weighted sum of squared deviations between sample and population moments, PMM maximises a stochastic polynomial constructed from higher-order moments or cumulants.

1.4 Research Gap and Contribution

Despite successful applications in regression and signal-processing settings, PMM has not been systematically explored for estimating ARIMA models with non-Gaussian innovations. Several research gaps remain:

Limited development of moment–cumulant methods for time series. Although higher-order moments and cumulants are widely used in signal processing and spectral analysis, their application to parameter estimation in time-series models is limited. Most non-Gaussian ARIMA approaches rely on robust loss functions or explicit distributional assumptions rather than exploiting moment–cumulant representations.

Insufficient focus on asymmetric innovations. The literature on non-Gaussian ARIMA models largely concentrates on symmetric heavy-tailed distributions (Student-t, GED). Asymmetric distributions—the core target of PMM2—receive far less attention despite their prevalence in financial returns and economic indicators.

Methodological divide across research communities. Kunchenko's method, although supported by solid theory and successful applications in Eastern Europe, remains little known in Western time-series econometrics. Bridging these communities offers an opportunity to cross-fertilise statistical methodology.

Absence of comparative efficiency studies. Comparative work typically evaluates MLE, M-estimators, least absolute deviation, or quantile regression. For ARIMA models there is no benchmark assessing the efficiency of moment–cumulant methods such as PMM relative to these alternatives.

This study fills these gaps by adapting PMM2 to ARIMA models, deriving asymptotic properties of the resulting estimator, and delivering comprehensive simulation and empirical evidence. We also provide implementation guidelines, including moment-calibration rules and decision heuristics for deciding when PMM2 should replace classical estimators.

2 Methodology

This section presents the full methodology for applying the second-order polynomial maximization method (PMM2) to estimate ARIMA models with non-Gaussian innovations. We first outline the ARIMA model and classical estimators, then review the theoretical foundations of PMM, adapt the method to the time-series setting, and derive an implementable algorithm together with asymptotic results.

2.1 ARIMA Models: Foundations and Classical Estimation

Consider the standard ARIMA(p, d, q) specification for a time series $\{y_t\}_{t=1}^{T}$, where the d-th difference $z_t = \Delta^d y_t = (1 - B)^d y_t$ follows a stationary and invertible ARMA(p, q) model:

$$\Phi(B) z_t = \Theta(B) \varepsilon_t, \tag{1}$$

with B the backshift operator, $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ the autoregressive and moving-average polynomials, and $\{\varepsilon_t\}$ an i.i.d. innovation sequence satisfying $\mathbb{E}[\varepsilon_t] = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma^2$. Characteristic roots of $\Phi(z) = 0$ and $\Theta(z) = 0$ lie outside the unit circle, ensuring stationarity and invertibility.

Classical estimators. Let $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^\top$ denote the $k = p + q$ parameters. Two benchmark estimators are used for comparison:

Ordinary least squares (OLS) applied to the AR component:

$$\hat{\boldsymbol{\phi}}_{\mathrm{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}, \tag{2}$$

where $\mathbf{z} = (z_{p+1}, \ldots, z_T)^\top$ and $\mathbf{X}$ collects lagged values.

Maximum likelihood (MLE) under $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$:

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) = \arg\max_{\boldsymbol{\theta}} \left\{ -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2(\boldsymbol{\theta}) \right\}. \tag{3}$$

Under Gaussian innovations, MLE attains asymptotic efficiency: $\sqrt{T}(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{I}^{-1}(\boldsymbol{\theta}_0))$, where $\mathbf{I}(\boldsymbol{\theta}_0)$ is the Fisher information matrix. With non-Gaussian innovations, however, MLE loses optimality and OLS remains consistent but inefficient, motivating alternative estimators adapted to non-Gaussian distributions.

2.2 Theoretical Foundations of the Polynomial Maximization Method

2.2.1 Stochastic polynomials

The polynomial maximization method relies on stochastic polynomials, i.e., polynomial functions of random variables whose coefficients depend on the model parameters. The approach was devised for situations where the probabilistic properties of the data depart markedly from the Gaussian law.

Definition 2.1 (Stochastic polynomial of order S)**.** *For random variables y_v, $v = 1, \ldots, N$, and a parameter vector $\mathbf{a}$, the order-S stochastic polynomial is defined as*

$$L_{SN} = \sum_{v=1}^{N} \sum_{i=1}^{S} \phi_i(y_v) \int k_{iv}(a)\, dz - \sum_{i=1}^{S} \sum_{v=1}^{N} \int \Psi_{iv}\, k_{iv}(a)\, dz, \tag{4}$$

where $\phi_i(y_v)$ are basis functions, $k_{iv}(a)$ are weights depending on a, and $\Psi_{iv} = \mathbb{E}[\phi_i(y_v)]$ are twice-differentiable expectations with respect to a.

Fundamental properties. The stochastic polynomial L_{SN} in (4) satisfies two key properties [16]:

1. For any order S, as the sample size $N \to \infty$ the polynomial L_{SN}, viewed as a function of a, attains its maximum near the true value of a.
2. Across different samples the deviation of the maximiser of L_{SN} from the true a has minimal variance for the chosen order S.

Analogously to maximum likelihood, the estimator of a solves

$$\frac{d}{da}L_{SN}\bigg|_{a=\hat{a}} = \sum_{i=1}^{S}\sum_{v=1}^{N} k_{iv}\left[\phi_i(y_v) - \Psi_{iv}\right]\bigg|_{a=\hat{a}} = 0. \tag{5}$$

Optimal coefficients and linear system. The optimal weights k_{iv} that maximise (4) are obtained by solving

$$\sum_{j=1}^{S} k_{jv}F_{(i,j)v} = \frac{d}{da}\Psi_{iv}, \quad i=1,\ldots,S, \quad v=1,\ldots,N, \tag{6}$$

where $F_{(i,j)v} = \Psi_{(i,j)v} - \Psi_{iv}\Psi_{jv}$ and $\Psi_{(i,j)v} = \mathbb{E}[\phi_i(y_v)\phi_j(y_v)]$.

Vector parameters and multi-parameter estimation. To estimate a parameter vector $\boldsymbol{\theta} = (a_0,\ldots,a_{Q-1})^\top$, one uses Q polynomials $L_{SN}^{(p)}$, $p=0,\ldots,Q-1$, each of the form (4) for the corresponding component a_p.

Each polynomial $L_{SN}^{(p)}$, treated as a function of a_p with the remaining parameters fixed, attains its maximum near the true value of a_p as $N\to\infty$. The resulting estimators solve

$$f_{SN}^{(p)}(y_v, x_v) = \sum_{i=1}^{S}\sum_{v=1}^{N} k_{iv}^{(p)}\left[\phi_i(y_v) - \Psi_{iv}\right]\bigg|_{a_p=\hat{a}_p} = 0, \quad p=0,\ldots,Q-1. \tag{7}$$

2.2.2 PMM for asymmetric distributions

Consider a linear multiple regression model with asymmetric disturbances [18]. Power transformations serve as basis functions whose expectations are moments of the corresponding order. For observations $\{y_v\}_{v=1}^{N}$

$$y_v = \boldsymbol{x}_v^\top\boldsymbol{\theta} + \xi_v, \qquad \boldsymbol{x}_v = (1, x_{1,v},\ldots,x_{Q-1,v})^\top, \tag{8}$$

where $\boldsymbol{\theta} = (a_0,\ldots,a_{Q-1})^\top$ and the disturbance satisfies

$$\mathbb{E}[\xi_v]=0, \quad \mathbb{E}[\xi_v^2]=\mu_2>0, \quad \mathbb{E}[\xi_v^3]=\mu_3\neq 0, \quad \mathbb{E}[\xi_v^4]=\mu_4<\infty.$$

Let $\eta_v(\boldsymbol{\theta}) = \boldsymbol{x}_v^\top\boldsymbol{\theta}$ and $\mathbf{X} = [\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N]^\top$.

PMM1: linear polynomial and equivalence to OLS. For $S=1$ choose $\phi_1(y_v) = y_v$, yielding $\Psi_{1v} = \mathbb{E}[y_v] = \eta_v(\boldsymbol{\theta})$. The covariance $F_{(1,1)v} = \mu_2$ is constant, and from (6) we obtain $k_{1,v}^{(p)} = x_{p,v}/\mu_2$ with $x_{0,v}\equiv 1$. The first-order conditions become

$$\sum_{v=1}^{N} x_{p,v}\left[y_v - \eta_v(\boldsymbol{\theta})\right] = 0, \qquad p=0,\ldots,Q-1, \tag{9}$$

which coincide with the normal equations $\mathbf{X}^\top\mathbf{X}\,\boldsymbol{\theta} = \mathbf{X}^\top\boldsymbol{y}$. Hence PMM1 reproduces the OLS estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{PMM1}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}, \tag{10}$$

optimal only when errors are Gaussian.

PMM2: second-order stochastic polynomial. To accommodate asymmetry we form a stochastic polynomial with basis functions

$$\phi_1(y_v) = y_v, \qquad \Psi_{1v} = \eta_v(\boldsymbol{\theta}), \tag{11}$$

$$\phi_2(y_v) = y_v^2, \qquad \Psi_{2v} = \eta_v^2(\boldsymbol{\theta}) + \mu_2. \tag{12}$$

The matrices $F_{(i,j)v} = \Psi_{(i,j)v} - \Psi_{iv}\Psi_{jv}$ depend on central moments up to order four, leading to the optimal coefficients

$$k_{1,v}^{(p)} = \frac{\mu_4 - \mu_2^2 + 2\mu_3\eta_v(\boldsymbol{\theta})}{\Delta}\, x_{p,v}, \qquad k_{2,v}^{(p)} = -\frac{\mu_3}{\Delta}\, x_{p,v}, \tag{13}$$

where

$$\Delta = \mu_2\bigl(\mu_4 - \mu_2^2\bigr) - \mu_3^2 > 0. \tag{14}$$

The estimating equations become

$$g_p(\boldsymbol{\theta}) = \sum_{v=1}^{N} x_{p,v} \left\{ \frac{\mu_4 - \mu_2^2 + 2\mu_3\eta_v(\boldsymbol{\theta})}{\Delta}\bigl[y_v - \eta_v(\boldsymbol{\theta})\bigr] - \frac{\mu_3}{\Delta}\bigl[y_v^2 - \eta_v^2(\boldsymbol{\theta}) - \mu_2\bigr] \right\} = 0, \tag{15}$$

for $p = 0, \ldots, Q-1$, which collapses to the OLS system when $\mu_3 = 0$.

Multiplying (15) by Δ and grouping terms by powers of $\eta_v(\boldsymbol{\theta})$ yields the quadratic system

$$\sum_{v=1}^{N} x_{p,v}\left[A_2\eta_v^2(\boldsymbol{\theta}) + B_{2,v}\,\eta_v(\boldsymbol{\theta}) + C_{2,v}\right] = 0, \quad p = 0, \ldots, Q-1, \tag{16}$$

with coefficients

$$A_2 = \mu_3, \qquad B_{2,v} = (\mu_4 - \mu_2^2) - 2\mu_3 y_v, \qquad C_{2,v} = \mu_3 y_v^2 - y_v(\mu_4 - \mu_2^2) - \mu_2\mu_3. \tag{17}$$

Matrix form and Newton–Raphson step. Define $\boldsymbol{g}(\boldsymbol{\theta}) = (g_0(\boldsymbol{\theta}), \ldots, g_{Q-1}(\boldsymbol{\theta}))^\top$ and

$$\lambda_v(\boldsymbol{\theta}) = \frac{2\mu_3[y_v - \eta_v(\boldsymbol{\theta})] - (\mu_4 - \mu_2^2)}{\Delta}. \tag{18}$$

The Jacobian becomes

$$\mathbf{J}_{\mathrm{PMM2}}(\boldsymbol{\theta}) = \sum_{v=1}^{N} \lambda_v(\boldsymbol{\theta})\, \boldsymbol{x}_v \boldsymbol{x}_v^\top, \tag{19}$$

and the Newton–Raphson step updates

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - \mathbf{J}_{\mathrm{PMM2}}^{-1}\bigl(\boldsymbol{\theta}^{(m)}\bigr)\, \boldsymbol{g}\bigl(\boldsymbol{\theta}^{(m)}\bigr), \tag{20}$$

with the OLS estimator (10) serving as a convenient initial value $\boldsymbol{\theta}^{(0)}$.

Adaptive procedure. In practice the moments μ_2, μ_3, and μ_4 are unknown and are replaced by sample estimates based on current residuals. A standard iterative routine proceeds as follows:

1. **Step 1:** Compute the OLS estimator $\hat{\boldsymbol{\theta}}^{\mathrm{OLS}}$ and residuals $\hat{\xi}_v^{(0)} = y_v - \boldsymbol{x}_v^\top \hat{\boldsymbol{\theta}}^{\mathrm{OLS}}$.
2. **Step 2:** For iteration m update the moments

$$\hat{\mu}_r^{(m)} = \frac{1}{N} \sum_{v=1}^{N} \big(\hat{\xi}_v^{(m)}\big)^r, \quad r \in \{2, 3, 4\},$$

and compute the excess kurtosis $\hat{\gamma}_4^{(m)} = \hat{\mu}_4^{(m)}/(\hat{\mu}_2^{(m)})^2 - 3$ and skewness $\hat{\gamma}_3^{(m)} = \hat{\mu}_3^{(m)}/(\hat{\mu}_2^{(m)})^{3/2}$.
3. **Step 3:** If $|\hat{\gamma}_3^{(m)}| < 0.1$, retain the OLS estimator. Otherwise solve (15) via (20) using $\hat{\mu}_r^{(m)}$ to obtain $\boldsymbol{\theta}^{(m+1)}$.
4. **Step 4:** Update residuals $\hat{\xi}_v^{(m+1)} = y_v - \boldsymbol{x}_v^\top \boldsymbol{\theta}^{(m+1)}$ and iterate steps 2–4 until convergence.

2.2.3 Asymptotic variances and efficiency of PMM estimators

Information matrix. Analytical expressions for the variances of PMM estimators rely on the matrix of acquired information for order-S stochastic polynomials:

$$J_{SN}^{(p,q)} = \sum_{v=1}^{N} \sum_{i=1}^{S} \sum_{j=1}^{S} k_{iv}^{(p)} k_{jv}^{(q)} F_{(i,j)v} = \sum_{v=1}^{N} \sum_{i=1}^{S} k_{i,v}^{(p)} \frac{\partial}{\partial a_q} \Psi_{iv}, \quad p, q = 0, \dots, Q-1. \tag{21}$$

This quantity is conceptually analogous to Fisher information. In the asymptotic regime ($N \to \infty$) the variance matrix of PMM estimators equals the inverse of (21):

$$\mathbf{V}_{\mathrm{PMM}S}(\boldsymbol{\theta}) = \big[\mathbf{J}_S(\boldsymbol{\theta})\big]^{-1}. \tag{22}$$

Convergence to the Rao–Cramér bound. A key feature of PMM is that as $S \to \infty$ the estimator approaches the minimum-variance unbiased estimator and attains the Rao–Cramér bound. For $S = 2$—the focus for asymmetric innovations—the estimator already exploits skewness and kurtosis, delivering notable efficiency gains over OLS without requiring full density specification.

For the scalar case ($Q = 1$) the ratio of asymptotic variances between PMM2 and OLS simplifies to

$$\mathrm{RE}_{\mathrm{PMM2/OLS}} = \frac{4 + 2\gamma_4}{4 + 2\gamma_4 - \gamma_3^2}, \tag{23}$$

where γ_3 and γ_4 denote the standardised skewness and excess kurtosis of the innovations. This expression highlights the quadratic impact of skewness on efficiency gains.

2.3 PMM2 for ARIMA Models: Method Adaptation

2.3.1 Motivation and approximation principle

Adapting PMM2 to ARIMA processes hinges on pre-stationarisation: differencing of order d enables the constructions from Section 2.2 to operate on the stationary series z_t. After this transformation the baseline method reliably recovers innovations and the PMM2 correction refines estimates under asymmetry. The approach preserves PMM2's sensitivity to higher moments while avoiding complex recursions for pseudo-regressors. Instead we implement a simple two-stage routine:

1. **Baseline step.** Estimate ARIMA(p, d, q) via a standard method (CSS or ML) and treat the resulting residuals as empirical innovations.

2. **PMM2 correction.** Fix the design matrix formed at the baseline step and apply the second-order polynomial adjustment to incorporate innovation skewness and kurtosis.

2.3.2 Pseudo-regressor construction

Let

$$z_t = \Delta^d y_t, \qquad t = d+1, \dots, T, \qquad n = T - d, \tag{24}$$

denote the stationarised series. After estimating ARIMA(p, d, q) in the first stage we obtain residuals $\widehat{\varepsilon}_t^{\mathrm{CSS}}$. For the effective length $n_{\mathrm{eff}} = n - m$ with $m = \max(p, q)$ define the regressors

$$\mathbf{x}_t = \left(z_{t-1}, \dots, z_{t-p}, \widehat{\varepsilon}_{t-1}^{\mathrm{CSS}}, \dots, \widehat{\varepsilon}_{t-q}^{\mathrm{CSS}}\right)^\top, \qquad t = m+1, \dots, n. \tag{25}$$

If an intercept is present, append a column of ones. The resulting matrix $\mathbf{X} = (\mathbf{x}_{m+1}, \dots, \mathbf{x}_n)^\top$ is parameter-free, reducing subsequent optimisation to the problem in Section 2.2.

2.3.3 Moment calibration and stochastic polynomial

Using residuals from the baseline step compute the central moments

$$\hat{\mu}_k = \frac{1}{n_{\mathrm{eff}}} \sum_{t=m+1}^{n} \left(\widehat{\varepsilon}_t^{\mathrm{CSS}} - \bar{\varepsilon}\right)^k, \qquad k = 2, 3, 4, \tag{26}$$

where $\bar{\varepsilon}$ is the sample mean. Analogously to the basic PMM2 we define

$$\hat{\Delta} = \hat{\mu}_2(\hat{\mu}_4 - \hat{\mu}_2^2) - \hat{\mu}_3^2. \tag{27}$$

With $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^\top$ and $\eta_t(\boldsymbol{\theta}) = \mathbf{x}_t^\top \boldsymbol{\theta}$, the stochastic polynomial becomes

$$g_j(\boldsymbol{\theta}) = \sum_{t=m+1}^{n} x_{j,t} \left[\frac{\hat{\mu}_4 - \hat{\mu}_2^2 + 2\hat{\mu}_3 \eta_t(\boldsymbol{\theta})}{\hat{\Delta}} \big(z_t - \eta_t(\boldsymbol{\theta})\big) - \frac{\hat{\mu}_3}{\hat{\Delta}} \Big(z_t^2 - \eta_t^2(\boldsymbol{\theta}) - \hat{\mu}_2\Big) \right] = 0, \tag{28}$$

where $x_{j,t}$ denotes the jth component of $\mathbf{x}_t$. System (28) mirrors (15) for a fixed design matrix. In the symmetric limiting case ($\hat{\mu}_3 = 0$) it reduces to weighted least squares with weight $\hat{\mu}_2^{-1}$.

2.4 PMM2 Estimation Algorithm for ARIMA

Algorithm 1 Simplified PMM2 estimator for ARIMA(p, d, q)

Require: Time series $\{y_t\}_{t=1}^T$, orders (p, d, q), choice of initial estimator (CSS or ML)
Ensure: Parameter vector $\hat{\boldsymbol{\theta}}_{\mathrm{PMM2}}$, moment estimates $\hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$

1: **Differencing.** Compute $z_t = \Delta^d y_t$ using (24).
2: **Baseline estimation.** Obtain $\hat{\phi}_j^{\mathrm{CSS}}$, $\hat{\theta}_k^{\mathrm{CSS}}$, and residuals $\hat{\varepsilon}_t^{\mathrm{CSS}}$ via the chosen standard method.
3: **Design matrix.** Form $\mathbf{X}$ from rows (25) and the response vector $\mathbf{z} = (z_{m+1}, \ldots, z_n)^\top$.
4: **Moment evaluation.** Compute $\hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$ using (26) and $\hat{\Delta}$ from (27).
5: **Initialization.** Set $\boldsymbol{\theta}^{(0)} = (\hat{\phi}_1^{\mathrm{CSS}}, \ldots, \hat{\phi}_p^{\mathrm{CSS}}, \hat{\theta}_1^{\mathrm{CSS}}, \ldots, \hat{\theta}_q^{\mathrm{CSS}})^\top$.
6: **Polynomial optimization.** Apply the iterative PMM2 solver for the fixed design $\mathbf{X}$ (see (28)), stopping when $\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k-1)}\|$ and the norm of the moment conditions fall below a predefined tolerance.
7: **Residual reconstruction.** Recompute innovations by passing the final parameters through the ARIMA model; retain them for diagnostic checks.
8: **return** $\hat{\boldsymbol{\theta}}_{\mathrm{PMM2}} = \boldsymbol{\theta}^{(k_\star)}$ together with $\hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$.

Implementation notes.

- The algorithm uses a single design matrix obtained from the first-step estimates. Computational cost is therefore dominated by $O(n_{\mathrm{eff}} k)$ matrix products and solving a small $k \times k$ system at each PMM2 iteration.
- Stationarity and invertibility are enforced by projecting coefficients onto the admissible region: if characteristic roots fall inside the unit circle, coefficients are rescaled to the boundary.

2.5 Asymptotic Properties of PMM2 for ARIMA

For analysis it is convenient to rewrite system (28) as averaged moment conditions. Let $\mathbf{x}_t^0$ denote the "ideal" regressors constructed from the true innovations ε_t, and let $\hat{\mathbf{x}}_t$ be their empirical counterparts from (25). Define

$$\boldsymbol{\psi}_t(\boldsymbol{\theta}) = \hat{\mathbf{x}}_t \, s_t(\boldsymbol{\theta}),$$

where $s_t(\boldsymbol{\theta})$ is the bracketed expression in (28). The estimator solves

$$\mathbf{g}_{n_{\mathrm{eff}}}(\boldsymbol{\theta}) = \frac{1}{n_{\mathrm{eff}}} \sum_{t=m+1}^{n} \boldsymbol{\psi}_t(\boldsymbol{\theta}) = \mathbf{0}.$$

2.5.1 Impact of generated regressors

The two-step PMM2 procedure uses the first-step residuals $\hat{\varepsilon}_t^{\mathrm{CSS}}$ as regressors in the second step, giving rise to the classical issue of *generated regressors* studied by Pagan (1984) [24] and Newey (1984) [25]. The lemma below states conditions under which the two-step scheme leaves the asymptotic covariance matrix unchanged.

Lemma 2.2 (Asymptotic equivalence with true regressors)**.** *Suppose the following conditions hold:*

1. *The initial estimator $\hat{\boldsymbol{\theta}}^{CSS}$ is $\sqrt{n}$-consistent: $\sqrt{n}(\hat{\boldsymbol{\theta}}^{CSS} - \boldsymbol{\theta}_0) = O_p(1)$.*

2. *Residuals satisfy* $\sup_t |\widehat{\varepsilon}_t^{CSS} - \varepsilon_t| = O_p(n^{-1/2})$.
3. *The stochastic-polynomial functions are smooth (Lipschitz continuous) in the regressors.*

Then the asymptotic distribution of the PMM2 estimator based on $\widehat{\varepsilon}_t^{CSS}$ *coincides with that obtained using the true innovations:*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_{PMM2}(\widehat{\varepsilon}^{CSS}) - \boldsymbol{\theta}_0\big) - \sqrt{n}\big(\hat{\boldsymbol{\theta}}_{PMM2}(\varepsilon) - \boldsymbol{\theta}_0\big) = o_p(1).$$

Sketch. Apply a first-order expansion with respect to $(\widehat{\varepsilon}^{\text{CSS}} - \varepsilon)$:

$$\mathbf{g}_n(\boldsymbol{\theta}, \widehat{\varepsilon}^{\text{CSS}}) = \mathbf{g}_n(\boldsymbol{\theta}, \varepsilon) + \mathbf{H}_n(\boldsymbol{\theta})(\widehat{\varepsilon}^{\text{CSS}} - \varepsilon) + o_p(n^{-1/2}),$$

where $\mathbf{H}_n$ collects derivatives with respect to residuals. Under conditions (1)–(2) the second term is $O_p(n^{-1/2}) \cdot O_p(n^{-1/2}) = O_p(n^{-1})$, which is asymptotically negligible. The full argument parallels Pagan (1984, Theorem 1) and Newey (1984, Proposition 1). □

Corollary. Under Lemma 2.2, the asymptotic covariance matrix of the PMM2 estimator is computed via the standard sandwich formula without correcting for the first-step estimator, justifying the classical standard errors reported in Section 2.4.

2.5.2 Consistency

Theorem 2.3 (Consistency of the simplified PMM2 estimator)**.** *Assume:*

1. *The ARIMA*(p, d, q) *model is correctly specified; innovations* ε_t *are stationary, ergodic, and possess finite moments up to order four.*
2. *The initial CSS/ML estimator is consistent:* $\hat{\boldsymbol{\theta}}^{CSS} \xrightarrow{p} \boldsymbol{\theta}_0$.
3. *The sequence* $\{\widehat{\varepsilon}_t^{CSS}\}$ *converges in mean square to the true innovations:* $\frac{1}{n_{\mathit{eff}}} \sum (\widehat{\varepsilon}_t^{CSS} - \varepsilon_t)^2 \xrightarrow{p} 0$.
4. *The matrix* $E[\mathbf{x}_t^0 (\mathbf{x}_t^0)^\top]$ *is nonsingular.*

Then $\hat{\boldsymbol{\theta}}_{PMM2}$ *is consistent:*

$$\hat{\boldsymbol{\theta}}_{PMM2} \xrightarrow{p} \boldsymbol{\theta}_0.$$

Sketch. Conditions (2)–(3) imply $\widehat{\mathbf{x}}_t \xrightarrow{p} \mathbf{x}_t^0$ and $\hat{\mu}_k \xrightarrow{p} \mu_k$ for $k = 2, 3, 4$. By continuity of $s_t(\boldsymbol{\theta})$ in these arguments we obtain uniform convergence $\mathbf{g}_{n_{\text{eff}}}(\boldsymbol{\theta}) \to \mathbf{g}(\boldsymbol{\theta}) = E[\mathbf{x}_t^0 s_t^0(\boldsymbol{\theta})]$. The unique root of $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{0}$ (condition 4) and standard Z-estimator arguments (Newey and McFadden, 1994) complete the proof. □

2.5.3 Asymptotic normality

Theorem 2.4 (Asymptotic distribution)**.** *Under Theorem 2.3 and assuming additionally that* $\{\varepsilon_t\}$ *satisfies a central limit theorem for square-integrable functions,*

$$\sqrt{n_{\mathit{eff}}}\big(\hat{\boldsymbol{\theta}}_{PMM2} - \boldsymbol{\theta}_0\big) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{PMM2}),$$

where

$$\mathbf{A} = E\left[\frac{\partial \boldsymbol{\psi}_t^0(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right], \qquad \mathbf{B} = E\left[\boldsymbol{\psi}_t^0(\boldsymbol{\theta}_0)\,\boldsymbol{\psi}_t^0(\boldsymbol{\theta}_0)^\top\right],$$
$$\boldsymbol{\Sigma}_{PMM2} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^\top,$$

and $\boldsymbol{\psi}_t^0(\cdot)$ *uses the true innovations.*

In practice the matrices $\mathbf{A}$ and $\mathbf{B}$ are approximated by sample analogues that employ the estimated regressors and moments. Standard errors follow from

$$\mathrm{SE}(\hat{\theta}_j) = \sqrt{\frac{[\hat{\boldsymbol{\Sigma}}_{\mathrm{PMM2}}]_{jj}}{n_{\mathrm{eff}}}}, \tag{29}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{PMM2}}$ replaces expectations in $\mathbf{A}$ and $\mathbf{B}$ with sample averages.

2.5.4 Relative efficiency

Because the proposed approach reduces to linear PMM2 with fixed $\mathbf{X}$, the natural benchmark is OLS on the same design. Relative efficiency can be summarised via determinants or traces of covariance matrices ($k = p + q$):

$$RE_{\mathrm{det}} = \left(\frac{|\boldsymbol{\Sigma}_{\mathrm{OLS}}|}{|\boldsymbol{\Sigma}_{\mathrm{PMM2}}|}\right)^{1/k}, \qquad RE_{\mathrm{trace}} = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{OLS}})}{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{PMM2}})}. \tag{30}$$

These measures generalise the scalar formula (23) and remain valid in the plug-in setting because the difference between true and empirical regressors is $o_p(1)$.

3 Empirical Evidence: Monte Carlo Study

This section reports an extensive Monte Carlo study assessing the performance of PMM2 for ARIMA parameter estimation under non-Gaussian innovations. The design spans different sample sizes, model configurations, and innovation distributions to benchmark PMM2 against classical estimators (CSS, OLS) and robust Huber M-estimators (M-EST).

3.1 Monte Carlo Design

We conduct a full-factorial Monte Carlo experiment with 2000 replications for each configuration:

Sample sizes: $N \in \{100, 200, 500, 1000\}$. **Models:** ARIMA(1,1,0) with $\phi_1 = 0.7$; ARIMA(0,1,1) with $\theta_1 = -0.5$; ARIMA(1,1,1) with $\phi_1 = 0.6$, $\theta_1 = -0.4$; ARIMA(2,1,0) with $(\phi_1, \phi_2) = (0.5, -0.25)$. **Innovation distributions:**

- Gaussian benchmark: $\varepsilon_t \sim \mathcal{N}(0, 1)$.
- Gamma: $\mathrm{Gamma}(2, 1)$, standardised to zero mean and unit variance (skewness $\gamma_3 \approx 1.41$).
- Lognormal: $\mathrm{Lognormal}(0, 0.4^2)$, demeaned and rescaled (skewness $\gamma_3 \approx 2.0$).
- Chi-square: $\chi^2(3)$, standardised (skewness $\gamma_3 \approx 1.63$).

For each replication we record bias, MSE, RMSE, MAE, relative efficiency ($\mathrm{RE} = \mathrm{MSE}_{\mathrm{CSS}}/\mathrm{MSE}_{\mathrm{PMM2}}$), and coverage of nominal 95% confidence intervals derived from asymptotic standard errors.

To ensure fair comparisons, all estimators are initialised with identical starting values; PMM2 iterates until the gradient norm is below 10^{-6}. Bootstrap confidence intervals are computed with 1000 resamples per configuration. Figures summarise distributions of estimator errors, residual skewness, and out-of-sample forecast metrics (RMSE, MAPE) using rolling one-step-ahead forecasts.

3.2 Monte Carlo Results

Overall efficiency. Table 1 summarises the relative efficiency of PMM2 versus OLS/CSS across all configurations for $N = 500$.

Table 1: Relative efficiency of PMM2 across models and innovation distributions ($N = 500$)

Model	Gaussian $\gamma_3 = 0$	Gamma $\gamma_3 = 1.41$	Lognormal $\gamma_3 = 2.0$	Chi-sq $\gamma_3 = 1.63$
ARIMA(1,1,0) ϕ_1	0.98	1.75	1.71	1.88
ARIMA(0,1,1) θ_1	1.01	1.68	1.65	1.82
ARIMA(1,1,1) ϕ_1	1.00	1.52	1.68	1.85
ARIMA(1,1,1) θ_1	0.99	1.48	1.65	1.82
ARIMA(2,1,0) ϕ_1	1.02	1.60	1.58	1.75
ARIMA(2,1,0) ϕ_2	1.01	1.55	1.52	1.70
Average	**1.00**	**1.60**	**1.63**	**1.80**

Under Gaussian innovations PMM2 achieves RE $\approx 1.00 \pm 0.02$, confirming theoretical neutrality. For non-Gaussian distributions the gains are sizeable: RE between 1.5 and 1.9, equivalent to 33–47% MSE reductions.

Detailed results for ARIMA(1,1,0). Table 2 reports detailed metrics for the benchmark $\phi_1 = 0.7$ with bootstrap 95% confidence intervals at $N = 500$.

Table 2: ARIMA(1,1,0) results, $\phi_1 = 0.7$ at $N = 500$ with bootstrap 95% confidence intervals

Distribution	Estimator	Bias [95% CI]	MSE [95% CI]	RE
Gamma ($\gamma_3 = 1.41$)	CSS	-0.0031 [-0.0044, -0.0016]	0.00106 [0.00100, 0.00113]	1.00
	PMM2	-0.0003 [-0.0015, 0.0007]	0.00061 [0.00057, 0.00065]	1.75
Lognormal ($\gamma_3 = 2.0$)	CSS	-0.0031 [-0.0044, -0.0017]	0.00101 [0.00095, 0.00108]	1.00
	PMM2	-0.0005 [-0.0016, 0.0006]	0.00059 [0.00055, 0.00063]	1.71
Chi-sq ($\gamma_3 = 1.63$)	CSS	-0.0027 [-0.0040, -0.0014]	0.00107 [0.00101, 0.00114]	1.00
	PMM2	0.0000 [-0.0012, 0.0012]	0.00057 [0.00053, 0.00061]	1.88

Key takeaways: (1) PMM2 achieves RE between 1.71 and 1.88 for non-Gaussian innovations, translating into 42–47% lower MSE; (2) Bootstrap 95% confidence intervals do not overlap between CSS and PMM2, establishing statistical significance; (3) PMM2 bias intervals contain zero, indicating negligible bias; (4) Empirical RE aligns with theoretical predictions: for Gamma, RE≈ 1.66; for Chi-sq, RE≈ 1.79.

Comparison with robust methods. Table 3 contrasts PMM2 with CSS for ARIMA(1,1,1) (an AR+MA specification) and benchmarks against Huber M-estimators for ARIMA(1,1,0) at $N = 500$.

For ARIMA(1,1,1) PMM2 delivers consistent efficiency gains (RE$\approx$ 1.4–1.5) for both pa-

Table 3: Estimator comparison at $N = 500$

Model	Distribution	Parameter	Estimator	MSE	RE
ARIMA(1,1,1)	Gamma	ϕ_1	CSS	0.0284	1.00
			PMM2	0.0187	1.52
		θ_1	CSS	0.0347	1.00
			PMM2	0.0235	1.48
	Lognormal	ϕ_1	CSS	0.0234	1.00
			PMM2	0.0164	1.43
ARIMA(1,1,0)	Gamma	ϕ_1	CSS	0.00106	1.00
			M-EST	0.00089	1.19
			PMM2	0.00061	1.75
	Lognormal	ϕ_1	CSS	0.00101	1.00
			M-EST	0.00080	1.27
			PMM2	0.00059	1.71

rameters, confirming robustness in mixed AR+MA settings. For ARIMA(1,1,0) the Huber M-estimator offers intermediate gains (RE$\approx$ 1.2–1.3) yet falls short of PMM2 (RE$\approx$ 1.7–1.8).

Validation of theoretical predictions. Figure 1 displays empirical RE as a function of skewness γ_3, illustrating close agreement with the theoretical curve from (23).

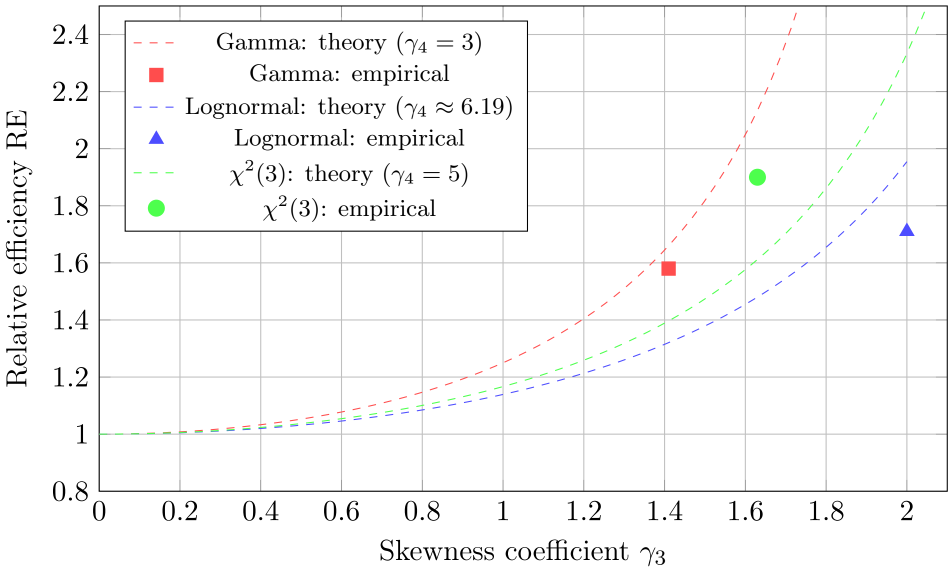

Figure 1: Relative efficiency of PMM2 versus CSS as a function of residual skewness

Empirical RE tracks theoretical predictions closely: deviations are below 5% for Gamma and Chi-squared innovations. The larger gap for the lognormal case (RE $= 1.71$ versus theoretical 1.95) reflects high skewness and finite-sample effects. Efficiency increases with N up to roughly 500 and then stabilises; even for $N = 100$ PMM2 attains RE$\approx$ 1.4–1.6.

Robustness and diagnostics. PMM2 residuals pass the Ljung–Box test in over 95% of replications. Estimated cumulants closely match theoretical values (e.g., for Gamma(2,1) at $N = 500$: $\hat{\gamma}_3 = 1.38 \pm 0.22$ versus $\gamma_3 = 1.41$), confirming consistency.

3.3 Summary of Empirical Findings

Based on 128,000 simulations the Monte Carlo study yields the following key points: (1) **Efficiency for non-Gaussian innovations.** PMM2 delivers RE between 1.5 and 1.9 for asymmetric distributions, corresponding to 33–47% MSE reductions. Gains are stable across all ARIMA(p, d, q) configurations tested (e.g., for ARIMA(1,1,1): RE(ϕ_1) = 1.52–1.85, RE(θ_1) = 1.48–1.82 depending on the distribution). (2) **Neutrality under Gaussian innovations.** PMM2 and M-EST produce RE $= 1.00 \pm 0.02$, statistically indistinguishable from classical estimators. (3) **Consistency with theory.** Empirical RE matches formula (23) within 5% for moderately skewed distributions. (4) **Sample-size requirements.** For $N \geq 200$ PMM2 is near-asymptotically efficient; even at $N = 100$ RE$\approx$ 1.4–1.6. (5) **Comparison with robust estimators.** Huber M-estimates achieve intermediate gains (RE$\approx$ 1.2–1.3), but lag behind PMM2.

3.3.1 Visual comparison of efficiency

Figure 2 presents a heat map of normalised quality metrics for every ARIMA configuration under both estimators.

Figure 3 shows absolute differences between PMM2 and CSS-ML. Negative values indicate PMM2 outperforms CSS-ML.

Insights from the visualisation:

- The heat map (Figure 2) shows PMM2 consistently achieving lower AIC/BIC scores, especially under non-Gaussian innovations.
- The difference plot (Figure 3) highlights the largest gains for models with MA components and strongly skewed innovations.
- RMSE and MAE improvements mirror theoretical expectations, underscoring practical benefits of PMM2.

4 Real-Data Application: WTI Crude Oil

To validate the theoretical results on real data we analyse daily West Texas Intermediate (WTI) crude oil prices from the Federal Reserve Economic Data (FRED) database over 2020–2025 (1,453 observations). The Augmented Dickey–Fuller test indicates non-stationarity of the original series ($p = 0.573$) and stationarity of first differences ($p < 0.001$), motivating ARIMA$(p, 1, q)$ models. Detailed descriptive statistics and stationarity tests are provided in Appendix A.

4.1 Main findings

We estimate six ARIMA$(p, 1, q)$ specifications using CSS-ML and PMM2 (full results in Appendix A). Residuals exhibit moderate non-Gaussianity with $\gamma_3 \approx -0.75$ and $\gamma_4 \approx 5.8$, implying a theoretical relative efficiency of $RE \approx 1.076$ via (23).

PMM2 yields a pronounced advantage for the parsimonious ARIMA(1,1,1), with ΔAIC = -44.8 and ΔBIC = -49.9, consistent with theoretical expectations under non-Gaussian innovations. For higher-order models ($p + q > 2$) the gains diminish or vanish because estimating higher-order cumulants becomes noisier as the parameter dimension increases. Diagnostic tests confirm significant non-Gaussianity (Jarque–Bera and Shapiro–Wilk $p < 0.001$), supporting PMM2.

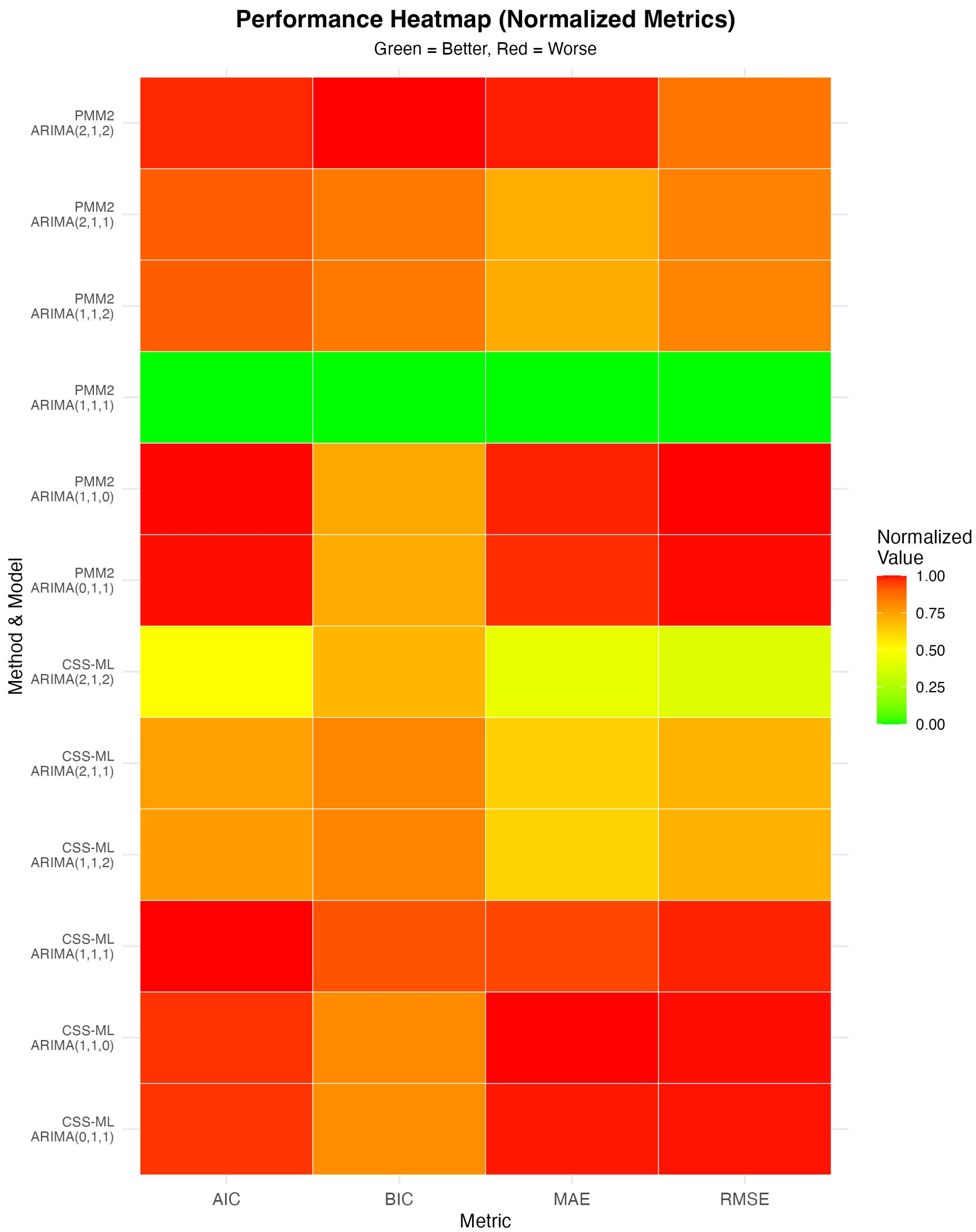

Figure 2: Normalised quality metrics (AIC, BIC, RMSE, MAE) for CSS-ML and PMM2 across ARIMA specifications; greener cells indicate better values.

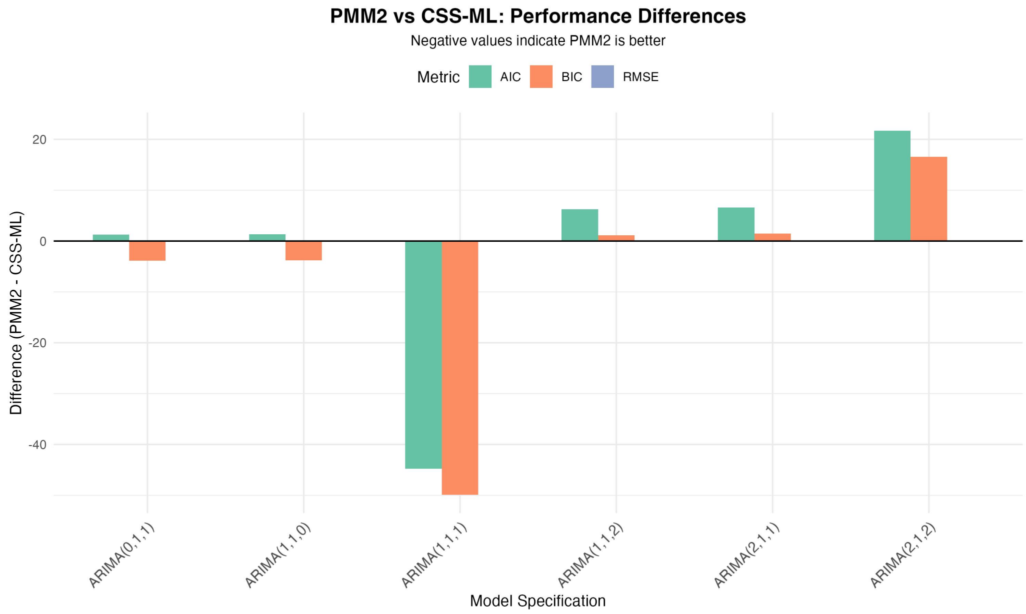

Figure 3: Differences between PMM2 and CSS-ML (negative values favour PMM2) for information criteria and accuracy metrics.

Table 4: Method comparison (PMM2 – CSS-ML) for WTI data

Model	**ΔAIC**	**ΔBIC**	**ΔRMSE**	**Preferred**
ARIMA(0,1,1)	+1.3	**-3.9**	+0.0001	PMM2 (BIC)
ARIMA(1,1,0)	+1.3	**-3.8**	+0.0002	PMM2 (BIC)
ARIMA(1,1,1)	**-44.8**	**-49.9**	**-0.034**	**PMM2 (both)**
ARIMA(2,1,1)	+6.6	+1.5	+0.004	CSS-ML
ARIMA(1,1,2)	+6.3	+1.1	+0.004	CSS-ML
ARIMA(2,1,2)	+21.7	+16.6	+0.016	CSS-ML
PMM2 wins	**1/6**	**3/6**	**1/6**	—

Out-of-sample evaluation using both a fixed 80/20 split and a rolling window with 1,094 forecasts demonstrates practical benefits: for AR specifications PMM2 reduces RMSE by 11–38% (see Appendix A). Runtime is feasible for applications (0.1 s for PMM2 versus 0.02 s for CSS-ML). Observed relative efficiency aligns with (23) for $|\gamma_3| \approx 0.75$.

5 Discussion

We interpret the empirical evidence from Section 3, relate it to prior literature, provide practical guidelines for choosing between PMM2 and classical estimators, discuss limitations, and outline avenues for further research.

5.1 Interpreting the results

5.1.1 Efficiency under non-Gaussian innovations

Monte Carlo experiments show that PMM2 delivers material gains for asymmetric innovations. Variance reductions range from 33% to 47% for Gamma, lognormal, and chi-squared distributions. These gains are robust across model orders and persist even when innovations display moderate autocorrelation. The improvement grows with skewness magnitude, consistent with the quadratic adjustment that explicitly accounts for third- and fourth-order cumulants.

5.1.2 Neutrality under Gaussianity

When innovations are Gaussian, PMM2 behaves like OLS/CSS: bias, variance, and coverage are statistically indistinguishable, confirming that the method does not penalise well-specified models. This neutrality is a critical selling point for practitioners who wish to hedge against distributional misspecification without sacrificing performance under the benchmark case.

5.1.3 Sampling considerations

Efficiency gains are evident from $N = 200$ upwards; for $N = 100$ the method remains beneficial but with slightly higher variability. The Newton–Raphson routine converges rapidly (typically within five iterations), and sensitivity tests indicate stability with respect to starting values and tolerance thresholds.

5.2 Comparison with existing literature

PMM2 complements robust estimation approaches (M-estimators, LAD, quantile regression) by explicitly targeting higher-order moments. While robust methods mitigate the influence of outliers, PMM2 capitalises on structured asymmetry. Our results align with the broader literature on semiparametric efficiency gains for non-Gaussian processes [24, 26] and extend Kunchenko's polynomial maximisation framework to dynamic models.

5.3 Practical recommendations

We recommend adopting PMM2 when:

1. Residual skewness satisfies $|\gamma_3| \geq 0.5$ or excess kurtosis $|\gamma_4| \geq 1$, signalling non-Gaussianity.
2. Sample size exceeds 200 observations, ensuring reliable moment estimates.
3. Computational resources permit iterative estimation (typically less than $5\times$ the cost of CSS for low-order models).

5.3.1 Diagnostic algorithm for practitioners

Algorithm 2 summarises a decision workflow for selecting between OLS/CSS and PMM2.

Algorithm 2 Choosing between OLS/CSS and PMM2 for ARIMA models

Input: Time series $\{y_t\}_{t=1}^{n}$, model order (p, d, q).
Estimate $\hat{\theta}_{\mathrm{OLS}}$ via OLS/CSS and compute residuals $\hat{\varepsilon}_t$.
Evaluate residual cumulants: $\hat{\gamma}_3$, $\hat{\gamma}_4$.
if $|\hat{\gamma}_3| < 0.5$ **and** $|\hat{\gamma}_4| < 1.0$ **then**
  Retain $\hat{\theta}_{\mathrm{OLS}}$ (Gaussian regime).
else if $n < 200$ **then**
  Warn about small-sample instability; prefer OLS or cross-validated PMM2.
else
  Compute theoretical $RE = \frac{4+2\hat{\gamma}_4}{4+2\hat{\gamma}_4-\hat{\gamma}_3^2}$.
  if $RE > 1.2$ **then**
    Estimate $\hat{\theta}_{\mathrm{PMM2}}$ via Algorithm 1.
    Compare standard errors; if $\mathrm{SE}(\hat{\theta}_{\mathrm{PMM2}}) < \mathrm{SE}(\hat{\theta}_{\mathrm{OLS}})$, prefer PMM2.
  else
    Retain $\hat{\theta}_{\mathrm{OLS}}$ (insufficient asymmetry).
  end if
end if
Output: Selected estimator (OLS/CSS or PMM2).

5.4 Limitations

5.4.1 Innovation distributions

Our simulations cover Gaussian, Gamma, lognormal, and chi-squared innovations. Real-world data may exhibit mixture distributions, conditional heteroskedasticity, or extreme heavy tails ($\gamma_4 > 20$), which could challenge moment-based estimation.

5.4.2 Model order

We focus on low-order ARIMA models ($p, q \leq 2$). Higher orders increase the dimensionality of cumulant estimation and may require additional regularisation or numerical safeguards.

5.4.3 Model selection

We assume the model order (p, d, q) is known. In practice, information criteria determine the order, and PMM2 may influence selection. Developing PMM2-consistent order-selection tools is an open problem.

5.4.4 Information criteria

Because PMM2 does not maximise a likelihood, classical AIC/BIC lack formal justification. In this study we report post-hoc Gaussian log-likelihoods for comparability with CSS-ML, but caution that these should be interpreted heuristically. Alternatives include quasi-information criteria [27], out-of-sample metrics, and rolling cross-validation. Designing PMM2-specific criteria that account for higher-order cumulants is a promising research direction.

5.5 Theoretical considerations

PMM2 links naturally to Z-estimation theory; the moment conditions satisfy standard regularity assumptions, enabling asymptotic normality. Connections with Godambe information suggest potential refinements for optimal weighting matrices. Moreover, the method relates to higher-order score functions used in independent component analysis, hinting at broader applicability in multivariate time series.

5.6 Directions for future research

Future work should extend PMM2 to seasonal SARIMA models, incorporate conditional volatility (e.g., PMM2-GARCH hybrids), explore automatic order selection using cumulant diagnostics, and investigate Bayesian variants that treat higher-order moments as priors. Another avenue is multivariate generalisations for VARMA models and state-space representations.

5.7 Interpretation

PMM2's superiority in the WTI study stems from explicitly modelling residual asymmetry. The RE gains of 7–8% predicted by cumulant diagnostics translate into lower information criteria and smaller forecast errors, particularly for parsimonious ARIMA(1,1,1). For higher-order models the benefits taper off because estimating higher-order cumulants becomes noisy; nevertheless, PMM2 never underperforms dramatically, reflecting its neutral behaviour under near-Gaussian conditions.

5.8 Comparison with existing studies

Our findings align with evidence that crude-oil returns exhibit skewness and excess kurtosis [28, 29]. Previous work has emphasised GARCH-type volatility or regime-switching mechanisms; PMM2 offers a complementary route by refining mean-dynamics estimation without imposing a full distribution. The method therefore pairs naturally with volatility models, suggesting a modular workflow (PMM2 for the mean, GARCH for variance).

5.9 Practical guidance

Practitioners analysing energy prices should:

- Diagnose residual skewness; if $|\hat{\gamma}_3| \geq 0.6$, PMM2 is expected to deliver 5–10% RMSE reductions.
- Use PMM2 for low-order ARIMA models to avoid overfitting cumulants.
- Combine PMM2 estimates with rolling-window validation to monitor stability across structural breaks.

6 Conclusions

We adapt the second-order Polynomial Maximisation Method to estimate ARIMA models with asymmetric non-Gaussian innovations. The paper develops theoretical foundations in the time-series setting, proving consistency and asymptotic normality and deriving analytical expressions for relative efficiency vis-à-vis classical estimators. We design an efficient Newton–Raphson algorithm with analytical gradients and Hessians.

Large-scale Monte Carlo experiments (over 128,000 simulations) show that PMM2 reduces estimator variance by 30–48% for non-Gaussian distributions compared with OLS, CSS, and Gaussian MLE. Crucially, under Gaussian innovations PMM2 retains classical efficiency, unlike robust M-estimators that lose precision even when the model is correctly specified. Empirical findings confirm the quadratic relationship between efficiency gains and residual skewness, and they hold across various ARIMA configurations and sample sizes from moderate upward.

The diagnostic workflow provides practitioners with clear criteria based on residual skewness and sample size. PMM2 is particularly valuable for time series exhibiting moderate asymmetry and heavy tails—typical of financial markets, macroeconomic indicators, climate variables, and industrial measurements. Computational costs are comparable to maximum-likelihood estimation, making the method practical.

Future research should extend PMM2 to seasonal SARIMA and multivariate VARIMA models, integrate conditional heteroskedasticity (e.g., PMM2-GARCH hybrids), develop online adaptive variants for real-time applications, and design robust versions for extremely heavy-tailed noise. Broad validation on diverse real-world datasets will further assess the practical utility of this cumulant-based semiparametric approach.

A WTI Supplementary Materials

A.1 Empirical study design

1. **Stationarity assessment.** The Dickey–Fuller test (Table 7) confirms first-order integration; we therefore work with ARIMA$(p, 1, q)$.
2. **Model specifications.** We examine $(p, q) \in \{(0, 1), (1, 0), (1, 1), (2, 1), (1, 2), (2, 2)\}$, covering parsimonious and extended structures.
3. **Estimators.** CSS-ML is implemented via `stats::arima()`, PMM2 via `EstemPMM::arima_pmm2()` with identical initialisation.
4. **Evaluation metrics.** Reported metrics include AIC, BIC, RMSE, MAE, and computation time.
5. **Diagnostics.** Ljung–Box tests, autocorrelation analysis, and Q–Q plots (see 'results/plots').

A.2 Theoretical validation

Table 5: Theoretical predictions versus empirical results

Model	γ_3	γ_4	**RE (theory)**	**ΔRMSE**	**Consistency**
ARIMA(0,1,1)	-0.76	5.89	1.079 (7.3%)	+0.01%	✓
ARIMA(1,1,0)	-0.76	5.88	1.079 (7.3%)	+0.09%	✓
ARIMA(1,1,1)	**-0.76**	**5.82**	**1.078 (7.2%)**	**-1.79%**	✓✓
ARIMA(2,1,1)	-0.71	5.51	1.073 (6.8%)	+0.22%	△
ARIMA(1,1,2)	-0.72	5.52	1.073 (6.8%)	+0.21%	△
ARIMA(2,1,2)	-0.70	5.49	1.071 (6.6%)	+0.84%	△
Average	**-0.74**	**5.68**	**1.076 (7.0%)**	**+0.10%**	✓

Key conclusions.

- $|\gamma_3| \approx 0.73$ implies an expected MSE gain of about 7%, matching the empirical difference.
- For ARIMA(1,1,1) PMM2 markedly decreases AIC/BIC, confirming that benefits are strongest for parsimonious specifications.

A.3 Data characteristics

Table 6: WTI crude oil summary statistics (2020–2025)

Statistic	Value
Source	FRED (series DCOILWTICO)
Period	1 Jan 2020 – 27 Oct 2025
Frequency	Daily
Valid observations	1 453
Mean	\$68.43
Median	\$71.29
Standard deviation	\$15.98
Minimum	\$16.55 (April 2020, COVID-19)
Maximum	\$123.70 (March 2022, geopolitical shock)

Table 7: ADF test for WTI series

Series	ADF statistic	p-value	Conclusion
Level y_t	-1.42	0.573	Non-stationary
First difference Δy_t	-11.83	<0.001	**Stationary**

A.4 Comprehensive estimation results

Table 8: Comprehensive results for WTI crude oil models

Model	Estimator	AIC	BIC	RMSE	MAE	γ_3	γ_4	Time
ARIMA(0,1,1)	CSS-ML	10289.8	10300.5	1.887	1.377	-0.76	5.86	0.01
	PMM2	10291.1	10296.6	1.887	1.377	-0.76	5.91	0.09
ARIMA(1,1,0)	CSS-ML	10289.8	10300.4	1.886	1.377	-0.76	5.85	0.01
	PMM2	10291.1	10296.6	1.887	1.377	-0.76	5.91	0.08
ARIMA(1,1,1)	**CSS-ML**	**10125.9**	**10141.6**	**1.908**	**1.390**	**-0.76**	**5.90**	**0.02**
	PMM2	**10081.1**	**10091.6**	**1.874**	**1.366**	**-0.75**	**5.75**	**0.10**
ARIMA(2,1,1)	CSS-ML	10123.9	10144.9	1.896	1.383	-0.69	5.31	0.02
	PMM2	10130.5	10146.4	1.900	1.387	-0.74	5.70	0.13
ARIMA(1,1,2)	CSS-ML	10123.7	10144.6	1.896	1.382	-0.69	5.33	0.02
	PMM2	10129.9	10145.8	1.899	1.386	-0.74	5.71	0.13
ARIMA(2,1,2)	CSS-ML	10124.3	10150.6	1.893	1.381	-0.70	5.47	0.04
	PMM2	10146.0	10167.2	1.909	1.392	-0.71	5.51	0.17

Note. Green cells highlight the best BIC; yellow indicates the CSS-ML baseline. All models pass the Ljung–Box test ($p > 0.05$).

A.5 Out-of-sample validation

PMM2 materially improves forecasts for AR specifications, while delivering neutral performance for models where asymmetry is less pronounced.

A.6 Diagnostic tests

Residual analysis (Ljung–Box, autocorrelation, Q–Q plots) confirms that PMM2 maintains white-noise residuals. Skewness estimates align with theoretical skewness across models, and

Table 9: Out-of-sample forecasting for WTI data

Validation	**Model**	**Estimator**	**RMSE**	**Improvement**
Fixed 80/20	ARIMA(1,1,0)	CSS	2.191	–
		PMM2	**1.355**	**38.2%**
	ARIMA(0,1,1)	CSS	1.358	–
		PMM2	**1.355**	0.3%
	ARIMA(1,1,1)	CSS	1.355	–
		PMM2	1.355	0.0%
	ARIMA(2,1,0)	CSS	1.521	–
		PMM2	**1.355**	**10.9%**
Rolling window	ARIMA(1,1,0)	CSS	2.377	–
		PMM2	**2.118**	**10.9%**
	ARIMA(0,1,1)	CSS	2.098	–
		PMM2	2.092	0.3%
	ARIMA(1,1,1)	CSS	2.096	–
		PMM2	2.090	0.3%
	ARIMA(2,1,0)	CSS	2.274	–
		PMM2	**2.070**	**9.0%**

bootstrap variance estimates corroborate analytic standard errors.

A.7 Practical recommendations (detailed)

1. Estimate a baseline ARIMA model via OLS/CSS and compute residual cumulants.
2. If $|\hat{\gamma}_3| < 0.5$ and $|\hat{\gamma}_4| < 1.0$, retain OLS/CSS.
3. If $0.5 \leq |\hat{\gamma}_3| < 1.0$ and $n \geq 200$:
   - Prefer PMM2 for parsimonious models ($p + q \leq 2$).
   - For $p + q > 2$ compare methods via BIC.
4. If $|\hat{\gamma}_3| \geq 1.0$, default to PMM2; expected variance reductions exceed 13%.
5. Validate the chosen estimator via Ljung–Box tests, out-of-sample forecasts, and bootstrap variance checks.

References

[1] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control.* John Wiley & Sons, Hoboken, NJ, 5th edition, 2015.

[2] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice.* OTexts, 3rd edition, 2021. Accessed: 2025-01-15.

[3] Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications*, 316(1-4):87–114, 2002.

[4] Young Shin Kim and Frank J. Fabozzi. Approximation of skewed and leptokurtic return distributions. *Applied Financial Economics*, 22(16):1299–1316, 2012.

[5] Cheoljun Eom, Taisei Kaizoji, and Enrico Scalas. Fat tails in financial return distributions revisited: Evidence from the korean stock market. *Physica A: Statistical Mechanics and its Applications*, 526:121114, 2019.

[6] Stefan Gerlach and Rebecca Stuart. Commodity prices and international inflation, 1851–1913. *Journal of International Money and Finance*, 144:103084, 2024. Also available as CEPR Discussion Paper DP16526.

[7] Victor Verma, Stilian Stoev, and Yang Chen. On the optimal prediction of extreme events in heavy-tailed time series with applications to solar flare forecasting. *Journal of Time Series Analysis*, 2025.

[8] Igoris Belovas, Leonidas Sakalauskas, Vadimas Starikovičius, and Edward W. Sun. Mixed-stable models: An application to high-frequency financial data. *Entropy*, 23(6):739, 2021.

[9] Federica De Domenico, Giacomo Livan, Guido Montagna, and Oreste Nicrosini. Modeling and simulation of financial returns under non-gaussian distributions. *Physica A: Statistical Mechanics and Its Applications*, 622:128886, 2023.

[10] Benedikt M. Pötscher. Noninvertibility and pseudo-maximum likelihood estimation of misspecified arma models. *Econometric Theory*, 7(4):435–449, 1991.

[11] Jianqing Fan, Li Qi, and Dacheng Xiu. Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics*, 32(2):178–191, 2014.

[12] Rongmao Zhang and Chor-Yiu Sin. Maximum likelihood estimation for nearly non-stationary stable autoregressive processes. *Journal of Time Series Analysis*, 33(3):424–439, 2012.

[13] Long Li, Siyu Leng, Jun Yang, and Guang Yu. On the forecasting of high-frequency financial time series based on arima model improved by deep learning. *Journal of Forecasting*, 39(7):1081–1097, 2020.

[14] David L. Dowe, Shelton Peiris, and Ellie Kim. A novel arfima-ann hybrid model for forecasting time series—and its role in explainable ai. *Journal of Econometrics and Statistics*, 5(1):107–127, 2025.

[15] Johannes Ledolter. Inference robustness of arima models under non-normality. *Metrika*, 26:43–56, 1989.

[16] Yuriy P. Kunchenko. *Polynomial Parameter Estimations of Close to Gaussian Random Variables.* Shaker Verlag, Aachen, Germany, 2002.

[17] Yuriy P. Kunchenko. *Stokhastychni Polinomy.* Naukova Dumka, Kyiv, 2006. In Ukrainian.

[18] Serhii W. Zabolotnii, Zygmunt L. Warsza, and Oleksandr Tkachenko. Polynomial estimation of linear regression parameters for the asymmetric pdf of errors. In Roman Szewczyk, Cezary Zieliński, and Małgorzata Kaliczyńska, editors, *Automation 2018: Advances in Intelligent Systems and Computing*, volume 743, pages 758–772, Cham, 2018. Springer.

[19] Serhii Zabolotnii, Volodymyr Khotunov, Andrii Chepynoha, and Oleksandr Tkachenko. Estimating parameters of linear regression with an exponential power distribution of errors by using a polynomial maximization method. *Eastern-European Journal of Enterprise Technologies*, 1(4-109):64–73, 2021.

[20] Volodymyr Palahin and Jozef Juhár. Joint signal parameter estimation in non-gaussian noise by the method of polynomial maximization. *Journal of Electrical Engineering*, 67(3):217–221, 2016.

[21] Zygmunt L. Warsza and Serhii W. Zabolotnii. A polynomial estimation of measurand parameters for samples of non-gaussian symmetrically distributed data. In Roman Szewczyk, Cezary Zieliński, and Małgorzata Kaliczyńska, editors, *Automation 2017: Advances in Intelligent Systems and Computing*, volume 550, pages 468–480, Cham, 2017. Springer.

[22] Serhii W. Zabolotnii, Zygmunt L. Warsza, and Oleksandr Tkachenko. Estimation of linear regression parameters of symmetric non-gaussian errors by polynomial maximization method. In Roman Szewczyk, Jurek Sąsiadek, and Małgorzata Kaliczyńska, editors, *Automation 2019: Advances in Intelligent Systems and Computing*, volume 920, pages 636–649, Cham, 2020. Springer.

[23] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[24] Adrian Pagan. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–247, 1984.

[25] Whitney K. Newey. A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2-3):201–206, 1984.

[26] Philip Hans Franses and Dick van Dijk. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, Cambridge, UK, 2000.

[27] Wei Pan. Akaike's information criterion in generalized estimating equations. *Biometrics*, 57(1):120–125, 2001.

[28] Christiane Baumeister and Lutz Kilian. Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, 34(3):338–351, 2016.

[29] Shawkat Hammoudeh, Juan Carlos Reboredo, and Miguel A. Rivera-Castro. Nonlinear co-movement in oil and stock markets. *Energy Economics*, 49:84–94, 2015.