Modern Journal of Statistics
2(2), 237–257
DOI:https://doi.org/10.64389/mjs.2026.022133
https://sphinxsp.org/journal/index.php/mjs

Research article

Dynamic Stability and Neural Residual Forecasting of an Inverse Lomax Autoregressive Model with Application to NSPO Stock Prices

Aida A. Mohammed

Ministry of Education, General Directorate of Education in Salah al-Din, Iraq

Correspondence: Ayeda.mohammed@st.tu.edu.iq

ARTICLE INFO

Keywords:
Dynamic stability
Neural residual
Inverse Lomax
Autoregressive model
NSPO stock prices

Mathematics Subject Classification:
62M10, 62M45, 37N40

Important Dates:
Received: 5 August 2026
Revised: 11 November 2026
Accepted: 11 November 2026
Online: 15 November 2026

Copyright © 2026 by the authors. Published under Creative Commons Attribution (CC BY) license.

ABSTRACT

This study investigates an inverse-Lomax smooth-transition autoregressive (ILAR) model integrated with neural residual forecasting, using 319 weekly NSPO stock-price observations. The ILAR(1) map has a positive deterministic fixed point of 2.3507 and local spectral radius of 0.9345, but its derivative is 1.0193. The fitted deterministic map does not satisfy the sufficient global-contraction condition over the examined state interval. The estimated transition-shape parameter reaches its imposed upper bound ($\alpha = 5$) and remains boundary-constrained when the bound is extended to 8. ADF and Phillips–Perron tests do not reject the unit-root null for price levels at 5%. The initial single-run ILAR-TCNA evaluation yielded an RMSE of 0.1427, compared with 0.1474 for the naive benchmark. A separate 20-seed experiment with trainable four-head attention produced mean test RMSE 0.1513 (SD 0.0063) and mean MAE 0.1215 (SD 0.0050). The validation procedure selected zero residual weight in 16 runs; the seed-4 Diebold–Mariano tests did not reject equal predictive accuracy against the naive and AR(1) forecasts. Penalized transition-parameter profiles also support weak identification. These findings indicate local deterministic stability but neither stochastic stationarity nor a reproducible neural forecasting advantage.

1. Introduction

Financial price series often combine strong persistence, abrupt level changes, asymmetric responses, and residual dependence. A linear autoregressive model can represent short-memory dependence, but its fixed coefficients cannot adapt to different price regimes. Nonlinear autoregressive models address this limitation by allowing the transition map to change with the state of the process. Dynamic analysis then determines whether the resulting paths approach a fixed point, oscillate around an attractor, or become unstable [1-3, 22].

The statistical component of the proposed model is in the so-called smooth transition autoregressive class. In a classic STAR model, regime weights change smoothly with an observable transition variable, where logistic and exponential transition functions are used. Teräsvirta [4] developed specification, nonlinear least-squares estimation, as well as diagnostic analysis of STAR models, and van Dijk, Teräsvirta, and Franses [5] looked at the STAR models and their applications. The proposed ILAR specification employs the inverse-Lomax cumulative distribution function as a transition mechanism within the smooth-transition autoregressive framework.

Recent work on nonlinear and positive-valued time series provides complementary perspectives. Guerouah et al. [25] developed an XLindley ARMA model with a positive conditional distribution and observation-driven conditional mean, whereas the present ILAR formulation uses an inverse-Lomax transition weight in an autoregressive mean equation. Laiche and Zeghdoudi [26]

studied bilinear time-series models driven by GARCH white noise, emphasizing nonlinear dependence and time-varying volatility in financial dynamics. Unlike a positivity-preserving conditional-distribution model, the Gaussian quasi-likelihood ILAR specification requires separate assumptions for positivity and stochastic invariance.

The methodological contribution consists of integrating an inverse-Lomax transition mechanism with Jacobian-based local stability analysis, monodromy-matrix analysis of periodic orbits, and neural residual forecasting. The empirical interpretation is not related to the formal two regimes, but if the estimated transition weights are close to zero, we can assume an effective lower-regime model in spite of the fact that the upper regime coefficient is used in the formal models. The nonlinear time series work of Ozaki [1, 2] and Tong [3] motivates the dynamical system approach and the STAR formulation follows Teräsvirta [4] and van Dijk et al. [5]. Inverse Lomax studies provide the distributional contribution [6-8] and temporal convolutional and attention-based studies motivate the residual learner [9-14]. The proposed framework combines an inverse-Lomax transition function with neural residual forecasting while distinguishing theoretical stability conditions from the properties established for the fitted model.

The proposed framework develops an inverse-Lomax smooth-transition autoregressive specification and examines its dynamical properties through fixed-point analysis, local Jacobian stability, global contraction conditions, stochastic Wasserstein contraction, and orbital stability. The empirical application establishes local stability and monotone convergence of the fitted ILAR(1) map over the examined positive interval, whereas the sufficient conditions for global contraction are not satisfied. The neural residual component is evaluated as a forecasting extension, with its predictive performance assessed independently of the theoretical stability properties. Comparisons between classical time-series forecasting and neural approaches in a different empirical setting were reported by Ibrahim et al. [21].

2. Inverse Lomax Distribution

Let Z be a positive random variable. The two-parameter inverse Lomax cumulative distribution function is

$$F_{IL}(z;\alpha,\lambda) = \left(\frac{z}{z+\lambda}\right)^{\alpha}, \qquad z > 0,$$

where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. Its probability density function is

$$f_{IL}(z;\alpha,\lambda) = \frac{\alpha\lambda z^{\alpha-1}}{(z+\lambda)^{\alpha+1}}, \qquad z > 0.$$

The survival function is

$$S_{IL}(z;\alpha,\lambda) = 1 - \left(\frac{z}{z+\lambda}\right)^{\alpha}.$$

The quantile function follows from $u = F_{IL}(z)$:

$$Q_{IL}(u) = \frac{\lambda u^{1/\alpha}}{1-u^{1/\alpha}}, \qquad 0 < u < 1.$$

The derivative of the cumulative distribution function is

$$F_{IL}{}'(z;\alpha,\lambda) = \frac{\alpha\lambda z^{\alpha-1}}{(z+\lambda)^{\alpha+1}}.$$

For all $z > 0$, $F_{IL}(z)$ lies strictly between zero and one. It approaches zero as z approaches zero and approaches one as z tends to infinity. Increasing α creates a sharper transition, while λ shifts the transition along the positive axis [6-8].

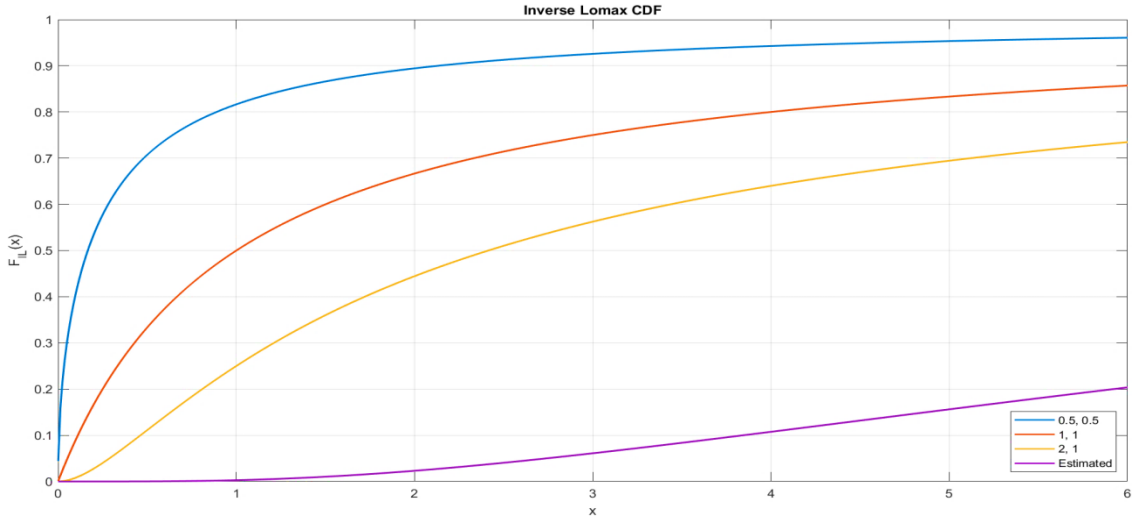

Figure 1. Inverse Lomax cumulative distribution function for selected parameter values.

3. The Proposed Inverse Lomax Autoregressive Model

Let $\{X_t\}$ be a positive time series and let p denote the autoregressive order. Define the scaled transition variable

$$Z_{t-1} = \frac{X_{t-1}}{s},$$

where $s > 0$ is a robust scale constant. In the empirical study, s equals the median of the training observations. The inverse Lomax transition weight is

$$G_t = F_{IL}(Z_{t-1}; \alpha, \lambda) = \left(\frac{X_{t-1}/s}{X_{t-1}/s + \lambda}\right)^{\alpha}.$$

The proposed ILAR(p) model is

$$X_t = (1 - G_t)\sum_{i=1}^{p} \phi_i X_{t-i} + G_t \sum_{i=1}^{p} \psi_i X_{t-i} + \varepsilon_t,$$

where $\phi = (\phi_1, \dots, \phi_p)^T$ and $\psi = (\psi_1, \dots, \psi_p)^T$ are the two regime coefficient vectors, and

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2).$$

The additive Gaussian innovation is used as a conditional quasi-likelihood device for estimation. As an unrestricted Gaussian error has support on the whole real line, it does not by itself preserve the positive state space. Positivity and stochastic invariance require additional assumptions on the admissible state region and innovation support, as neither property follows from unrestricted Gaussian innovations alone. The positivity preserving stochastic extension would be to restrict the innovation support or to use a positive link function. An equivalent coefficient-modulation form is

$$X_t = \sum_{i=1}^{p} [\phi_i + (\psi_i - \phi_i)G_t] X_{t-i} + \varepsilon_t.$$

When G_t is close to zero, the dynamics are governed by ϕ. When G_t is close to one, the dynamics are governed by ψ. For intermediate values, the model creates a smooth probabilistic transition between the two regimes.

Therefore, ILAR(p) can be regarded as a STAR model with an inverse-Lomax transition function. To understand the lower regime and upper regime, we use the term limiting coefficient vectors to mean that the real regime activation must be checked from the range of G_t rather than from the parameterization [4, 5]. Define the state vector

$$V_t = (X_t, X_{t-1}, \dots, X_{t-p+1})^T.$$

The model can be represented as

$$V_t = \Phi(V_{t-1}) + e_t,$$

where

$$e_t = (\varepsilon_t, 0, \dots, 0)^T,$$

and the first component of Φ is

$$\Phi_1(v) = [1 - G(v_1)]\phi^T v + G(v_1)\psi^T v.$$

The remaining components shift the state:

$$\Phi_j(\mathbf{v}) = v_{j-1}, \qquad j = 2, \dots, p.$$

4. Parameter Estimation and Order Selection

For fixed α and λ, the model is linear in ϕ and ψ. This separability reduces a high-dimensional nonlinear optimization problem to a two-dimensional bounded search.

For $t = p + 1, \dots, n$, define

$$x_t = (X_{t-1}, \dots, X_{t-p})^T$$

and

$$w_t(\alpha, \lambda) = \begin{bmatrix} (1 - G_t)x_t \\ G_t x_t \end{bmatrix}.$$

Stacking the vectors gives the design matrix W. The regime coefficients are estimated by a lightly regularized least-squares solution:

$$\hat{\beta} = (W^T W + \tau I)^{-1} W^T y,$$

where

$$\hat{\beta} = \left(\hat{\phi}^T, \hat{\psi}^T\right)^T$$

The ridge constant is fixed at 10^{-6} only to prevent numerical singularity. It is computational regularization, not an estimated model parameter. The value 10^{-6} matches the ridge term.

$$\mathcal{J}(\alpha, \lambda) = \log\left(\frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2\right) + 10^{-4}\|\hat{\beta}\|_2^2$$

on the compact parameter region

$$0.2 \le \alpha \le 5, \qquad 0.1 \le \lambda \le 5.$$

The innovation variance is estimated by

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2.$$

For model-order selection, the residual sum of squares defines a profiled Gaussian conditional quasi-likelihood criterion. Since the regime coefficients are obtained by least squares and the transition parameters are obtained by bounded SSE minimization, the criterion can be regarded as quasi-likelihood based and not as free exact maximum likelihood.

$$\hat{\ell} = -\frac{n-p}{2}[\log(2\pi\hat{\sigma}^2) + 1].$$

For k fitted quantities, the corresponding information criteria are computed from this profiled Gaussian quasi-likelihood [17, 18].

$$AIC = -2\hat{\ell} + 2k,$$
$$BIC = -2\hat{\ell} + k\log(n-p),$$
$$CAIC = -2\hat{\ell} + k[\log(n-p) + 1],$$

and

$$HQIC = -2\hat{\ell} + 2k\log\log(n-p).$$

A candidate order is retained only when its estimated fixed point is admissible and the local spectral radius is below one. The final order minimizes the quasi-BIC (QBIC) among these candidates; the resulting criterion is denoted by quasi-BIC (QBIC), reflecting the profiled Gaussian quasi-likelihood formulation adopted for parameter estimation.

5. Fixed Point, Local Stability, and Geometric Stability

Let

$$A = \sum_{i=1}^{p}\psi_i, \qquad C = \sum_{i=1}^{p}\phi_i.$$

The deterministic skeleton of the ILAR(p) model is obtained by setting $\varepsilon_t = 0$. A constant positive solution has the form

$$X_t = X_{t-1} = \cdots = X_{t-p} = \mu > 0.$$

Substitution into the deterministic model gives

$$\mu = \left[\left(1 - G(\mu)\right)C + G(\mu)A\right]\mu.$$

Because $\mu > 0$, division by μ yields

$$1 = C + (A - C)G(\mu).$$

When $A \neq C$, define

$$\kappa = \frac{1-C}{A-C}.$$

The zero state is always a fixed point because $G(0) = 0$ and the deterministic map sends the zero vector to itself. The following theorem completely characterizes the isolated positive fixed point.

Theorem 1. Complete existence and uniqueness result for the positive nonzero fixed point: Assume $A \neq C$. The deterministic ILAR(p) model possesses an isolated finite positive fixed point if and only if

$$0 < \kappa = \frac{1-C}{A-C} < 1.$$

When this condition holds, the positive fixed point is unique and is given by

$$\mu = s\frac{\lambda\kappa^{1/\alpha}}{1-\kappa^{1/\alpha}}.$$

If $A = C$, then no isolated positive fixed point exists. More precisely, if $A = C \neq 1$, no positive constant solution exists, whereas if $A = C = 1$, every positive constant is a fixed point and none of these points is isolated.

Proof

Let $\mu = (\mu, \dots, \mu)^T$ with $\mu > 0$. At a fixed point, every lag equals μ, so

$$\sum_{i=1}^{p} \phi_i X_{t-i} = C\mu, \qquad \sum_{i=1}^{p} \psi_i X_{t-i} = A\mu.$$

Therefore, the fixed-point equation becomes

$$\mu = [C + (A - C)G(\mu)]\mu.$$

Since $\mu > 0$, this equation is equivalent to

$$G(\mu) = \frac{1 - C}{A - C} = \kappa.$$

The scaled inverse Lomax transition function is

$$G(x) = \left(\frac{x/s}{x/s + \lambda}\right)^{\alpha} = \left(\frac{x}{x + s\lambda}\right)^{\alpha}, \qquad x > 0.$$

It is continuous and strictly increasing because

$$G'(x) = \frac{1}{s}\frac{\alpha\lambda(x/s)^{\alpha-1}}{(x/s + \lambda)^{\alpha+1}} > 0.$$

Moreover,

$$\lim_{x \downarrow 0} G(x) = 0, \qquad \lim_{x \to \infty} G(x) = 1.$$

Hence the equation $G(\mu) = \kappa$ has one and only one positive solution exactly when $0 < \kappa < 1$. Taking the positive αth root gives

$$\kappa^{1/\alpha} = \frac{\mu}{\mu + s\lambda}.$$

Consequently,

$$\mu\left(1 - \kappa^{1/\alpha}\right) = s\lambda\kappa^{1/\alpha},$$

and therefore

$$\mu = s\frac{\lambda\kappa^{1/\alpha}}{1 - \kappa^{1/\alpha}}.$$

This value is positive and finite because $0 < \kappa^{1/\alpha} < 1$. Conversely, substituting this value into G gives $G(\mu) = \kappa$, and substitution into the deterministic ILAR equation returns μ. This proves necessity, sufficiency, and uniqueness when $A \neq C$.

If $A = C$, the fixed-point equation reduces to $1 = \mathrm{C}$. Thus $\mathrm{C} \neq 1$ excludes every positive constant solution, while $C = 1$ makes the equation valid for every $\mu > 0$. In the latter case, the fixed points form a continuum and are not isolated. This completes the proof.

5.1 Jacobian matrix and local asymptotic stability

Write the deterministic state map as

$$\Phi(v) = \begin{pmatrix} \Phi_1(v) \\ v_1 \\ \vdots \\ v_{p-1} \end{pmatrix},$$

where $\Phi_1(\mathbf{v}) = \boldsymbol{\phi}^{\mathrm{T}}\mathbf{v} + \mathrm{G}(\mathrm{v}_1)(\boldsymbol{\psi} - \boldsymbol{\phi})^{\mathrm{T}}\mathbf{v}$.

At the positive fixed point, define $g_\mu = G(\mu) = \kappa$ and

$$g_\mu{}' = G'(\mu) = \frac{1}{s}\frac{\alpha\lambda(\mu/s)^{\alpha-1}}{(\mu/s + \lambda)^{\alpha+1}}.$$

Differentiation of the first component gives

$$\frac{\partial \Phi_1}{\partial v_1} = \phi_1 + (\psi_1 - \phi_1)G(v_1) + G'(v_1)(\psi - \phi)^T v,$$

and, for $j = 2, \dots, p$,

$$\frac{\partial \Phi_1}{\partial v_j} = \phi_j + \left(\psi_j - \phi_j\right)G(v_1).$$

Since $(\boldsymbol{\psi} - \boldsymbol{\phi})^T\boldsymbol{\mu} = \mu(A - C)$, the first row of the Jacobian $\boldsymbol{J}(\mu) = D\boldsymbol{\Phi}(\boldsymbol{\mu})$ is

$$h_1 = \phi_1 + (\psi_1 - \phi_1)\kappa + \mu g_\mu{}'(A - C),$$

and

$$h_j = \phi_j + \left(\psi_j - \phi_j\right)\kappa, \qquad j = 2, \dots, p.$$

Thus

$$J(\mu) = \begin{pmatrix} h_1 & h_2 & \cdots & h_{p-1} & h_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Its characteristic polynomial is

$$P(r) = r^p - h_1 r^{p-1} - h_2 r^{p-2} - \cdots - h_p.$$

Theorem 2. Complete local-stability result: If $\rho\big(\boldsymbol{J}(\mu)\big) < 1$, then the positive fixed point is locally asymptotically stable and the convergence is locally geometric. Specifically, there exist a neighborhood $\mathcal{U}$ of $\boldsymbol{\mu}$, a constant $K \geq 1$, and a number $q \in (0,1)$ such that

$$\|V_t - \mu\| \leq K q^t \|V_0 - \mu\|, \qquad V_0 \in \mathcal{U}.$$

If $\rho\big(J(\mu)\big) > 1$, the fixed point is unstable. Therefore, when no eigenvalue lies on the unit circle, local asymptotic stability is equivalent to all roots of $P(r) = 0$ lying strictly inside the unit circle.

Proof

Let

$$u_t = V_t - \mu.$$

Since Φ is continuously differentiable near μ, Taylor's formula gives

$$u_t = J(\mu)u_{t-1} + R(u_{t-1}),$$

where

$$\lim_{u \to 0} \frac{\| R(u) \|}{\| u \|} = 0.$$

Assume $\rho\big(J(\mu)\big) < 1$. Choose q_0 such that

$$\rho\big(J(\mu)\big) < q_0 < 1.$$

A standard matrix-norm result guarantees the existence of an equivalent vector norm $\|\cdot\|_e$ whose induced matrix norm satisfies

$$\| J(\mu) \|_e < q_0.$$

Because $D\Phi$ is continuous, there is a radius $\delta > 0$ and a number $q \in (q_0, 1)$ such that, for every v satisfying $\| v - \mu \|_e \leq \delta$,

$$\| D\Phi(v) \|_e \leq q < 1.$$

For any v in this neighborhood, the mean-value inequality gives

$$\| \Phi(v) - \Phi(\mu) \|_e \leq q \, \| v - \mu \|_e.$$

Since $\Phi(\mu) = \mu$, repeated application yields

$$\| V_t - \mu \|_e \leq q^t \, \| V_0 - \mu \|_e.$$

All norms in a finite-dimensional space are equivalent, so constants $c_1, c_2 > 0$ exist with

$$c_1 \, \| z \| \leq \| z \|_e \leq c_2 \, \| z \|.$$

Therefore,

$$\| V_t - \mu \| \leq \frac{c_2}{c_1} q^t \, \| V_0 - \mu \|,$$

which proves local asymptotic stability and geometric convergence with $K = c_2/c_1$.

Now suppose $\rho\big(J(\mu)\big) > 1$. Then $J(\mu)$ has an eigenvalue whose modulus exceeds one. A sufficiently small perturbation with a component in the corresponding invariant direction is enlarged by the linear part. The remainder is of smaller order than the perturbation, so it cannot cancel this expansion in a sufficiently small neighborhood. Hence arbitrarily small initial perturbations leave that neighborhood, and the fixed point is unstable. The case of an eigenvalue on the unit circle is non-hyperbolic and requires higher-order analysis. This completes the proof.

5.2 Global geometric convergence and forgetting of initial values

The following result states the strongest direct condition under which all admissible initial state vectors converge to one fixed point at a uniform geometric rate.

Theorem 3. Global contraction on an invariant state region

Let $\mathcal{D} \subset \mathbb{R}^p$ be a nonempty closed and convex region such that $\boldsymbol{\Phi}(\mathcal{D}) \subseteq \mathcal{D}$. Assume that, for some norm $\|\cdot\|_e$ and some $q \in (0,1)$, the following derivative bound holds for every $\mathbf{v} \in \mathcal{D}$:

$$\| D\boldsymbol{\Phi}(\mathbf{v}) \|_e \leq q.$$

Here the norm on states is fixed, the matrix norm is its induced operator norm, and the deterministic map $\boldsymbol{\Phi}$ is continuously differentiable on an open neighborhood of $\mathcal{D}$. Positive invariance $\boldsymbol{\Phi}(\mathcal{D}) \subseteq \mathcal{D}$ is required, rather than inferred from local stability.

Then $\mathbf{\Phi}$ has exactly one fixed point $\mathbf{v}^{\mathrm{fp}}$ in $\mathcal{D}$, and for every pair of initial states $\mathbf{u}_0, \mathbf{v}_0 \in \mathcal{D}$,

$$\| \mathbf{\Phi}^t(\mathbf{u}_0) - \mathbf{\Phi}^t(\mathbf{v}_0) \|_e \leq q^t \| \mathbf{u}_0 - \mathbf{v}_0 \|_e.$$

In particular,

$$\| \Phi^t(v_0) - v^{\mathrm{fp}} \|_e \leq q^t \| v_0 - v^{\mathrm{fp}} \|_e,$$

so every admissible initial state converges to the same point.

Proof

For $\mathbf{u}, \mathbf{v} \in \mathcal{D}$, the mean-value inequality and the derivative bound imply

Indeed, for $\mathbf{u}, \mathbf{v} \in \mathcal{D}$, convexity ensures that the entire segment $\mathbf{v} + \theta(\mathbf{u} - \mathbf{v}),\ 0 \leq \theta \leq 1$, is contained in $\mathcal{D}$. The integral form of the mean-value theorem therefore applies along this segment:

$$\mathbf{\Phi}(\mathbf{u}) - \mathbf{\Phi}(\mathbf{v}) = \int_0^1 \mathrm{D}\mathbf{\Phi}\big(\mathbf{v} + \theta(\mathbf{u} - \mathbf{v})\big)(\mathbf{u} - \mathbf{v})\, \mathrm{d}\theta.$$

$$\| \mathbf{\Phi}(\mathbf{u}) - \mathbf{\Phi}(\mathbf{v}) \|_e \leq q \| \mathbf{u} - \mathbf{v} \|_e.$$

Thus Φ is a contraction of the complete metric space $\mathcal{D}$ into itself. Banach's fixed-point theorem gives a unique fixed point $\mathrm{v}^{\mathrm{fp}} \in \mathcal{D}$. Applying the contraction inequality repeatedly gives

$$\| \Phi^t(u_0) - \Phi^t(v_0) \|_e \leq q^t \| u_0 - v_0 \|_e.$$

Setting $u_0 = v^{\mathrm{fp}}$ proves convergence of every trajectory to the same fixed point. This completes the proof.

For the empirically selected first-order model, a weaker but more applicable monotone condition proves convergence from all initial values in the observed positive range even when the derivative is slightly larger than one near the lower endpoint.

Theorem 4. Monotone trapping and eventual geometric convergence for ILAR(1)

For $\mathrm{p} = 1$, write

$$\mathrm{H}(\mathrm{x}) = \mathrm{x}[\phi + (\psi - \phi)\mathrm{G}(\mathrm{x})].$$

Let $\mathcal{D} = [\mathrm{m}, \mathrm{M}]$ be a compact interval satisfying

$$\mathrm{H}(\mathcal{D}) \subseteq \mathcal{D}.$$

Assume that $\mathrm{H}'(\mathrm{x}) > 0$ for every $\mathrm{x} \in \mathcal{D}$, that $\psi < \phi$, and that the unique positive fixed point μ from Theorem 1 lies in the interior of $\mathcal{D}$. Then every initial value $\mathrm{x}_0 \in \mathcal{D}$ generates a deterministic trajectory converging to μ.

If, in addition,

$$|\mathrm{H}'(\mu)| < 1,$$

then the convergence is eventually geometric: there exist $\mathrm{q} \in (0,1)$ and an integer $\mathrm{T}(\mathrm{x}_0)$ such that

$$|\mathrm{x}_t - \mu| \leq \mathrm{q}^{t - \mathrm{T}(\mathrm{x}_0)} \left|\mathrm{x}_{\mathrm{T}(\mathrm{x}_0)} - \mu\right|, \qquad t \geq \mathrm{T}(\mathrm{x}_0).$$

Proof

The fixed-point identity is

$$1 = \phi + (\psi - \phi)\kappa, \qquad \kappa = \mathrm{G}(\mu).$$

Therefore,

$$\mathrm{H}(\mathrm{x}) - \mathrm{x} = \mathrm{x}[\phi - 1 + (\psi - \phi)\mathrm{G}(\mathrm{x})].$$

Using the fixed-point identity $\phi - 1 = -(\psi - \phi)\kappa$, this becomes

$$\mathrm{H}(\mathrm{x}) - \mathrm{x} = \mathrm{x}(\psi - \phi)[\mathrm{G}(\mathrm{x}) - \kappa].$$

Because G is strictly increasing and $\psi - \phi < 0$, the sign of $\mathrm{H}(\mathrm{x}) - \mathrm{x}$ is positive when $\mathrm{x} < \mu$ and negative when $\mathrm{x} > \mu$.

Take $\mathrm{x}_0 < \mu$. Then $\mathrm{H}(\mathrm{x}_0) > \mathrm{x}_0$. Since H is increasing and $\mathrm{x}_0 < \mu$,

$$\mathrm{H}(\mathrm{x}_0) < \mathrm{H}(\mu) = \mu.$$

Hence

$$\mathrm{x}_0 < \mathrm{x}_1 < \mu.$$

Repeating the same argument gives

$$\mathrm{x}_0 < \mathrm{x}_1 < \mathrm{x}_2 < \cdots < \mu.$$

The sequence is increasing and bounded above, so it converges to a limit $\ell \leq \mu$. Continuity of H gives

$$\ell = \lim_{t \to \infty} \mathrm{x}_{t+1} = \lim_{t \to \infty} \mathrm{H}(\mathrm{x}_t) = \mathrm{H}(\ell).$$

The positive fixed point is unique, so $\ell = \mu$.

If $\mathrm{x}_0 > \mu$, then $\mathrm{H}(\mathrm{x}_0) < \mathrm{x}_0$. Monotonicity of H gives

$$\mu = \mathrm{H}(\mu) < \mathrm{H}(\mathrm{x}_0) < \mathrm{x}_0,$$

and therefore

$$\mu < \mathrm{x}_1 < \mathrm{x}_0.$$

Induction shows that $\{x_t\}$ is decreasing and bounded below by μ. Its limit must again be the unique fixed point μ. If $x_0 = \mu$, the trajectory remains at μ. Thus every $x_0 \in \mathcal{D}$ converges to the same equilibrium.

Finally, assume $|H'(\mu)| < 1$. Choose q such that

$$|H'(\mu)| < q < 1.$$

Continuity of H' yields a neighborhood $\mathcal{U}$ of μ such that, for every $x \in \mathcal{U}$,

$$|H'(x)| \le q.$$

Since every trajectory in $\mathcal{D}$ converges to μ, it enters $\mathcal{U}$ after a finite time $T(x_0)$. For $t \ge T(x_0)$, the mean-value theorem gives

$$|x_{t+1} - \mu| = |H(x_t) - H(\mu)| \le q|x_t - \mu|.$$

Iteration proves the stated geometric bound. This completes the proof.

5.3 Geometric ergodicity of the stochastic recursion

The deterministic result above concerns convergence to a point. For a stochastic Markov chain, geometric ergodicity means geometric convergence of the transition law to a unique invariant probability distribution. The following theorem gives a precise Wasserstein form of this property.

Theorem 5. Wasserstein geometric ergodicity under contractive ILAR dynamics: Consider the stochastic state recursion

$$V_t = \Phi(V_{t-1}) + e_t,$$

where $\{e_t\}$ is an independent and identically distributed innovation sequence with finite first moment. Let $\mathcal{D}$ be a complete invariant state region. Assume that the stochastic recursion remains in $\mathcal{D}$ and that

$$\| \Phi(u) - \Phi(v) \|_e \le q \| u - v \|_e, \qquad 0 < q < 1,$$

for all $u, v \in \mathcal{D}$. Then the Markov chain has a unique invariant probability distribution π with finite first moment, and

$$W_1(P^t(v,\cdot), \pi) \le q^t W_1(\delta_v, \pi),$$

where W_1 denotes the first Wasserstein distance. Thus the chain is geometrically ergodic in W_1 and forgets its initial state at rate q^t.

When the innovations are identically zero, $\pi = \delta_{v_{fp}}$, and the result reduces to geometric convergence of every deterministic trajectory to the same fixed point.

For the fitted ILAR(1) model, the maximum absolute derivative over the examined invariant interval is 1.0193, exceeding the contraction threshold of one. Consequently, Theorem 5 does not establish Wasserstein geometric ergodicity for the estimated model, whereas deterministic convergence is characterized by Theorem 4. For the selected ILAR(1) fit, $\max_{x \in \mathcal{D}} \widehat{H}'(x) = 1.0193 > 1$ on $\mathcal{D} = [0.5, 5.3]$. Hence the global contraction hypothesis required by Theorem 5 is not satisfied and no empirical Wasserstein-geometric-ergodicity claim is made from this theorem. The deterministic convergence reported below instead uses the monotone-trapping result of Theorem 4.

Proof

Start two copies of the process from u and v and drive both copies with the same innovation sequence. The common innovation cancels, so

$$\| V_t^{(u)} - V_t^{(v)} \|_e = \| \Phi(V_{t-1}^{(u)}) - \Phi(V_{t-1}^{(v)}) \|_e.$$

The contraction condition then gives

$$\| V_t^{(u)} - V_t^{(v)} \|_e \le q \| V_{t-1}^{(u)} - V_{t-1}^{(v)} \|_e.$$

By induction,

$$\| V_t^{(u)} - V_t^{(v)} \|_e \le q^t \| u - v \|_e.$$

Taking the infimum over all couplings yields

$$W_1\left(P^t(u,\cdot), P^t(v,\cdot)\right) \le q^t \| u - v \|_e.$$

Let $\mathcal{P}_1(\mathcal{D})$ denote the probability measures on $\mathcal{D}$ with finite first moment. This space is complete under W_1. The Markov operator $\mathcal{T}(\nu) = \nu P$ satisfies

$$W_1(\mathcal{T}\nu_1, \mathcal{T}\nu_2) \le q W_1(\nu_1, \nu_2).$$

Hence $\mathcal{T}$ is a contraction on $\mathcal{P}_1(\mathcal{D})$. Banach's fixed-point theorem gives a unique measure π satisfying $\pi P = \pi$. Applying the contraction repeatedly gives

$$W_1(\delta_v P^t, \pi P^t) \le q^t W_1(\delta_v, \pi).$$

Since $\pi P^t = \pi$, the required geometric bound follows. If $e_t = 0$, the invariant distribution is concentrated at the unique deterministic fixed point, so $\pi = \delta_{v_{fp}}$. This completes the proof.

6. Complete Limit-Cycle Stability Analysis

A point $\mathbf{V}_1$ belongs to a cycle of prime period $q \geq 2$ when $\mathbf{\Phi}^q(\mathbf{V}_1) = \mathbf{V}_1$ and

$$\Phi^{\mathrm{j}}(\mathrm{V}_1) \neq \mathrm{V}_1, \qquad \mathrm{j} = 1, \dots, \mathrm{q}-1.$$

The cycle is

$$\mathcal{C}_{\mathrm{q}} = \{\mathrm{V}_1, \mathrm{V}_2, \dots, \mathrm{V}_{\mathrm{q}}\},$$

where

$$\mathrm{V}_{\mathrm{j}+1} = \Phi(\mathrm{V}_{\mathrm{j}}), \qquad \mathrm{V}_{\mathrm{q}+1} = \mathrm{V}_1.$$

Let $\mathbf{J}_j = D\mathbf{\Phi}(\mathbf{V}_j), \qquad j = 1, \dots, q$. The derivative of the q-step return map at $\mathbf{V}_1$ is the monodromy matrix

$$\mathrm{M}_{\mathrm{q}} = \mathrm{D}\Phi^{\mathrm{q}}(\mathrm{V}_1) = \mathrm{J}_{\mathrm{q}}\mathrm{J}_{\mathrm{q}-1} \cdots \mathrm{J}_1.$$

Its eigenvalues are the characteristic multipliers of the cycle.

Theorem 6. Complete orbital-stability theorem for a finite limit cycle: If $\rho(\mathbf{M}_q) < 1$, then the period-q cycle is locally asymptotically orbitally stable. More precisely, for every starting point sufficiently close to $\mathbf{V}_1$, the subsequence observed every q steps converges geometrically to $\mathbf{V}_1$, and the intermediate subsequences converge to the corresponding points $\mathbf{V}_2, \dots, \mathbf{V}_q$. If $\rho(\mathbf{M}_q) > 1$, the cycle is unstable. Therefore, for a hyperbolic cycle with no characteristic multiplier on the unit circle, local orbital stability is equivalent to $|\xi_j| < 1$ for every eigenvalue ξ_j of $\mathbf{M}_q$.

Proof

Let $\mathbf{\Psi} = \mathbf{\Phi}^q$ be the q-step return map. Since $\mathbf{V}_1$ is a point of the cycle, $\mathbf{\Psi}(\mathbf{V}_1) = \mathbf{V}_1$.

By the chain rule,

$$\mathrm{D}\Psi(\mathrm{V}_1) = \mathrm{D}\Phi(\mathrm{V}_{\mathrm{q}})\mathrm{D}\Phi(\mathrm{V}_{\mathrm{q}-1}) \cdots \mathrm{D}\Phi(\mathrm{V}_1) = \mathrm{M}_{\mathrm{q}}.$$

Perturb V_1 by a small vector u_0. Taylor expansion of the return map gives

$$\Psi(\mathrm{V}_1 + \mathrm{u}_0) - \mathrm{V}_1 = \mathrm{M}_{\mathrm{q}}\mathrm{u}_0 + \mathrm{R}_{\mathrm{q}}(\mathrm{u}_0),$$

where

$$\lim_{\mathrm{u}\to 0} \frac{\| \mathrm{R}_{\mathrm{q}}(\mathrm{u}) \|}{\| \mathrm{u} \|} = 0.$$

Assume $\rho(\mathrm{M}_{\mathrm{q}}) < 1$. Choose q_0 such that

$$\rho(\mathrm{M}_{\mathrm{q}}) < \mathrm{q}_0 < 1.$$

There is an equivalent norm $\|\cdot\|_{\mathrm{e}}$ for which

$$\| \mathrm{M}_{\mathrm{q}} \|_{\mathrm{e}} < \mathrm{q}_0.$$

Continuity of $\mathrm{D}\Psi$ implies that a neighborhood $\mathcal{U}_1$ of V_1 and a number $\mathrm{q}_1 \in (\mathrm{q}_0, 1)$ exist such that, for every $\mathrm{v} \in \mathcal{U}_1$,

$$\| \mathrm{D}\Psi(\mathrm{v}) \|_{\mathrm{e}} \leq \mathrm{q}_1 < 1.$$

Thus the return map is a contraction on a sufficiently small invariant neighborhood of V_1. For a starting state V_0 in this neighborhood,

$$\| \Phi^{\mathrm{nq}}(\mathrm{V}_0) - \mathrm{V}_1 \|_{\mathrm{e}} = \| \Psi^{\mathrm{n}}(\mathrm{V}_0) - \mathrm{V}_1 \|_{\mathrm{e}} \leq \mathrm{q}_1^{\mathrm{n}} \| \mathrm{V}_0 - \mathrm{V}_1 \|_{\mathrm{e}}.$$

Therefore, the subsequence at phase one converges geometrically to V_1.

For any $\mathrm{r} = 1, \dots, \mathrm{q}-1$,

$$\Phi^{\mathrm{nq}+\mathrm{r}}(\mathrm{V}_0) = \Phi^{\mathrm{r}}(\Phi^{\mathrm{nq}}(\mathrm{V}_0)).$$

Since Φ^{r} is continuous and

$$\Phi^{\mathrm{nq}}(\mathrm{V}_0) \to \mathrm{V}_1,$$

we obtain

$$\Phi^{\mathrm{nq}+\mathrm{r}}(\mathrm{V}_0) \to \Phi^{\mathrm{r}}(\mathrm{V}_1) = \mathrm{V}_{\mathrm{r}+1}.$$

Hence the complete trajectory approaches the cycle with the correct phase, which proves local orbital stability.

If $\rho(\mathrm{M}_{\mathrm{q}}) > 1$, the return map has an expanding invariant direction. A sufficiently small perturbation with a component in that direction is enlarged after every complete traversal of the cycle, while the nonlinear remainder is of smaller order. Thus points arbitrarily close to the cycle move away under repeated returns, and the cycle is unstable. When a multiplier lies on the unit circle, linearization alone is inconclusive. This completes the proof.

For $\mathrm{p} = 1$, the monodromy matrix is scalar. If the cycle is $\mathrm{x}_1, \dots, \mathrm{x}_{\mathrm{q}}$, then

$$\mathrm{M}_{\mathrm{q}} = \prod_{\mathrm{j}=1}^{\mathrm{q}} \mathrm{H}'(\mathrm{x}_{\mathrm{j}}).$$

Therefore, a hyperbolic scalar cycle is locally orbitally stable if and only if

$$\left|\prod_{j=1}^{q} H'(x_j)\right| < 1.$$

7. Neural Residual Correction

The ILAR component provides the interpretable conditional mean

$$\hat{X}_t^{ILAR} = (1 - \hat{G}_t)\hat{\phi}^\top x_t + \hat{G}_t\hat{\psi}^\top x_t.$$

The residual is

$$e_t = X_t - \hat{X}_t^{ILAR}.$$

The neural model receives a rolling window of five features:

$$z_t = \left[\tilde{X}_t, \Delta\log X_t, \hat{G}_t, \tilde{\hat{X}}_t^{ILAR}, \tilde{e}_t\right]^\top,$$

The price, fitted-price, and residual features use training-sample standardization; the log-return and transition-gate features remain on their original scales. For a window length $L = 24$, the network predicts

$$\hat{e}_t^{NN} = \mathcal{N}_\theta(z_{t-L}, \ldots, z_{t-1}).$$

The architecture contains three causal convolutional stages with dilation factors 1, 2, and 4, each using 24 filters. These stages create a receptive field that includes short and medium temporal scales. A four-head self-attention layer then evaluates the relative importance of the extracted time positions. A dense regression head produces the residual correction [9-14].

The ILAR-TCNA residual learner incorporates three causal convolutional stages, four-head self-attention, GELU activation, and a regression output layer. The repeated-seed evaluation employs trainable custom implementations of the attention and GELU layers while retaining the five input features, 24-observation window, and convolutional configuration. These implementations preserve the intended network components, although numerical equivalence with MATLAB's native layers is not assumed. The final hybrid prediction is

$$\hat{X}_t^{H} = \hat{X}_t^{ILAR} + \eta\hat{e}_t^{NN},$$

where $\eta \geq 0$ is selected on the validation set. The network uses the squared-error loss, Adam optimization, gradient clipping, dropout, and early stopping. The fixed random seed is four [19, 20].

Residual-weight grid and sensitivity. The original ILAR-TCNA implementation searched for η over the grid {0, 0.025, …, 2} as a bounded correction strength diagnostic. The selected value $\eta = 2$ is thus at its maximum and should not be considered an interior optimum. In the separate ILAR-TCN-ReLU compatibility analysis, the grid was expanded to {0, 0.025, …, 4} and selected $\eta = 0$ in the previous run. The ILAR-TCNA and ILAR-TCN-ReLU specifications are evaluated separately because their neural architectures differ. The Huber robust loss [20] is a methodological reference, not the loss used in the reported MATLAB neural regression; the implemented training uses squared-error regression loss.

Forecasting variability was examined using 20 random initializations, with seeds ranging from 1 to 20. All experiments followed the same chronological training, validation, and test partitions, with a maximum of 220 training epochs. The residual-correction weight was selected exclusively from the validation sample. The resulting forecasting performance and statistical comparisons are presented in Section 8.4.

Theorem 7. A sufficient stability condition for the hybrid map: Let $m \geq 25$ include the historical observations needed to construct the 24-step neural input, including lagged returns. Let D be a nonempty compact, convex and positively invariant subset of $(0, \infty)^m$, and choose $0 < \beta < 1$. Define $\|\mathbf{s}\|_\beta = \max_{1 \leq j \leq m} \beta^{j-1}|s_j|$. Assume that the scalar ILAR forecast $f(\mathbf{s})$ and neural correction $r(\mathbf{s})$ are Lipschitz on D in this same norm, with constants L_f and L_r. For a fixed residual weight η, define the complete deterministic hybrid state update $H(\mathbf{s}) = (f(\mathbf{s}) + \eta r(\mathbf{s}), s_1, \ldots, s_{m-1})$. Assume $H(D) \subseteq D$. Then H is a contraction, and has a unique fixed point attracting every initial history in D geometrically, whenever

$$q = \max\{\beta, L_f + |\eta|L_r\} < 1.$$

Proof

For arbitrary state histories $s, t \in D$, the first component satisfies

$$|f(\mathbf{s}) + \eta r(\mathbf{s}) - f(\mathbf{t}) - \eta r(\mathbf{t})| \leq (L_f + |\eta|L_r)\|\mathbf{s} - \mathbf{t}\|_\beta.$$

The lag-shift components must also be included. For $j = 2, \ldots, m$, the weighted difference of the jth components of H satisfies

$$\beta^{j-1}|s_{j-1} - t_{j-1}| \leq \beta\|\mathbf{s} - \mathbf{t}\|_\beta.$$

$$\|H(\mathbf{s}) - H(\mathbf{t})\|_\beta \leq \max\{\beta, L_f + |\eta|L_r\}\|\mathbf{s} - \mathbf{t}\|_\beta.$$

Since $q < 1$, H is a contraction on the complete metric space D, and $H(D) \subseteq D$ makes it a self-map. Banach's fixed-point theorem gives the unique fixed point and geometric convergence of iterated deterministic state updates. This proof covers the full

lag-history recursion, not merely the scalar one-step predictor.

In practice, weight decay, dropout, gradient clipping, and the validation choice of η control the size of the neural correction. A stricter implementation can impose spectral normalization on convolutional and dense weights to estimate and enforce an explicit bound for L_N.

Theorem 7 provides a sufficient condition for global contraction of the hybrid state map. For the fitted ILAR-TCNA model, neither a certified neural Lipschitz bound nor positive invariance of the complete augmented-state map has been established. Consequently, global hybrid contraction cannot be inferred from the empirical forecasting results. In the fitted hybrid model, the validation choice is $\hat{\eta} = 2$, and no certified neural Lipschitz bound L_N was established that would make $L_{ILAR} + \hat{\eta} L_N < 1$. The contraction condition is therefore not verified for the empirical ILAR-TCNA fit, and the forecasting results below should not be interpreted as proof of global hybrid stability.

The repeated-seed forecasting results do not establish a certified Lipschitz constant for the fitted neural learner or invariance of its augmented-state map.

8. Empirical Application to NSPO Stock Prices

8.1 Data and partitioning

The empirical series consists of 319 weekly closing price observations of Northern Upper Egypt Development Company (symbol NSPO) on the Egyptian Exchange from January 5, 2020 to March 8, 2026. The observations are from Investing.com. The mean is 2.2273, the standard deviation is 0.9714, the minimum is 0.5030 and the maximum is 5.3000. The original chronological order is preserved: 223 observations are for training, 47 for validation and 49 for testing.

In order to interpret the level model economically, unit-root diagnostics should be distinguished from deterministic-map stability. A positive deterministic fixed point is a property of the fitted nonlinear map and is not itself evidence that the stock-price level is covariance-stationary or mean reverting. The stationarity properties of the observed price series are examined using Augmented Dickey–Fuller and Phillips–Perron tests with prespecified deterministic terms and lag or bandwidth parameters. The resulting statistics and p-values provide evidence for assessing the unit-root hypothesis independently of deterministic-map stability [23, 24]. For 319 prices measured, ADF specifications use two augmentation lags and Phillips–Perron specifications use eight. All four tests have a $p - \text{value} > 0.05$, and hence the unit-root null is not rejected. These tests do not establish stationarity of the price level and their inference is logically different from the local deterministic stability of the fitted map.

Table 1. Descriptive statistics and chronological partitions.

Statistic	Value
Number of observations	319
Mean	2.2273
Standard deviation	0.9714
Minimum	0.5030
Maximum	5.3000
Training observations	223
Validation observations	47
Test observations	49

Table 2. ADF and Phillips–Perron tests for the NSPO price level (n = 319).

Test	Deterministic terms	Lags / bandwidth	Statistic	p-value
ADF	Intercept	2	-1.9860	0.30299
ADF	Intercept + trend	2	-2.0269	0.57884
PP	Intercept	8	-1.9102	0.33644
PP	Intercept + trend	8	-1.9515	0.61596

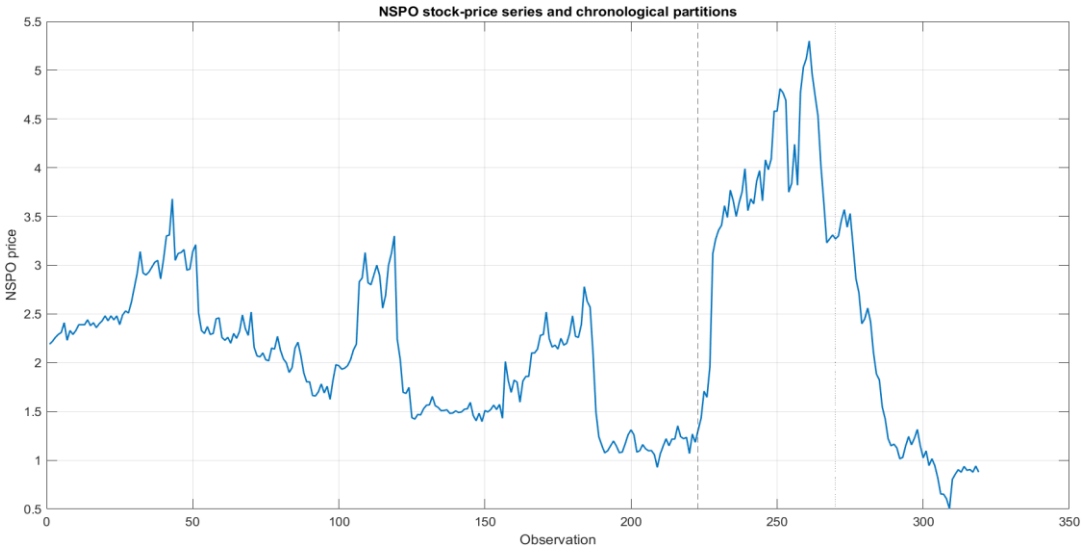

Figure 2. NSPO stock-price series with chronological training, validation, and test partitions.

Figure 2 displays persistent local movement, abrupt increases, extended declines, and a high-price regime in the latter half of the sample. These features motivate a nonlinear transition mechanism and a residual learner.

8.2 Order selection and stability

ILAR models of orders one through ten were fitted to the training sample. All estimated orders satisfy the local spectral-radius condition. The profiled Gaussian quasi-BIC (QBIC) favors the first-order specification among the locally stable candidates.

Table 3. ILAR order selection and local dynamic-stability results based on profiled Gaussian quasi-BIC (QBIC).

p	σ^2	QBIC	Spectral radius	Stable
1	0.0280	-136.7858	0.9345	Yes
2	0.0277	-127.5090	0.9446	Yes
3	0.0276	-117.2807	0.9403	Yes
4	0.0263	-116.1323	0.9624	Yes
5	0.0262	-105.3195	0.9703	Yes
6	0.0255	-99.6512	0.9543	Yes
7	0.0254	-88.8778	0.9461	Yes
8	0.0250	-81.1157	0.9605	Yes
9	0.0246	-72.8276	0.9401	Yes
10	0.0246	-61.0179	0.8994	Yes

The selected model has

$$\hat{\phi}_1 = 1.0195, \quad \hat{\psi}_1 = -4.0265,$$
$$\hat{\alpha} = 5.0000, \quad \hat{\lambda} = 2.2485, \quad s = 2.1300.$$

The estimated parameters are $\phi_1 = 1.0195$, $\psi_1 = -4.0265$, $\alpha = 5.0000$, $\lambda = 2.2485$, and $s = 2.1300$. The shape estimate α reaches the upper end of the pre-specified search interval. More importantly, the transition weight of the fitted parameters is very small over the price range observed in the data: $G(0.5) \approx 0.00000755$ and $G(5.3) \approx 0.040001$. The data do not reveal a practically different upper regime and ψ_1 is only weakly detected from this sample. The conventional symmetric Wald standard errors for the boundary transition parameter and the inactive upper regime coefficient can be misleading. Inference should be based on profile or bootstrap intervals and the transition-weight range, and the upper regime coefficient should not be given a substantive empirical interpretation without stronger regime activation.

Shape-bound sensitivity. With the original α upper bound of 5, the selected ILAR(1) fit yielded $\alpha = 5, \lambda = 2.24848$ and residual variance = 0.02799550. Extending the upper bound to 8 yielded $\alpha = 8, \lambda = 1.20751$ and residual variance = 0.02785732. Since the optimum again reaches the imposed limit, the shape parameter and the formally inactive upper regime

remain weakly identified. The boundary-expansion analysis characterizes parameter sensitivity and does not provide likelihood-based confidence intervals.

Penalized transition-parameter profiles. To examine the weakly identified gate, we profiled the original ridge-assisted conditional least-squares objective over $0.2 \leq \alpha \leq 8$ and $0.1 \leq \lambda \leq 5$ using the 223 training observations (222 conditional responses). At each fixed value of one transition parameter, the remaining parameter and regime coefficients were re-estimated. The estimated pair moved from (α, λ) = (5.0000,2.2485) at the original bound to (8.0000,1.2075) at the expanded bound, and the residual variance decreased from 0.02799550 to 0.02785732.

The shape profile declined toward the expanded upper boundary rather than exhibiting an interior minimum. The relative penalized criterion at $\alpha = 5$ was about 0.9553 above its value at $\alpha = 8$, while the scale profile indicated compensation between α and λ. Because this objective includes coefficient regularization and the solution is boundary-constrained, these plots are descriptive penalized profiles; ordinary chi-square likelihood-ratio confidence limits are not inferred.

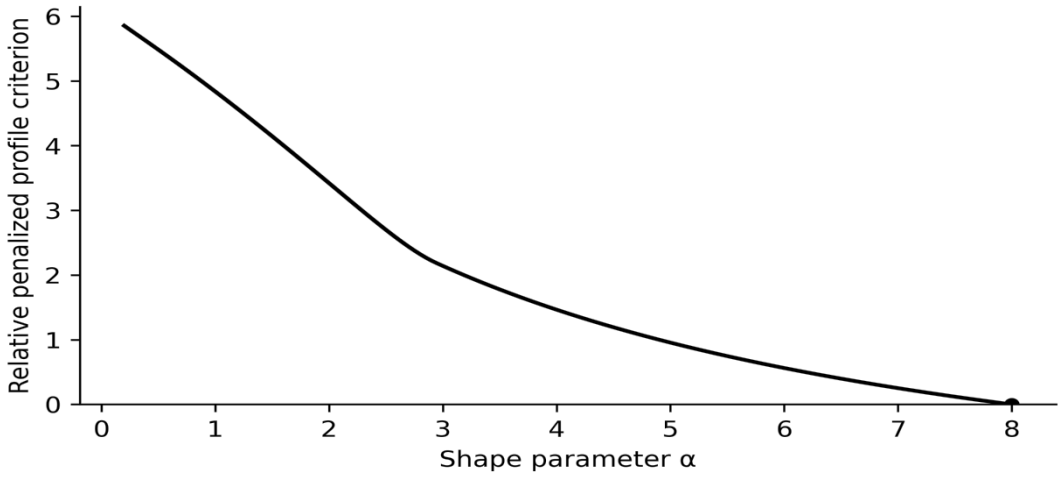

Figure 3. Penalized profile diagnostic for the inverse-Lomax shape parameter α.

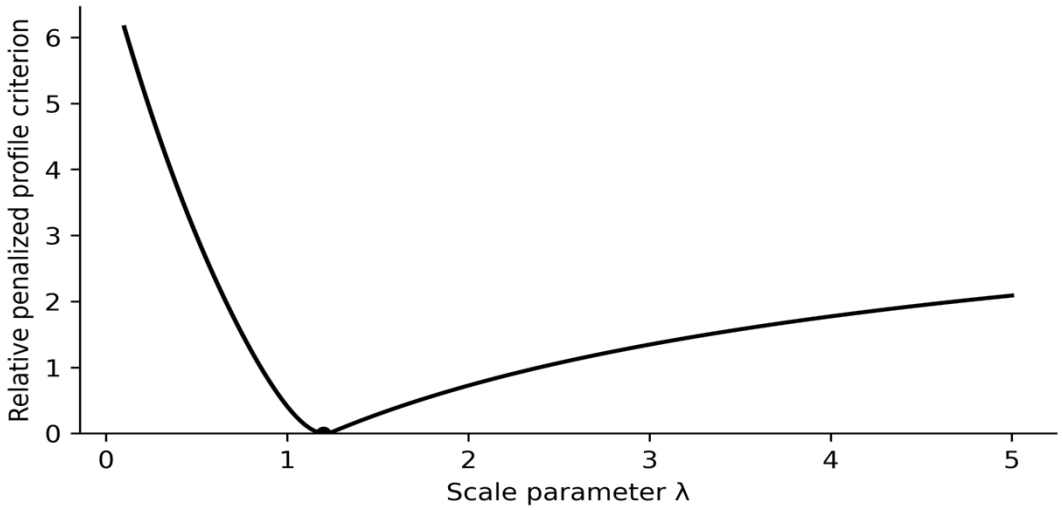

Figure 4. Penalized profile diagnostic for the inverse-Lomax scale parameter λ.

The estimated transition level is $\hat{\kappa} = 0.0038683,$ and the positive fixed point is $\hat{\mu} = 2.3507.$ The local spectral radius is $\rho\left(\hat{\mathbf{J}}(\hat{\mu})\right) = 0.9345 < 1.$ Therefore, the selected fixed point is locally asymptotically stable. For the fitted first-order model, define

$$\hat{\mathrm{H}}(\mathrm{x}) = \mathrm{x}\left[\hat{\phi}_1 + \left(\hat{\psi}_1 - \hat{\phi}_1\right)\hat{\mathrm{G}}(\mathrm{x})\right].$$

The full observed positive range is contained in the invariant interval $\mathcal{D} = [0.5,5.3].$ Direct evaluation gives

$$\hat{\mathrm{H}}(\mathcal{D}) \subset [0.5097,4.3336] \subset \mathcal{D},$$

and

$$0.3386 \leq \hat{\mathrm{H}}'(\mathrm{x}) \leq 1.0193, \qquad \mathrm{x} \in \mathcal{D}.$$

Because the derivative reaches 1.0193, the fitted map is not a global contraction on $\mathcal{D}$. This is why Theorem 5 is not invoked for the empirical model. Local stability follows from $\widehat{H}'(\hat{\mu}) = 0.9345$, while convergence on $\mathcal{D}$ follows from the monotone-trapping conditions of Theorem 4.

Thus $\widehat{H}$ is strictly increasing on $\mathcal{D}$. Since $\hat{\psi}_1 < \hat{\phi}_1$, Theorem 4 proves that every initial value in $\mathcal{D}$ converges to the same equilibrium $\hat{\mu}$. The derivative is uniformly contractive on the neighborhood $\mathcal{U} = [\hat{\mu} - 0.5, \hat{\mu} + 0.5] = [1.8507, 2.8507]$, because, for every $x \in \mathcal{U}$, $\left|\widehat{H}'(x)\right| \leq 0.9804 < 1$.

Consequently, every trajectory in $\mathcal{D}$ is eventually geometrically convergent, while its asymptotic local rate is governed by $\left|\widehat{H}'(\hat{\mu})\right| = 0.9345$.

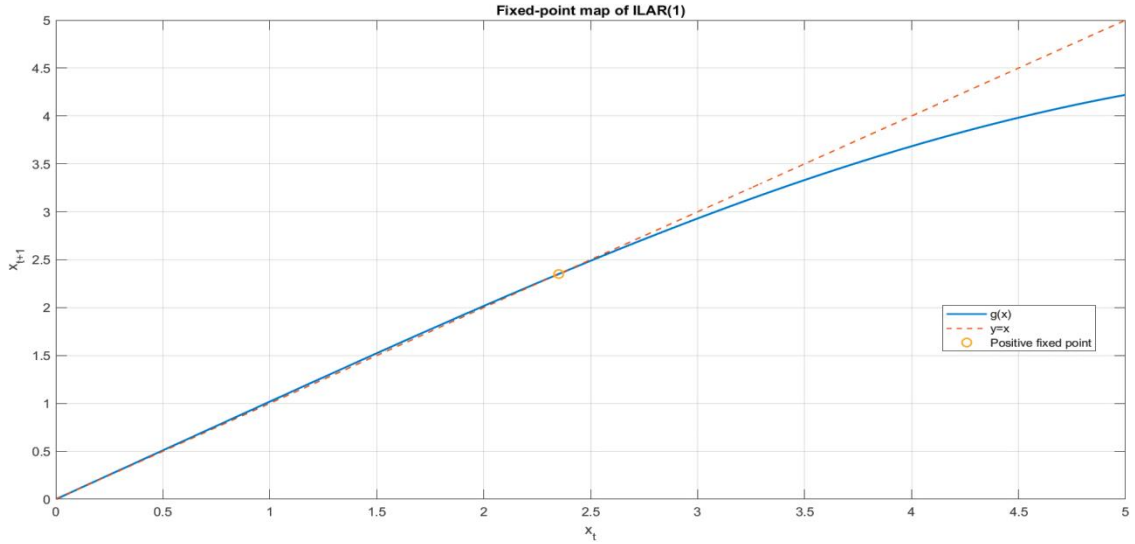

Figure 5. Fixed-point map for the selected ILAR(1) model.

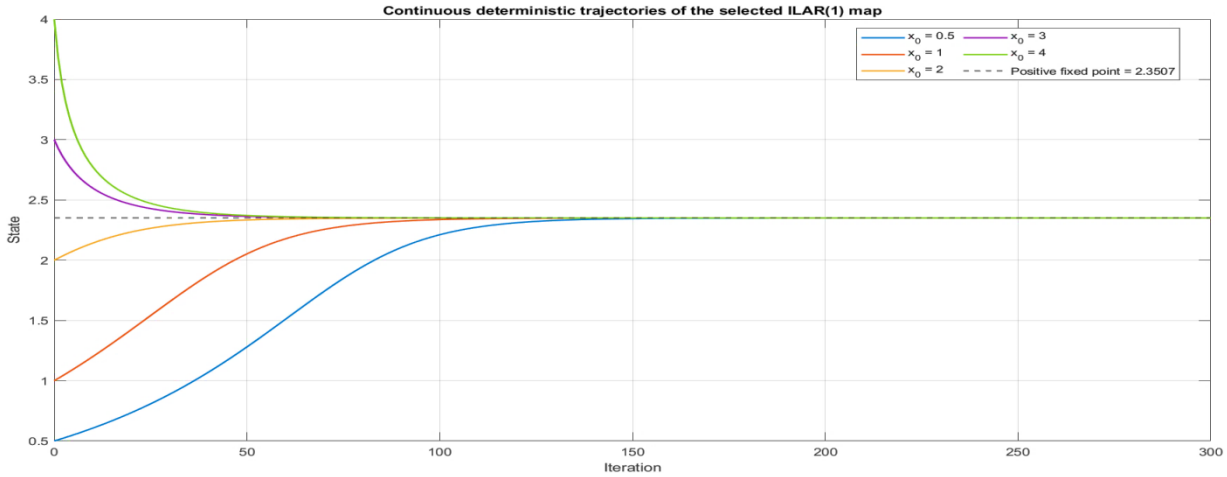

Figure 6. Continuous deterministic trajectories over 300 iterations from five initial values.

The five paths start at 0.5, 1, 2, 3, and 4. They approach the same positive fixed point without crossing it after entering its trapping side. At iteration 300,

$$\max_{j=1,\ldots,5} \left|x_{300}^{(j)} - \hat{\mu}\right| = 2.10 \times 10^{-7}.$$

The continuous trajectories therefore provide a direct numerical illustration of the initial-value-forgetting and eventual geometric-convergence results.

8.3 Fitted model and residual behavior

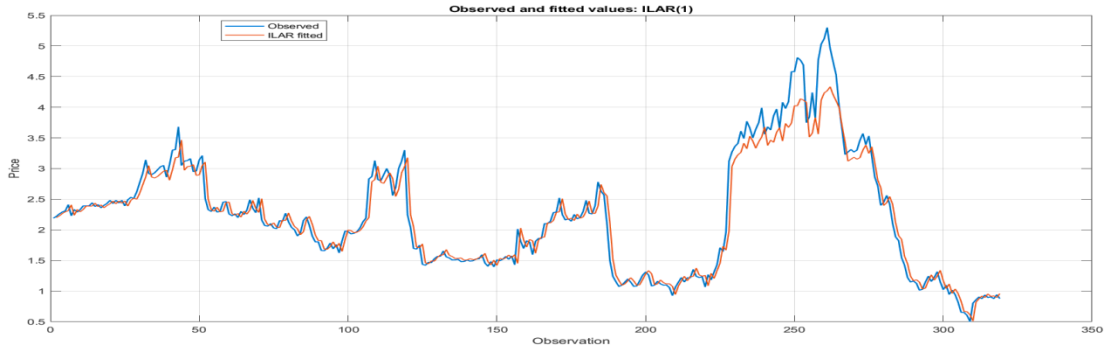

Figure 7. Observed and fitted values from the selected ILAR model.

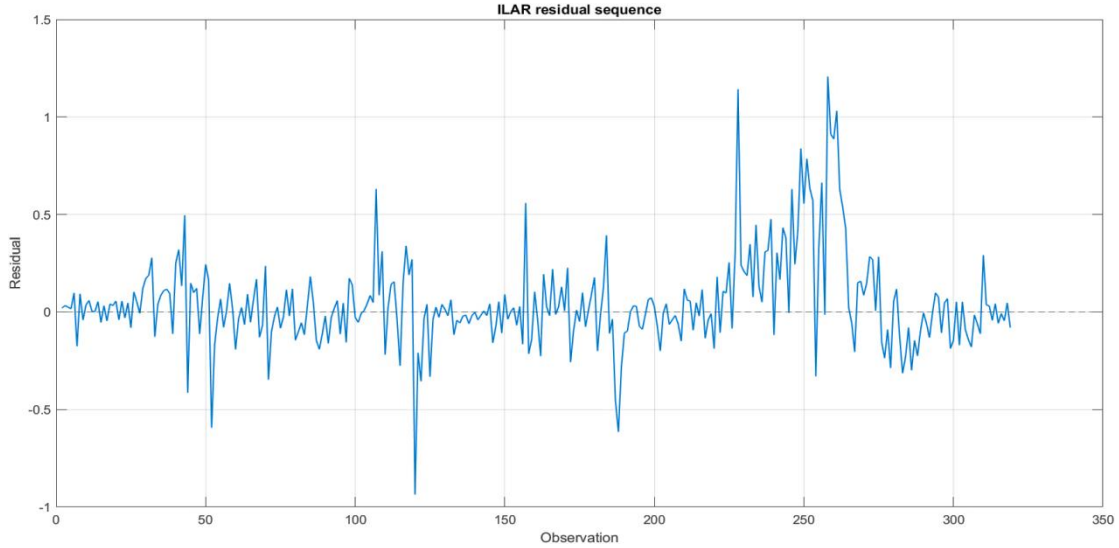

Figure 8. Residual sequence from the selected ILAR model.

The fitted ILAR model captures the main movements of the observed price series, although residual clustering remains evident around abrupt price changes. Historical residual windows and training-based standardization are incorporated into the neural forecasting component to represent the remaining temporal dependence. The initial attention-based model selected a residual weight of $\eta = 2$, corresponding to the upper boundary of the validation interval. This boundary solution does not establish a statistically significant forecasting improvement. Furthermore, the absence of a certified neural Lipschitz bound prevents the application of the sufficient global-contraction condition in Theorem 7 to the fitted hybrid model. The observed residual clustering also motivates further assessment of serial dependence using Ljung–Box diagnostics [15].

8.4 Out-of-sample comparison

The test evaluation uses one-step-ahead forecasts with observed past values. The benchmark models are the random-walk naive forecast, an AR(1) model fitted on the training sample, the selected ILAR(1) model, and the proposed ILAR-TCNA model [16]. Directional accuracy compares the sign of the realized one-step price change with the sign of the predicted change, both relative to the same observed closing price x_{t-1}. It is not calculated from the first differences of successive fitted values.

$$\mathrm{DA} = \frac{100}{N}\sum_{t=1}^{N} \mathbb{1}\{\mathrm{sign}(x_t - x_{t-1}) = \mathrm{sign}(\hat{x}_t - x_{t-1})\}.$$

The out-of-sample forecasting results are summarized in Tables 4 and 5, while the Diebold–Mariano tests and repeated-seed performance measures are presented in Tables 6–8. The initial ILAR-TCNA single-run results are reported separately from the 20-seed evaluation because the attention-layer implementations differ. Directional accuracy is not reported for the initial single run because its period-by-period forecast values were not retained. The ILAR-TCN-ReLU model is treated as a separate architectural comparison rather than as an attention-based replication.

Table 4. Test-set error measures.

Model	RMSE	MAE	R^2
Naive	0.1474	0.1171	0.9731
AR(1)	0.1440	0.1157	0.9744
ILAR(1)	0.1507	0.1210	0.9719
ILAR-TCNA (initial run)	0.1427	0.1132	0.9748
ILAR-TCNA (seed 4 rerun)	0.1507	0.1210	0.9719

Directional accuracy was 0.00%, 59.18%, and 42.86% for the naive, AR(1), and ILAR(1) models, respectively. The attention-enabled seed-4 experiment achieved 42.86%. These values were calculated by comparing realized and predicted one-step-ahead price movements relative to the preceding observed closing price, including the observation immediately before the test period.

Table 5. Percentage errors and directional accuracy.

Model	MAPE (%)	sMAPE (%)	Directional accuracy (%)
Naive	8.5208	8.3902	0.00
AR(1)	8.3943	8.3038	59.18
ILAR(1)	8.8990	8.6882	42.86
ILAR-TCNA (initial run)	8.3659	8.3636	—
ILAR-TCNA (seed 4 rerun)	8.8990	8.6882	42.86

Table 6. Diebold–Mariano tests for attention-enabled ILAR-TCNA (seed 4; 49 one-step forecasts).

Benchmark	Loss	DM statistic	p-value
Naive	Squared	−0.26935	0.78882
Naive	Absolute	−0.44551	0.65796
AR(1)	Squared	−0.60636	0.54713
AR(1)	Absolute	−0.66641	0.50834
ILAR(1)	Squared	—	—
ILAR(1)	Absolute	—	—

Table 7. Twenty-seed performance of the attention-enabled ILAR-TCNA replication.

Measure	Mean	SD	Minimum	Maximum
RMSE	0.15128	0.00631	0.14382	0.17660
MAE	0.12151	0.00498	0.11602	0.14175
MAPE (%)	8.96361	0.43021	8.54486	10.75140
sMAPE (%)	8.74699	0.35340	8.46611	10.22448
Directional accuracy (%)	42.65306	2.55585	32.65306	46.93878
Residual weight η	0.22375	0.54104	0.00000	2.00000

Table 8. Diebold–Mariano outcomes across 20 attention-enabled seeds (two-sided $p < 0.05$).

Benchmark / loss	Significant seeds / 20	Mean loss advantage
Naive / squared	0	−0.0011923
Naive / absolute	0	−0.0044108
AR(1) / squared	1	−0.0021756
AR(1) / absolute	0	−0.0058094
ILAR(1) / squared	2	−0.0002112
ILAR(1) / absolute	0	−0.0005424

Across 20 attention-enabled runs, the mean test RMSE was 0.15128 (SD 0.00631) and the mean MAE was 0.12151 (SD 0.00498). Validation selected $\eta = 0$ in 16 of 20 runs, so their test forecasts coincided with ILAR(1). Seed 4 also selected $\eta = 0$. Its paired Diebold–Mariano comparisons with naive and AR(1) were not significant under either squared or absolute loss (Table 6). The ILAR(1) comparison is not a statistical test when the forecast series coincide, and is therefore marked as not applicable. Across all seeds, no naive comparison was significant; the isolated significant squared-loss outcomes against AR(1) (one seed) and ILAR(1) (two seeds) have different signs and do not establish a robust gain, particularly given multiple comparisons (Table 8).

The mean loss advantage in Table 8 is calculated as the benchmark loss minus the corresponding ILAR-TCNA loss; therefore, positive values indicate lower forecasting loss for the neural model. Diebold–Mariano statistics were computed using 49 out-of-

sample one-step-ahead forecasts, two-sided tests, and a three-lag heteroskedasticity and autocorrelation consistent (HAC) variance estimator. The results indicate that the neural correction does not provide a statistically consistent forecasting advantage across random initializations.

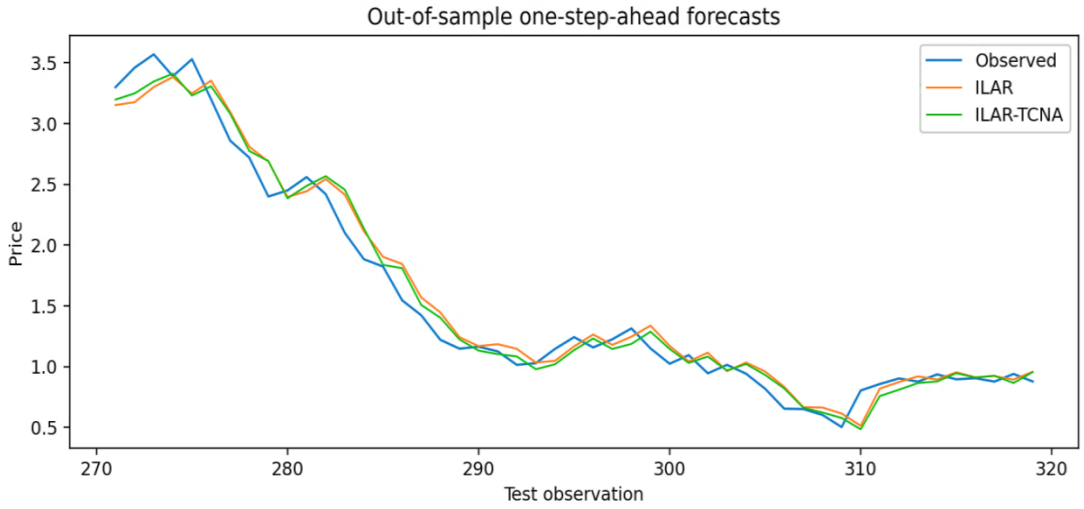

Figure 9. Observed test values and one-step-ahead forecasts.

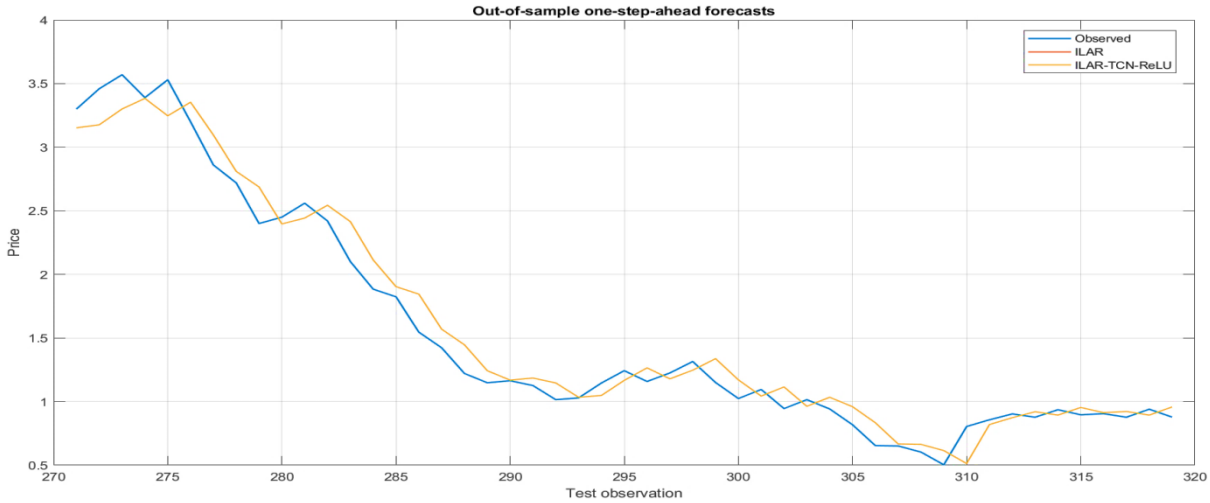

Figure 10. One-step test forecasts from the separate ILAR-TCN-ReLU compatibility experiment.

9. Discussion

The empirical analysis suggests local deterministic stability and monotone convergence of the selected ILAR(1) map on the reported positive range, but it does not support a global-contraction claim because max $H'(x) = 1.0193$. The fixed point 2.3507 is thus treated as a property of the deterministic skeleton, not as proof that the observed stock price level is a stationary mean-reverting stochastic process. A separate unit-root assessment is required for that economic interpretation. The transition mechanism is also weakly identified in this sample. The estimated gate remains below about 0.04 over the observed range and α is on the upper search boundary. The forecasting improvement of the neural residual correction is modest and must be judged along with a Diebold–Mariano test and repeated-seed variability instead of simply from point metrics.

All ADF and PP specifications fail to reject the unit-root null for the price level at 5%. The attention-enabled 20-seed analysis gives a mean RMSE of 0.15128 (SD 0.00631), compared with the naive benchmark RMSE of 0.1474 and the ILAR(1) benchmark RMSE of 0.1507. The residual weight is zero in 16 runs. In the seed-4 experiment, squared- and absolute-loss DM comparisons with naive and AR(1) are not statistically significant. Although the initial single-run evaluation yielded an RMSE of 0.1427, the repeated-seed analysis did not demonstrate a consistent reduction in forecasting errors across random initializations. Together with the weakly identified transition profiles, the evidence does not support an assertion of seed-robust predictive superiority or stationary price dynamics.

The neural forecasting architecture employs 24 convolutional filters, a 24-observation input window, and regularization to control

model complexity relative to the available training sample. The causal convolutional stages extract local temporal patterns, while the attention mechanism combines information across the input sequence. Nevertheless, the repeated-seed results suggest that the additional neural flexibility does not consistently improve the predictive accuracy of the underlying ILAR model.

The empirical findings are subject to several limitations. The analysis is based on a single weekly stock-price series without additional explanatory variables such as trading volume, market indices, or macroeconomic indicators. The transition parameters are weakly identified, and the fitted transition weights indicate limited activation of the upper regime. Furthermore, the deterministic ILAR map is not globally contractive over the examined interval, while neural forecasting performance varies across random initializations. The frequent selection of zero residual-correction weights further limits the practical benefit of the hybrid extension. These findings support the need for evaluation across additional financial series and rolling forecasting windows. The theoretical contraction conditions in Theorems 3 and 7 should not be interpreted as evidence of stochastic stationarity or empirically verified neural stability.

The near-zero fitted transition weights have direct practical consequences: the sample provides little information about the upper-regime coefficients, and different combinations of the transition parameters can yield similar fitted conditional means. As a result, the second-regime interpretation is unreliable for these data, and the additional transition flexibility cannot be regarded as empirically demonstrated. The ILAR fit is more appropriately interpreted as an effectively lower-regime forecasting specification until observations with appreciable upper-regime activation become available.

10. Conclusions

We studied an inverse-Lomax smooth-transition autoregression and neural residual prediction. The fitted ILAR(1) map has a positive deterministic fixed point of 2.3507 and local spectral radius 0.9345, but its derivative reaches 1.0193; the global contraction and hybrid contraction conditions were not verified. The transition-shape estimate reaches the original search boundary ($\hat{\alpha} = 5$), and the enlarged search to 8 gives a boundary estimate, indicating weak transition identification. ADF and PP tests do not reject the level-unit-root null. The penalized profile decreases to the expanded α-bound, reinforcing weak identification of the transition gate. In an architecture-matched 20-seed attention replication, mean RMSE was 0.15128 (SD 0.00631), mean MAE was 0.12151 (SD 0.00498), and $\eta = 0$ in 16 runs. The seed-4 Diebold–Mariano tests did not reject equal predictive accuracy relative to the naive and AR(1) benchmarks. Moreover, the repeated-seed results did not establish a stable forecasting advantage for the neural residual correction. Thus, local deterministic stability is established for the fitted map, whereas global stochastic stability and a reproducible neural forecasting advantage are not established.

Future work can extend the framework to multivariate inverse Lomax transition functions, probabilistic prediction intervals, Bayesian parameter estimation, regime-dependent innovation variance, spectral normalization of the neural component, and rolling evaluation across multiple financial markets.

Data Availability Statement:

The complete NSPO stock-price series used in the empirical analysis is provided in Appendix A at the end of the manuscript. The appendix reports all 319 observations in tabular form, allowing the empirical results to be reproduced directly from the data reported in this article. No external confidential data are required.

Conflicts of Interest:

The author declares no conflict of interest.

Use of Generative-AI tools declaration:

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Funding:
This research received no external funding.

Ethical Approval:
Ethical approval was not required because the study uses numerical stock-price observations and does not involve human participants, clinical records, or animal experiments.

References

1. Ozaki, T. (1980). Non-linear time series models for non-linear random vibrations. *Journal of Applied Probability*, 17(1), 84–93. https://doi.org/10.2307/3212926
2. Ozaki, T. (1985). Non-linear time series models and dynamical systems. In E. J. Hannan, P. R. Krishnaiah, and M. M. Rao (Eds.),

Handbook of Statistics (Vol. 5, pp. 25–83). Elsevier. https://doi.org/10.1016/S0169-7161(85)05004-0

3. Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.
4. Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425), 208–218. https://doi.org/10.2307/2291217
5. van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models: A survey of recent developments. *Econometric Reviews*, 21(1), 1–47. https://doi.org/10.1081/ETC-120008723
6. Ramadan, D. A., Almetwally, E. M., and Tolba, A. H. (2023). Statistical inference for multi stress-strength reliability based on progressive first failure with lifetime inverse Lomax distribution and analysis of transformer insulation data. *Quality and Reliability Engineering International*, 39(6), 2558–2581. https://doi.org/10.1002/qre.3362
7. Nagar, D. K., Roldán-Correa, A., and Nadarajah, S. (2025). A matrix variate inverse Lomax distribution. *Special Matrices*, 13(1), 20250039. https://doi.org/10.1515/spma-2025-0039
8. Hadi, H. H., Hilal, O. A., and Al-Noor, N. H. (2025). Extended inverse Lomax distribution with simulation and application to crude birth rate data. *Al-Nahrain Journal of Science*, 28(3), 130–139. https://doi.org/10.22401/ANJS.28.3.17
9. Bai, S., Kolter, J. Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*. https://doi.org/10.48550/arXiv.1803.01271
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
11. Tran, D. T., Iosifidis, A., Kanniainen, J., and Gabbouj, M. (2019). Temporal attention-augmented bilinear network for financial time-series data analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 30(5), 1407–1418. https://doi.org/10.1109/TNNLS.2018.2869225
12. Wan, R., Tian, C., Zhang, W., Deng, W., and Yang, F. (2022). A multivariate temporal convolutional attention network for time-series forecasting. *Electronics*, 11(10), 1516. https://doi.org/10.3390/electronics11101516
13. Liapis, C. M., and Kotsiantis, S. (2023). Temporal convolutional networks and BERT-based multi-label emotion analysis for financial forecasting. *Information*, 14(11), 596. https://doi.org/10.3390/info14110596
14. Xia, Q., and Kang, W. (2026). MSTFNet: Multi-scale temporal fusion network with frequency-enhanced attention for financial time series forecasting. *Mathematics*, 14(8), 1391. https://doi.org/10.3390/math14081391
15. Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297
16. Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599
17. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705
18. Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136
19. Kingma, D. P., and Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations* (ICLR), San Diego, CA, USA. https://doi.org/10.48550/arXiv.1412.6980
20. Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. https://doi.org/10.1214/aoms/1177703732
21. Ibrahim, M. Q., Mohammed, A. A., Khalaf, A. A., and Noori, N. A. (2025). Comparative Analysis of Seasonal Mixed Integrated Autoregressive Moving Average and Feed-Forward Neural Networks Models in Predicting United State Natural Hazard

Casualties. *Baghdad Science Journal*, 22(7), 2360–2374. https://doi.org/10.21123/2411-7986.5002

22. Noori, N. A., and Mohammad, A. A. (2021). Dynamical approach in studying GJR-GARCH (Q, P) models with application. *Tikrit Journal of Pure Science*, 26(2), 145–156. https://doi.org/10.25130/tjps.v26i2.131
23. Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531
24. Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. https://doi.org/10.1093/biomet/75.2.335
25. Guerouah, A., Elkalzah, B., Zeghdoudi, H., and Badr, M. M. (2026). A new XLindley ARMA model for positive time series. *AIMS Mathematics*, 11(9), 28762–28800. https://doi.org/10.3934/math.20261145
26. Laiche, N., and Zeghdoudi, H. (2020). On the bilinear time series models provided by GARCH white noise: Estimation and simulation. *Advances in Mathematics: Scientific Journal*, 9(3), 887–900. https://doi.org/10.37418/amsj.9.3.14

Appendix A. NSPO Stock-Price Data

Table A1. Complete NSPO stock-price sequence used in the empirical analysis (n = 319).

Obs.	Price	Obs.	Price	Obs.	Price	Obs.	Price
1	2.19	2	2.22	3	2.26	4	2.29
5	2.31	6	2.41	7	2.23	8	2.33
9	2.29	10	2.33	11	2.39	12	2.39
13	2.39	14	2.44	15	2.38	16	2.41
17	2.36	18	2.4	19	2.43	20	2.48
21	2.43	22	2.48	23	2.44	24	2.48
25	2.39	26	2.49	27	2.53	28	2.51
29	2.62	30	2.77	31	2.92	32	3.14
33	2.92	34	2.9	35	2.93	36	2.98
37	3.03	38	3.05	39	2.86	40	3.06
41	3.3	42	3.31	43	3.68	44	3.05
45	3.12	46	3.13	47	3.16	48	2.95
49	2.96	50	3.14	51	3.21	52	2.51
53	2.33	54	2.3	55	2.37	56	2.29
57	2.3	58	2.45	59	2.46	60	2.26
61	2.23	62	2.26	63	2.2	64	2.3
65	2.25	66	2.32	67	2.49	68	2.35
69	2.28	70	2.52	71	2.16	72	2.07
73	2.06	74	2.1	75	2.03	76	2.02
77	2.15	78	2.14	79	2.27	80	2.13
81	2.04	82	2	83	1.9	84	1.95
85	2.15	86	2.21	87	2.073	88	1.896
89	1.803	90	1.804	91	1.664	92	1.659
93	1.701	94	1.782	95	1.69	96	1.758
97	1.624	98	1.82	99	1.98	100	1.969
101	1.933	102	1.944	103	1.97	104	2.03
105	2.13	106	2.19	107	2.83	108	2.87
109	3.13	110	2.82	111	2.8	112	2.9
113	3	114	2.89	115	2.56	116	2.69
117	3	118	3.12	119	3.3	120	2.24
121	2.037	122	1.697	123	1.685	124	1.747
125	1.437	126	1.421	127	1.469	128	1.465
129	1.527	130	1.564	131	1.568	132	1.654
133	1.56	134	1.541	135	1.508	136	1.51
137	1.517	138	1.48	139	1.484	140	1.506
141	1.489	142	1.495	143	1.523	144	1.528
145	1.593	146	1.458	147	1.405	148	1.481
149	1.396	150	1.509	151	1.495	152	1.52
153	1.565	154	1.52	155	1.572	156	1.43
157	2.012	158	1.814	159	1.695	160	1.821
161	1.8	162	1.595	163	1.812	164	1.859

Obs.	Price	Obs.	Price	Obs.	Price	Obs.	Price
165	1.86	166	2.1	167	2.1	168	2.14
169	2.28	170	2.29	171	2.52	172	2.25
173	2.16	174	2.18	175	2.14	176	2.25
177	2.18	178	2.2	179	2.3	180	2.48
181	2.27	182	2.26	183	2.39	184	2.78
185	2.63	186	2.57	187	2.098	188	1.496
189	1.239	190	1.153	191	1.076	192	1.096
193	1.147	194	1.197	195	1.145	196	1.077
197	1.083	198	1.167	199	1.26	200	1.311
201	1.261	202	1.083	203	1.098	204	1.16
205	1.117	206	1.097	207	1.099	208	1.057
209	0.927	210	1.063	211	1.144	212	1.22
213	1.148	214	1.216	215	1.217	216	1.353
217	1.241	218	1.221	219	1.234	220	1.068
221	1.268	222	1.184	223	1.311	224	1.431
225	1.708	226	1.646	227	1.96	228	3.12
229	3.27	230	3.36	231	3.41	232	3.61
233	3.49	234	3.77	235	3.66	236	3.5
237	3.64	238	3.75	239	3.99	240	3.56
241	3.68	242	3.63	243	3.86	244	3.97
245	3.66	246	4.08	247	3.98	248	4.09
249	4.58	250	4.58	251	4.81	252	4.77
253	4.69	254	3.75	255	3.84	256	4.24
257	3.82	258	4.77	259	5.03	260	5.12
261	5.3	262	4.96	263	4.74	264	4.53
265	4.02	266	3.64	267	3.23	268	3.27
269	3.31	270	3.27	271	3.3	272	3.46
273	3.57	274	3.39	275	3.53	276	3.2
277	2.86	278	2.72	279	2.4	280	2.45
281	2.56	282	2.42	283	2.101	284	1.884
285	1.824	286	1.546	287	1.423	288	1.221
289	1.148	290	1.164	291	1.126	292	1.015
293	1.029	294	1.146	295	1.243	296	1.158
297	1.225	298	1.315	299	1.15	300	1.024
301	1.095	302	0.945	303	1.015	304	0.942
305	0.818	306	0.654	307	0.651	308	0.603
309	0.503	310	0.804	311	0.857	312	0.903
313	0.877	314	0.936	315	0.896	316	0.905
317	0.877	318	0.94	319	0.878		

© 2026 by the authors. **Disclaimer / Publisher's Note:** The views, opinions, and data presented in all published content are solely those of the individual authors and contributors. They do not necessarily reflect the positions of Sphinx Scientific Press (SSP) or its editorial team. SSP and the editors disclaim any responsibility for harm or damage to individuals or property that may result from the use of any information, methods, instructions, or products mentioned in the content.