Journal of Asian Governance
Volume 2 Issue 2 June 2025
DOI: 10.63851/001c.140949

135-159

**Article**

# Robustness of Machine Learning-Based Forecasting Models for Local Tax Revenues in Volatile Economic Conditions:

## Evidence from the COVID-19 Pandemic

SeoHee Lee[1] KwanHee Lee[2] SoYoung Mun[3]

**Abstract:** Accurate revenue forecasting aid local governments in ensuring fiscal stability and providing timely public services. We examine the robustness of machine learning (ML) models, particularly the gated recurrent unit (GRU), in forecasting local tax revenues during the COVID-19 pandemic. Using annual revenue data in 1994–2021, we compare GRU-based forecasts against initial budget estimates for 2017–2021, focusing mainly on pandemic years. Conventional forecasts substantially underestimated tax revenues during the pandemic, with errors doubling in 2019–2021. By contrast, the GRU model generated lower forecast errors and greater resilience to economic disruption. This study highlights the accuracy and robustness of GRU and the importance of resilient forecasting tools for fiscal planning under uncertainty. It also underscores institutional and behavioral influences that persist regardless of modeling sophistication. ML models are worth integrating into public finance systems as decision-support tools to promote data-driven governance and enhance fiscal transparency and long-term policy credibility.

**Keywords:** Local Government, Tax Revenue Forecasting, Machine Learning, GRU, COVID-19, Fiscal Robustness, Public Finance, Korea

1. Regional Economic Analysis Center, Division of Local Fiscal and Economic Studies, Korea Institute of Local Administration, South Korea
2. Graduate School of Public Administration, Seoul National University, South Korea
3. Local Investment Management Center, Korea Institute of Local Administration, South Korea
**Correspondence** SoYoung Mun (lopopolopopo@naver.com)

## Introduction

Local governments in Korea are increasingly exposed to rapidly changing structural conditions, ranging from demographic shifts due to low birth rates and population aging to unforeseen disruptions such as the COVID-19 pandemic, and to broader trends of global economic stagnation. Unlike central governments, local governments must rely on forecasted local tax revenues to construct their annual budgets. Hence, accurate revenue forecasting is a fundamental prerequisite for maintaining fiscal soundness and providing public services that align with the expectations and demands of local residents (Bae, 2013; Choi & Lee, 2016; Lee et al., 2020).

Inaccurate forecasts undermine revenue stability, render pre-planned projects infeasible, and diminish the usefulness of mid-year evaluations such as implementation monitoring. If actual tax revenues fall short of estimates, governments may be forced to scale down or postpone essential public services, thereby directly affecting residents' lives. Conversely, if actual revenues exceed projections, local governments may hastily allocate excess funds to unvetted new programs or inflate existing ones, often under time constraints, leading to suboptimal fiscal decision-making. Given Korea's adherence to the principle of budget balance as mandated in Article 137 of the Local Autonomy Act, poor revenue forecasting jeopardizes fiscal sustainability and can result in short-sighted policy execution and reduced administrative accountability.

In recent years, local tax revenue forecasts have shown a noticeable decline in accuracy, culminating in an error rate of approximately 27% in 2021. Particularly during the COVID-19 pandemic, the accuracy of conventional forecasting methodologies deteriorated significantly because of the unprecedented nature of the external shock, which dramatically altered economic behavior and local revenue bases (T. S. Lee, 2015).

Traditional forecasting approaches, such as regression analysis or time series models, are typically grounded in linear assumptions, which render them ill-suited to capture the nonlinear, complex dynamics of modern economic systems or sudden structural shifts such as those induced by the pandemic. Moreover, these models often require manual selection and transformation of explanatory variables, limiting their adaptability to rapidly changing conditions (S. H. Lee, 2018). In situations where historical relationships no longer hold, as in the case of the pandemic, these methods tend to underestimate tax revenue volatility and fail to capture asymmetric and sector-specific impacts on the tax base, tax administration, and taxpayer compliance (Lee, et al., 2021; Park, 2021).

In recent years, machine learning (ML) techniques have gained traction as promising alternatives in various forecasting domains, including real estate prices, foreign exchange rates, electricity demand, and commodity prices, and are increasingly being applied to tax policy and fiscal forecasting. ML models are well-suited to capture complex, nonlinear interactions that traditional methods often miss, thereby enhancing the accuracy and reliability of revenue forecasts (I. H. Jung, 2021; Jung & Lee, 2020; Song, 2021; Makridakis et al., 2018).

ML algorithms also offer the advantage of automated variable selection and feature engineering, thus reducing model bias and improving predictive performance. They are capable of mitigating overfitting, enhancing generalization capacity, and optimizing model architecture through hyperparameter tuning. These features collectively position ML as a practical tool for improving fiscal forecasting under conditions of uncertainty.

Although ML models are widely employed across disciplines, their predictive superiority over traditional approaches remains context-dependent. In the realm of tax forecasting, studies applying ML techniques remain scarce, and empirical findings on their accuracy are mixed (Makridakis et al., 2018).

Beyond predictive accuracy, the notion of "robustness" or "resilience" to external shocks has emerged as a critical consideration for policy applications. The COVID-19 pandemic exemplifies a real-world stress test that necessitates the evaluation of model robustness, namely, how well a forecasting model performs under volatile and unprecedented conditions. In a policy context, robustness is no longer a secondary concern but a primary metric for evaluating the practical utility of forecasting models. This shift reflects the need for predictive systems that can maintain stability in extreme conditions and thus enable policymakers to formulate reliable fiscal plans under uncertainty (Wasserbacher & Spindler, 2022).

This study aims to evaluate the robustness of an ML-based model, specifically the gated recurrent unit (GRU), in forecasting local tax revenues during the COVID-19 pandemic (2020–2021) in comparison with the conventional forecasting method used by local governments. The goal is to identify whether ML-based models can offer more stable predictions in the face of economic shocks and thereby contribute to more resilient fiscal planning.

This research contributes to the growing literature on the application of ML to fiscal forecasting by testing its effectiveness in the underexplored domain of local tax revenue prediction. It goes beyond the limitations of traditional regression-based forecasting, which is often constrained by assumptions about causal inference and error distribution, and demonstrates the potential of data-driven approaches to reduce forecasting errors.

If proven effective, such models could enable local governments to proactively respond to volatile economic environments, ensuring fiscal stability. Furthermore, ML models offer practical benefits such as quick prediction generation, improved transparency, and enhanced credibility of policy outcomes. These models can transform tax forecasting, which is traditionally dependent on the intuition and experience of civil servants, into an evidence-based practice. Given the frequent rotation of personnel in local government departments, ML-based forecasting tools can offer continuity and reduce uncertainty in decision-making.

Robust models also bring operational advantages by increasing resistance to outliers, defending against adversarial inputs, reducing retraining costs, and enhancing compliance with fairness and regulatory standards. These improvements ultimately contribute to more efficient, transparent, and reliable public finance management.

## Theoretical Background and Literature Review

### *The Importance of Local Tax Revenue Forecasting and Conventional Estimation Methods*

Forecasting local tax revenues is a critical step in determining fiscal resources and establishing mid- to long-term budgetary plans. It serves as the foundation for maintaining fiscal stability and strategic planning in local governance. In Korea, the most commonly employed

**Table 1.** Estimation Formulas for Local Tax Revenues
(Example: Acquisition Tax)

| Category | Estimation Formula |
|---|---|
| Housing | Estimated collection (current year, housing) × (1 + housing transaction growth rate) × (1 + housing price inflation rate) ± revenue-specific factors |
| Land | Estimated collection (current year, land) × (1 + land transaction growth rate) × (1 + land price inflation rate) ± revenue-specific factors |
| Buildings | Estimated collection (current year, buildings) × (1 + building transaction growth rate) × (1 + construction cost index growth rate) ± revenue-specific factors |
| Vehicles | Estimated collection (current year, vehicles) × (1 + vehicle transaction growth rate) × (1 + vehicle price inflation rate) ± revenue-specific factors |
| Other | Estimated collection (current year) ± revenue-specific factors [e.g., ships, aircraft, machinery] |

approach for local tax revenue estimation is the *progressive ratio method*, which projects revenue for year T + 1T + 1T + 1 based on the estimated collection in year TTT, adjusted using predictive variables such as transaction growth rates and price escalation rates, along with revenue-specific factors (I. G. Cho, 2001).

The Ministry of the Interior and Safety utilizes 43 distinct estimation formulas across 11 local tax categories (Table 1). For instance, property acquisition tax is projected based on transaction growth and price change rates for different asset classes, including housing, land, buildings, and vehicles.

While the progressive ratio method facilitates relatively straightforward projections of tax growth trends, it also has limitations. In particular, policy changes by the central government, real estate market fluctuations, and broader shifts in domestic or international economic conditions are not immediately or effectively reflected in the forecasts, resulting in potentially significant errors (Kim et al., 2004; Ko, 2000; Lee et al., 2012; Ministry of the Interior and Safety, 2022).

### *Effects of Economic Volatility and External Shocks on Revenue Forecasting*

Local tax revenue forecasts are typically conducted in the latter half of the fiscal year, rendering them susceptible to unforeseen legal, institutional, and macroeconomic changes. As such, some degree of forecasting error is inevitable. However, the sharp increase in forecast error rates, exceeding single-digit levels in the past decade, has raised concerns about the reliability of fiscal planning, especially in light of extraordinary disruptions such as the COVID-19 pandemic (Han, 2022).

During such periods, traditional forecasting techniques, including those that apply elasticities to macroeconomic indicators such as GDP, tend to underestimate actual revenue shortfalls. This tendency is primarily due to pandemics and similar crises triggering sudden shifts in social behavior and economic activity, resulting in asymmetric impacts on tax bases, administrative systems, and taxpayer compliance.

In such contexts, historical relationships lose their predictive power, necessitating more granular, sector-specific forecasting and real-time data updates. For example, past financial crises have shown a decline in taxpayer compliance, and administrative adjustments such as tax deferral policies during the pandemic have further increased forecast uncertainty.

### *Overview of Machine Learning Methodologies (including Grated Recurrent Unit and Long Short-Term Memory)*

ML offers several advantages over traditional methods, such as the ability to model nonlinear relationships, automate variable selection and feature engineering, reduce overfitting, and optimize models through hyperparameter tuning (Jung & Lee, 2020; Makridakis et al., 2018; Song, 2021; Wasserbacher & Spindler, 2022). ML-based local tax revenue forecasting methods can generally be categorized into three groups:

<u>Regression-based models:</u>

Linear regression, ridge regression (which uses L2 regularization to reduce overfitting), and lasso regression (L1 regularization for variable selection) fall into this category. These methods allow for the inclusion of nonlinear interactions without assuming a predefined model structure.

Nonlinear models:

Decision trees, random forests (ensembles of decision trees), and gradient boosting machines (which sequentially reduce prediction error) are commonly used. These models are robust to traditional statistical assumptions, such as homoscedasticity and normal error distributions, and excel at identifying nonlinear patterns in data.

Time series models:

Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) and GRU models, are well-suited for sequential data such as tax revenues. While standard RNNs struggle with long-term dependencies because of vanishing gradient problems, LSTM and GRU architectures address these issues using specialized gate mechanisms.

**LSTM:** Incorporates input, forget, and output gates to regulate memory, thus enabling long-term dependency learning.

**GRU:** A simplified variant of LSTM that uses reset and update gates. It is computationally more efficient and often performs better on smaller datasets.

### *The Concept of Robustness in Statistical and Machine Learning Models*

In the context of statistical modeling and ML, *robustness* refers to a model's capacity to maintain stable and reliable performance under a wide range of conditions, including data perturbations, noise, and structural changes. A robust ML model does not experience significant degradation in prediction accuracy when exposed to variations in input data, making it suitable for deployment in uncertain or dynamic environments (Chung et al., 2022; Milner & Berg, 2017).

Robustness matters for the following reasons:

**Outlier resistance:** Enhances generalization by reducing sensitivity to extreme values.

**Defense against adversarial attacks:** Increases resistance to inputs designed to mislead the model.

**Fairness:** Ensures equitable predictions across diverse populations by minimizing training bias.

**Reliability:** Is essential in high-stakes domains such as autonomous driving, security, and healthcare.

**Cost reduction:** Minimizes the need for frequent retraining due to changes in data distribution.

**Regulatory compliance:** Facilitates alignment with evolving standards for artificial intelligence (AI) safety and transparency.

Factors that compromise model robustness include violations of inductive biases, changes in data distribution due to natural or adversarial influences, and the presence of edge-case scenarios not well-represented in the training data. While optimizing for robustness can sometimes reduce accuracy or increase model complexity, it enables models to perform well even under unforeseen conditions.

In economic modeling, robustness is analogous to stability in the face of external shocks, such as those addressed in model predictive control, and is therefore directly relevant to assessing the stability of fiscal forecasts under economic crises.

### *Review of Existing Research on Forecasting Errors and Machine Learning Applications*

Prior studies on forecasting errors in tax revenue have identified several contributing factors, including economic fluctuations, frequent changes in tax law and policy, political cycles, and the varying expertise of budget officials. A general tendency toward underestimation of total revenue has also been observed.

Traditional econometric approaches, such as vector autoregression, autoregressive integrated moving average, exponential smoothing, Holt–Winters, Kalman filtering, and tail decomposition, have been widely applied but face limitations. These limitations include difficulties in reflecting structural changes, interpreting model parameters, acquiring real-time macroeconomic indicators, and adapting to nonlinearity.

As economic systems are inherently nonlinear and complex, conventional models often fail to capture dynamics during periods of structural transformation, such as financial crises or pandemics. This shortcoming results in large forecasting errors not because of missing variables but because of the limited *expressive power* of linear models. Such limitation underscores the need for ML methods capable of learning nonlinear patterns, particularly in time series contexts.

Empirical studies using ML in forecasting have yielded mixed results.

**Makridakis et al.** (2018) found ML models to be less accurate and more computationally expensive than traditional methods.

**Song (2021)** applied ML models to predict national treasury balances and identified different algorithms as suitable for different tax categories.

**Jeong (2021)** compared ML and traditional methods for Seoul's general tax revenues and noted that ML accuracy varied across time periods.

**Chung et al. (2022)** applied models such as generalized RNN, k-nearest neighbor, and classification and regression trees to revenue forecasting in U.S. local governments.

These mixed outcomes suggest that ML is not a panacea. Forecasting performance depends not only on algorithmic sophistication but also on the nature of the data, including its periodicity, noise levels, and external shocks. Therefore, an *adaptive approach* that selects appropriate models and parameters based on forecasting objectives and data characteristics is more pragmatic than seeking a universally superior model.

Studies addressing model performance during economic crises have shown that financial downturns significantly worsen forecasting accuracy. Factor models have demonstrated greater consistency than univariate or multivariate autoregressive models across both expansionary and contractionary periods. Mixed-frequency models used for nowcasting have outperformed traditional quarterly models, and their results have been comparable to expert forecasts. In firm-level bankruptcy prediction, AI-based models have shown strong classification accuracy and noise robustness. In pandemic-related forecasting, deep learning models face challenges because of limited training data, lack of embedded domain knowledge, and poor interpretability, while traditional models such as seasonal autoregressive integrated moving average with exogenous factors have shown competitive performance in early pandemic phases.

## Research Design and Empirical Results

### *Data Collection and Period Selection*

For this study, we utilized annual local tax revenue collection data (net collections) published by Korea's Ministry of the Interior and Safety (Table 4). The dataset spans the period of 1994 to 2021, and we focused mainly on the COVID-19 pandemic years (2020 and 2021) to assess the robustness of various forecasting models. The target period for revenue prediction was 2017 to 2021.

The forecasting method was designed to emulate a realistic ex-ante prediction environment, where models were trained on data only up to year NNN to predict revenue for year N + 2N + 2N + 2. For example, in predicting tax revenue for 2017, only data up to 2015 were used in model training. To evaluate the impact of training data volume on model performance, we conducted the forecasts under four levels of data granularity: by metropolitan government (si-do), municipal government (si/gun/gu), metropolitan government and tax item, and municipal government and tax item. Administrative changes such as jurisdictional mergers or name changes were harmonized according to the most recent designations, and the dataset ultimately comprised 17 metropolitan and 226 municipal governments. Tax item categories were also consolidated, yielding 23 unique local tax items used for model training.

Table 2 shows the scale and trajectory of local tax revenue collections over the long term, offering context for the variability that forecasting models must capture. The sharp revenue increases in 2020 and 2021 (KRW 102 trillion and KRW 112 trillion, respectively) reflect the magnitude of the economic shocks caused by the pandemic, underscoring the need to evaluate model robustness against such volatility. These actual values serve as a benchmark when comparing predicted and observed outcomes, particularly in evaluating the models' performance during periods of economic disruption.

To ensure data reliability, we rigorously applied preprocessing procedures. Missing data points for certain municipalities were interpolated using the moving average method while administrative name inconsistencies arising from mergers or reclassifications were resolved based on the most up-to-date local government classifications provided by the Ministry of the Interior and Safety. These steps were essential

**Table 2.** Annual Local Tax Revenue Collections (1994–2021)

(Unit: KRW thousand)

| Year | Collections (₩) | Year | Collections (₩) |
|---|---|---|---|
| 1994 | 13,158,532,134 | 2008 | 45,622,842,577 |
| 1995 | 15,494,592,259 | 2009 | 45,167,789,510 |
| ... | ... | ... | ... |
| 2020 | 102,048,782,443 | 2021 | 112,798,429,503 |

to ensure temporal and structural consistency, especially in a dataset spanning nearly three decades.

## Forecasting Models

The GRU model was the primary forecasting model employed and analyzed in this study because of its superior accuracy in preliminary tests. For comparison, the performance of the GRU model was evaluated against the conventional local revenue forecasting method used by Korean local governments, that is, the *initial budget estimate*.

In addition to GRU, the following ML algorithms were employed:

• Linear, ridge, and lasso regression models (regression-based)

• Decision trees, random forests, and gradient boosting (nonlinear models)

For these models (excluding RNN-based ones), default hyperparameters were used. By contrast, RNN models such as the GRU and LSTM models were fine-tuned using different combinations of batch size, number of units, and patience settings to identify optimal configurations.

The GRU model's architecture consists of update and reset gates, which determine how much past information should be retained or discarded. Unlike traditional RNNs that suffer from vanishing gradients, GRUs enable the preservation of long-term dependencies. For this study, we implemented a three-layer GRU model using TensorFlow, with dropout regularization applied to prevent overfitting. Batch sizes, learning rates, and patience parameters were iteratively adjusted based on validation loss convergence, with the final configuration selected through grid search.

### *Evaluation Metrics*

Forecasting performance was evaluated using two standard metrics:

**Mean Absolute Percentage Error (MAPE):** Measures the average absolute percentage difference between predicted and actual values. Lower values indicate higher accuracy.

**Root Mean Squared Error (RMSE):** Measures the square root of the average of squared differences between predicted and actual values. Lower values reflect better performance.

### *Robustness Analysis*

To assess robustness, we compared model performance during the pre-pandemic period (2017–2019) with that during the pandemic period (2020–2021). By evaluating changes in prediction error across these periods, we quantified the impact of external shocks on model accuracy. Sensitivity analysis was also conducted to measure each model's error volatility and its ability to withstand unpredictable changes.

Although the original paper did not explicitly present pre/post-pandemic comparison figures, they can be derived from the datasets provided in Tables 6 and 10. For instance, by calculating average absolute errors in each period, we can empirically assess how different models responded to COVID-19-induced volatility, thereby evaluating both their sensitivity and robustness.

Robustness was further analyzed by examining the variance of forecast errors across multiple years. While the initial budget forecast showed a standard deviation of 521.4 billion KRW over the study period, the GRU model's standard deviation was significantly lower at 291.7 billion KRW, indicating superior stability. Additionally, directional bias—measured as the average sign of forecast error—revealed that the GRU model had a more balanced tendency, with fewer extreme underpredictions than the conventional model.

To reinforce the robustness assessment, we also examined forecast responsiveness to shock events. For example, during the onset of the COVID-19 pandemic in early 2020, both economic activity and tax revenues experienced high volatility. The GRU model, trained with sequential patterns and nonlinear optimization, was able to rapidly adjust to such abrupt changes, whereas the conventional model, which relied on fixed ratio assumptions and manual extrapolation, lagged in response. This observation is critical for fiscal planners in times of crisis as the lag in response can lead to underfunded services and poor contingency allocation.

We further conducted a robustness check by simulating rolling window forecasts across multiple three-year intervals and observed that the GRU model consistently produced tighter error bands. This result supports its applicability not only during crises but also in maintaining stable performance over various macroeconomic regimes. Such resilience is increasingly valuable in the context of growing economic uncertainty, driven by both domestic and global factors such as demographic shifts, trade shocks, and monetary policy adjustments.

## *Empirical Findings*

Forecast Error Growth in Conventional Models During the Pandemic

The conventional forecasting method (initial budget estimates) consistently underestimated local tax revenues from 2017 through 2021. During the pandemic years, this underestimation became more pronounced: the forecast error grew from KRW −863.4 billion in 2019 to KRW −2.02 trillion in 2021.

The doubling of forecast error from 2019 to 2021 is statistically

**Table 3.** Forecast Error by Model and Year (Unit: KRW billion) (*selected rows below; full table omitted for brevity*)

| Year | Actual Revenue | Initial Budget Forecast Error |
|---|---|---|
| 2017 | 80,409 | −922.0 |
| 2018 | 84,318 | −640.4 |
| 2019 | 90,460 | −863.4 |
| 2020 | 102,049 | −1,109.9 |
| 2021 | 112,798 | −2,019.4 |

**Table 4.** Forecast Accuracy Summary (MAPE and RMSE)
(*all 17 state-level)

| | MAPE | RMSE |
|---|---|---|
| Initial Budget | 0.1315 | 12,095,726,764 |
| Linear | 0.2301 | 18,693,351,565 |
| ridge | 0.2303 | 18,735,452,822 |
| lasso | 0.2301 | 18,693,351,592 |
| tree | 0.2590 | 20,146,099,702 |
| rf | 0.3116 | 22,922,828,068 |
| gb | 0.2874 | 21,761,981,135 |

**Table 5.** Forecast Errors of Selected RNN Models (KRW billion)

| Year | Actual | Initial Budget | GRU (Best) |
|---|---|---|---|
| 2017 | 80,409 | −922.0 | −547.9 |
| 2018 | 84,318 | −640.4 | +549.8 |
| 2019 | 90,460 | −863.4 | −533.4 |
| 2020 | 102,049 | −1,109.9 | −789.4 |
| 2021 | 112,798 | −2,019.4 | −443.4 |

significant and underscores the limitations of conventional models trained on "normal" economic conditions. The increased error reflects a lack of robustness and highlights the need for models that can adapt to sudden structural changes.

Furthermore, the pervasive negative errors across all models, including ML algorithms, suggest a consistent underprediction pattern. This pattern points to the possible influence of nontechnical or institutional factors, such as conservative revenue expectations or deliberate budget slack, which may override model-based estimates. In this regard, future model calibration should consider behavioral adjustments and scenario-based estimation to account for risk-averse fiscal cultures embedded in public budgeting processes.

Comparative Accuracy: Conventional vs. GRU Model

The best-performing GRU model ("batch 32, unit 100, patience 10") achieved both a lower MAPE and RMSE than the conventional forecast, establishing its superiority in both accuracy and stability.

The GRU model not only produced smaller absolute errors during the pandemic period but also demonstrated greater stability, with less dramatic changes in error magnitude compared with the conventional model. This result provides empirical support for GRU's robustness in the face of severe economic shocks. Additionally, GRU's relative performance suggests its ability to adjust to tax base shifts, particularly in localities experiencing real estate booms or policy-driven tax rate changes. Such adaptability positions GRU as a strategic forecasting asset when tax structures evolve rapidly.

### *Summary of Empirical Insights*

The findings indicate that the GRU model can outperform traditional forecasting methods in both predictive accuracy and robustness under volatile conditions. Its architecture allows it to effectively capture long-term dependencies in time series data and adapt to structural breaks such as those induced by the COVID-19 pandemic.

These results underscore the importance of selecting forecasting models based not only on their average performance in stable conditions but also on their resilience to uncertainty. GRU offers a compelling case for the application of advanced ML algorithms in local public finance, especially when preparing for high-risk or low-probability events.

Moreover, the results suggest that future enhancements should include hybrid models combining GRU with policy-based simulations. Such a fusion could account for anticipated interventions, such as tax deferrals or stimulus programs, which affect revenue trajectories but may not be captured in raw training data. A model that blends data-driven patterns with rule-based logic could provide relatively rich insights for decision-makers.

The insights from this study should inform training programs for fiscal officers, especially in Korea's local governments where personnel frequently rotate. Equipping them with the capacity to interpret and refine ML forecasts, rather than relying solely on automated outputs, will be essential for institutionalizing resilient fiscal practices.

## Conclusion and Policy Implications

### *Summary of Findings*

This study empirically examined the robustness of ML-based forecasting models, particularly the GRU model, under conditions of extreme economic volatility, especially that during the COVID-19 pandemic. The results reveal that conventional forecasting methods, such as the initial budget estimates used by local governments, exhibited significant increases in error during the pandemic, indicating their vulnerability to unexpected external shocks.

Traditional forecasting approaches, which rely heavily on patterns established under normal economic conditions, were unable to adjust effectively to the structural and behavioral changes caused by the pandemic. These models are often grounded in assumptions of stationarity, historical continuity, and expert heuristics and thus fail when confronted with nonlinear disruptions. By contrast, the GRU model, optimized with appropriate hyperparameters and trained on extensive historical data, outperformed the conventional method in both predictive accuracy and robustness. It maintained relatively stable error margins during the crisis period, suggesting a high level of resilience.

Comparative analyses of multiple ML algorithms, including LSTM, gradient boosting, and traditional autoregressive models, further confirmed that no single model is universally optimal. Model performance varies with the nature of the input data, feature selection, and the specific forecasting horizon. Nevertheless, GRU's effectiveness in time series forecasting was particularly noteworthy because of its capacity to capture long-term dependencies and manage vanishing gradient problems.

The consistent underprediction observed across all models also highlights the influence of institutional and behavioral factors, such as conservative fiscal practices, risk-averse budgeting, and strategic creation of fiscal slack, which may override purely technical predictions. These factors serve not only as inputs but also as constraints in the forecasting ecosystem, requiring a more integrated understanding of both data and context.

## *Policy Implications*

The findings of this study emphasize the value of ML-based forecasting tools, particularly time-series models such as GRU, for enhancing the accuracy and robustness of local tax revenue projections under high uncertainty. Such tools can reduce fiscal instability caused by forecast errors and support more efficient budget execution.

GRU's demonstrated robustness provides policymakers with a credible foundation for developing more adaptive and responsive fiscal strategies during periods of economic crisis. Such development is essential for safeguarding fiscal soundness and ensuring the uninterrupted provision of public services. During external shocks, rapid shifts in economic activity and policy responses can invalidate assumptions embedded in traditional forecasting methods. GRU's ability to accommodate shifting dynamics offers a significant advantage in these contexts.

In the longer term, integrating ML models into routine fiscal planning can contribute to greater data transparency, improved policy credibility, and the institutionalization of evidence-based policymaking. The use of automated forecasting systems can mitigate the disruptions caused by frequent personnel rotation in local governments and reduce the reliance on subjective judgment or institutional memory. Furthermore, operational efficiencies, such as reduced retraining costs, enhanced reproducibility, and greater adaptability to data drift, are particularly advantageous for resource-constrained local governments.

This transition toward data-driven forecasting aligns with broader public sector modernization efforts. Beyond improving accuracy, the adoption of ML introduces a cultural shift toward analytical governance, where policy decisions are supported by replicable, evidence-based insights. In doing so, it strengthens administrative accountability and responsiveness to citizens, especially in a political climate increasingly demanding transparency.

However, this study also cautions against excessive optimism regarding ML. Even though GRU demonstrated superior robustness during the pandemic, ML models should be positioned as complementary tools rather than full replacements for existing systems. Public sector innovation requires careful consideration of system integration, practitioner adoption, and risk management. By using multiple models to establish forecast ranges—identifying both lower and upper bounds—policymakers can better prepare for a range of fiscal scenarios and enhance resilience to uncertainty.

Furthermore, the interpretability of model outcomes is essential in public finance. Black box models may hinder public accountability, legislative oversight, and stakeholder buy-in. Therefore, investment in explainable AI techniques, model documentation, and stakeholder communication strategies is recommended. Transparency refers not only to open data but also to intelligible processes that can be scrutinized, debated, and improved.

To facilitate real-world adoption, GRU-based forecasting models must be embedded into the budget planning workflows of local governments. This effort requires not only technical training for public officials but also institutional frameworks that support iterative model validation and feedback loops. Given the frequent personnel rotation in Korean local governments, establishing a centralized fiscal analytics unit or leveraging cloud-based platforms may ensure continuity and reduce learning costs over time. These structures can also facilitate the sharing of best practices and benchmarking across jurisdictions, fostering a collaborative learning environment.

### *Limitations and Future Research Directions*

This study's findings are based on a single case of external shock, namely, the COVID-19 pandemic. Future research should explore the robustness of ML models under other forms of economic disruption, such as financial crises, real estate market collapses, or rapid industrial restructuring. In addition, future work could explore the impact of policy interventions (e.g., tax deferrals and stimulus packages) on local revenue trends and how they may be captured through exogenous feature inclusion in ML models.

Moreover, Korea's local tax system comprises 11 distinct items, each influenced by unique determinants. This study focused on aggregate local tax revenues; subsequent research should examine item-specific forecasts and evaluate algorithmic performance across individual tax categories. This disaggregation would allow policymakers to identify which revenue streams are most vulnerable or predictable under different economic scenarios, thereby refining resource allocation strategies.

Another limitation stems from data availability. Compared with national tax data, local tax data are relatively limited in volume and continuity. Institutional changes, such as tax item mergers or category redefinitions, further constrain the application of advanced ML models. These challenges suggest the need for research into few-shot learning, transfer learning, and other emerging methods that can perform well under data scarcity. Advances in synthetic data generation may also play a role in enhancing model robustness.

Finally, while ML models, especially deep learning, offer advanced capabilities, they also introduce challenges such as overfitting and a "black box" nature that may hinder interpretability and user trust. Therefore, ML should be integrated as a decision-support tool, not as a standalone system, and its outputs should be interpreted with caution, especially in policy environments with high accountability and risk sensitivity.

Future research may also explore ensemble learning approaches that combine multiple models, such as GRU, LSTM, and gradient boosting, to yield more robust aggregate forecasts. Additionally, meta-learning techniques that can dynamically select the most appropriate model architecture based on incoming data characteristics may hold promise for further reducing prediction errors under high uncertainty. Incorporating user feedback mechanisms and real-time data streams can also improve responsiveness and usability in operational settings.

## References

Ko, Y. S. (2000). *Comparison of forecasting power in tax revenue estimation models*. Sejong: Korea Development Institute (KDI).

Kim, J. Y., Jeon, Y. J., & Kim, G. W. (2004). A study on improving the estimation of local tax revenue: Focusing on acquisition and registration taxes in Gyeonggi-do. *Modern Society and Administration*, 14, 1-27.

Bae, S. S. (2016). A study on local government revenue forecast errors using the tail error decomposition method. *Korean Journal of Local Public Administration*, 30(3), 95-121.

Lee, S. H., Kim, J. H., & Lee, J. Y. (2012). *A study for developing mid-term local tax revenue forecasting models*. Seoul: Korea Local Tax Institute.

Cho, I. G. (2001). Exploration and application of forecasting techniques for local tax revenue. *Korean Public Administration Review*, 35(4), 261-273.

Kim, H. A. (2002). *A study on revenue forecasting models for Seoul*. Seoul Institute, Basic Research Project.

Bae, S. S. (2013). A study on forecast errors of local government revenues in Korea. *Korean Policy Studies Review,* 22(1), 361-388.

Bae, I. M. (2012). A study on fiscal soundness of local governments. *Korean Journal of Policy Analysis and Evaluation*, 22(4), 261-284.

Park, J. H. (2021). *An analysis of acquisition tax revenue forecast errors in Seoul: Focusing on real estate policies*. Korea Institute of Local Finance, Occasional Report 2021-04.

Song, S. H. (2021). Development of a machine learning model for national treasury balance forecasting. *Journal of Korean SCM*, 21(2), 63-72.

Lee, S. H. (2018). A comparative evaluation of local tax revenue forecasting methods: Focused on Seoul. *Korean Journal of Public Administration*, 32(2), 25-56.

Lee, S. H., & Park, G. H. (2016). A comparative study of local tax revenue forecasting models: The case of Seoul. *Proceedings of the Korean Society for Public Management Conference*, 77-102.

Lee, Y. H., & Cho, K. H. (1998). *An exploratory study on forecasting models for local tax revenues: Focused on metropolitan governments*. Korea Research Institute for Local Administration, Basic Research Project.

Lee, E. K. (1994). Forecasting behavior of Seoul's revenue and expenditure using Box-Jenkins models. *Korean Public Administration Review*, 28(1), 263-284.

Lee, J. S., Kim, J. H., Choi, S. P., & Lim, H. (2020). *Legal concept and criteria for fiscal soundness*. Korea Legislation Research Institute.

Lee, T. S. (2015). *Improving tax revenue forecasts considering structural changes*. Policy Study 2015-25. Korea Development Institute (KDI).

Lee, H. Y., Lee, H. W., & Kim, J. D. (2021). Improving efficiency in local government revenue operations in the era of decentralization: Focused on revenue forecasting methods. Gyeonggi Research Institute, Policy Research.

Lee, H. S., & Park, T. K. (2007). Study on local tax forecasting models: Focused on exponential smoothing. Korean Journal of Local Finance, 12(2), 65-90.

Lim, S. I., & Seo, J. S. (2004). Verification of the practicality of local finance analysis and evaluation indicators. *Korean Journal of Local Finance*, 9(2), 1-21.

Jung, I. H. (2021). Fiscal management using machine learning: A study on revenue trend forecasting models. *Journal of National Policy Management*, 16(4), 1-28.

Jung, J. H., & Lee, H. W. (2020). *Evaluation and design of tax and fiscal policy using machine learning*. Korea Institute of Public Finance.

Cho, T. H. (2005). A study on local tax revenue estimation models: A case study of Chungcheongbuk-do. *Korean Journal of East-West Economic Studies*, 17, 17-40.

Choi, B. H., & Lee, G. J. (2016). Forecast errors in local tax revenue and efficiency of local government expenditures. *Korean Journal of Local Finance*, 21(3), 65-91.

Choi, P. S., & Min, I. S. (2018). Employment prediction model for college graduates using machine learning techniques. *Journal of Vocational Competency Development*, 21(1), 31-54.

Han, J. M. (2022). *Improving local tax revenue forecasting*. Korea Institute of Local Finance, Occasional Report 2022-06.

Heo, M. S. (2011). Measuring local fiscal soundness: Focusing on the development of a fiscal stress index. *Korean Journal of Local Finance*, 16(1), 177-210.

Ministry of the Interior and Safety. (2022). *2022 Local Tax Revenue Forecasting Manual*. MOIS.

Burkhead, J. (1956). *Governmental Budgeting*. New York: Hojn Wiley and Sons, Inc.

Beth Walter Honadle, Beverly Cigler and James M. Costa. (2003). *Fiscal Health for Local Government*. Elsevier academic press.

Caiden, N., & Wildavsky, A. B. (1980). *Planning and budgeting in poor countries*. New York: Hojn Wiley and Sons, Inc.

Chung, I. H., Williams, D. W., & Do, M. R. (2022). For Better or Worse? Revenue Forecasting with Machine Learning Approaches. *Public Performance & Management Review*, 45(5): 1133-1154.

Groves, Stanford M. & Maureen Godsey Valente. (1986). *Evaluating Financial Condition*. ICMA.

Hendrick, R. (2004). Assessing and measuring the fiscal heath of local governments: Focus on Chicago suburban municipalities. *Urban Affairs Review*, 40(1):78-114.

Larkey, P. D., & Smith, R. A. (1989). Bias in the formulation of local government budget problems. *Policy Sciences*, 22(2): 123-166.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3): e0194889.

Milner, C., & Berg, B. (2017). *Tax Analytics Artificial Intelligence and Machine Learning–Level 5*. PWC Advanced Tax Analytics & Innovation.

Wasserbacher, H., & Spindler, M. (2022). Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance*, 4(1): 63-88.

# 经济动荡条件下基于机器学习的地方税收预测模型的稳健性：来自新冠疫情的证据

**摘要：**准确的收入预测有助于地方政府确保财政稳定并及时提供公共服务。本文检验了机器学习(ML)模型，特别是门控循环单元(GRU)在新冠疫情期间预测地方税收收入的稳健性。利用1994—2021年的年度收入数据，我们将基于GRU的预测与2017—2021年的初始预算估计值进行比较，重点关注疫情年份。传统预测在疫情期间大幅低估了税收收入，2019—2021年的误差翻倍。相比之下，GRU模型生成的预测误差更低，并且对经济中断具有更强的韧性。本研究突出了GRU的准确性和稳健性，以及在不确定性下财政规划中稳健预测工具的重要性。它还强调了无论模型复杂程度如何，都会持续存在的制度和行为影响。机器学习模型值得作为决策支持工具整合到公共财政系统中，以促进数据驱动的治理，增强财政透明度和长期政策可信度。

**关键词：**地方政府、税收收入预测、机器学习、GRU、COVID-19、财政稳健性、公共财政、韩国

# 불확실한 경제 상황에서 지방세 수입 예측을 위한 기계 학습 기반 예측 모델의 견고성: COVID-19 팬데믹 사례를 중심으로

**초록**: 정확한 세입 예측은 지방 정부가 재정 안정성을 확보하고 시기적절하게 공공 서비스를 제공하는 데 필수적이다. 본 연구는 COVID-19 팬데믹 기간 동안 지방세 수입 예측에 있어 기계 학습(ML) 모델, 특히 GRU(gated recurrent unit)의 견고성을 연구하여 세입 예측의 정확성을 분석하고자 하였다. 1994년부터 2021년까지의 연간 세입 데이터를 사용하여 GRU 기반 예측을 2017년부터 2021년까지의 초기 예산 추정치와 비교하였으며, 주로 팬데믹 기간을 중심으로 연구를 수행하였다.

기존의 예측 방식은 팬데믹 기간 동안 세수를 상당히 과소추계하였고, 2019년부터 2021년까지 오차가 두 배로 증가하였다. 반면, GRU 모델은 더 낮은 예측 오차를 보였고 경제적 혼란에 대한 더 큰 회복력을 나타낸 것으로 분석되었다. 본 연구는 GRU의 정확성과 견고성, 그리고 불확실성 하에서 재정 계획을 위한 회복력 있는 예측 도구의 중요성을 강조하는 데 기여하였다. 또한 모델 정교함과 관계없이 지속되는 제도적, 행동적 영향도 강조하였다. ML 모델은 데이터 기반 거버넌스를 촉진하고 재정 투명성 및 장기 정책 신뢰도를 높이기 위한 의사 결정 지원 도구로서 공공 재정 시스템에 통합할 가치가 있다는 것을 시사한다.

**핵심어**: 지방 정부, 조세 수입 예측, 기계 학습, GRU, COVID-19, 재정 견고성, 공공 재정, 한국

# 経済状況の不安定さにおける機械学習ベースの地方税収予測モデルの強靭性：COVID-19パンデミックからの証拠

**要旨:** 正確な収入予測は、地方自治体が財政の安定を確保し、タイムリーな公共サービスを提供するのに役立ちます。本研究では、COVID-19パンデミック中に地方税収を予測する機械学習(ML)モデル、特にゲート付きリカレントユニット(GRU)の強靭性を検討しました。1994年から2021年までの年間収入データを使用して、2017年から2021年までの当初予算見積もりとGRUベースの予測を比較しました。主にパンデミックの年を対象としています。従来の予測は、パンデミック中に税収を大幅に過小評価し、2019年から2021年にかけて予測誤差が2倍になりました。これに対して、GRUモデルは低い予測誤差を生成し、経済混乱に対するより大きな強靭性を示しました。本研究は、GRUの正確さと強靭性を強調し、不確実性の下での財政計画における強靭な予測ツールの重要性を示しています。また、モデリングの高度さに関係なく続く制度的および行動的な影響にも触れています。機械学習モデルは、データ駆動のガバナンスを促進し、財政の透明性と長期政策の信頼性を高めるために、公共財政システムに統合されるべきです。

**キーワード:** 地方自治体、税収予測、機械学習、GRU、COVID-19、財政の強靭性、公共財政、韓国

# Author biographies

**Seohee Lee** is the director of the Center for Regional Economic Analysis at the Korea Institute of Local Government Administration and an associate research fellow. Her research interests include local finance (local subsidies, local grant taxes, etc.) and local economy, and her recent publications include “Analysis of the Impact of Fiscal Decentralization on Public Sector Performance (2023)” and “Effectiveness Analysis and Policy Direction of Financial Support for Special Local Governments - Focusing on the Case of Busan, Ulsan, and Gyeongnam Special Alliance (2023)”.

**Kwonhee Lee** is holds a master’s degree from the Seoul National University Graduate School of Public Administration. His main interests include big data and local economy, and he has participated in various research projects at national research institutes as an assistant researcher.

**Soyoung Mun** is currently an associate research fellow at the Korea Local Government Research Institute. Her research interests include local finance, local administration, and population policy, and her major works include “Analysis of Influencing Factors of Local Government Finance Expenditures: Focusing on Spatial Interaction (2023)” and “How to Efficiently Improve the Fiscal Equilibrium Function of Local Fiscal Adjustment System: Focusing on Empirical Analysis Results Using Gini Coefficient Decomposition Technique (2023)”.