Does Money Help Forecast Nominal GDP in Deep Learning Models?

Kenean Yemane Kejela
Software Engineer, Google, Seattle, WA
kkejela@oberlin.edu

John V. Duca
Oberlin College, Dept. of Economics, 223 Rice Hall, Oberlin, OH 44074, jduca@oberlin.edu
Emeritus (Retired) Economist, Research Department, Federal Reserve Bank of Dallas

June 15, 2025

Abstract

This study analyzes whether adding one of three broad Divisia money measures (Divisia M3, M4, or M4-) constructed by the Center for Financial Stability (CFS) and a determinant of their long-run demand—stock mutual fund loads—improves deep learning model forecasts of nominal GDP. We train a long short-term memory (LSTM) neural network on data from 1985q1 to 2013q4, using 8 datasets: a baseline pair of 1,500 FRED variables and a version adding three COVID-19 variables, and three other pairs that add stock mutual fund loads plus one of the three broad Divisia money measures. Adding any one of the Divisia variables plus stock mutual fund loads significantly improved one-quarter- and four-quarter-ahead forecasts of nominal GDP, particularly in capturing long-run trends. Findings imply that broad Divisia money should be among variables used to monitor and forecast nominal GDP, with results slightly favoring adding the broadest measure, Divisia M4, for forecasting nominal GDP four quarters ahead.

Key Words: Forecasting, Deep Learning, LSTM, RNN, Nominal GDP

*The views expressed are those of the authors and are not necessarily those of the Federal Reserve Bank of Dallas or the Federal Reserve System or Google. Any errors are our own.

1. Introduction

Macroeconomists and policymakers have dismissed measures of money as macroeconomic indicators owing to instability in the demand for traditional, simple-sum money aggregates. Many economists overlooked the surge in money growth during the pandemic and were surprised by the rebound in aggregate demand and inflation. However, recent research finds that the demand for the three broad Divisia measures of monetary services (Divisia M3, M4-, and M4) constructed by the Center for Financial Stability (CFS) can be well modeled, and that broad Divisia money measures have valuable information about aggregate demand (nominal GDP) in recent years (Bordo and Duca, 2025). Other research finds that Divisa M3 can help explain recent rise and fall in inflation (Bordo, Duca, and Jones, forthcoming). Even before the COVID Recession, Keating et al. 2019 showed that Divisia measures are better indicators of output and price than the federal funds rate.

Such findings suggest that Divisia money indexes may capture key macroeconomic trends that would otherwise remain undetected. Specifically, Divisia aggregates are hypothesized to reflect trends emerging from the broader mechanisms and channels through which monetary and fiscal policies influence the economy—trends not captured by other variables. For instance, Bordo and Duca (2025) argue that the central bank's implementation of quantitative easing and other unconventional policy actions lead to shifts in reserves and money growth that manifest in Divisia M3, providing information about nominal GDP. This paper examines the role of money as an economic indicator, testing the hypothesis that incorporating Divisia data can provide marginal explanatory power for nominal GDP (NGDP) beyond that in other variables. Specifically, we investigate whether adding one of the three broad Divisia money measures produced by the CFS (M3, M4, or M4-) plus a main determinant of their long-run demand—stock mutual fund loads (hereafter, stock loads)— to a baseline FRED dataset enhances the accuracy of NGDP forecasts generated by a deep learning model. This study uses broad Divisia and stock load data jointly to better capture the long-run relationship between money and nominal GDP. Bordo and Duca (2025) and Bordo, et al. (forthcoming) show that controlling for stock loads significantly strengthens the links between broad Divisia money and nominal GDP over time. Because our objective is to extract as much information as possible from broad Divisia money to forecast NGDP, we include stock load data to increase the aggregate's explanatory power within our deep-learning model.

Recent research has highlighted machine-learning techniques, particularly Recurrent Neural Networks (RNNs), in economic time series forecasting (e.g., Stoneman and Duca, 2024). Our study trains a variant of RNN—specifically, a long short-term memory (LSTM) model—using four distinct datasets to forecast nominal GDP. The baseline dataset (A1) includes 1,500 FRED variables from 1985q1 to 2024q4., to which the second dataset (A1C) adds three COVID-related variables. We create three more pairs of models that add one of the three broad Divisia money measures plus their primary long-run money demand determinant, stock mutual fund loads, to each of the baseline datasets. We train the model to predict one-quarter and four-quarter-ahead NGDP across the eight datasets and find that adding a broad Divisia money measure and stock load data significantly improves trend fit and forecasting accuracy compared to baseline models omitting these variables. We also find that among the three broad Divisia monetary services measures, results slightly favor adding the broadest monetary measure, Divisia M4, for forecasting nominal GDP four quarters ahead. Consistent with recent research highlighting the relevance of Divisia money as an economic indicator (e.g., Bordo and Duca, 2025, Bordo et al., forthcoming, and Keating, et al., 2019), we find that Divisia money should be considered among other macroeconomic indicators.

2. Literature Review on Money, Output, and Prices

The use of money as an indicator and a policy target has evolved and sparked considerable debate. In the early 1990s, empirical research began to challenge the importance of monetary aggregates in explaining post-1980 output and price trends. Friedman and Kuttner (1992) and Estrella and Mishkin (1997) found that the traditional relationship between simple-sum money and key macroeconomic variables had weakened beginning in the mid-1980s. Such findings led policymakers to view money measures as unreliable indicators and/or policy targets.

Later in the late 1990s, several researchers challenged studies that maintained little to no link between money and the broader economy. Belongia (1996), in particular, argued that the measurement of simple-sum monetary aggregates used in those studies was flawed and that Divisia measures of monetary services, which Barnett (1980) had introduced earlier, was strongly linked to real output and prices, which Schunk (2001), also found. As Barnett (1980) has stressed, the primary problem with simple-sum monetary aggregates is that they are measured in ways that are not consistent with underlying consumer theory. Not surprisingly, their demand functions are unstable, which makes it difficult to identify strong and consistent relationships between these aggregates and macroeconomic variables. Much of the underlying problems arise because simple-sum aggregates treat all financial assets as equally liquid and equally effective in facilitating transactions, which is empirically rejected (see Jones et al. (2008, Fleissig and Jones (2015), Jadidzadeh and Serletis (2019), and Xu and Serletis (2022)).

Addressing this shortcoming, Barnett (1980) introduced Divisia monetary aggregates, which measure the flow of nonpecuniary money services across different monetary components. For each component, services are inferred from their user cost, which depends on the gap between a return on a non-liquid benchmark asset and the pecuniary rate earned on the component. The growth rate of an overall Divisia measure is calculated by weighing the growth rate of balances in each component by its share of overall liquidity services. This weighting method produces indexes that have a more stable link to nominal GDP by filtering out artificial volatility caused by funds shifting between asset categories. Starting with Barnett (1982) and more recently Haron (2024) and Darret et al., (2024), research has found that Divisia money measures have a stronger and more stable relationship with output and prices than simple-sum M2.

In addition, Bordo and Duca (2025) have found that broad Divisia monetary aggregates have a stronger relationship with nominal GDP than do simple-sum or narrow measures of Divisia money. They also find that Divisia M3 tends to have the closest relationship, consistent with Dery and Serletis (2021). Divisia M3 measures the liquidity services of currency, bank deposits, and household money market mutual funds included in M2 plus large time deposits and institutional money market funds. As such, it is the broadest measure of monetary services from currency and the monetary liabilities of financial intermediaries, the latter of which transform their asset holdings into money-like liabilities. To Divisia M3, Divisia M4 minus adds the liquidity services of commercial paper not held by banks or money funds, and Divisia M4 further adds the services of all Treasury bills held outside of the Federal Reserve, banks and money funds. These additions to Divisia M3 assume that assets that have not undergone liquidity transformation provide the same services as assets covered by Divisia M3. This assumption is challenged by Bordo, Duca, and Jones (forthcoming), who using a traditional, more structural econometric approach, show that Divisia M3 has a tight connection to nominal spending, unlike Divisia M4 minus or Divisia M4.

More recently, Bordo and Duca (2025) provide evidence of both short- and long-run stability in the demand for Divisia money, using a sample that extends through the COVID pandemic to the first quarter of 2023. They find that once one controls for stock mutual fund loads which affect the long-run demand for Divisia M3, that aggregate is closely linked to nominal GDP in the long-run. Their subsequent analysis reveals that growth in Divisia M3 signaled a rise in nominal aggregate demand, which contributed significantly to the inflation surge observed in 2021–2022. They argue that such signals might have been overlooked if monetary growth had been ignored. Consequently, they emphasize that disregarding information from broad Divisia money indexes could result in missing critical indicators of underlying economic conditions. Further strengthening the case for Divisia money measures as valuable economic indicators, Keating et al. (2019) find that Divisia M4 serves as a more effective monetary policy indicator than the federal funds rate.

This paper contributes to the existing literature in three ways. First, it extends previous findings on money as a macroeconomic indicator by testing whether including broad Divisia money and stock loads enhances the accuracy and overall fit of a deep learning model. Second, in contrast to most prior studies that use traditional methods (e.g., Vector Auto-regressions, VARs) to examine the links between money and macroeconomic variables, this paper employs deep learning techniques that better capture complex, non-linear relationships that traditional approaches may overlook. Third, in contrast to the more structural and traditional econometric approaches used by Bordo and Duca (2025) and Bordo, Duca, and Jones (forthcoming) which favor Divisia M3 over Divisia M4- and Divisia M4, our LSTM approach finds that the broadest of the broad Divisia money measures, Divisia M4, provides slightly more information than either Divisia M3 or Divisia M4-, for forecasting nominal GDP four quarters ahead..

Recent research highlights the difficulty of capturing economic trends during the COVID-19 Recession, as the pandemic was a non-economic shock with pervasive economic effects. The sudden stop in economic activity in 2020q2 is widely regarded as an outlier event, not reflective of typical behavior. Ng (2021) shows that post-COVID observations often yield uninterpretable estimates, significantly distorting the pre-COVID model fits. She attributes this instability to the introduction of large, unforeseen non-economic shocks in the post-COVID period. Bordo and Duca (2025) emphasize the need to account for the dynamic effects of COVID-19 when modeling the demand for Divisia monetary aggregates. Accordingly, we anticipate some level of difficulty in modeling economic patterns in the post-COVID period in our deep learning model as well.

3. Machine Learning, Deep Learning & Economic Forecasting

Before delving into the specifics of the deep learning model used in this paper, this section provides a brief overview of deep learning and its applications in macroeconomic forecasting.

Machine Learning (ML) encompasses a range of algorithms and models that learn patterns from data to make informed decisions or predictions. Among the most advanced ML techniques are Artificial Neural Networks (ANNs), which are inspired by the structure and function of the human brain and are designed to capture complex, non-linear relationships in data (Stoneman and Duca 2024). Deep learning is a powerful subset of ANNs that has been successful in areas such as image and speech recognition, natural language processing (NLP), and complex time series forecasting.

Within deep learning, Recurrent Neural Networks (RNNs) stand out for their ability to process sequential data through built-in memory mechanisms. These networks retain information from previous inputs to influence future outputs, making them particularly effective at learning temporal dependencies. As a result, RNNs have become a popular choice for time series forecasting tasks (IBM, 2024).[1] Compared to traditional models like VAR's or ARIMA, RNNs are better equipped to capture non-linear relationships, trends, and seasonality (Chen, 2020). Although deep learning models are often criticized for their "black box" nature—making it difficult to interpret the role of

[1] For a step-by-step workflow on training RNNs for time series prediction, see Stoneman and Duca (2024).

individual input variables—research has shown that they can capture trends and dynamics that conventional models often miss (Wang et al., 2024).

One of the most widely used and effective variants of RNNs in macroeconomic forecasting is the Long Short-Term Memory (LSTM) model. For example, Zhang et al. (2022) combine LSTM with hidden Markov models (HMMs) to forecast Chinese GDP using pre-COVID consumer price index data. Longo et al. (2022) use an ensemble method based on LSTM to forecast U.S. GDP and offer insights into the contribution of various economic indicators during the pandemic period.

4. Methodology

This section outlines the data cleaning and preparation process, the model architecture, and the rationale for our methodological choices.

4.1 Data Fetching & Cleaning

We start by fetching 8,321 unique data series from FRED using the FRED API. The filtering criteria included selecting only monthly and quarterly series with start and end dates between 1984 and 2024. As discussed in Bordo and Duca (2025) and Bordo, et al. (forthcoming), we exclude regional, international, and academic data categories to maintain a consistent focus on national macroeconomic indicators (which still include exchange rate and other open-economy U.S. macro indicators). Additional filtering removes series having missing values, avoiding the need for interpolation or imputation. After this step, 7,491 data series remained. Note that the 7,491 variables include simple-sum, conventional measures of money, such as M1 and M2, which studies find are inferior to broad Divisia measures of money (see Bordo and Duca, 2025, and Bordo, Duca, and Jones, forthcoming). After ensuring that the dataset included critical variables such as our nominal GDP, we convert monthly to quarterly frequency to align with quarterly data already fetched from FRED.

We augment the dataset by adding COVID-19-related variables from Bordo and Duca (2025), including the government stringency index (*GSTR*), the share of fully vaccinated individuals (*VaxFull*), and an interaction term $GSTR \times (1-0.01 VaxFull_{t-1})$. We also add stock load data and Divisia M3, applying the same cleaning process. Adding these five variables raises the number of quarterly series increased to 7,496.

Throughout our study, we jointly use Divisia measures of money and stock loads to better capture the long-run relationship between money and nominal GDP. Bordo and Duca (2025) show that controlling for stock mutual fund loads, which affects the long-run demand for Divisia M3, significantly strengthens *DivM3*'s relationship to nominal GDP over time. Accordingly, because our aim is to extract as much information as possible from Divisia measures of money for forecasting NGDP, we include stock-loads in our deep-learning model.

4.2 Feature Engineering and Normalization

Initial attempts to train the model using all 7,491 features resulted in overfitting—a condition where the model performs well on training/validation data but poorly on the test set. This was evident in degraded test metrics, including a negative R^2, and high MSE, MAE, and MAPE scores. Performance did not improve using different model architectures. This reflects Stoneman and Duca's (2024) findings that although LSTMs are less sensitive to multicollinearity than linear models, including many statistically irrelevant features can hinder performance and convergence.

To mitigate this problem, we used Scikit-learn's Mutual Information Regression to select the top 1,500 variables with the highest information about the target variable. This feature selection significantly improved model performance, and the resulting set is our baseline dataset (A1). To the 1,500 variable dataset A1, we add three COVID control variables (*GSTR*, *VaxFull, GSTR*×(1-0.01 $Vaxfull_{t-1}$) to form a COVID-augmented baseline dataset, A1C, which has 1,503 variables.

Off this pair of baseline datasets, we create three pairs of broad Divisia money datasets, each of which adds stock mutual fund loads and one of the three broad Divisia monetary services measures. This results in the following 8 datasets, reflecting (a) whether they include or do not include three COVID control variables (two options), and (b) whether they omit any broad Divisia money measure and stock mutual fund loads or include *one* of the three broad Divisia money measures and stock mutual fund loads (four options):

A1: 1,500 top macroeconomic variables

A1C: A1 + three COVID variables (*GSTR*, *VaxFull, GSTR*×(1- 0.01 $Vaxfull_{t-1}$), 1,503 var.

A1M3: A1 + Divisia M3 + stock loads, totaling 1,502 variables

A1CM3: A1 + Divisia M3 + stock loads + the three covid variables, totaling 1,505 variables

A1M4: A1 + Divisia M4 + stock loads, totaling 1,502 variables

A1CM4: A1 + stock loads + Divisia M4 + three Covid variables, totaling 1,505 variables

A1M4-: A1 + Divisia M4- + stock loads, totaling 1,502 variables

A1CM4-: A1 + stock loads + Divisia M4- + three Covid Variables, totaling 1,505 variables

To mitigate any remaining noise and outliers, we normalized each dataset, including nominal GDP, using the StandardScaler function from the Scikit-learn library (Pedregosa et al., 2011).

4.3 Training, Test, and Validation Sets

We employed a time-based split for our datasets, based on Bordo and Duca (2025) and Bordo, Duca, and Jones (forthcoming) who note that the methodology used to construct Divisia indices miss-measures liquidity services before full deposit deregulation in 1983.[2] To avoid using lags of data affected by this regime shift, our training period spans from 1985q1 to 2013q4, while the

[2] This point is acknowledged by Jones and Anderson, who created the basic empirical approach (described in Anderson, et al., 1997, later adopted by the CFS to construct Divisia indexes.

test/prediction period covers 2014q1 to 2024q4. A subset of the training data was used as the validation set, applying the expanding window cross-validation technique, which is particularly well-suited for time series data (Cerqueira et al., 2020).

The period from 1985 to 2024 was selected to allow the model ample time to learn and forecast key patterns and trends in the U.S. economy. The training period of 1985-2013 spans the Great Moderation and significant events such as the Great Recession, while the test period captures the 2020 recession and the inflation surge of 2021–2022. The period since 2020 is noteworthy for covering the dramatic rise in money growth and inflation, and the Covid-19 pandemic and government-imposed economic shutdowns—phenomena not observed in prior years.

4.4 Model Architecture and Parameters

The architecture of an RNN is a crucial determinant of its performance, encompassing factors such as the number of neurons, the number of hidden layers, and the types of layers utilized. While increasing the number of neurons can enhance the model's ability to learn complex patterns, an excessive number can lead to overfitting by making the model overly sensitive to noise in the training data (Stoneman and Duca, 2024). We performed hyperparameter tuning to arrive at the best combination of parameters to yield better accuracy scores.

We adopt an incremental approach in training the models, beginning with a baseline model that used only the1500 FRED variables and gradually incorporating additional variables, such as those related to Covid-19 and Divisia money, to assess performance improvements. Our model learns from data spanning the past three quarters (i.e., sequence length = 3) to make one-quarter-ahead and four-quarter-ahead predictions. Attempts to train the model using longer time windows led to unstable and suboptimal results, prompting us to adopt a three-quarter lookback period as the most effective. Table 1 outlines the configuration of other parameters for the LSTM models developed in this study, offering insights into the model's setup and optimization strategies.

Parameter Name	Value
Layers (4 layers in the model)	1. Input (# of input variables per time stamp for each dataset) 2. LSTM (16 units) 3. Dropout (0.2) 4. Dense (1)
Learning Rate	0.001
Batch Size	32
Epoch	200
Training Function	[Relu, linear]
Performance Validation Function	MSE
Weight Initializer	HE initialization
Optimizer	Adam
Sequence Length	3 Quarters

Table 1: LSTM Model Parameters and Their Values

5. Results and Discussion

This section focuses on the results for one-quarter-ahead forecasts presented in subsection 5.1, which perform better than corresponding four-quarter-ahead forecasts in Section 5.2. This emphasis is due to the model's stronger overall fit in the one-quarter-ahead predictions across all datasets. Nonetheless, results from both forecasting horizons consistently imply that adding a broad Divisia monetary measure and stock loads adds significant and economically meaningful information for forecasts, with results slightly favoring the addition of Divisia M4-, for forecasting nominal GDP one quarter ahead and the addition of Divisia M4, for forecasting nominal GDP four quarters ahead.

Additionally, in Figures 1 through 4, the COVID-19 dip in NGDP predicted by the model appears slightly delayed relative to the actual NGDP decline. This likely owes to both the inherent lag in economic cycles and the unusual nature of the COVID-19 Recession—a sudden-stop unlike any slowdown encountered in the training period. Consequently, the model may have struggled to anticipate the downturn in advance, resulting in a delayed forecast. However, since this pattern is consistent across all datasets, it does not affect our comparative assessment of model performance.

5.1 One-Quarter-Ahead Forecasts

Among the eight datasets compared, A1M4-, which includes the top 1,500 relevant FRED variables, plus *DivM4-* and stock loads, performed the best. It achieved an impressive 99% R^2, along with the lowest MSE and MAE values (Table 2). The forecast also aligns with the trend almost perfectly (Figure 1), and the model's MAPE of 1.1% indicates that, on average, it diverged from actual one-quarter ahead GDP values by approximately 1%. In contrast, the baseline model, which was trained solely on the A1 FRED dataset, performed worse, with an R^2 of 88.8% and a

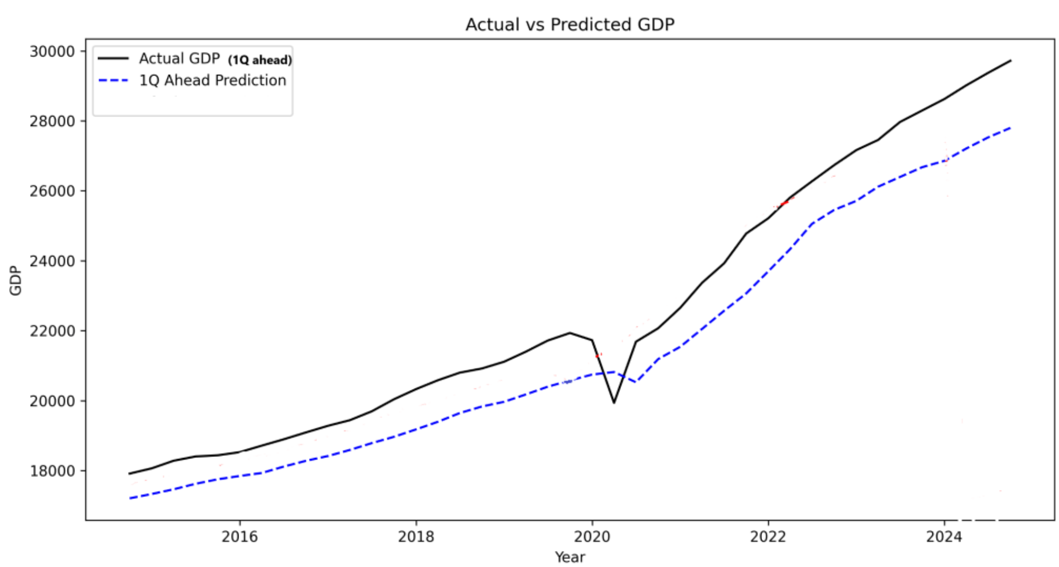

Figure 1: One-quarter-ahead GDP for A1: Baseline Model has Lower Trend Fit but Slightly Captures the Sharpness (not depth) of the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

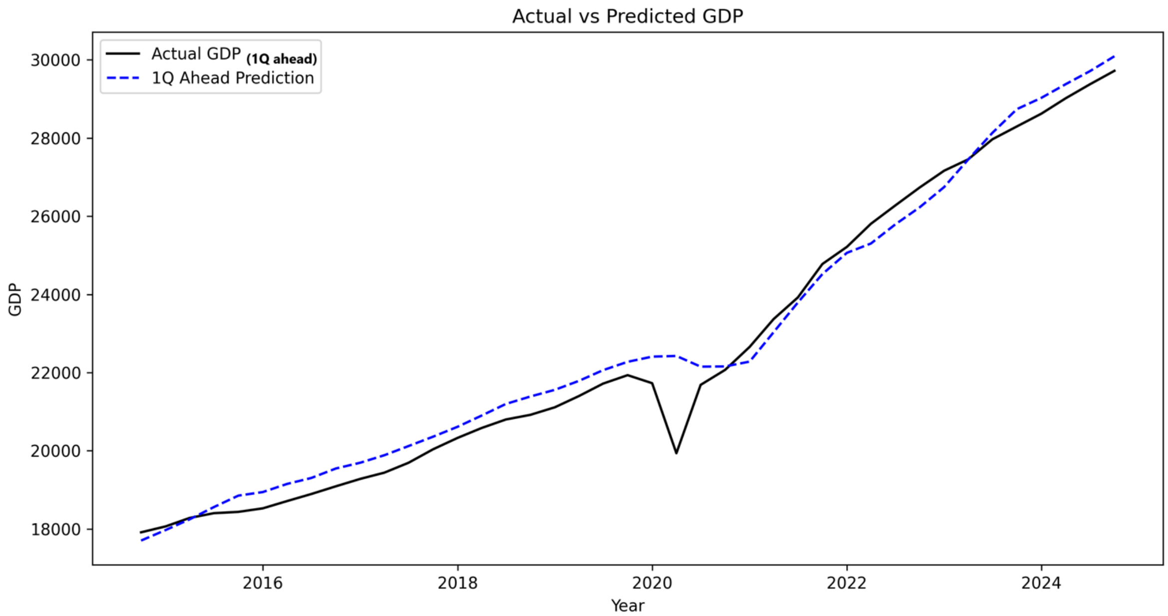

Figure 2: One-quarter-ahead GDP for A1M3: Adding Divisia M3 and Stock Loads to the Baseline Data Improves Overall Trend Fit but Understates the COVID-19 Dip
(Sources: CFS, FRED, and authors' calculations)

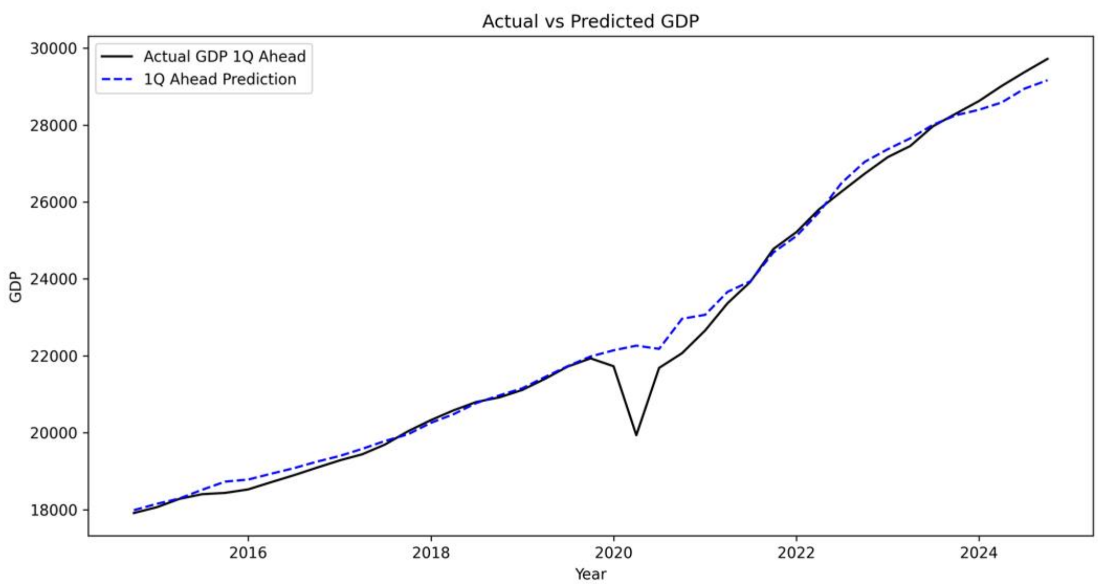

Figure 3: One-quarter-ahead GDP for A1M4-: Adding Divisia M4- and Stock Loads to the Baseline Data Improves Overall Trend Fit but Understates the COVID-19 Dip
(Sources: CFS, FRED, and authors' calculations)

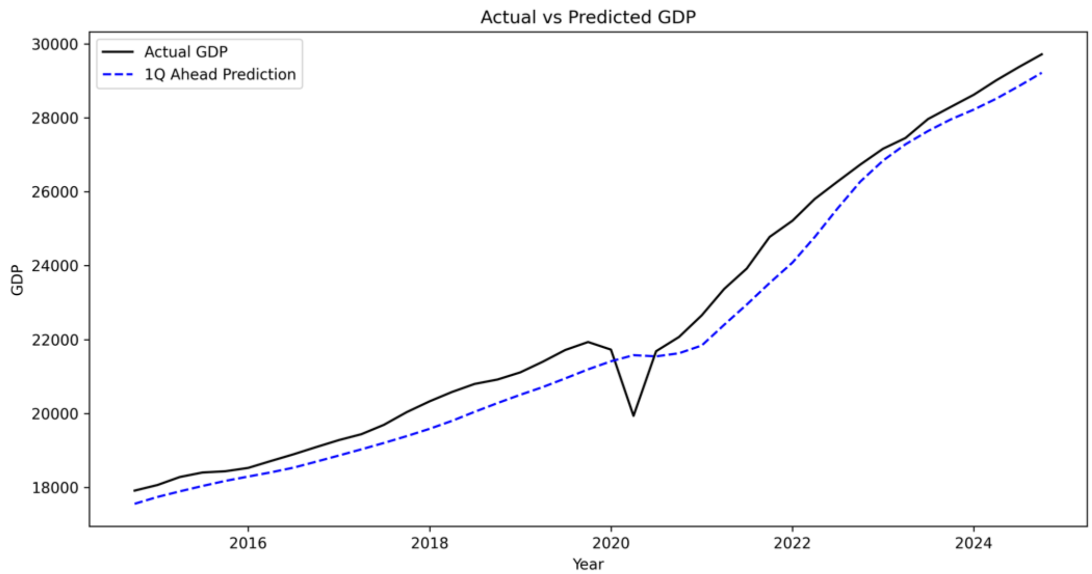

Figure 4: One-quarter-ahead GDP for A1M4: Adding Divisia M4 and Stock Loads to the Baseline Data Improves Overall Trend Fit but Understates the COVID-19 Dip
(Sources: CFS, FRED, and authors' calculations)

Dataset	R^2	MAPE	MSE	RMSE	MAE
A1	0.89	5.13%	1511085	1229.26	1179.01
A1C	0.81	6.2%	2532962.50	1591.53	1434.90
A1M3	0.97	1.82%	286079.77	534.86	395.26
A1CM3	0.88	4.06%	1609454.82	1268.64	986.00
A1M4-	0.99	1.1%	200746.25	448.05	243.87
A1CM4-	0.91	4.52%	1154851.87	1074.64	986.71
A1M4	0.97	2.59%	429963.58	655.72	576.74
A1CM4	0.93	3.19%	949862.69	974.61	771.51

Table 2: Model Performance Comparison (One-Quarter-Ahead Nominal GDP Forecasts)
(Sources: CFS, FRED, and authors' calculations)

MAPE of 5.1%. This represents a 78% reduction in error ((5.1−1.1)/5.1×100). While MAPE is scale-independent and measures relative error, the fact that the target variable (*NGDP*) is in billions of dollars highlights the practical significance of this improvement. A 4.0 percentage point decrease in MAPE (from 5.1% to 1.1%) indicates a significantly more accurate model. For context, MAPE values under 10% are generally considered strong in many fields, making a value of 1.1% exceptional (Table 2).

As hypothesized, adding a broad Divisia money measure and stock loads to the FRED dataset improved prediction accuracy, ostensibly, because their addition better tracks aggregate liquidity and nominal aggregate demand. One plausible reason is that Divisia money aggregates may capture broader channels and mechanisms through which conventional and unconventional monetary and fiscal actions influence the economy. This could include tracking the effects of not only quantitative easing and forward guidance, but also how Treasury-backed credit easing programs created by the Fed prevented the COVID recession from being amplified by financial accelerator effects (see Bordo and Duca (2021, 2022). Another is that broad Divisia money also tracks the pace of financial intermediation and liquidity transformation in both the traditional and shadow banking sectors. These results are consistent with Bordo and Duca's (2025) more conventional forecasting model in which Divisia M3 and stock mutual fund loads have information about future nominal GDP.

However, returning to Figures 1 through 4, the forecasts do not fully match the sharp dip in GDP observed in 2020q2. While both the baseline A1 model and those that augment the baseline datasets with Divisia money and stock loads fail to capture the depth of the dip, the baseline model did capture the sharpness of the decline. Thus, adding Divisia money and stock loads improved the alignment of the trend but resulted in a smoother dip during the COVID-19 recession. This discrepancy can be attributed to several economic and modeling factors. One reason is that the abrupt halt in economic activity caused by pandemic-related shutdowns is an outlier event that is difficult to model, as it deviates significantly from typical economic behavior (Bordo and Duca, 2025). The other is that the impact of stimulus can have long and variable lags (Friedman and

Schwartz (1963, 1982), especially in the midst of a pandemic, that Divisia money provides information about.

The sudden halt in spending imposed by the shutdowns is not tracked by Divisia money, which explains the understatement of the 2020 Q2 dip in the NGDP forecast when money was added. During the COVID-19 shock, rapid fiscal and monetary interventions—such as stimulus checks, quantitative easing, and financial support for firms—injected significant amounts of money into the economy, leading to sharp increases in Divisia money growth (see Figure 5 for the example of Divisia M3 which is similar to using Divisia M4- or Divisia M4). Despite this monetary expansion, the shutdowns abruptly curtailed much of consumer and business spending, causing a sharp drop in velocity. This decline in velocity was of greater magnitude than the rise in money supply reflecting a fall in nominal GDP relative to money (see Figure 5). Divisia money fails to capture this temporary decline in NGDP that was caused by a non-economic shock leading to the model's understatement of the NGDP dip in datasets containing Divisia money.

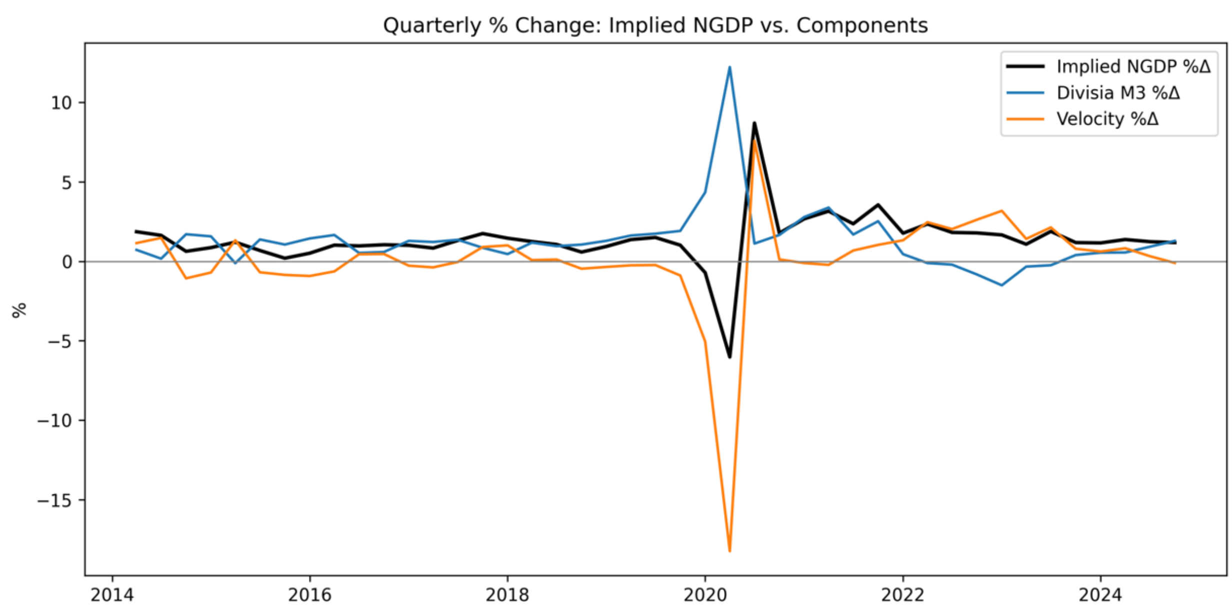

Figure 5: Percentage Change in Divisia M3 and Its Velocity Over Time
(Sources: CFS, FRED, and authors' calculations)

A similar pattern occurred during World War II, when goods rationing limited spending despite continued growth in the money supply. Anderson, et al. (2017) show that wartime rationing reduced nominal GDP relative to money, leading to a temporary decline in nominal spending that was not captured by monetary aggregates. Specifically, during WWII, the U.S. government expanded the money supply and debt issuance to support the war effort while restraining consumption to a lot more to defense production. However, the spending cap imposed by rationing temporarily suppressed NGDP relative to the growing money supply. As a result, monetary aggregates failed to reflect this transitory decline in nominal GDP—analogous to Divisia money's inability to solely track the 2020q2 and 2020q3 COVID-19-induced shifts in nominal GDP.

The smoothing of the COVID-19 dip may partly arise from the model itself, highlighting the inherent "black box" nature of deep learning models—where we cannot fully discern which components and features the model prioritizes for learning. For instance, Divisia money plus stock mutual fund loads might function as trend components, enhancing the long-term fit while potentially overshadowing short-term cyclical shocks. If the model places more emphasis on trend variables, it could downplay transient shocks, such as the 2020 dip.

Results from the other datasets suggest that the model may have traded off precision in capturing the COVID-19 dip (an unexpected shock) in favor of better capturing the broader trend when incorporating the money measures. In the A1+Covid variables dataset, predicted nominal GDP more accurately reflects the depth and sharpness of the dip, but at the cost of reduced accuracy in capturing long-term trends, as anticipated (see Table 2 and, Figure 1). The overall accuracy of the A1C model, as measured by a MAPE of 6.2% and a lower R^2 of 81%, decreased despite the addition of three new variables to the baseline dataset. Comparisons of the models that include stock loads and a Divisia money measure with or without the three COVID controls also favor omitting the three COVID controls. These results are similar to those of Stoneman and Duca (2024), who reported that including COVID-19 variables worsened their model's overall performance, resulting in a higher RMSE. A potential explanation is that our model may have interpreted the unprecedented nature of the COVID-19 recession—along with the sudden dip—as a short-term shock or noise, which accounts for the loss in overall fit accuracy. This is likely due to the model not being trained on similar sharp dips before, and the inclusion of these COVID-related variables—such as government restrictions and vaccines—representing atypical, non-standard features. This aligns with findings from Ng (2021), who describes the post-pandemic period as one where non-economic shocks distort the model fit established with pre-COVID data, effectively "messing up" its performance.

Finally, among the Divisia money augmented datasets, the one which includes Divisia M4- and stock loads (A1M4-), demonstrated strong overall trend accuracy with a MAPE of 1.1% and an R^2 of 99%, easily outperforming the baseline (A1) and slightly outperforming the other two broad Divisia money measures. Regarding the 2020q2 dip in nominal GDP, the baseline dataset's performance is similar to that of the Divisia money datasets, with a smoother dip than the baseline and the dataset with only COVID variables (Figure 6 versus Figure 7, 8, and 9). However, the dip is now slightly sharper and longer than in the models that add the money but not the COVID variables, which is expected given the inclusion of COVID-related variables. This pattern confirms that adding money measures is associated with a smoothing of the COVID-19 dip (consistent with the result from A1M3, A1M4-, and A1M4), while the addition of COVID variables better captures the sharp dip in nominal GDP, sacrificing overall model fit. Our results from this larger dataset also confirm the trends observed earlier, showing that adding money measures is linked to improved trend accuracy, yielding the better MAPE and RMSE values.

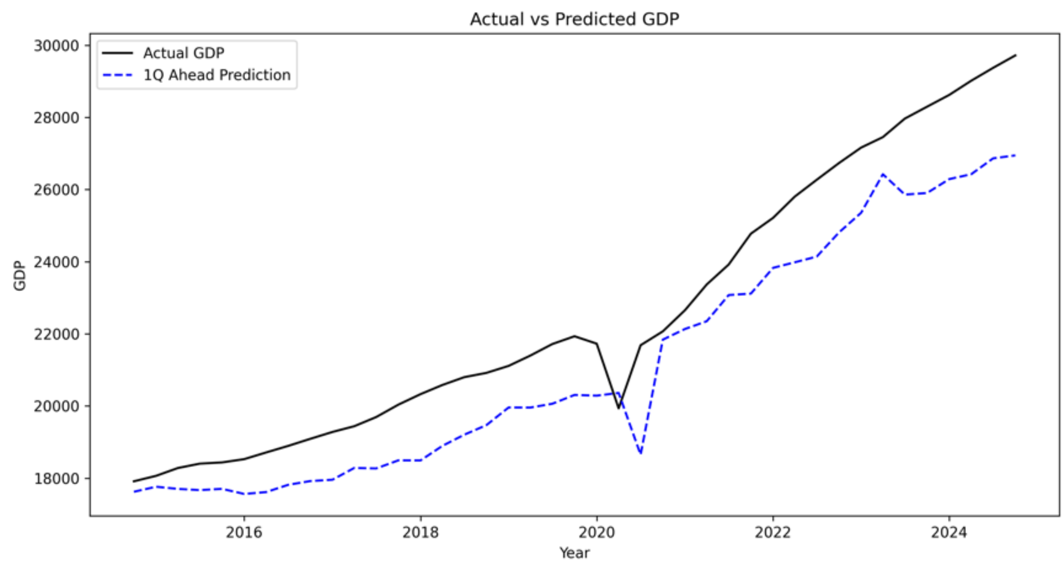

Figure 6: Actual vs Predicted One-quarter-ahead GDP for A1C: Adding COVID Variables Sacrifices Trend Fit and Accuracy but Perfectly Captures the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

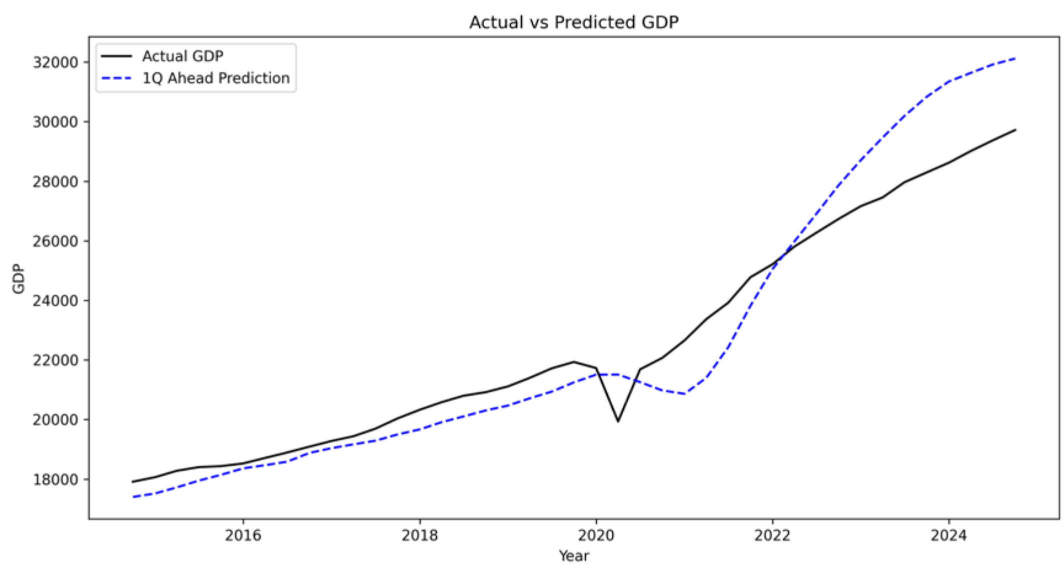

Figure 7: One-quarter-ahead GDP for A1CM3: Divisia M3 and COVID Variables Show a Tradeoff between the Model's Ability to Capture Long-term vs Short-term Trends.
(Sources: CFS, FRED, and authors' calculations)

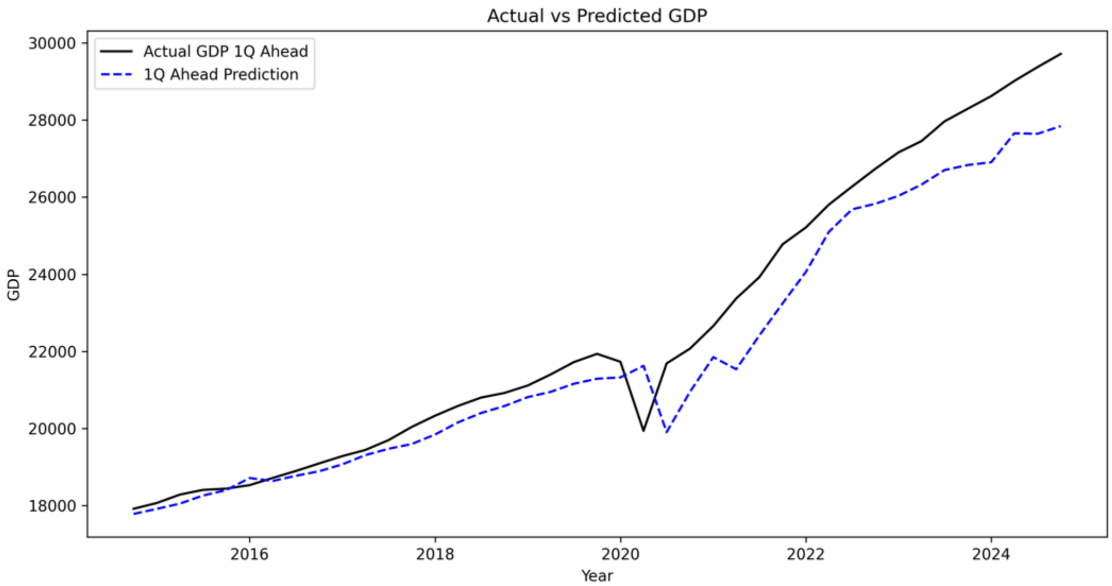

Figure 8: One-quarter-ahead GDP for A1CM4: Divisia M4 and COVID Variables Show a Tradeoff between the Model's Ability to Capture Long-term vs Short-term Trends.
(Sources: CFS, FRED, and authors' calculations)

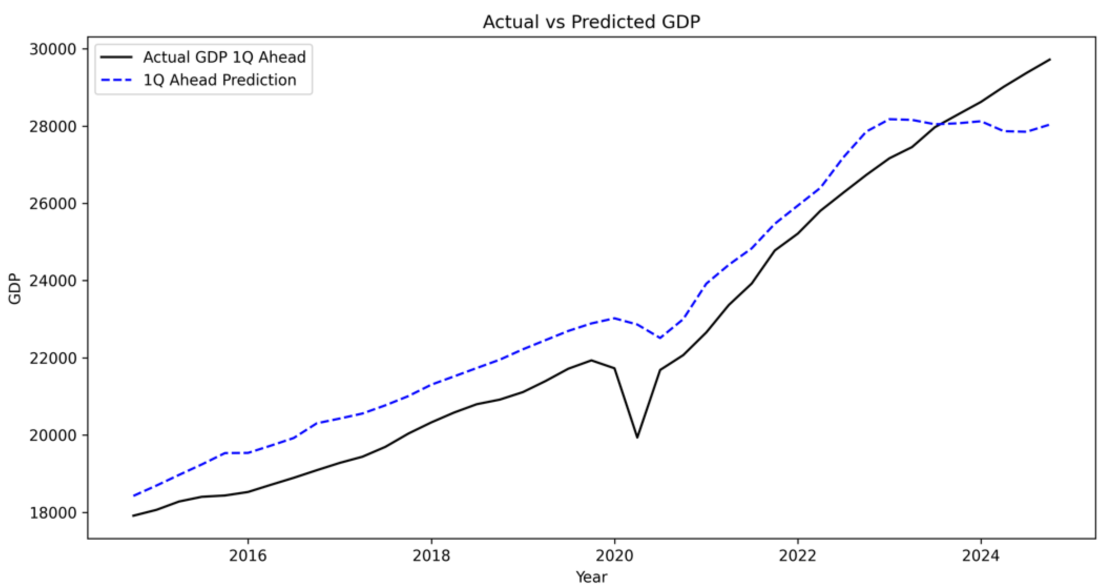

Figure 9: One-quarter-ahead GDP for A1CM4-: Divisia M4- and COVID Variables Show a Tradeoff between the Model's Ability to Capture Long-term vs Short-term Trends.
(Sources: CFS, FRED, and authors' calculations)

5.2 Four-Quarter-Ahead Forecasts

Overall, our four-quarter-ahead forecasts had lower accuracy across the datasets compared to the one-quarter-ahead forecasts. This is expected, as forecasting four quarters ahead is inherently more difficult: the model must capture stable, predictive relationships that persist over a longer horizon. To mitigate this challenge, we experimented with progressively generating multi-step forecasts—modeling the economy's evolution step-by-step to help the model grasp gradual economic adjustments. However, under our current model architecture, this reduced accuracy, leading us to adopt the method that directly forecasts the four-quarter-ahead value. In the future we could explore alternative model architectures that are specifically designed for longer-horizon forecasts. The current study uses a single architecture and only varied the prediction horizon, so the model was not fine-tuned for optimizing four-quarter-ahead forecast performance.

Despite these limitations, our four-quarter-ahead forecast results lead to a consistent conclusion: incorporating a Divisia money measure plus stock load data significantly enhances overall forecast accuracy compared to the baseline dataset (see Table 3). One slight difference is that results slightly favored using Divisia M4 among the broad Divisia money measures. For this broadest Divisia measure, the R^2 increased from the baseline dataset's 79% to 94%, while the MAPE dropped from 5.93% to 2.5%—a reduction of about 2.4 percentage points or 58%. The difficulty of the baseline model to track trend nominal GDP is evident in Figure 10.

The challenge of capturing the COVID-19 dip in 2020q2 when incorporating Divisia data is also evident in the four-quarter-ahead forecasts (see Figures 11-13). As discussed earlier, this may partly owe to model-related factors, but primarily to the sudden halt in spending that Divisia money measures failed to track (see Section 5.1 for more details).

Dataset	R^2	MAPE	MSE	RMSE	MAE
A1	0.79	5.93%	2693216.50	1641.10	1415.34
A1C	0.76	5.12%	3032630.86	1741.45	1296.85
A1M3	0.94	2.90%	701048.63	837.29	692.34
A1CM3	0.84	4.71%	2014251.01	1419.24	1097.94
A1M4-	0.91	3.06%	1094940.56	1046.39	749.06
A1CM4-	0.88	4.10%	1542854.23	1242.12	1008.74
A1M4	0.94	2.5%	734758.62	857.18	615.89
A1CM4	0.84	4.41%	2061053.31	1435.64	1054.49

Table 3: Model Performance Comparison (Four-Quarter-Ahead Forecasts)
(Sources: CFS, FRED, and authors' calculations)

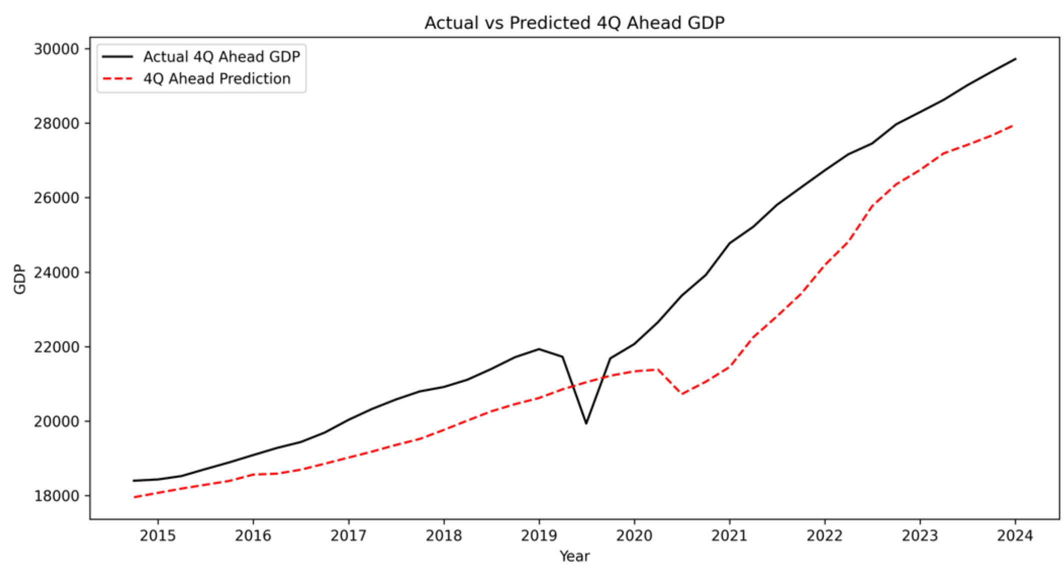

Figure 10: Four-quarter-ahead GDP for A1: Baseline Model has Lower Trend Fit but Slightly Captures the Sharpness (not depth) of the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

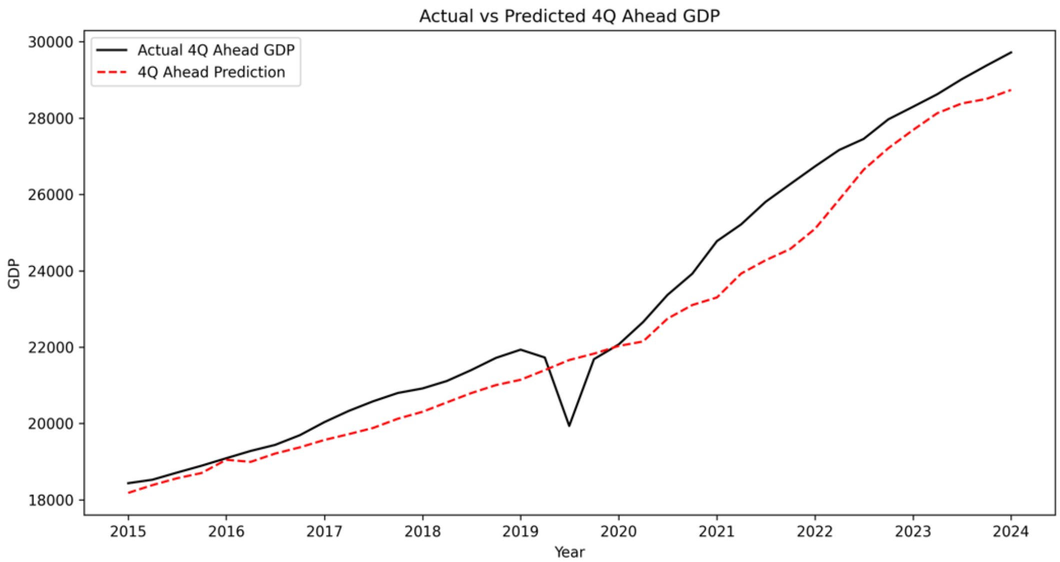

Figure 11: Four-quarter-ahead Nominal GDP for A1M3: Adding Divisia M3 and Stock Loads to the Baseline Data Improves Overall Trend Fit but Misses the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

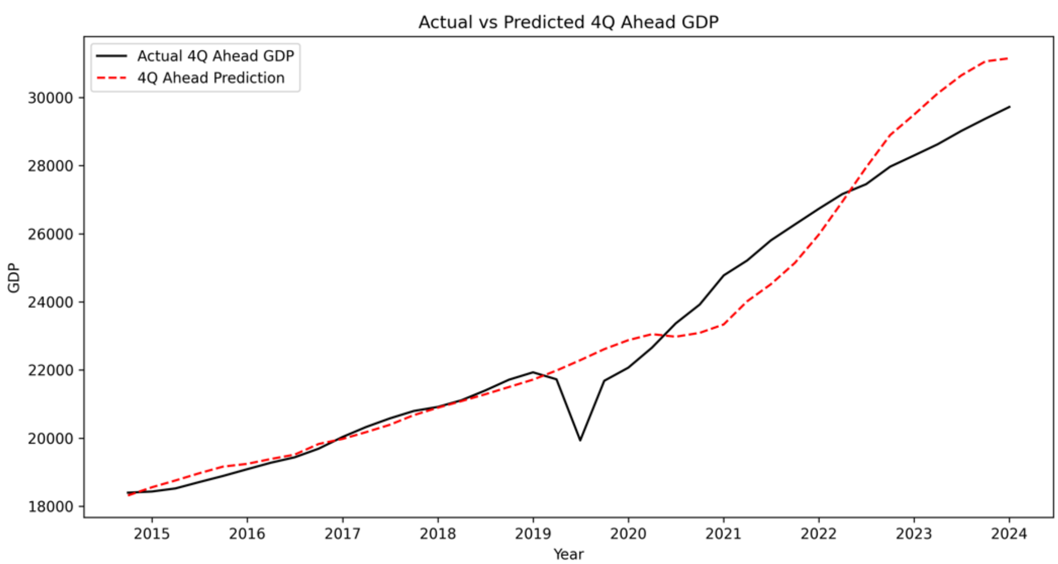

Figure 12: Four-quarter-ahead Nominal GDP for A1M4: Adding Divisia M4 and Stock Loads to the Baseline Data Improves Overall Trend Fit but Misses the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

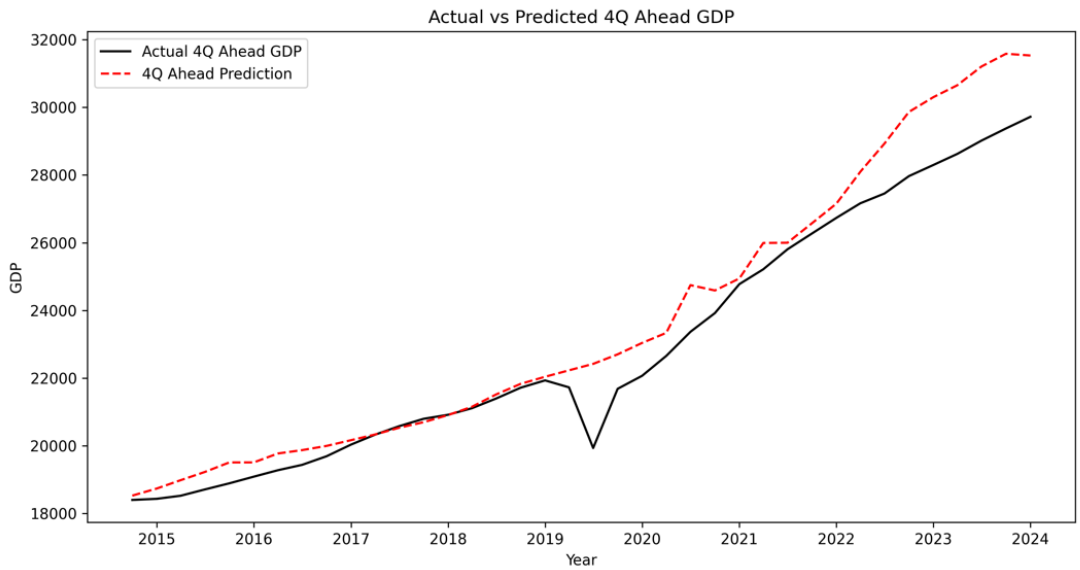

Figure 13: Four-quarter-ahead Nominal GDP for A1M4-: Adding Divisia M4- and Stock Loads to the Baseline Data Improves Overall Trend Fit but Misses the COVID-19 Dip.
(Sources: CFS, FRED, and authors' calculations)

Furthermore, similar to the one-quarter-ahead forecasts, incorporating COVID-19 variables compromises the overall trend fit. This can be seen by comparing pairs of models that have corresponding no Divisia or Divisia money variables: Figures 10 vs. 14, Figures 11 vs. 15, Figures 12 vs. 16, and Figures 13 vs.17.[3] The models using the largest datasets, A1CM3, A1CM4, and A1CM4-, illustrate this tradeoff clearly: adding both COVID-19 and money variables results in a reduced overall fit, but a slightly better forecast of the sharpness of the dip compared to datasets omitting the three COVID variables. However, as in the one-quarter-ahead forecasts, the inclusion of money improves the overall trend fit. The only notable difference is that Divisia M4 performs best for forecasting nominal GDP four quarters ahead whereas Divisia M4- performs best in one-quarter ahead forecasts (see Tables 2 and 3). Overall, the significant improvement in the accuracy of four-quarter-ahead forecasts using any of the Divisia money

[3] Unlike in the one-quarter-ahead forecasts, incorporating COVID-19 variables did not enhance the model's ability to predict the NGDP dip in the four-quarter-ahead forecasts.

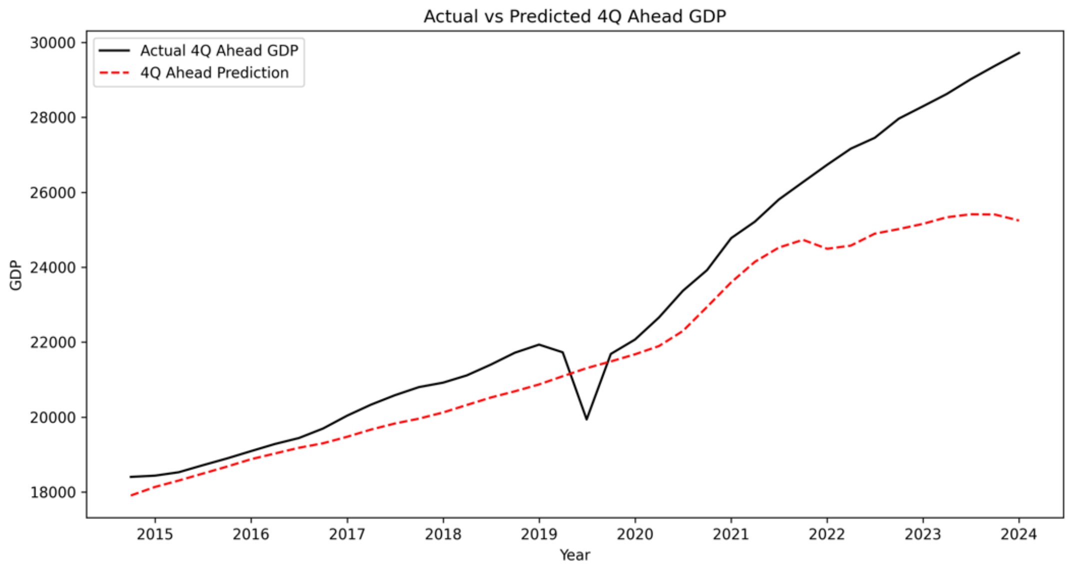

Figure 14: Actual vs Predicted Four-quarter-ahead GDP for A1C: Adding COVID Variables Sacrifices Trend Fit and Accuracy
(Sources: CFS, FRED, and authors' calculations)

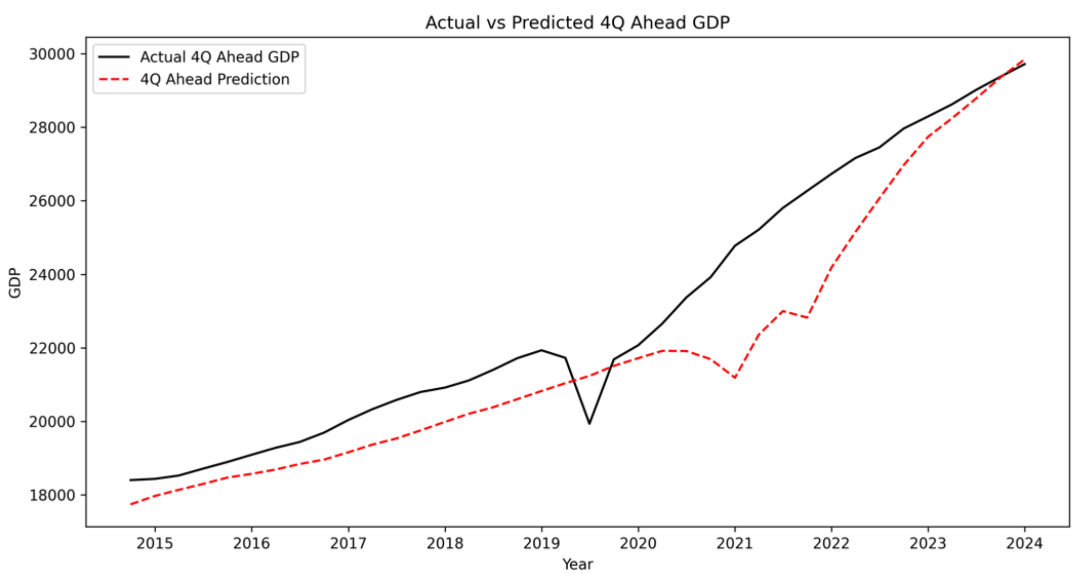

Figure 15: Actual vs Predicted Four-quarter-ahead GDP Using A1CM3 (Divisia M3 and COVID) Variables Show a Tradeoff between Capturing Long-term vs Short-term Trends
(Sources: CFS, FRED, and authors' calculations)

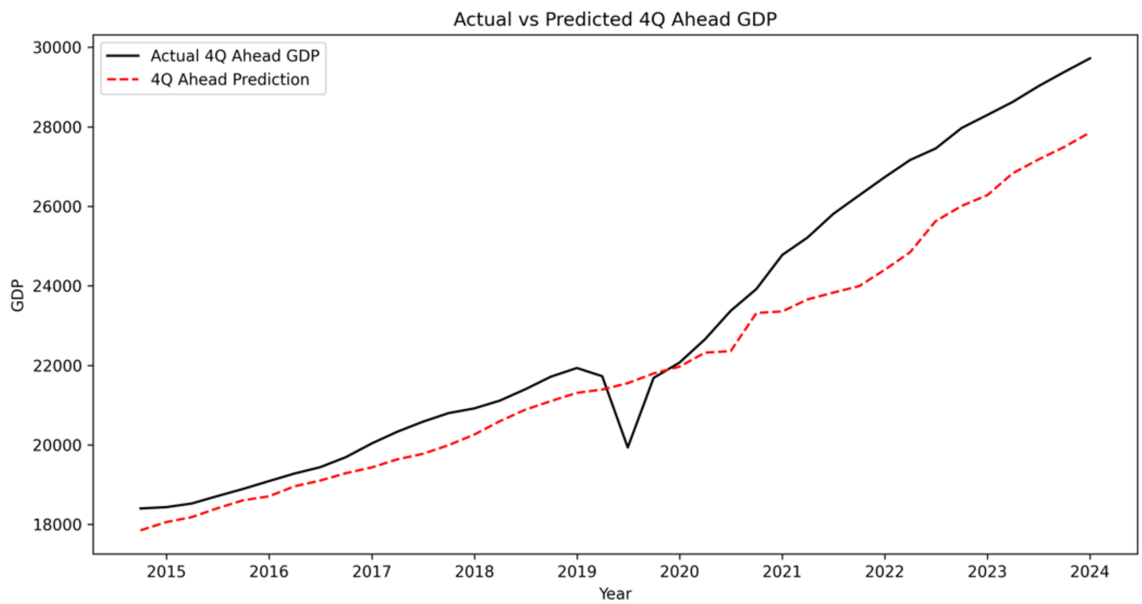

Figure 16: Actual vs Predicted Four-quarter-ahead GDP Using A1CM4- (Divisia M4- and COVID) Variables Show a Tradeoff between Capturing Long-term vs Short-term Trends
(Sources: CFS, FRED, and authors' calculations)

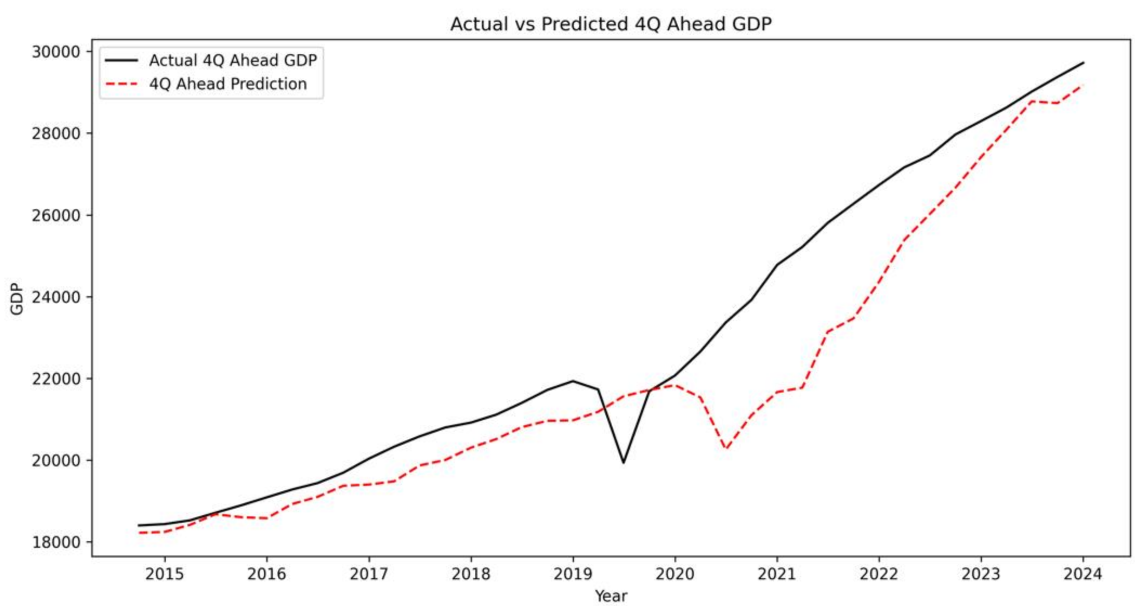

Figure 17: Actual vs Predicted Four-quarter-ahead GDP Using A1CM4 (Divisia M4 and COVID) Variables Show a Tradeoff between Capturing Long-term vs Short-term Trends
(Sources: CFS, FRED, and authors' calculations)

datasets reveals that broad Divisia money contains valuable long-horizon information about NGDP— beyond mere momentum—enabling the model to make more accurate predictions up to a year ahead.

6. Conclusion

This study demonstrates that adding a broad Divisia measure of monetary services plus stock mutual fund load data to a baseline FRED dataset significantly enhances the overall accuracy of nominal GDP predictions at one-quarter-ahead and four-quarter-ahead horizons. While some precision was lost in capturing the depth of the recession, the trade-off—sacrificing short-term ability to track the 2020q2 drop in nominal GDP for better long-term trend accuracy—is less severe than the trade-off involved in excluding Divisia money measures altogether, which led to a noticeable loss in overall trend accuracy. The reason is that the models with Divisia money still manage to capture the recession trend with improved overall accuracy, while on both the baseline and A1C datasets, despite better capturing short-term cycles (such as the dip), struggle to achieve a strong overall fit.

This improved performance likely reflects that broad Divisia money measures can capture broader channels through which expansionary monetary and fiscal policies affect the economy and through which money is linked to economic activity. This supports our main hypothesis that incorporating money measures enhances the performance of deep learning models, particularly in terms of trend fit. It also is consistent with evidence from more traditional, structural models that Divisia money measures provide more marginal information than do conventional, simple-sum measures of money (see Bordo and Duca, 2025, and Bordo, Duca, and Jones, forthcoming).

Nevertheless, using an LSTM approach yields somewhat different results concerning which measure of Divisia money provides the most information about nominal GDP, an indicator of aggregate nominal demand. The dataset adding Divisia M4- and stock mutual fund loads to FRED scraped data yielded the best overall performance in one-quarter-ahead forecasts of nominal GDP, while adding Divisia M4 and stock fund loads performed best in four-quarter ahead forecasts. This differs from recent findings using more traditional, structural methods that Divisia M3 is preferable to Divisia M4- and Divisia M4 (Bordo and Duca, 2025, and Bordo, et al., forthcoming). Although the difference in forecasting performance across the three main broad measures of Divisia money is not large, our findings indicate that deep learning models have an ability to surpass traditional, more structural modeling approaches, highlighting the usefulness of deep learning approaches.

Our findings for forecasting nominal GDP also suggest that Divisia money measures may provide similar information for forecasting inflation, something which we plan to research in the next version of this study. Future work could also focus on improving the model architecture and testing whether alternative deep learning models could better capture both long-term trends and short-term cycles in inflation when incorporating money measures.

References

Anderson, Richard G., Michael Bordo, and John V. Duca V. 2017. "Money and Velocity During Financial Crises: From the Great Depression to the Great Recession." *Journal of Economic Dynamics and Control* 81(C): 32–49.

Anderson, Richard G., Barry E. Jones, and Travis D. Nesmith, T.D., 1997. "Special Report on the Monetary Services Index Project of the Federal Reserve Bank of St. Louis." *Federal Reserve Bank of St. Louis Rev*iew 79 (1): 25-82.

Barnett, William A. 1980. "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory." *Journal of Econometrics* 14(1): 11–48.

Barnett, William A. 1982. "The Optimal Level of Monetary Aggregation." *Journal of Money, Credit, and Banking* 14(4): 687-710.

Belongia, Michael T. 1996. "Measurement Matters: Recent Results From Monetary Economics Reexamined." *Journal of Political Economy* 104(5): 1065–1083.

Bordo, Michael D. and John V. Duca. 2022. "How New Fed Corporate Bond Programs Cushioned the Covid-19 Recession," *Journal of Banking and Finance* 136: 1-22.

Bordo, Michael D. and John V. Duca. 2021. "An Overview of The Fed's New Credit Policy Tools and Their Cushioning Effect on the COVID-19 Recession." *Journal of Government Economics* 3: 1-10.

Bordo, Michael and John V. Duca. 2025. "Money Matters: Broad Divisia Money and the Recovery of the U.S. Nominal GDP From the COVID-19 Recession." *Journal of Forecasting* 44(4): 1071–1096.

Bordo, Michael D., John V. Duca, and Barry Jones. Forthcoming." Broad Divisia Money, Supply Pressures, and U.S. Inflation Following the COVID-19 Recession." *Macroeconomic Dynamics*.

Cerqueira, Vítor, Luis Torgo, and Igor Mozetič. 2020. "Evaluating Time Series Forecasting Models: An Empirical Study on Performance Estimation Methods." *Machine Learning* 109: 1997-2028.

Chen, Jing Min. (2020). "Economic Forecasting With Autoregressive Methods and Neural Networks." *SSRN*, accessed at https://doi.org/10.2139/ssrn.3521532 on April 20, 2025.

Darrat, Ali F., Michel C. Chopin, and Bento J. Lobo. 2005. "Money and Macroeconomic Performance: Revisiting Divisia Money." *Review of Financial Economics* 14(1): 93–101.

Dery, Craig and Apostolos Serletis. 2021. "Interest Rates, Money, and Economic Activity." *Macroeconomic Dynamics* 25(7): 1842–1891.

Estrella, Arturo and Frederic S. Mishkin. 1997. "Is There a Role for Monetary Aggregates in the Conduct of Monetary Policy?" *Journal of Monetary Economics* 40(2): 279–304.

Fleissig, Adrian R. and Barry E. Jones 2015, “The Impact of Commercial Sweeping on the Demand for Monetary Assets during the Great Recession.” *Journal of Macroeconomics* 45: 412-422.

Friedman, Benjamin M. and Kenneth N. Kuttner. 1992. "Money, Income, Prices, and Interest Rates." *American Economic Review* 82(3): 472–492.

Friedman, Milton and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867-1960*. Princeton University Press, Princeton.

Friedman, Milton and Anna J. Schwartz. 1982. *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867-1975*. University of Chicago Press for the NBER, Chicago.

Horan, Patrick J. 2024. "Money, Output, and Prices: 1967–2022." *Quarterly Review of Economics and Finance* 97: 1-18.

IBM. 2024. "What Is a Recurrent Neural Network (RNN)?" *IBM Think Blog*, accessed at https://www.ibm.com/think/topics/recurrent-neural-networks on April 20, 2025.

Jadidzadeh, Ali and Apostolos Serletis. 2019. “The Demand for Assets and Optimal Monetary Aggregation.” *Journal of Money, Credit, and Banking* 51(4): 929-52.

Jones, Barry E., Dutkowsky, Donald H., Elger, Thomas and Adrian R. Fleissig 2008a. “Monetary Policy and Monetary Asset Substitution.” *Economics Letters* 99 (1): 18-22.

Keating, John W., Victor J. Valcarcel, and Pooya Amir Ahmadi. 2019. "A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions." *Journal of Money, Credit and Banking* 51(1): 227–259.

Longo, Luigi, Massimo Riccaboni, and Armando Rungi. 2022. "A Neural Network Ensemble Approach for GDP Forecasting." *Journal of Economic Dynamics and Control* 134(1): 1-29.

Ng, Serena. 2021. "Modeling Macroeconomic Variations After COVID-19." *NBER Working Paper* 29060.

Pedregosa, Fabian et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.

Schunk, Donald L. 2001. "The Relative Forecasting Performance of the Divisia and Simple Sum Monetary Aggregates." *Journal of Money, Credit and Banking* 33(2): 272–283.

Stoneman, David and John V. Duca. 2024. "Using Deep (Machine) Learning to Forecast US Inflation in the COVID-19 Era." *Journal of Forecasting* 43(4): 894–902.

Thornton, Daniel L. and Phoebe Yue. 1992. "An Extended Series of Divisia Monetary Aggregates." *Federal Reserve Bank of St. Louis Review* November: 35-52.

Wang, Tianyu, Ryan Beard, Jared Hawkins, and Rohit Chandra. 2024. "Recursive Deep Learning

Framework for Forecasting the Decadal World Economic Outlook." *IEEE Access*, accessed at https://arxiv.org/html/2301.10874v2 on April 20, 2025.

Xu, L., and A. Serletis. 2022. "The Demand for Assets: Evidence from the Markov Switching Normalized Quadratic Model." *Journal of Money, Credit, and Banking* 54 (4): 989-1025.

Zhang, Jichao, Jiaqi Wen, and Zhen Yang. 2022. "China's GDP Forecasting Using Long Short Term Memory Recurrent Neural Network and Hidden Markov Model." *PLOS ONE* 17(6), 1-26.