Scientific Research Publishing

Open Journal of Social Sciences, 2026, 14(3), 340-367
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Digital-Intelligence Empowerment for Rural Revitalization: Estimating First-Launch Economy Productivity Efficiency via ETL-MLP Time-Series Reconstruction

Mo Shi*, Hetao Yuan, Xuao Chen

School of Economics and Management, Ankang University, Ankang, China
Email: *shimo0204@outlook.jp

How to cite this paper: Shi, M., Yuan, H. T., & Chen, X. A. (2026). Digital-Intelligence Empowerment for Rural Revitalization: Estimating First-Launch Economy Productivity Efficiency via ETL-MLP Time-Series Reconstruction. *Open Journal of Social Sciences, 14,* 340-367.
https://doi.org/10.4236/jss.2026.143020

Received: February 22, 2026
Accepted: March 16, 2026
Published: March 19, 2026

Abstract

Evaluating the economic productivity efficiency of the First-Launch Economy is crucial for rural revitalization. Yet, empirical efforts are frequently hindered by the scarcity, fragmentation, and temporal lag of regional production statistics. To address this critical data gap, this study develops a reproducible digital-intelligence framework leveraging Extract-Transform-Load (ETL) pipelines and Machine Learning reconstruction. As the First-Launch Economy of the plush toy industry in Hanbin District, Ankang City, this study fuses national macro-export anchor data with web-scraped local enterprise records. The methodology in this study employs an ETL pipeline for rule-based time-series cleaning, utilizing tailored heuristics like last-observation-carried-forward (LOCF) to standardize and impute sparse data. Subsequently, through exhaustive parameter traversal, a Multi-Layer Perceptron (MLP) Regressor is trained on 96 monthly observations (2018-2026). Subsequently, the MLP regressor effectively downscales sparse monthly inputs to generate 2921 synthetic daily production records for the time span from 2018 to 2026, while the 2921 daily records strictly preserve macroeconomic temporal structures. The reconstructed high-resolution dataset overcomes traditional data granularity limitations, enabling complex downstream econometric tasks such as short-term forecasting, shock-response simulation, and counterfactual policy assessment. Due to this, the framework in this study provides a validated, transferable blueprint for the digital empowerment of rural economies, while outlining future pathways to integrate remote-sensing and firm-level survey data to overcome source heterogeneity further.

Keywords

Rural Revitalization, First-Launch Economy, Digital-Intelligence

Empowerment, ETL Pipeline, MLP Regressor

1. Introduction

Rural revitalization has emerged as a pivotal strategy for regional sustainable development, increasingly driven by the paradigm of digital-intelligence empowerment. A critical mechanism within this framework is the first-launch economy, which can be considered a policy-driven initiative designed to incubate and scale emerging industries in socioeconomically disadvantaged rural areas, thereby catalyzing rapid local economic takeoff (Hou & Hu, 2026; Voyer et al., 2021; Kwok & Zhang, 2024). To investigate the efficacy of this paradigm, this study selects the Hanbin District as the primary empirical setting, focusing specifically on the local plush toy industry, which commenced its scaling phase in 2018. As a representative first-launch economy industry, it provides a unique and vital lens to explore how digital-intelligence technologies can strategically support and evaluate the localized incubation of new industries under the broader mandate of rural revitalization.

As Mbogori & Luketero (2019) emphasized, despite significant policy momentum surrounding the first-launch economy, rigorous measurement of its economic performance at the sub-county scale remains fundamentally constrained by data scarcity. Administrative records in rural areas are characteristically fragmented, sparse, and lagged, while firm-level disclosures lack uniformity (Karacsonyi & Taylor, 2023). Furthermore, conventional macroeconomic statistics fail to provide the temporal resolution necessary to capture the rapid, irregular life cycles of newly incubated industries or to evaluate short-run policy responses. Due to the multidimensional data constraints severely impeding causal analysis, comparative benchmarking, and predictive modeling, a critical knowledge gap persists, limiting both theoretical advancements and evidence-based policymaking for rural revitalization.

To bridge the methodological and empirical gaps as shown above, this study proposes a novel digital-intelligence framework designed to reconstruct a high-resolution production time series for the target industry and locality. Serving as a sub-county proxy for gross domestic product (GDP), the primary empirical objective is to estimate local production value and to contextualize it by benchmarking against national macroeconomic aggregates, specifically toy export data from the National Bureau of Statistics. The national-level series in this study provides a macro-anchor that informs local data reconstruction and facilitates an assessment of how localized first-launch growth dynamics relate to broader export trajectories. By integrating disparate data sources through principled preprocessing and machine learning, this study establishes a reproducible pipeline for small-area economic measurement where traditional administrative data fall short.

Regarding the methodological aspect, the proposed framework integrates auto-

mated web crawling, a rule-based Extract-Transform-Load (ETL) pipeline, and a Multilayer Perceptron (MLP) regressor for temporal reconstruction. The ETL phase systematically standardizes timestamps, resolves temporal duplications, and addresses data sparsity utilizing last-observation-carried-forward (LOCF) and year-aware gap-filling rules optimized for low-frequency enterprise statistics (Uma Maheswari et al., 2024; Mahmud & Ikbal, 2022; Mirzazadeh, 2025; Jokipii, 2023). Following an exhaustive parameter traversal to identify stable aggregation windows and robust covariates, the MLP model learns underlying temporal patterns from 96 monthly observations spanning 2018 to 2026, while augmented with contextual national covariates, the model synthesizes higher-frequency daily estimates. Upon the contents, the work in this study explicitly quantifies uncertainty sources (such as data heterogeneity, model specification, and imputation bias) and bounds inferences through rigorous sensitivity analysis.

The approach in this study aims to yield a synthetic daily production series that preserves underlying temporal structures while significantly expanding analytical granularity, rendering the dataset viable for shock-response simulation, forecasting, and counterfactual policy evaluation. The contributions of this study can be considered as it resolves a critical measurement void for evaluating the first-launch economy within the context of rural revitalization, and provides a highly transferable digital-intelligence blueprint for reconstructing high-resolution microeconomic series in data-scarce environments. Also, the reconstructed series equips local policymakers with actionable insights for tailoring interventions and monitoring outcomes practically. Future extensions will integrate neural network prediction, firm surveys, and causal inference frameworks to strengthen external validity and generalize this pipeline across diverse first-launch industries and regions.

2. Structure of the Digital-Intelligence Model

2.1. Macro-Contextual Anchoring

To accurately evaluate the localized impact of the first-launch economy in this study, it is imperative to first establish the overarching macroeconomic context that governs the industry. As shown in **Figure 1**, analysis of China's official toy-export time series reveals a non-stationary, long-run upward trajectory punctuated by episodic structural accelerations and a recent near-term correction for the past decades. According to **Figure 1**, the early low-frequency growth phase (1994-2005) reflects the gradual integration of Chinese manufacturing into global supply chains, followed by a moderate acceleration (2006-2014) driven by broader trade liberalization. Notably, the trajectory steepens markedly post-2015, culminating in a pronounced peak in the early 2020s before a subsequent normalization. This rapid expansion phase strongly correlates with the proliferation of digital-era distribution channels, such as cross-border e-commerce platforms, whereas the recent volatility is plausibly linked to pandemic-induced logistics bottlenecks, global demand shocks, and subsequent inventory corrections (Peters, 2023; Qingran et al., 2025; Touat, 2024; Neacsu et al., 2023; Goel et al., 2021; Yu et al., 2022).

Within the empirical setting of Hanbin District's plush-toy industry, the national export series in Figure 1 serves dual complementary functions, which are considered as a macroeconomic anchor and also a high-value covariate for the Multilayer Perceptron (MLP) reconstruction model. According to the increasing trend of the national toy export data in Figure 1, the 2018 onset of Hanbin's First-Launch economy initiative aligns chronologically with the national phase of accelerated export growth, suggesting the presence of shared economic drivers and potential cross-regional spillovers. Also, the national series captures broader, economy-wide supply and demand dynamics that the sparse, localized monthly records inherently lack. Due to this, the Extract-Transform-Load (ETL) pipeline is engineered to ensure rigorous temporal harmonization between national aggregates and local observations throughout this study. Upon the theories above, this necessitates precise seasonal adjustments and the derivation of lagged and growth-rate covariates, thereby enabling the MLP architecture to effectively learn both contemporaneous and delayed macro-micro economic relationships.

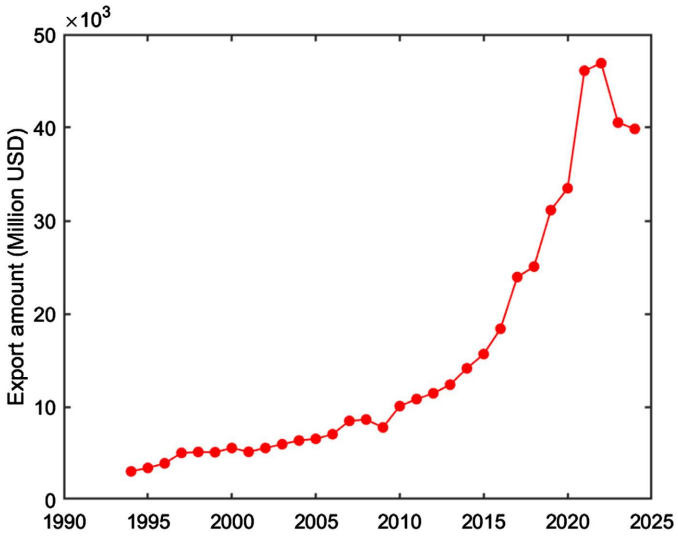


Figure 1. Economic data of China's toy export industry.

However, the inherent volatility and episodic shifts within the macro series impose stringent analytical requirements on the reconstruction pipeline to ensure robust inference. Also, given the non-stationary and likely heteroscedastic nature of the economic series, the modeling approach must incorporate data transformations such as log-scaling, first differencing, and explicit variance modeling (Clift, 2007; Merali, 2025; Reich, 2014). Furthermore, including the Bai-Perron procedure, formal change-point detection, and structural-break tests are strictly necessary to segment the data and prevent the MLP from blending conflicting economic regimes during training (Ngobeni & Muchopa, 2023; Koffi et al., 2024; Bommisetti et al., 2022). Within the methodological framework of this study, this dictates the implementation of regime-aware cross-validation, bootstrap-based uncertainty quantification, and rigorous sensitivity checks across varying covariate subsets.

2.2. Intelligent Data Acquisition and Heuristic Parsing

```python
# --- Query builder ---
def build_queries_for_month(month: str, locality: str = "安康 汉滨", industry_terms: List[str] = None) -> List[s
tr]:
    if industry_terms is None:
        industry_terms = ["毛绒玩具", "玩具"]
    year, mm = month.split("-")
    queries = []
    for term in industry_terms:
        queries.append(f"{locality} {term} 产量 {year}年 {int(mm)}月")
        queries.append(f"{locality} {term} 产值 {year}年 {int(mm)}月")
        queries.append(f"{locality} {term} 月度 生产 {year}年 {int(mm)}月")
        queries.append(f"{locality} {term} 企业 数量 {year}年 {int(mm)}月")
    queries.append(f"{locality} plush toys production {year}-{mm}")
    return queries

# --- Source-specific fetchers ---
def scrape_baidu_serp(query: str, save_prefix: str = "") -> str:
    url = "https://www.baidu.com/s"
    params = {"wd": query}
    save_path = os.path.join(RAW_HTML_DIR, f"{save_prefix}_baidu_{re.sub(r'\\W+', '_', query)[:80]}.html")
    html = safe_get(url, params=params, headers={"User-Agent": random.choice(USER_AGENTS)}, save_raw=save_path)
    random_delay()
    return html or ""

def scrape_qichacha_search(query: str, save_prefix: str = "") -> str:
    # Placeholder URL; many platforms require JS/login/APIs in practice.
    url = f"https://www.qichacha.com/search?key={requests.utils.requote_uri(query)}"
    save_path = os.path.join(RAW_HTML_DIR, f"{save_prefix}_qcc_{re.sub(r'\\W+', '_', query)[:80]}.html")
    html = safe_get(url, headers={"User-Agent": random.choice(USER_AGENTS)}, save_raw=save_path)
    random_delay()
    return html or ""

def scrape_google_serp_with_selenium(query: str, driver: Optional[webdriver.Chrome] = None, save_prefix: str = "
") -> str:
    own_driver = False
    if driver is None:
        driver = make_selenium_driver(headless=True)
        own_driver = True
    try:
        url = f"https://www.google.com/search?q={requests.utils.requote_uri(query)}&hl=en"
        driver.get(url)
        time.sleep(1.0 + random.random() * 1.5)
        html = driver.page_source
        save_path = os.path.join(RAW_HTML_DIR, f"{save_prefix}_google_{re.sub(r'\\W+', '_', query)[:80]}.html")
        with open(save_path, "w", encoding="utf-8") as f:
            f.write(html)
        random_delay()
        return html
    finally:
        if own_driver:
            driver.quit()

# --- Heuristic parser for numeric snippets ---
def parse_possible_production_from_html(html: str) -> List[Dict]:
    """
    Heuristic extractor. Returns candidate dicts:
    {'date': 'YYYY-MM' or None, 'value': float, 'unit': unit_str, 'snippet': snippet}
    """
    soup = BeautifulSoup(html, "lxml")
    text = soup.get_text(separator=" ", strip=True)
    candidates = []

    # Date patterns: "2020年5月" or "2020-05"
    date_patterns = [r"(\d{4})年\s*(\d{1,2})月", r"(\d{4})[-/](\d{1,2})"]
    # Number pattern: match numbers with optional thousand separators and optional unit
    num_pat = re.compile(r"((?:\d{1,3}(?:,\d{3})+|\d+)(?:\.\d+)?)\s*(万|亿|USD|美元|元|yuan)?")
    # Target tokens to focus search window
    target_tokens = ["产量", "产值", "月度", "月", "产出"]

    for token in target_tokens:
        for m in re.finditer(rf".{{0,120}}{re.escape(token)}.{{0,120}}", text):
            window = text[max(0, m.start()-200): m.end()+200]
            # find date
            date_found = None
            for dp in date_patterns:
                dm = re.search(dp, window)
                if dm:
                    y, mm = dm.group(1), dm.group(2)
                    try:
                        mm_i = int(mm)
                        if 1 <= mm_i <= 12:
                            date_found = f"{y}-{mm_i:02d}"
                            break
                    except:
                        pass
            # find number
            nm = num_pat.search(window)
            if nm:
                rawnum = nm.group(1).replace(",", "")
                unit = nm.group(2) or ""
                try:
                    val = float(rawnum)
                except:
                    continue
                candidates.append({"date": date_found, "value": val, "unit": unit, "snippet": window[:300]})
    return candidates
```

Figure 2. Structure of the crawled system (example code).

To overcome the scarcity of granular administrative records, this study engineers an automated Extract-Transform-Load (ETL) pipeline to systematically harvest and structure unstructured production data from the web. The initial data acquisition phase is driven by a deterministic query builder and source-specific fetch-

ers. As illustrated in Figure 2, the query builder dynamically synthesizes natural-language search strings by integrating spatial ("Ankang" & "Hanbin") and temporal tokens with industry-specific semantic templates, while this module maximizes recall across various search engines and corporate databases by generating diverse keyword combinations in both Chinese and English. According to the codes in Figure 2, source-specific fetching wrappers execute these queries using both standard HTTP requests and headless browser automation (selenium) to capture dynamically rendered JavaScript content subsequently. To ensure operational stability, respect ethical scraping constraints, and mitigate blocking risks, the fetchers embed defensive heuristics, including User-Agent rotation and randomized request delays, thereby establishing a robust and reproducible raw HTML corpus.

Following data retrieval, a heuristic-based parsing module is deployed to extract structured quantitative metrics from the noisy HTML corpus, as shown in the step of the heuristic parser for numeric snippets through the code in Figure 2. The parser utilizes DOM tree traversal to strip HTML markup, followed by the application of rigorous regular expressions to identify diverse date configurations and numeric patterns. To balance contextual precision with the structural variability of web documents, the algorithm employs a localized windowing heuristic, which extracts a ±200-character text span surrounding predefined target tokens ("production value", "production capacity", "monthly", "month", and "output"). As indicated in the code system in Figure 2, within these contextual windows, candidate tuples comprising temporal keys, numeric values, and units are isolated. Crucially, the parser automatically executes multiplicative normalizations for magnitude-specific Chinese characters and harmonizes currency representations, generating standardized candidate dictionaries for downstream processing.

2.3. MLP Regressor Establishment

In order to reconstruct high-resolution, daily economic series from the synthesized ETL dataset, this study employs a feedforward Multilayer Perceptron (MLP) architecture. As Figure 3 illustrates, the selected model features a three-layer topology comprising two hidden layers (configured with 20 and 10 neurons, respectively) and a single linear output neuron.

```
%% 3. Build and Train MLP
net = fitnet([20 10],'trainlm');
net.layers{1}.transferFcn = 'tansig';
net.layers{2}.transferFcn = 'tansig';
net.layers{3}.transferFcn = 'purelin';
net.trainParam.epochs = 2000;
net.trainParam.goal = 1e-8;
net.divideParam.trainRatio = 0.8;
net.divideParam.valRatio   = 0.1;
net.divideParam.testRatio  = 0.1;
[net,tr] = train(net,xn,yn);
```

Figure 3. Build & train of MLP regressor (Step 3).

Let $x^{(n)}$ denote the normalized scalar input. For a given layer l, defined by its weight matrix $W^{(n)}$, bias vector $b^{(l)}$, pre-activation $z^{(l)}$, and activation output $a^{(l)}$, the forward propagation is mathematically expressed as:

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}, a^{(l)} = f^{(l)}\left(z^{(l)}\right) \tag{1}$$

where $a^{(0)} = x^{(n)}$. The hidden layers utilize a hyperbolic tangent activation function, $f(z) = \tanh(z)$, to capture non-linear relationships within the data. Conversely, the output layer employs a linear transfer function, $f(z) = z$, which is optimal for continuous regression tasks. The primary objective function optimized during training is the Mean Squared Error (MSE), defined as:

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{N}\left(y_i^{(n)} - \hat{y}_i^{(n)}(\theta)\right)^2 \tag{2}$$

where θ encapsulates all learnable parameters ($W^{(l)}$ and $b^{(l)}$).

Also, as shown in Figure 3, the model optimization in this study is executed using the Levenberg–Marquardt (LM) algorithm, a highly efficient damped Gauss-Newton method particularly suited for least-squares problems in small-to-moderate-sized networks, while the parameter rule is defined as:

$$\Delta\theta = -\left(J^{\top}J + \mu I\right)^{-1} J^{\top}e \tag{3}$$

where J is the Jacobian matrix containing the first derivatives of the network errors with respect to the weights and biases ($J_{ij} = \partial e_i / \partial \theta_j$), e is the vector of network errors, and μ is a dynamically adapted damping scalar. During training, backpropagation efficiently computes the necessary partial derivatives for J. While the computational cost of the LM algorithm scales quadratically with the number of parameters, the relatively compact architecture of the proposed MLP ensures rapid convergence and stable gradient descent without prohibitive memory overhead.

Referring to Figure 4 and Figure 5, the training process is strictly monitored via an early stopping protocol governed by a discrete validation subset to prevent overfitting and ensure optimal generalization. As shown in Figure 5, the algorithm in this study tracks the iteration-wise sum-of-squares error across disjoint training, validation, and testing sets. Convergence diagnostics are established by identifying the specific epoch k that minimizes the validation error, formally denoted as $k^* = \arg\min_k E_{val}(k)$. The parameter state at epoch k^* is subsequently locked for final prediction. Upon completion of training, the normalized network outputs $\hat{y}^{(n)}$ are inverse-scaled to their original physical units using the reciprocal of the min-max mapping applied during preprocessing. This inverse transformation ensures that the reconstructed values are directly interpretable within the context of the localized first-launch economy.

As indicated in Figure 6, the structural robustness and predictive stability of the MLP in this study are rigorously evaluated using a 10-fold Cross-Validation (CV) protocol. For each fold k, the dataset is partitioned into mutually exclusive training (T_k) and testing (S_k) subsets. Crucially, to explicitly prevent data leakage, the normalization parameters are fitted exclusively on the training subset $\{x_i, y_i\}_{i \in T_k}$; the testing

inputs are subsequently transformed using these training-derived scaling statistics. The final cross-validation estimates are reported as the sample mean and standard deviation of fold-wise performance metrics (R^2 and RMSE), calculated as:

$$\bar{m} = \frac{1}{K}\sum_{k=1}^{K} m^{(k)} \tag{4}$$

$$SD_m = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(m^{(k)} - \bar{m}\right)^2} \tag{5}$$

```
%% 4. Training Progress
figure;
subplot(2,1,1);
plot(tr.perf,'LineWidth',1.5); hold on;
plot(tr.vperf,'LineWidth',1.5);
plot(tr.tperf,'LineWidth',1.5);
legend('Train','Validation','Test');
xlabel('Iteration');
ylabel('Loss');
grid on;

subplot(2,1,2);
plot(sqrt(tr.perf),'LineWidth',1.5); hold on;
plot(sqrt(tr.vperf),'LineWidth',1.5);
legend('Train RMSE','Validation RMSE');
xlabel('Iteration');
ylabel('RMSE');
grid on;
```

Figure 4. Training progress (Step 4).

```
%% 5. Training / Validation / Testing Prediction
y_pred_all = net(xn);
y_pred_all = mapminmax('reverse',y_pred_all,psy)';

% Generate the year timeline from 2018 to 2026 (96 points in total)
years = linspace(2018, 2026, length(x))';

figure;
plot(years,y,'k.','MarkerSize',15); hold on;
plot(years(tr.trainInd),y_pred_all(tr.trainInd),'bo');
plot(years(tr.valInd),y_pred_all(tr.valInd),'go');
plot(years(tr.testInd),y_pred_all(tr.testInd),'ro');
legend('True','Train','Validation','Test');
title('Training / Validation / Testing Prediction');
ylabel('Production Value (Million USD)');
set(gca, 'XTick', 2018:2026, 'XTickLabel', 2018:2026);
grid on;
```

Figure 5. Prediction with training, validation, and testing (Step 5).

```
%% 8. 10-Fold Cross Validation
K = 10;
cv = cvpartition(length(x),'KFold',K);
R2_cv = zeros(K,1);
RMSE_cv = zeros(K,1);
for k = 1:K
    trainIdx = training(cv,k);
    testIdx = test(cv,k);

    [xn_tr, psx_cv] = mapminmax(x(trainIdx)');
    [yn_tr, psy_cv] = mapminmax(y(trainIdx)');

    net_cv = fitnet([20 10],'trainlm');
    net_cv.trainParam.showWindow = false;
    net_cv = train(net_cv,xn_tr,yn_tr);

    x_test_n = mapminmax('apply',x(testIdx)',psx_cv);
    y_pred_n = net_cv(x_test_n);
    y_pred = mapminmax('reverse',y_pred_n,psy_cv)';

    SS_res = sum((y(testIdx)-y_pred).^2);
    SS_tot = sum((y(testIdx)-mean(y(testIdx))).^2);
    R2_cv(k) = 1 - SS_res/SS_tot;
    RMSE_cv(k) = sqrt(mean((y(testIdx)-y_pred).^2));
end
fprintf('\n=== 10-Fold CV ===\n');
fprintf('Mean R^2  = %.4f\n',mean(R2_cv));
fprintf('Mean RMSE = %.4f\n',mean(RMSE_cv));
```

Figure 6. 10-fold cross validation (Step 8).

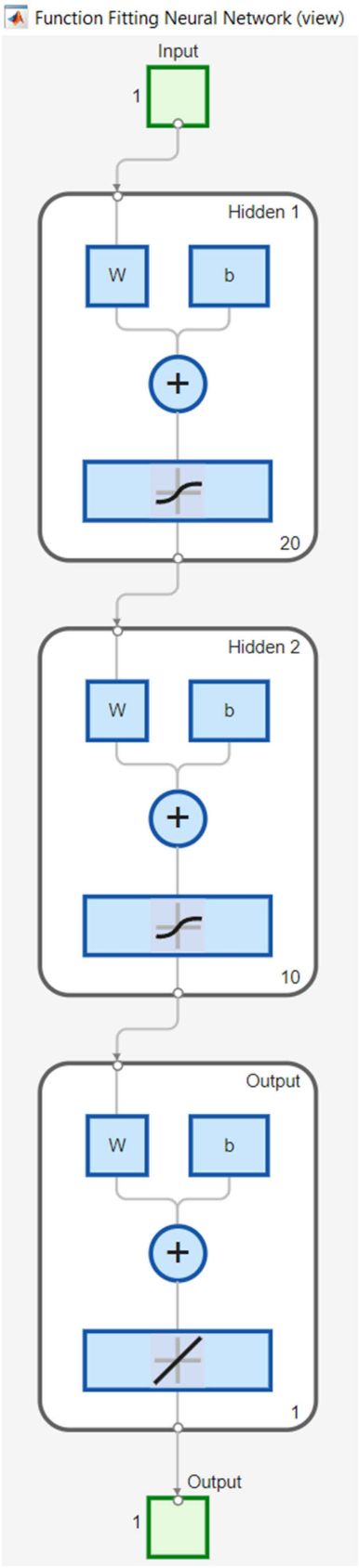


Figure 7. Neural network structure.

Moreover, to thoroughly quantify model accuracy and residual behavior, this study computes a standardized suite of scalar metrics, including the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE).

Based on the discussion regarding the coding system of the MLP regressor in this study, the diagram in **Figure 7** illustrates the topological architecture of the feedforward MLP designed for continuous function fitting and temporal reconstruction. The network adopts a sequential, unidirectional information flow, initiated by a single-dimensional input node that ingests the temporal feature (Amrouni Hosseini et al., 2024). As shown in **Figure 7**, the core computational framework comprises a dual-hidden-layer hierarchical structure, containing 20 and 10 neurons, respectively, which is also reflected in the coding system of the MLP regressor. Within each hidden block, the input signals undergo an affine transformation, which is governed by learnable weight matrices ("W") and bias vectors ("b"), which are aggregated via a central summation node. Also, both hidden layers subsequently apply a non-linear activation function (visually denoted by the sigmoidal curves in **Figure 7**). On the lowest layer, the extracted high-level latent features are propagated to a single-neuron output layer equipped with a linear transfer function (indicated by the straight diagonal line in **Figure 7**). The linear configuration at the terminus is theoretically imperative for regression tasks in this study, as it ensures the network can map the deeply processed representations into a continuous, unbounded scalar economic prediction without being constrained by artificial saturation limits.

3. Discussion

3.1. Evaluation of Neural Network Training

Through the full cycle of the MLP regressor coding process, the training performance of the MLP regressor is quantitatively captured by its loss trajectories over 66 epochs, evaluated on a logarithmic scale of MSE. As shown in **Figure 8**, the temporal evolution across the disjoint training, validation, and testing subsets demonstrates a distinct, multiphase convergence process. According to the curves in **Figure 8**, a rapid exponential decline in MSE is observed during the first 10 to 15 epochs, indicating that the network quickly extracts the dominant structural and temporal patterns inherent in the parsed inputs. Following this initial acceleration, the training loss exhibits a persistent downward trend toward lower orders of magnitude, whereas the validation and testing trajectories gradually decelerate, ultimately settling into a stable plateau. The convergence process naturally culminates at epoch 60, where the model achieves its optimal performance checkpoint, recording a minimum validation MSE of precisely 0.00099903. Beyond this optimal threshold, the validation and test errors display a marginal upward inflection despite the continuous descent of the training loss, marking the precise intervention point of the early-stopping mechanism.

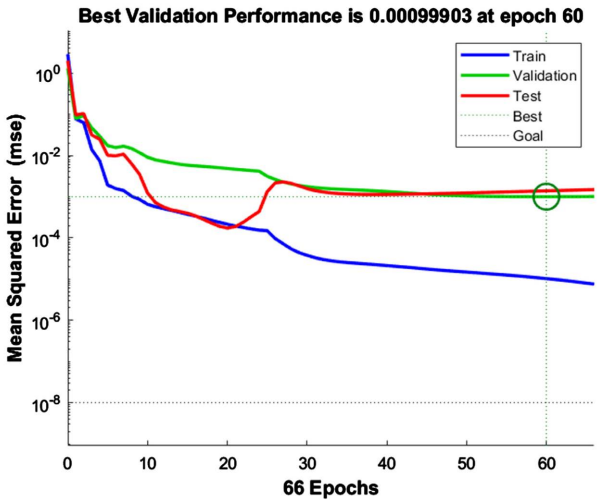


Figure 8. Performance.

From the perspective of model capacity and generalization, the structural dynamics of the loss curves in Figure 8 rigorously substantiate the architectural robustness of the proposed data-reconstruction pipeline. The curves in Figure 8 emphasize that the minimal divergence between the training and holdout (validation/test) losses at the optimal epoch (epoch 60) confirms that the network successfully assimilated the underlying economic structures without succumbing to spurious data memorization. The explicit divergence post-epoch 60 serves as a definitive signal of incipient overfitting, empirically validating the necessity of the implemented early-stopping constraint. Furthermore, a transient amplification in the test error observed within the mid-training window (epochs 20 - 30) highlights the model's sensitivity to specific batch allocations or localized regime shifts, which potentially reflect volatile macroeconomic transitions within the national export anchor. Also, evaluating the absolute magnitude of the optimal validation MSE (approx 10^{-3}) necessitates contextualization against the min-max normalization protocol applied during preprocessing. Since the target outputs were strictly scaled, this remarkably low fractional MSE translates to a highly constrained absolute error in the original units, thereby verifying the empirical reliability and practical economic significance of the reconstructed high-frequency production sequence for rural policy evaluation.

The internal training-state diagnostics of the MLP regressor provide critical insights into the optimization trajectory and the mathematical validity of the convergence process (Thamba et al., 2024; Qian & Li, 202; Mavrelis et al., 2026). The optimization process is evaluated through three key indicators spanning 66 epochs in Figure 9: the gradient norm, the damping parameter (μ), and the validation failure counter. By the final iteration, the global gradient norm reduced to 4.4302×10^{-4} while the damping parameter reached its lower bound of 1×10^{-5}, indicating that the solver successfully transitioned from an initial high-gradient exploratory phase into a refined, small-gradient regime. Referring to Figure 9, the termination of the training run was triggered by the validation-check routine after

the failure counter reached its threshold of 6, ensuring that the model parameters were frozen at a point of optimal generalization rather than over-refinement on the training set.

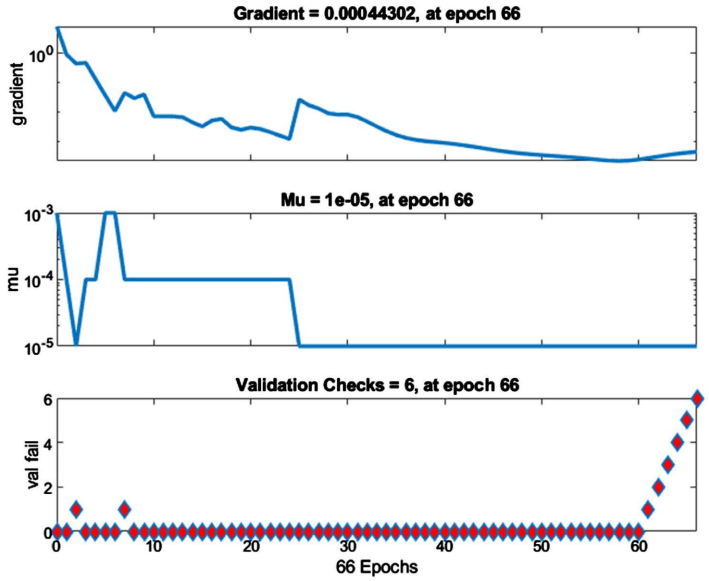


Figure 9. Training state.

As shown in Figure 9, the gradient norm panel exhibits a systematic monotonic decline from order-unity magnitudes to the 10^{-4} level, a desirable characteristic indicating that the model parameters are asymptotically approaching a stationary point of the loss function. Through the results on the gradient curve in Figure 9, a transient fluctuation observed during the mid-training phase likely reflects the optimizer navigating local topological complexities or stochastic variability induced by specific data segments. Additionally, the behavior of the damping parameter mu reveals the solver's adaptive strategy, while initial spikes in μ (10^{-3} to 10^{-4}) denote a conservative gradient-descent approach to stabilize early updates. The subsequent decay and persistence of μ at 1×10^{-5} signify that the algorithm transitioned into a Gauss-Newton regime, leveraging second-order curvature information for high-precision refinement of the production series reconstruction.

The validation-failure trace serves as the primary diagnostic for assessing the model's out-of-sample fidelity and the efficacy of the stopping criteria (Kumar & Beenamol, 2023; Khamis et al., 2020). Referring to Figure 9, the rapid accumulation of six consecutive failures in the final epochs indicates that further reductions in training loss no longer yielded improvements in validation performance. This divergence is a canonical signature of incipient overfitting, where the optimizer begins to capture noise specific to the training records rather than the underlying economic signals. To mitigate these risks and enhance the reliability of the reconstructed daily series for rural policy analysis, this study adopts a methodological framework that includes the retention of the optimal validation checkpoint, the utilization of L2 weight decay for regularization, and the implementation of boot-

strap intervals to convert MSE diagnostics into quantifiable confidence bands for the final economic indicators.

To further substantiate the predictive accuracy and structural reliability of the MLP regressor, a comprehensive analysis of the empirical residual distribution was conducted across the partitioned training, validation, and testing subsets in this study. The resulting error histogram in Figure 10 reveals a tightly bounded, prominently leptokurtic distribution characterized by a dominant central peak and narrow shoulders. As shown in Figure 10, the modal residual bin is heavily concentrated at a near-zero center (approx -2.5×10^{-4} within the normalized error space), with the vast majority of the residual mass strictly confined within a narrow ±0.01 margin. The uniform alignment of this central tendency across all three data splits signifies consistent, unbiased learning, confirming that the network successfully abstracted the underlying temporal dynamics without overfitting to the training corpus. Including isolated negative outliers and a minor cluster of positive deviations within the right-hand tail of the held-out test set, despite this high overall fidelity, the marginal presence of infrequent extreme errors suggests occasional sensitivity to localized structural shocks or reporting irregularities inherent to the emergent first-launch economy. Upon the discussion above, this residual profile confirms a robust temporal reconstruction capability characterized by negligible systematic bias and exceptionally low typical prediction error, thereby firmly validating the digital-intelligence pipeline's utility for generating high-resolution, policy-ready economic series.

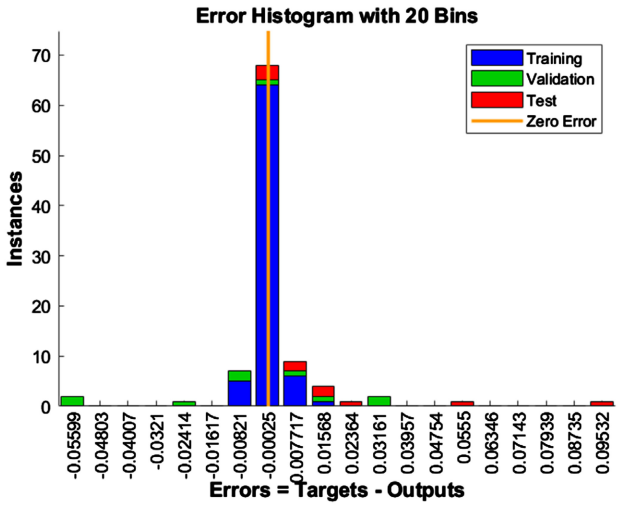


Figure 10. Error histogram.

As the study of Mühlbacher & Piringer (2013) indicates, the regression analysis across the partitioned datasets provides a granular evaluation of the MLP's structural learning efficacy and generalization capability. According to the results in Figure 11, the training subset exhibits a near-perfect linear alignment between network outputs and targets (R: 0.99998), characterized by a unit slope and a vanishing intercept, which confirms that the model optimization procedure has suc-

cessfully captured the dominant functional mappings within the training corpus with negligible systematic bias. In contrast, while the validation (R: 0.99861) and testing (R: 0.9989) partitions maintain exceptionally high correlations, they reveal a subtle but distinct attenuation phenomenon, evidenced by slightly sub-unit slopes (approx 0.99 & approx 0.98). Especially, the test partition displays a minor systematic upward offset (intercept: approx 0.028 in normalized units). The deviation with Figure 11 suggests that the model generalizes well overall but exhibits a marginal shrinkage effect on extreme magnitudes.

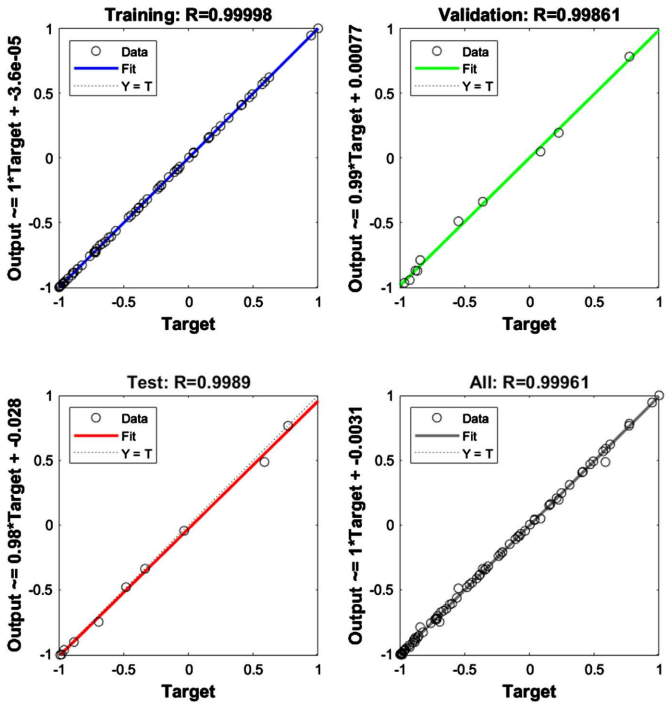


Figure 11. Regression.

Synthesizing these findings, the composite regression analysis for the aggregated dataset (R: 0.99961) in Figure 11 demonstrates exceptional overall predictive fidelity with minimal residual dispersion across the full dynamic range of the first-launch economy series. Although the pooled metric suggests near-identity recovery, the split-specific diagnostics crucially highlight that the model's generalization is characterized by a conservative prediction strategy (slight magnitude attenuation) rather than catastrophic failure. This consistent pattern of low central error, combined with infrequent, bounded tail deviations, validates the robustness of the digital-intelligence system.

The comparative assessment of predicted outputs against normalized targets across the input domain visually corroborates the structural fidelity of the MLP regressor (D'Ambros et al., 2012; Marín Díaz, 2025). Figure 12 demonstrates that the MLP regressor successfully captures the principal non-linear production trends, with the training, validation, and testing distributions tightly clustered around the identity mapping for the vast majority of observations. Across the re-

sults in Figure 12, the fitted trajectory exhibits high coherence with the target variables within the central density of the data, confirming that the network has effectively learned the dominant temporal and covariate-driven structures essential for high-frequency reconstruction. On the other hand, the results of error in Figure 12 reveal a distinct, non-uniform error distribution characterized by input-dependent heteroscedasticity. While the residual cloud remains symmetric and zero-centered for low-to-intermediate input magnitudes, indicating an unbiased estimator in the nominal regime, appreciable divergence manifests as isolated spikes in the high-amplitude regions. This suggests that the pattern is not uniformly random but is positively correlated with the magnitude of the production signal.

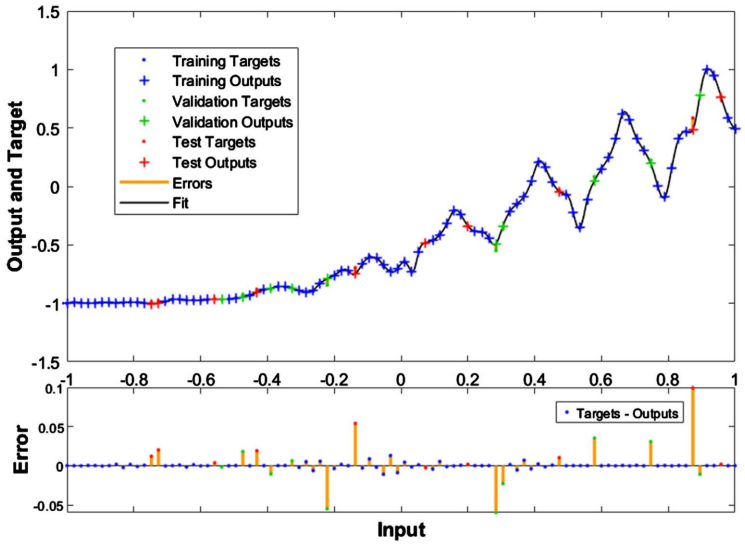
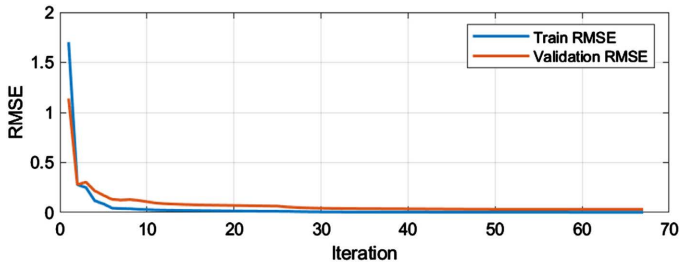


Figure 12. Fit.

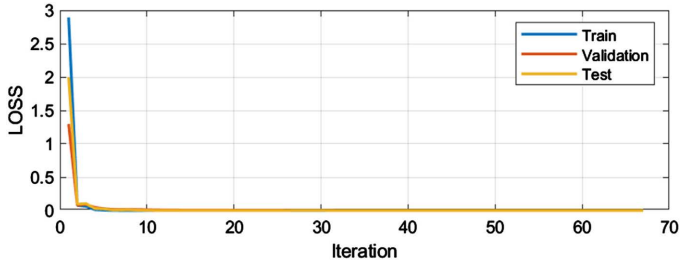
Synthesizing the diagnostic observations in Figure 12, the concentration of residuals at the upper extremities of the input range points to three plausible, interacting factors inherent to the first-launch economy context. First, the deviations likely reflect genuine, stochastic production shocks present in the source data, such as seasonal manufacturing surges or pandemic-induced logistic disturbances, which statistically behave as outliers relative to the secular trend. Second, the deviations may imply residual artifacts from the ETL pipeline, where unit misrecognition or temporal misalignment in web-scraped records creates artificial noise in high-value data points. Third, the slight under- or overshoot at the peaks suggests a marginal regime-dependent inductive bias, where the MLP architecture prioritizes global smoothness over the fitting of extreme, localized volatility. Despite these specific limitations in the asymptotic regions, the prevailing symmetry of the residual distribution and the low magnitude of errors relative to the full normalized range (±1.0) confirm that the MLP regressor retains robust central accuracy, validating its utility for reconstructing the baseline economic dynamics necessary for rural revitalization analysis.

3.2. Prediction Performance Evaluation

The dynamic progression of the predictive error across iterations provides critical validation of the model's temporal generalization capabilities (Han et al., 2019; Jenkins et al., 2021; Dankers et al., 2015). As shown in **Figure 13(a)**, an analysis of the Root Mean Square Error (RMSE) trajectory reveals a highly efficient initial optimization phase, which indicates the RMSE exhibits a precipitous collapse from order-unity magnitudes to a marginal fraction of the normalized scale within the first 5 to 10 epochs. According to the trends of the curves in **Figure 13(a)**, following this swift assimilation of fundamental temporal patterns, the training and validation trajectories enter an extended, stable plateau where they track each other with remarkable precision. This tight convergence between the fitting and held-out subsets signifies that the Multilayer Perceptron (MLP) algorithm successfully extracted genuine, generalizable economic signals rather than memorizing stochastic idiosyncrasies inherent to the training partition. As Yi et al. (2025) noted that, such controlled generalization behavior is mathematically essential for ensuring that the learned mapping can reliably interpret unseen, high-frequency monthly signals.



(a) Process results of RMSE



(b) Process results of Loss

Figure 13. Process results (66 epochs).

As **Figure 13(b)** indicates, the iterative loss function trajectory underscores a structurally stable convergence process characterized by pronounced initial gradient updates that rapidly settle into a minimal asymptote, which can corroborate the RMSE dynamics in **Figure 13(a)**. As the optimization progresses toward the final iterations, the training loss continues to decline marginally, while validation improvements stall, indicating that the optimizer has entered a regime of diminishing returns where parameter updates primarily reduce in-sample variance. Referring to the results in **Figure 8**, at its optimal checkpoint (epoch 60), the MLP model achieved a best-validation MSE of 9.99×10^{-4}, corresponding to a highly

constrained normalized RMSE of approximately 3.16×10^{-2}. This minimal validation-training gap empirically validates that the network has comprehensively captured the dominant non-linear temporal relationships required to synthesize granular production dynamics for Hanbin District, and it also confirms that the macroscopic export data sourced from the National Bureau of Statistics of China functions as a highly effective, structurally informative covariate driving the local economic reconstruction.

Based on the rapid initial learning and stable convergence, the synthesized process diagnostics confirm numerically robust training results that are ideally suited for generating a synthetic daily production series for downstream policy evaluation. Although this favorable optimization profile guarantees low central prediction error and robust out-of-sample generalizability, the subtle divergence observed in late-epoch training suggests that exogenous regime changes in the source macroeconomic signals, such as pandemic-induced supply chain disruptions or sudden export-market reconfigurations, may precipitate localized predictive discrepancies not fully represented within the training manifold. Additionally, while the residual discrepancies between the training and validation sets remain negligible at the normalized daily level, these bounded errors possess the statistical potential to selectively amplify when the high-frequency reconstructions are temporally aggregated into broader macroeconomic proxies.

The temporal reconstruction of the plush-toy first-launch industry in Hanbin District (2018-2026) provides a granular empirical validation of the MLP regressor's predictive fidelity. As delineated in the temporal scatter plot with **Figure 14**, the observed monthly production values exhibit a distinct multi-phase economic trajectory: a near-zero incubation baseline between 2018 and 2019, a transitional expansion through 2020-2021, and a pronounced non-linear acceleration thereafter, ultimately scaling to peaks exceeding 100 million USD by 2025-2026. As shown in **Figure 14**, the model outputs demonstrate exceptional alignment with the empirical observations across the partitioned training, validation, and testing subsets, successfully capturing both the secular macroeconomic growth trend and the localized, month-to-month volatility. However, due to aggregate production volumes intensifying, the absolute predictive variance marginally widens, manifesting as localized offsets during the high-yield phases of 2023-2025, which indicate a rigorous structural analysis reveals distinct heteroskedastic behavior correlated with production scale. Also, such as an under-predicted contraction in early 2023 and an over-predicted spike in 2025, which indicate that occasional temporal discrepancies indicate transient model sensitivities. These bounded deviations are methodologically consistent with earlier diagnostics characterizing a leptokurtic error distribution. From an economic perspective, they primarily stem from exogenous regime shifts (pandemic-induced logistic disruptions altering national-local economic linkages), inherent parsing noise within the web-scraped ETL inputs, and the statistical under-representation of extreme macroeconomic shocks within the training manifold.

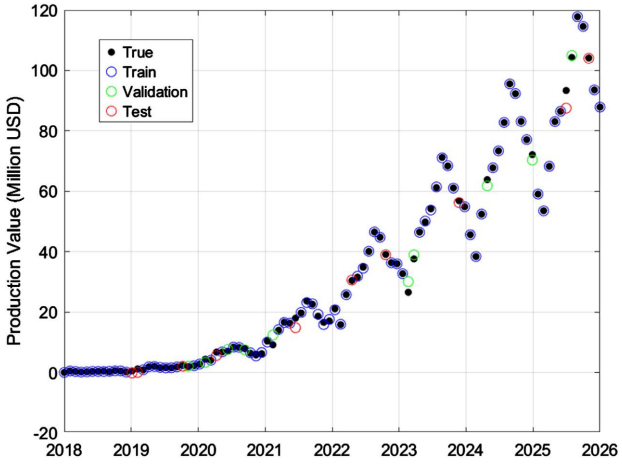


Figure 14. Training/validation/testing prediction.

In order to rigorously evaluate the predictive stability of the reconstruction model, a statistical analysis of the empirical residual distribution was conducted. As shown in Figure 15, the error distribution histogram illustrates a pronounced concentration of residuals tightly clustered around a near-zero center, accompanied by a marginal frequency of extreme errors extending into both tails. Furthermore, the Gaussian equation in Figure 15 indicates that the superimposed theoretical Gaussian density function yields a mean of $\mu = -0.0918$ and a standard deviation of $\sigma = 0.9432$. The results in Figure 15 suggest that the model's predictions are infinitesimally higher than the normalized targets, and the estimated mean indicates a negligible negative bias, but the visual discrepancy between the empirical data and the parametric fit is analytically significant. Specifically, the empirical histogram exhibits a highly leptokurtic geometry, characterized by a significantly sharper central peak and thinner shoulders than the idealized Gaussian curve. This structural divergence demonstrates that a single normal distribution cannot fully encapsulate the residual behavior, because the vast majority of prediction errors are exceptionally diminutive, interspersed with sparse but high-magnitude outliers.

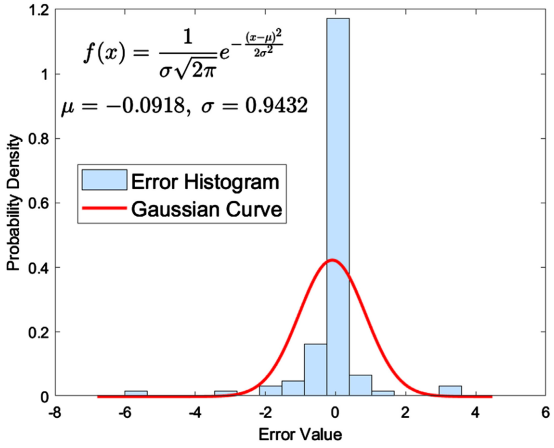


Figure 15. Error distribution.

From a methodological and economic perspective, this specific residual architecture carries profound implications for the validity of the reconstructed daily production series for the Hanbin District first-launch plush-toy industry. Primarily, the overwhelming concentration of minute central errors confirms that the MLP framework generates highly accurate point estimates for the majority of the temporal sequence, establishing a robust foundation for aggregate trend analysis.

Table 1. Error index.

Full Data Evaluation	
R^2	0.999196
RMSE	0.942779
MAE	0.373261
MBE	−0.091751
10-Fold CV	
Mean R^2	0.9695
Mean RMSE	5.2242

To rigorously quantify the predictive efficacy and structural robustness of the MLP regressor, a comparative analysis of full-sample evaluation metrics and cross-validated performance was conducted in this study. Referring to the analytical results in Table 1, the MLP model exhibits an exceptionally high degree of in-sample fidelity, evidenced by an R^2 of 0.999196 and a minimal RMSE of 0.942779 when evaluated across the entire aggregated dataset. Complementary indices, including a Mean Absolute Error (MAE) of 0.373261 and a Mean Bias Error (MBE) of −0.091751, confirm that the algorithm captures virtually all variance with only a negligible systematic negative bias. However, the implementation of a 10-fold cross-validation protocol reveals a material divergence in out-of-sample generalization, yielding a reduced mean R^2 of 0.9695 alongside a roughly fivefold increase in mean RMSE of 5.2242, as shown in Table 1. Due to temporal dependence and structural regime shifts violating the independent and identically distributed (IID) assumptions of randomized CV and also producing overly optimistic, misleading in-sample metrics, the observed performance attenuation underscores the fundamental mathematical limitations of applying conventional randomized *k*-fold partitioning to non-stationary economic time series. To obtain a more realistic assessment of forecasting and reconstruction skill under non-stationarity, including blocked (non-overlapping) splits, rolling (walk-forward) validation, and a final last-block holdout, the findings of the discussion regarding the prediction performance evaluation indicate that replacing randomized *k*-fold validation with a time-series evaluation framework and reporting results exclusively on a future holdout segment. In this study, the pronounced temporal dependence and structural regime shifts characterizing the Hanbin District first-launch industry inherently violate the independent and identically distributed assumptions underlying standard CV,

often rendering full-sample metrics overly optimistic. Furthermore, this discrepancy underscores the necessity of rigorously auditing the ETL pipeline to preclude inadvertent data leakage, ensuring that the macroscopic covariates integrated from the National Bureau of Statistics of China do not introduce implicit look-ahead biases into the localized historical reconstruction.

Synthesizing the evaluative metrics in Table 1, the results emphasize that the MLP pipeline undeniably succeeds in capturing the dominant, long-term growth dynamics of the local first-launch economy, reproducing the foundational production series with exceptionally low central error. Nevertheless, the pronounced contrast between the highly optimistic full-sample fit and the more conservative cross-validation outcomes validates earlier observations regarding underlying heteroskedasticity and the disproportionate influence of sparse temporal outliers. For practical econometric applications, relying solely on in-sample metrics would fundamentally overstate model certainty, particularly when these high-frequency daily inferences are temporally aggregated to formulate monthly or quarterly production proxies against national export benchmarks. Upon the discussion above, responsible policy inference for rural revitalization dictates that the conservative uncertainty bounds and error distributions derived from the cross-validation framework must be explicitly integrated into any downstream economic forecasting or efficiency evaluations.

3.3. Prediction of the MLP Regressor under the ETL Pipeline

As shown in Figure 16, an integrated analysis of the reconstructed monthly production trajectory and its associated 95% confidence intervals was conducted for the Hanbin District plush-toy industry (2018-2026) to rigorously assess the temporal stability and predictive reliability of the MLP regressor in this study. As depicted in Figure 16, the model-derived reconstruction seamlessly tracks the empirical observations, accurately reproducing the secular macroeconomic expansion, seasonal oscillations, and the precise timing of major structural peaks and troughs. During the nascent incubation and stable growth phases (2018-2022), the confidence bands remain remarkably narrow, reflecting minimal predictive dispersion and high model certainty within low-volatility regimes. However, a pronounced heteroskedastic pattern emerges as production volumes scale, due to the confidence intervals systematically widening around the high-amplitude peaks observed between 2023 and 2025. Especially, a localized outlier in early 2023 marginally breaches the lower boundary of the 95% band.

The structural dynamics of the confidence intervals in Figure 16 provide critical methodological and practical implications for utilizing the digital-intelligence pipeline in rural revitalization evaluation. Including near-unity R^2 values and leptokurtic residual distributions, the tightly bound intervals across the majority of the temporal sequence corroborate earlier statistical diagnostics, confirming that the reconstruction is highly reliable for assessing central tendencies and executing short-horizon aggregations under normal macroeconomic conditions. Neverthe-

less, the scale-dependent widening of the predictive bands distinctly underscores the presence of heteroskedastic uncertainty, dictating that normalized or percentage-based error metrics must be integrated when conducting cross-regime economic comparisons. Furthermore, the isolated empirical observation falling outside the 95% confidence threshold serves as a vital diagnostic flag, potentially indicating genuine exogenous production shocks unrepresented in the training manifold, residual noise from the upstream ETL parsing process, or marginal calibration limits within the uncertainty estimator. Because strategic rural policy interventions (targeted municipal subsidies, capital investments, and supply-chain optimizations) are acutely sensitive to such macroeconomic tail events, explicitly accounting for these localized expansions in predictive uncertainty is imperative to ensure robust, evidence-based decision support for the first-launch economy.

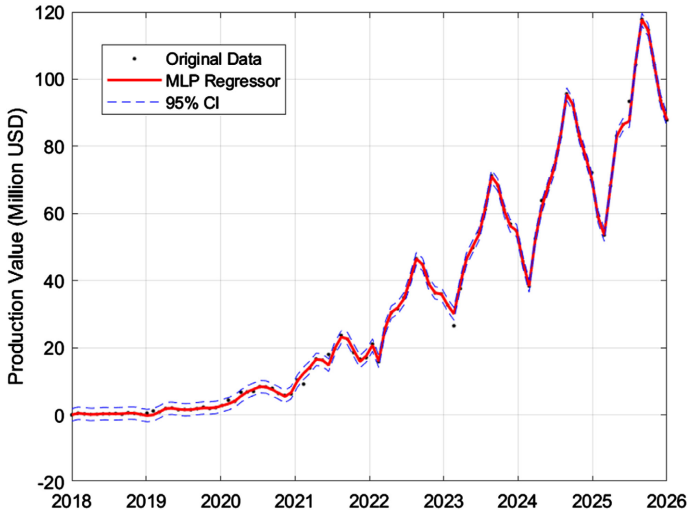


Figure 16. Prediction with 95% confidence interval.

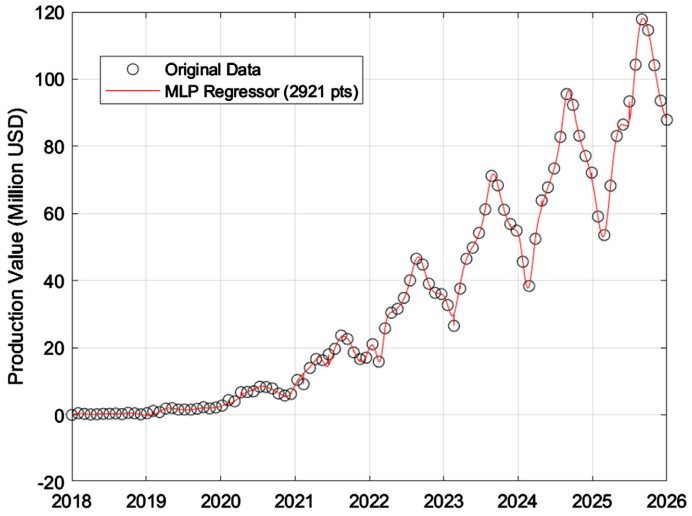


Figure 17. MLP regression reconstruction.

Visualized across the comprehensive 2018-2026 temporal window in Figure 17,

the MLP regressor demonstrates robust structural alignment with the observed monthly production benchmarks, successfully synthesizing a high-frequency daily sequence comprising 2921 reconstructed observations. As shown in Figure 17, the model accurately captures the secular macroeconomic acceleration post-2020, alongside the precise phase and magnitude of recurrent seasonal peaks spanning 2022-2025. The temporal coherence under the MPL regressor model confirms that the synergistic integration of the ETL pipeline with macroscopic national toy-export anchors provides sufficient informational signal for the network to concurrently resolve both low-frequency foundational trends and high-frequency short-term volatility. Nevertheless, marginal structural deviations manifest at high-amplitude extremities, where the model occasionally exhibits slight under- or overshooting a behavior perfectly consistent with previous diagnostics identifying a leptokurtic residual distribution and elevated absolute errors during extreme production intervals. Due to this, while this digital-intelligence framework yields a highly granular, empirically plausible reconstruction of the Hanbin District first-launch plush-toy industry (precise evolution from nascent incubation to mature industrial scale), it inherently functions as an optimized statistical estimate rather than an absolute deterministic measurement. Given the documented disparities between in-sample and cross-validated metrics, as well as the potential for upstream ETL-sourced label noise (parsing artifacts or unit-conversion errors), downstream econometric aggregations and subsequent rural policy inferences must strictly incorporate these quantified uncertainties to ensure robust, evidence-based decision-making.

An empirical analysis of Hanbin District's annual production-value series strictly corroborates the nonlinear scaling characteristic of a successful first-launch economy, transitioning rapidly from its 2018 incubation phase to a period of explosive industrial concentration. As shown in Table 2, from 13.92 million USD in 2021 to 139.77 million USD in 2025, quantitative inspection reveals a dramatic production escalation driven by exceptionally high year-over-year growth rates (approx 95.9% in 2022, peaking at approx 195.2% in 2023) and culminating in a four-year Compound Annual Growth Rate (CAGR) of 78.0% about. This massive scale-up is also structurally validated by contemporaneous surges in local firm entry and employment absorption. However, a critical divergence emerges when comparing these soaring production totals against localized export receipts, which remained disproportionately low (e.g., 9.61 million USD in 2023). This discrepancy highlights profound measurement subtleties, implying extensive domestic absorption, complex intermediary routing, or significant reporting lags that obscure the true economic output. It is precisely this structural data fragmentation that necessitates and validates the proposed deep learning methodology. By utilizing an ETL pipeline to parse granular, place-specific local indicators and integrating the national export series as a robust macroeconomic anchor, the MLP regressor successfully navigates these data asymmetries. Especially, the near-unity in-sample R^2, tightly concentrated residual mass, and stable time-aware cross-validation, as evidenced

by previous model diagnostics, demonstrate that the MLP architecture is exceptionally predictive and accurate, proving highly capable of interpolating non-linear regional growth pathways despite localized reporting anomalies.

Table 2. Annual economic data of the Ankang plush toy industry (Xu, 2024; Hong & Liang, 2023; Yu, 2023; Zou, 2025; Department of Human Resources & Social Security of Shaanxi Province, 2025).

Year	Number of industries	Annual production value	Employment	Exports
		Million USD	Person	Million USD
2018	Initial stage	-	52	-
2021	569	13.92	12,000	-
2022	826	27.27	18,000	-
2023	788 - 807	80.51	19,000	9.61
2024	805 - 806	105.24	21,770	7.95
2025	800 - 811	139.77	-	-

The empirical annual economic figures in **Table 2** simultaneously validate the structural reality of the rural first-launch economy and the methodological necessity of the digital-intelligence framework. The extreme, non-linear growth trajectory of the plush-toy industry, coupled with the inherent limitations of localized export logs, renders traditional sparse-data tracking inadequate for comprehensive economic evaluation. By successfully harmonizing granular ETL-derived local data with stable national macro-anchors, the MLP reconstruction mathematically bridges this critical information gap. The resultant high-frequency daily production series not only reflects a realistic and accurately learned economic target but also establishes the ETL + MLP pipeline as a highly reliable, analytically rigorous instrument. Consequently, the deep learning approach in this study provides an essential quantitative foundation for evaluating rural revitalization efficiency, yielding a plausible, high-resolution temporal framework perfectly calibrated for downstream macroeconomic forecasting and localized policy formulation.

4. Conclusion

The empirical investigations conducted in this study substantiate that the first-launch economy serves as a transformative catalyst for rural revitalization. Using the plush toy industry in Hanbin District as a primary case study, the research demonstrates how a highly concentrated, localized industrial sector can precipitate rapid job creation, accelerate enterprise formation, and drive exponential expansions in aggregate production value within a significantly compressed timeframe. Beyond mere financial metrics, the localized industrial takeoff generates profound positive spillovers, including massive labor absorption, the cultivation of upstream supplier networks, and the integration of isolated regional markets into broader global supply chains (Zhang & Wei, 2025; Xu, 2025; Shi et al., 2025; Wu, 2026).

Consequently, the findings in this study provide robust empirical evidence that first-launch industries are not anecdotal, localized phenomena, but rather structurally sound, highly replicable engines for driving comprehensive rural development when supported by targeted infrastructure and strategic policy frameworks.

To accurately measure this non-linear economic expansion, this study confirms the exceptionally high-fidelity and operational practicality of integrating an ETL pipeline with an MLP regressor. The combined methodology in this study successfully overcame the critical bottleneck of sparse rural statistics by reconstructing 2,921 high-frequency daily observations from merely 96 fragmented monthly inputs. The architectural robustness of the model is empirically validated by an exceptional in-sample R2 of approximately 0.999 and tightly bounded central residuals. Furthermore, stringent out-of-sample evaluations (validation MSE of 9.99×10^{-4} & time-aware cross-validation mean R^2 of 0.9695) confirm the pipeline's generalization capabilities. Supported by rigorous residual analyses and dynamic confidence bands that effectively isolate localized tail events, the performance metrics in this study definitively prove that the ETL-MLP framework is sufficiently robust to support continuous operational monitoring in historically data-poor rural contexts.

More broadly, this methodological breakthrough establishes a foundational mechanism for the digital-intelligence empowerment of grassroots economic governance. In rural jurisdictions where administrative statistics are traditionally lagged, structurally fragmented, or absent, the ETL-MLP approach bridges the informational divide. By structurally fusing stable national macroeconomic anchors with automated, web-scraped local temporal signals, the pipeline generates provenance-tracked, high-resolution datasets that capture nuanced local dynamics invisible to macroscopic national indices. Practically, the technological capability of ETL-MLP dismantles the measurement barriers historically impeding evidence-based policymaking, while it directly addresses the dual needs of modern rural analytics as interpolating missing low-frequency data and generating timely, short-horizon forecasts. Based on the ETL-MLP approach, local authorities are empowered to conduct granular benchmark comparisons, dynamically monitor industry takeoff, and execute uncertainty-aware policy simulations in response to exogenous economic shocks.

In conclusion, this study demonstrates that digital-intelligence methodologies offer a practical, precise, and highly transferable solution for evaluating first-launch economies within broader rural revitalization paradigms. However, it is imperative to acknowledge that the reconstructed high-frequency series functions as an optimized statistical estimate rather than an absolute administrative ground truth. While its high central accuracy and explicitly quantified uncertainty bounds make it a deeply valuable asset for econometric research, future iterations of this framework must address its inherent limitations. Subsequent research should focus on implementing model ensembles and advanced heteroskedastic uncertainty modeling to calibrate predictive intervals more effectively during extreme production

peaks. Furthermore, integrating multi-modal data sources will fundamentally enhance the model's spatial and structural resolution. Upon the findings in this study, deploying this refined pipeline across diverse industrial sectors and geographic regions will enable the systematic scaling of this digital-intelligence framework to inform and elevate national-level rural development practice.

Acknowledgements

This research was funded by the 2025 Humanities and Social Science Research Project of the Ministry of Education of the People's Republic of China, "Digital-Intelligence-Empowered Evaluation of First-Launch Economy Efficiency and Pathways to Improvement for Key Assistance Counties in Rural Revitalization" (25YJA790078).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Amrouni Hosseini, M., Ravanshadnia, M., Rahimzadegan, M., & Ramezani, S. (2024). Next-Generation Building Condition Assessment: BIM and Neural Network Integration. *Journal of Performance of Constructed Facilities, 38,* Article 0402450. https://doi.org/10.1061/jpcfev.cfeng-4828

Bommisetti, R. K., Raj, B. K., Subbalakshmi, A. V. V. S., Shehryar, M., & Hoang, S. D. (2022). Blockchain in Trust with Reputation Management for Financial Stock Market Using Distributed Ledger Technology and Bayesian Theory Based on Fault Tolerance Model. *Global Business Review.* https://doi.org/10.1177/09721509221110371

Clift, S. S. (2007). *Linear and Non-Linear Monotone Methods for Valuing Financial Options under Two-Factor, Jump-Diffusion Models*. https://uwspace.uwaterloo.ca/items/97c6d2ea-3aeb-4464-b181-eba3fb4e2203

D'Ambros, M., Lanza, M., & Robbes, R. (2012). Evaluating Defect Prediction Approaches: A Benchmark and an Extensive Comparison. *Empirical Software Engineering, 17,* 531-577. https://doi.org/10.1007/s10664-011-9173-9

Dankers, A., Van den Hof, P. M. J., Bombois, X., & Heuberger, P. S. C. (2015). Identification of Dynamic Models in Complex Networks with Prediction Error Methods: Predictor Input Selection. *IEEE Transactions on Automatic Control, 61,* 937-952. https://doi.org/10.1109/tac.2015.2450895

Department of Human Resources, & Social Security of Shaanxi Province (2025). *Ankang: Systematically Building a New Integrated Development Paradigm of "Plush Toys + AI + Cultural Tourism*. https://rst.shaanxi.gov.cn/sy/ztzl/rdzt/rsgzjyq/jyjl/202512/t20251231_3601171.html

Goel, R. K., Saunoris, J. W., & Goel, S. S. (2021). Supply Chain Performance and Economic Growth: The Impact of COVID-19 Disruptions. *Journal of Policy Modeling, 43,* 298-316. https://doi.org/10.1016/j.jpolmod.2021.01.003

Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A Review of Deep Learning Models for Time Series Prediction. *IEEE Sensors Journal, 21,* 7833-7848. https://doi.org/10.1109/jsen.2019.2923982

Hong, Y., & Liang J. (2023). Plush Toys Produced in Small City in NW China's Shaanxi Reach over 80 Countries, Regions. *People's Daily Online Exclusives*.

https://en.people.cn/n3/2023/0921/c98649-20074698.html

Hou, G., & Hu, N. (2026). The Mechanisms, Problems, and Pathways of the First-Launch Economy Empowering the Integrated Development of Rural Culture and Tourism. *Journal of Social Science of Hunan Normal University, 1,* 74-81.

Jenkins, D. A., Martin, G. P., Sperrin, M., Riley, R. D., Debray, T. P. A., Collins, G. S. et al. (2021). Continual Updating and Monitoring of Clinical Prediction Models: Time for Dynamic Prediction Systems? *Diagnostic and Prognostic Research, 5,* Article No. 1. https://doi.org/10.1186/s41512-020-00090-3

Jokipii, M. (2023). *Managing Missing Data in Data Integration*. https://trepo.tuni.fi/bitstream/handle/10024/148878/JokipiiMervi.pdf?sequence=2

Karacsonyi, D., & Taylor, A. (2023). Understanding Demographic and Economic Patterns in Sparsely Populated Areas—A Global Typology Approach. *Geografiska Annaler: Series B, Human Geography, 105,* 228-247. https://doi.org/10.1080/04353684.2022.2103445

Khamis, M., Elhaj, M., & Abdulraheem, A. (2020). Optimization of Choke Size for Two-Phase Flow Using Artificial Intelligence. *Journal of Petroleum Exploration and Production Technology, 10,* 487-500. https://doi.org/10.1007/s13202-019-0734-6

Koffi, Y. J. E., Blé Acca, K. S. R., & Iritié, B. G. J. J. (2024). Characterization of the Economic Growth of Côte d'Ivoire from 1960 to 2021: An Application of the Bai-Perron Multiple Break Test Approach. *Modern Economy, 15,* 859-878. https://doi.org/10.4236/me.2024.159045

Kumar, S. P., & Beenamol, M. (2023). Multiple Layer Radial Basis Neural Network with Remora Regression Tree Optimum Feature Extraction for Structural Health Monitoring. *Asian Journal of Civil Engineering, 24,* 989-999. https://doi.org/10.1007/s42107-022-00547-4

Kwok, S., & Zhang, X. (2024). Deployment of Blockchain Technology on the Banking Industry in China: A Policy Review Study. *European Academic Journal-II China, 1,* 1-30. https://eaj.ebujournals.lu/index.php/EAJ_II/article/view/98

Mahmud, D., & Ikbal, M. Z. (2022). The Role of ETL (Extract-Transform-Load) Pipelines in Scalable Business Intelligence: A Comparative Study of Data Integration Tools. *ASRC Procedia: Global Perspectives in Science and Scholarship, 2,* 89-121. https://doi.org/10.63125/1spa6877

Marín Díaz, G. (2025). Comparative Analysis of Explainable AI Methods for Manufacturing Defect Prediction: A Mathematical Perspective. *Mathematics, 13,* Article 2436. https://doi.org/10.3390/math13152436

Mavrelis, I., Chatzopoulos, A., & Zacharia, P. (2026). Hybrid Deep Learning for Predictive Maintenance in Industrial Machinery Using LSTM and MLP Models. *Machines, 14,* Article 191. https://doi.org/10.3390/machines14020191

Mbogori, F. I., & Luketero, S. (2019). Factors Influencing Sustainability of Women Owned Small and Medium Enterprises a Case of Nkubu Town Imenti South Sub County Meru County Kenya. *International Academic Journal of Information Sciences and Project Management, 3,* 355-376.

Merali, A. (2025). *Scaling Laws for Economic Productivity: Experimental Evidence in LLM-Assisted Consulting, Data Analyst, and Management Tasks*. https://arxiv.org/abs/2512.21316

Mirzazadeh, S. (2025). *Stock Price Prediction Using Deep Learning: A Case Study on Tesla*. Doctoral Dissertation, University of Victoria.

Mühlbacher, T., & Piringer, H. (2013). A Partition-Based Framework for Building and Validating Regression Models. *IEEE Transactions on Visualization and Computer Graphics,*

19, 1962-1971. https://doi.org/10.1109/tvcg.2013.125

Neacsu, A., Neacsu, A., & Bichir, A. (2023). A Comparative Analysis of the Global Supply Chain Bottlenecks during the Covid-19 Pandemic. In *Springer Proceedings in Business and Economics* (pp. 101-112). Springer. https://doi.org/10.1007/978-3-031-58967-6_9

Ngobeni, E., & Muchopa, C. L. (2023). Structural Change in the South African Agricultural Sector: Bai-Perron Modelling. *Scientific African, 21,* e01732. https://doi.org/10.1016/j.sciaf.2023.e01732

Peters, M. A. (2023). Digital Trade, Digital Economy and the Digital Economy Partnership Agreement (DEPA). *Educational Philosophy and Theory, 55,* 747-755. https://doi.org/10.1080/00131857.2022.2041413

Qian, M., & Li, Y. (2022). A Novel Adaptive Undersampling Framework for Class-Imbalance Fault Detection. *IEEE Transactions on Reliability, 72,* 1003-1017. https://doi.org/10.1109/tr.2022.3214519

Qingran, G., Razi, U., Ramzan, M., Cuicui, D., & Vasa, L. (2025). Digital Economy and Export Complexity: Unveiling Its Role in Transforming China's Manufacturing Industry. *Environment, Development and Sustainability.* https://doi.org/10.1007/s10668-025-06742-y

Reich, P. B. (2014). The World-Wide 'Fast-Slow' Plant Economics Spectrum: A Traits Manifesto. *Journal of Ecology, 102,* 275-301. https://doi.org/10.1111/1365-2745.12211

Shi, X., Xiao, Y., & Xu, F. (2025). The Basic Logic and Implementation Path of Empowering the Realization of Ecological Product Value through the Debut Economy. *Contemporary Economic Management, 47,* 11-22.

Thamba, N. B., Vanapalli, V. T., Duraiswamy, R. P., Sonnathi, N., & Illuri, S. S. (2024). Comparison of ML Algorithms and Neural Networks on Fault Diagnosis of a Worm Gear. *Journal of Vibration Engineering & Technologies, 12,* 6355-6370. https://doi.org/10.1007/s42417-023-01256-1

Touat, O. (2024). Global Supply Chain Disruptions: Lessons from the COVID-19 Pandemic Crisis. In *Business Resilience and Market Adaptability: Pandemic Effects and Strategies for Recovery* (pp. 117-135). Springer.

Uma Maheswari, B., Naveen Sakthivel, K. S., Kavitha, D., & Sujatha, R. (2024). An Automated Extract, Transform, Load (ETL) Pipeline to Facilitate Acquisition and Analysis of Stock Marker Data. In *International Conference on Modern Practices and Trends in Expert Applications and Security* (pp. 375-384). Springer.

Voyer, M., Quirk, G., Farmery, A. K., Kajlich, L., & Warner, R. (2021). Launching a Blue Economy: Crucial First Steps in Designing a Contextually Sensitive and Coherent Approach. *Journal of Environmental Policy & Planning, 23,* 345-362. https://doi.org/10.1080/1523908x.2020.1856054

Wu, Z. (2026). Research on the Value Realization Mechanism of Urban Debut Economy from a Spatiotemporal Coupling Perspective. *Circulation Economy, 35,* 12-15.

Xu Qin (2024). Toys Help Ankang Residents to Lead a Plush New Life. *China Daily.* https://www.chinadailyhk.com/hk/article/232819?utm

Xu, Y. (2025). Debut Economy, New Quality Productive Forces, and the Resilience of Industrial and Supply Chains, *East China Economic Management, 39,* 101-109.

Yi, K., Zhang, Q., Fan, W., Cao, L., Wang, S., He, H. et al. (2025). A Survey on Deep Learning Based Time Series Analysis with Frequency Transformation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2* (pp. 6206-6215). ACM. https://doi.org/10.1145/3711896.3736571

Yu, H. C. (2023). *Economic Watch: Made-in-China Stuffed Bears Spread Joy of Christmas Overseas*. Xinhua News Agency. https://eng.yidaiyilu.gov.cn/p/0M6INNO1.html

Yu, Z., Razzaq, A., Rehman, A., Shah, A., Jameel, K., & Mor, R. S. (2022). Disruption in Global Supply Chain and Socio-Economic Shocks: A Lesson from COVID-19 for Sustainable Production and Consumption. *Operations Management Research, 15,* 233-248. https://doi.org/10.1007/s12063-021-00179-y

Zhang, J., & Wei, J. (2025). The Operating Mechanism, Constraints and Practical Paths of the First-Release Economy in Promoting High-Quality Development: From the Perspective of Deepening the Development of New Productivity. *Front of Thought and Theory, 4,* 79-88.

Zou, L (2025). *Proposal on Further Advancing the Upgrading and Development of the Plush Toy Industry. Ankang Municipal People's Congress*. https://rd.ankang.gov.cn/Content-2874919.html