Monetary and Capital Markets Department
TECHNICAL ASSISTANCE HANDBOOK

Liquidity Forecasting—Part II: The Statistical Component

Prepared by the Central Bank Operations Division (CO)

Zhuohui Chen, Nikos Kourentzes, Romain Lafarguette, Anastasios Panagiotelis, and Romain Veyrune

THIS ONLINE HANDBOOK

This handbook aims to distill, document, and make widely available the lessons learned from MCM technical assistance over a long period, while also incorporating lessons learned globally. It covers a wide range of central banking topics pertaining to governance and risk management, monetary policy, monetary and foreign exchange operations, and financial market development and infrastructures, while highlighting, where relevant, specific issues for low-income, resource-rich countries. The handbook is intended to document and promote good practices and to support consistency of advice over time. It is, however, stressed that one-size solutions cannot fit all, and all advice therefore needs to be tailored to country-specific circumstances. The handbook comprises self-contained, issue-specific chapters with cross-references on overlapping issues where needed. It is targeted at those individuals who provide technical assistance (both IMF and non-IMF personnel), and practitioners in central banks and other relevant institutions.

THIS CHAPTER: LIQUIDITY FORECASTING—PART II: THE STATISTICAL COMPONENT

This chapter elucidates liquidity forecasting within the context of technical assistance. The audience for this chapter is central bank staff with a strong quantitative background. Liquidity forecasting entails a process of estimating the near-term path of a bank's reserves using a centralized framework. Short-term liquidity forecasts are used to calibrate the volume of central bank monetary operations to align liquidity with the announced stance of monetary policy, whether expressed as an interest rate or as a quantity. The best practice would be for the central bank to receive accurate information for counterparties that have accounts in its books, including its monetary counterparties (banks) or non-monetary counterparties, such as the government. However, the central bank may not have direct access to some counterparties (e.g., the public which demands banknotes) or the information could include significant errors. This chapter presents the statistical methods that have been used in technical assistance to forecast liquidity factors and the demand for liquidity. It also proposes solutions to select the best models, measure forecast accuracy, and reconcile forecasts. Some liquidity factors are relatively easy to forecast due to regular patterns (currency in circulation) while others require more sophisticated models, such as the government account. There is a tradeoff between the cost of implementing complex models and the accuracy gains.

CHAPTER AUTHORS

Zhuohui Chen (zhuohuichen@hotmail.com); Nikos Kourentzes (nikolaos@kourentzes.com); Romain Lafarguette (romain.lafarguette@gmail.com); Anastasios Panagiotelis (anastasios.panagiotelis@sydney.edu.au); Romain Veyrune (RVeyrune@imf.org)

Contents **Page**

Glossary

AGG	Aggregate
AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
CB	Central Bank
CIC	Currency in Circulation
CO	Central Bank Operations Division (IMF)
ETS	Exponential Smoothing
GAB	Government Account Balance
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MAE	Mean Absolute Error
MCM	Monetary and Capital Markets Department (IMF)
ME	Mean Error
MIS	Mean Interval Score
MSE	Mean Squared Error
NFA	Net Foreign Assets
OMO	Open Market Operations
RegARIMA	ARIMA with additional regressors
RMSE	Root Mean Squared Error
TBATS	Trigonometric Seasonality BOX-Cox, ARMA errors, Trend, and Seasonal Components
TA	Technical Assistance

Executive Summary

In most circumstances, central banks must manage liquidity to implement monetary policy. As an integral part of liquidity management, liquidity forecasting provides an outlook on the changes in liquidity-impacting items of central banks' balance sheets. By understanding the forecasted gap between liquidity supply and demand, central banks can effectively calibrate open market operations (OMO) to control the cost of refinancing (short-term interest rate), which is a common operational target of monetary policy. Forecasting is also important for central banks that have reserve money as their operational target.

Statistical forecasts are necessary if the development of liquidity factors is not certain and if accurate forecasts cannot be obtained otherwise. Some factors impacting liquidity require no "prediction" because they are predetermined or known in advance by the central banks. However, several factors, such as changes in Currency in Circulation (CIC) and net flows in the Government Account Balance (GAB), are not directly controlled by central banks. Fluctuations in Net Foreign Assets (NFA) are also not under the direct control of the central bank in fixed exchange rate arrangements. These factors require a more systematic forecasting approach. Forecasts could be obtained directly for the counterparties, e.g., the government could provide the central bank with its forecast or give prior notice of transactions on its account. Similarly, foreign exchange transactions could be known exactly at the two-day horizon if it settles on a T+2. (Considerations regarding the institutional framework for the sharing and centralization of information are considered in another handbook chapter.) Statistical forecasting is necessary when “qualitative” information is not available or accurate enough.

Many researchers have developed and experimented with various liquidity forecasting methods. However, most existing models in the literature rely on sophisticated specifications or focus primarily on the easiest-to-forecast item, i.e., the CIC. A more comprehensive framework that can forecast not only autonomous factors but, more importantly, the aggregate liquidity supply is desired.

This handbook proposes a framework for predicting short-term liquidity supply generated by three autonomous factors: CIC, GAB, and NFA. The framework is designed to cross-validate a family of time series models and identify the best-performing model for forecasts, while allowing for generic yet flexible country-specific customizations. Furthermore, reconciliation techniques are employed to enhance the estimation accuracy of aggregate liquidity supply, which is crucial for the calibration of OMO. While most central banks use point forecasts of the mean to calibrate their operations, this handbook argues that the central bank should forecast the predictive distribution and target the percentile of this distribution that reflects its risk preference. It would likely be different from the mean (point) forecast. If forecasts are published, the predictive distribution should be published as well, rather than have only the point forecast and to leave the users to factor the forecast uncertainty into their decisions.

I. Introduction

A key component of liquidity management in central banks is accurate forecasting of short-term changes in liquidity. For commercial banks, liquidity management is a process by which they ensure sufficient cash or liquid assets to meet their short-term obligations and operating needs. In central banks, it is the mechanism by which to steer the dynamics of short-term interbank interest rates through controlling the provision of reserves to commercial banks. Reserves are a highly liquid type of central bank liability held by commercial banks, deposited to fulfil the required reserve and/or payment settlement. When there is a shortage of reserves, interbank refinancing rates soar as commercial banks scramble for funding in the interbank market to avoid penalties for unmet reserve requirements. Conversely, surplus liquidity may drive short-term rates down, depressing interbank market transactions volumes, lowering banks' net interest margin, and flattening the yield curve.

Central banks need to proactively monitor and manage liquidity through open market operations (OMO) to achieve their mandates of price stability. Hence, central banks require an accurate forecast of the current and future changes of liquidity impacting items on their balance sheets (Gray 2008). The net of all changes will represent the gap between liquidity supply and demand and will determine the amount the central bank should absorb or inject through OMO. Table 1 presents a stylized central bank balance sheet.

Table 1. Stylized Central Bank Balance Sheet

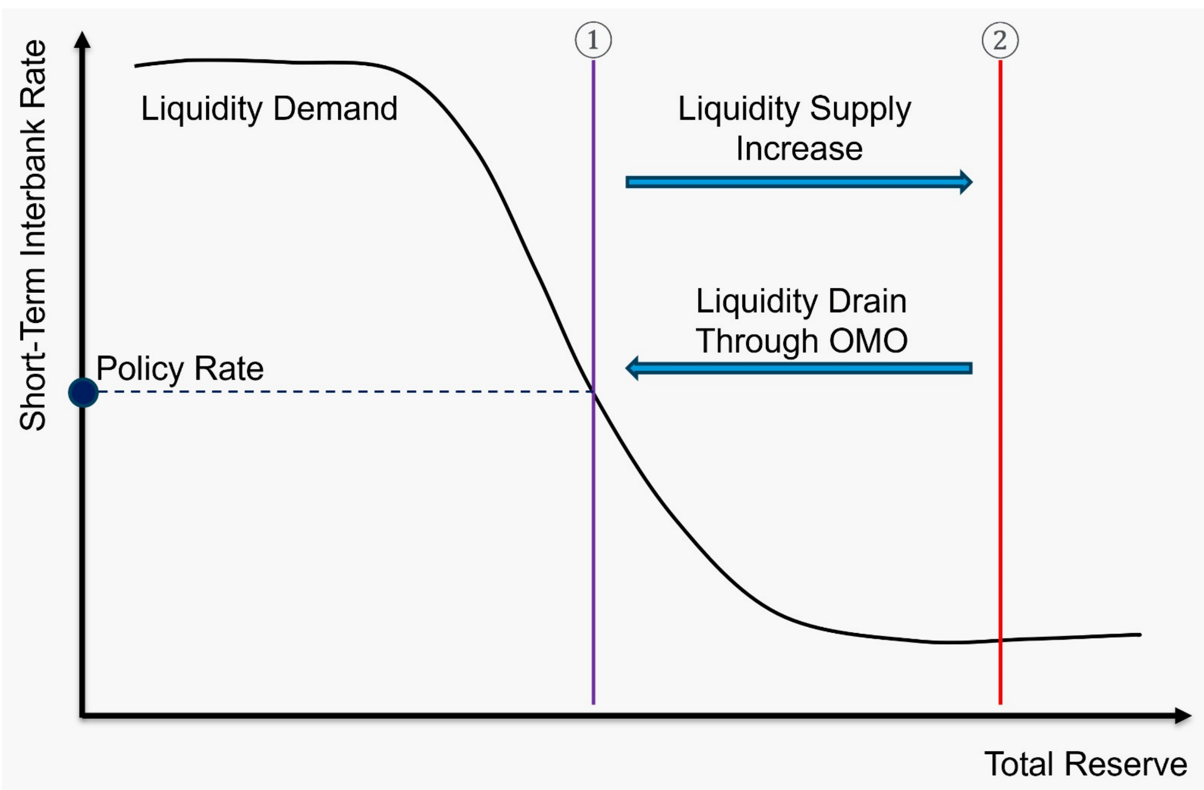
Assets	Liabilities
Net Foreign Assets (NFA) (+)	Currency in Circulation (CIC) (-)
OMO: Liquidity Supply (+)	Net Government Account Balance (GAB) (-)
Lending Facility (+)	OMO: Liquidity Drain (-)
Others	Deposit Facility (-)
	Commercial Banks' Required Reserves (+)
	Commercial Banks' Excess Reserves (+)
	Capital and Reserves
Note: (+) indicates increase in liquidity; (-) indicates decrease in liquidity	

The equilibrium equation between demand and supply is:

$$RR + ER = NFA - CIC - GAB + Net\ OMO\ Supply + Net\ Credit\ Facilities. \qquad (1)$$

The required reserve (RR) and excess reserve (ER) at the current account of the central bank's monetary policy[1] can be viewed as the demand for reserves, while the remaining impacting factors form the basis of reserve supply. Figure 1 plots an idealized liquidity demand curve, connecting the reserves with the short-term interbank rate, and illustrates a scenario in which liquidity supply increases, causing a shift from position (1) to (2). The short-term interbank rate will be drawn to the floor, away from the stable point. The central bank can intervene with OMO to relieve the downward pressure and maintain the level of the short-term interbank rate. Therefore, estimating changes in liquidity supply is essential as central banks need to absorb equivalent amounts of liquidity to return the short-term rate to the policy target, or, likewise, inject liquidity when the pressure is in the opposite direction.

Figure 1. Equilibrium of Reserve Demand and Supply through OMO



Source: IMF staff.

Some of the accounts in Table 1 are easier to predict because central banks can determine or have advance information about certain movements. For example, the required reserve ratio is predetermined

[1] The monetary policy counterparties of the central bank with whom the central bank implement monetary policy are usually commercial banks. Other financial actors could be included if they are critical for the transmission of monetary policy.

by central banks, making these "forecasts" straightforward. However, several accounts are under the control of non-monetary policy institutions,[2] and not under the direct control of central banks. Estimating these factors therefore becomes a priority in liquidity management. These accounts, known as autonomous factors, typically include the public's demand for currency in circulation (CIC), the government's position at the central bank (GAB), and the volatility of the net foreign assets (NFA) account.

Many central banks have established bespoke statistical models for liquidity forecasting, typically focusing on CIC. In practice, these models can be difficult to maintain, due to their specialized and manual model specifications. Maintenance can be constrained by software limitations and the modeling preferences and skillset of the responsible analysts, who often change over time. Moreover, forecasting CIC alone is insufficient for OMO calibration, with quantitative forecasting models for the remaining autonomous factors being less common in central banks. Although these limitations are not universal across central banks, there is a need for a systematic modeling methodology for liquidity forecasting.

In this work, we propose a modeling framework for developing daily forecasts for all autonomous factors. Key benefits of the framework for the central bank users are: (i) a unified approach for all autonomous factors, thereby reducing modeling overheads; (ii) automatic model calibration and selection; (iii) detailed reporting to facilitate individualized interventions from analysts; and (iv) leverage hierarchical forecasting to further improve forecasting performance of the autonomous factors. The modeling framework has been extensively tested in various central banks and is flexible for adjustment to different market and operational conditions. The audience for this chapter is central bank staff with a strong quantitative background.

The chapter is organized as follows. Section 2 summarizes previous studies on autonomous factors and aggregate liquidity forecasting. Section 3 provides illustrative examples of the liquidity autonomous factors that we aim to model. Section 4 introduces the forecasting models and methodologies used to generate the forecasts, with Section 5 detailing how the forecast evaluation and selection is done. Section 6 elaborates on the technical assistance (TA) approach, explaining the liquidity forecasting mission process and briefly describing the implementation of the framework into a toolbox. Section 7 discusses limitations and future avenues for research, followed by concluding remarks.

II. Literature Review

Literature on forecasting aggregate liquidity or autonomous factors is scarce. Most of the quantitative modeling on the autonomous factors centers on specifying an Autoregressive Integrated Moving Average (ARIMA) to forecast CIC. Following Bell et al. (1983) and Harvey et al. (1997) non-linear approaches to quantify calendar and seasonal effects, Cabrero et al. (2009) used ARIMA and Structural Time Series models to forecast CIC in the euro area, finding that a combination forecast was best. El Hamiani-Khatat (2018) investigated short- and long-term money demand forecasting models. For long-term money demand, macroeconomic factors were important, while short-term demand is dominated by seasonal patterns or recurring events that are relevant for calibrating OMO. ARIMA with additional regressors

[2] Which non-monetary policy institutions are allowed to bank with the central bank depend on the circumstances but usually include the government and the public (demand for banknotes via intermediaries).

(RegARIMA) is widely used to forecast the demand for currency and has been applied in many countries, including Poland (Koziński et al., 2015), Ghana (Nasiru et al., 2013), and Nigeria (Ikoku, 2014). The regressors are often used to model recurring events, seasonal patterns, and holidays. Hlaváček et al. (2005) compared a neural network model with a RegARIMA on a task to predict Fed CIC and concluded that the neural network model was slightly better. However, they also noted that the neural network model will overlearn sparse holiday events.

The existing literature does not touch much upon the detailed modeling for the GBA and NFA. Gray (2008), in his liquidity forecasting handbook, elaborated on the factors that influence aggregate liquidity and each autonomous factor, respectively. However, there are no quantitative models discussed in detail. For GAB, Williams (2010) acknowledged that it was a universal challenge to obtain good cash flow forecasting and mentioned the importance of intelligence input from relevant spending or revenue departments to improve the forecast quality of government cash balance. Iskandar et al. (2018) investigate ARIMA, neural networks, and hybrid models to forecast expenditures for the Indonesian government, constituting the only time series model investigation for non-CIC autonomous factors that we identified.

III. Characteristics of the Autonomous Factors

To better understand the challenges and requirements for forecasting autonomous factors we provide a brief exploratory analysis of typical CIC, GAB, NFA, and their aggregate (AGG), which captures the total contribution of the autonomous factors on liquidity. Following Equation (1) the aggregate is calculated as:

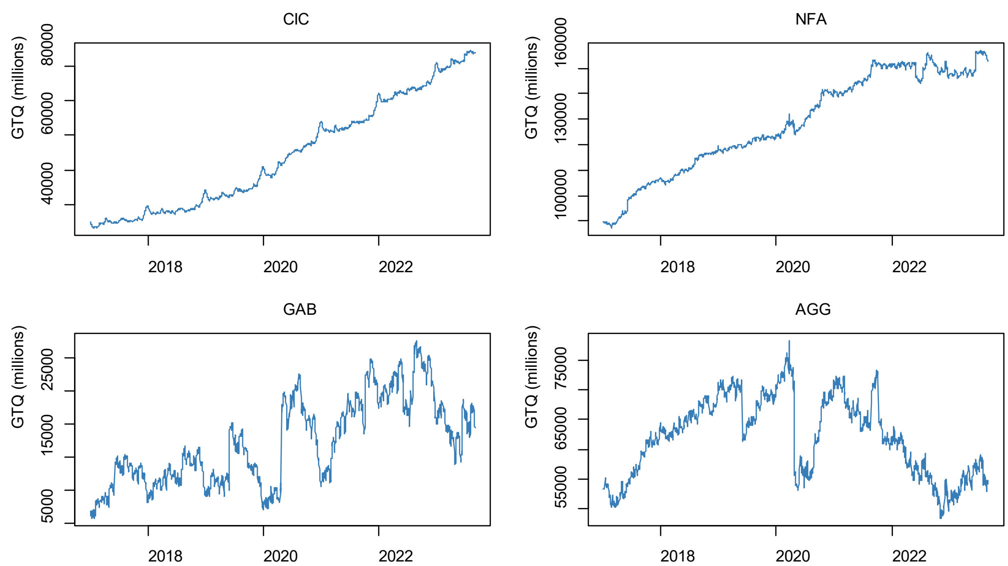
$$AGG = NFA - CIC - GAB. \qquad (2)$$

Given the typical operations of central banks, and specifically OMO, for the modeling of liquidity, daily data are used and often there are long time series for each of the autonomous factors. Some central banks opt for data organized in a five-day week, while others use complete seven-day weeks.

Figure *2* provides examples of CIC, NFA, GAB, and AGG (for details, see Gallardo et al., 2024). Note that all time series are non-stationary. Daily time series can exhibit multiple seasonal patterns as well as day of the week, day of the month, dominant events like payday, and day of the year. We show this in Figure 3, which plots the various seasonal cycles for CIC. The trend of the time series is first removed using classical decomposition, and for each period of the seasonal cycle, the minimum, 10 percent, 25 percent, median (50 percent), 75 percent, 90 percent empirical quantiles, and maximum are plotted (Kourentzes, 2023). Observe that in all cases, a seasonal pattern emerges. As the number of days in a month is not constant, it is more convenient to consider the seasonal cycle of days in a quarter.

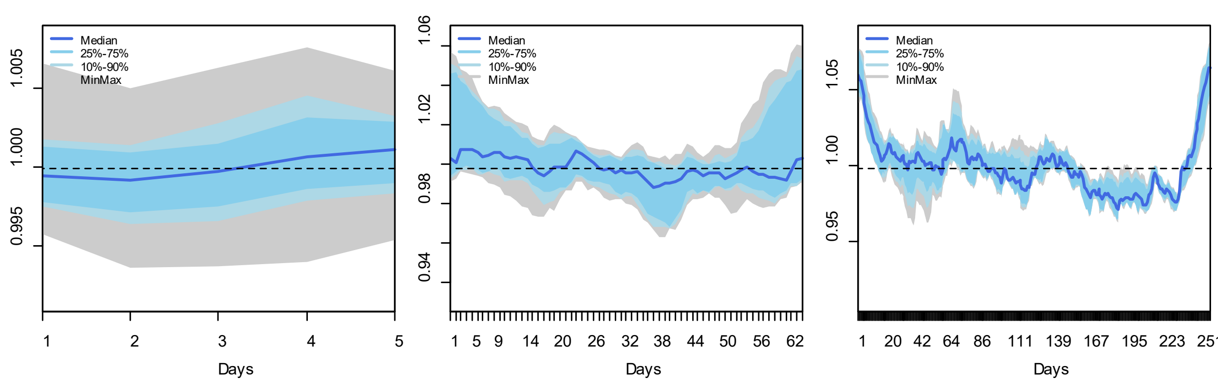
The CIC, in agreement with the literature, exhibits fairly canonical time series patterns, which should be possible to approximate well with time series models. NFA and GAB demonstrate a mix of standard time series components and additional exogenous effects. For instance, government actions may affect them, and this information cannot be retrieved from past observations before the government action. This necessitates high-quality exogenous information (Williams, 2010), with the net government balance often considered by central bankers to be the most challenging series to forecast. The most significant factor impacting liquidity in the NFA account is unsterilized foreign exchange intervention. In countries with a floating exchange rate regime, foreign exchange operations are less common, leading to less drastic changes in the NFA series. NFA rarely exhibits seasonality. GAB sometimes exhibits strong seasonal patterns in both expenditure and revenue decomposition, such as biweekly payrolls for civil servants and quarterly corporate tax collections (Gray, 2008). CIC is not immune to these effects, but typically we observe smaller effects. Moreover, there may be additional effects due to special calendar events, such as holidays and celebrations, as well as structural breaks that may be due to policy effects or extraordinary long-lasting effects, such as the COVID-19 pandemic.

Figure 2. Examples of Autonomous Factors Series



Source: Gallardo et al., (2024).
Note: Guatemalan quetzal (GTQ).

Figure 3. Plots of Seasonal Cycles in CIC



Source: Gallardo et al., (2024).

IV. Forecasting Models

In this section we briefly describe the various modeling alternatives that are provided by the proposed liquidity forecasting framework. As the objective is to provide forecasts of the short-term liquidity movements to support OMO, exogenous information such as macroeconomic variables are not considered, as changes in the short-term are dominated by stochastic trends, seasonal components, and other factors such as holidays, payroll dates, subsidy dates, regular foreign exchange auctions, etc. (El-Hamiani Khatat 2018). Macroeconomic variables are expected to change much slower than the usual range of forecast horizons that are required by OMO and, likewise, such variables are not typically sampled at high frequencies. Nonetheless, this may impact the quality of the models, and we discuss this further below.

To support OMO, the forecasting horizon shall align with the actual OMO schedule. For instance, if OMO is scheduled on Wednesday every week, the forecasting exercise should be performed on Tuesday, generating 7-day (or 5-day if the weekend is skipped) forecasts ahead from Wednesday to next Tuesday. If the OMO happens biweekly, the forecast horizon should be 14 days (or 10 days if weekends are skipped). However, the forecasters may need to be flexible with the forecast horizon due to special factors such as delayed delivery of data or ad-hoc fine-tuning at the end of the reserve maintenance period. When setting an appropriate forecasting horizon, one should also bear in mind that forecasting errors increase over time.

The various models can be grouped into two categories: (i) extrapolative time series models; and (ii) volatility models. OMO requires point forecasts, nonetheless, in all cases we provide predictive distributions to help analysts weigh the uncertainty that the different forecasts entail. Moreover, we argue that there are benefits in communicating the predictive distribution to counterparties and that it can lead to better decisions when costs may be asymmetric. These points are expanded upon in later sections. Similarly, volatility models are useful to capture any structure in the predictive distribution beyond the mean, particularly when the latter exhibits minimal or no structure, as is often the case for NFA. Below we provide a brief overview of the models, with relevant references. The reader is recommended to look at Ord et al. (2017) and Hyndman and Athanasopoulos (2021) for a more thorough treatment of the various models.

Equation (2) describes the connection between autonomous factors. Observationally, it will always hold. However, when we forecast the autonomous factors separately there is no expectation that this will be the case, and we anticipate that reconciliation errors, i.e., deviations from that equation, will occur. As these reconciliation errors are connected with the forecast errors, we can take advantage of these to improve forecasts further. Conversely, if there were no forecast errors, the forecasts of the autonomous factors would be coherent with no reconciliation errors. This is done with the use of forecast reconciliation methods (Athanasopoulos et al., 2023).

4.1 The Naive Method (Random Walk)

The naive method assumes that the time series has no structure, while at the same time requiring no parameter estimation or any other modeling choices. The forecast is generated as:

$$\hat{y}_{t+h} = y_t,$$

where y_t is the observation of the period t, $\hat{y}_{t+h}$ is the forecast for the period $t + h$, and h is the forecast horizon. To understand why the model assumes that there is no structure in the data, we can refer to the underlying random walk model:

$$y_{t+1} = y_t + \varepsilon_t,$$

$$\varepsilon_t = y_{t+1} - y_t,$$

where ε_t are i.i.d. normally distributed innovations, and therefore the change in the value of y_t is only due to the random ε_t. As the future random innovations are unknown, these are assumed to be zero when we generate forecasts, resulting in all forecasted values to be equal to the last observation of the time series. Arguably, this is an inappropriate model to forecast liquidity, but it is a useful benchmark. Any more complex models must outperform the naive in forecasting performance to be considered predictively valuable. This can guard us against overfit models and demonstrates an active preference toward simpler models that are more transparent. Therefore, within the proposed forecasting framework the naive is used as an upper bound of acceptable forecasting errors.

A helpful modification of the naive is its seasonal counterpart, where instead of repeating the last observation, the last seasonal period is repeated:

$$\hat{y}_{t+h} = y_{t-s+h},$$

where s is the seasonal period, corresponding to the number of days in the week, for instance, 5 when weekends are excluded. The seasonal naive is a useful benchmark for highly seasonal time series.

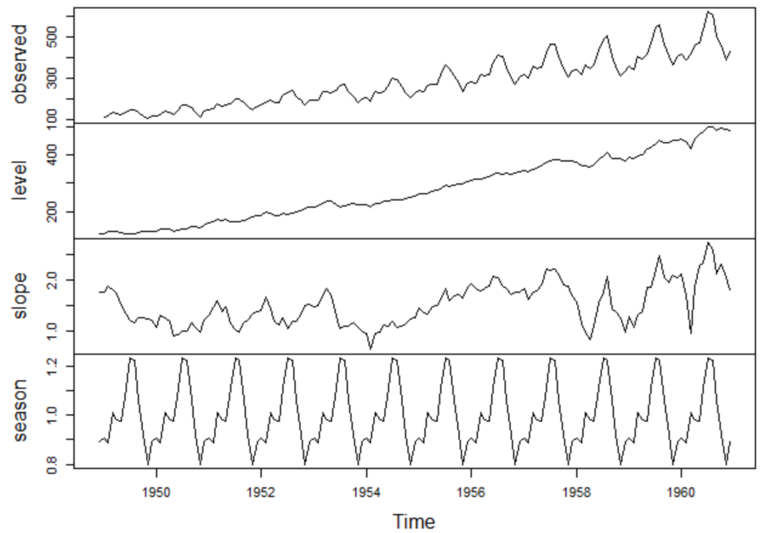
4.1.2. Exponential Smoothing Family of Models

Exponential smoothing models operate by modeling the time series as a collection of patterns, namely level, trend, and seasonality. Usually, exponential smoothing (ETS) is framed within a state-space model, where each component of the time is a state, and together they produce the forecast $\hat{y}_i$, as:

$$y_i = f(\mu_i, \varepsilon_i)$$
$$\mu_i = g(level_i, slope_i, season_i)$$

The functions $f(\cdot)$ and $g(\cdot)$ can be either additive, multiplicative, or have some mixed form. Figure 4 provides an example of the decomposition of a time series into separate components by exponential smoothing. Observe that the level, slope, and season components together can explain most of the time series, with any unexplained part attributed to the noise component. The level tracks the local mean of the time series, while the slope models how the level increases or decreases over time (e.g., a slope of +2 suggests an upward movement by two units per period). Finally, the season component models any periodic patterns in the data. Not all time series require all components to be modeled, as some may be absent.

Figure 4. Decomposition of an Observed Time Series by Exponential Smoothing



Source: IMF staff calculations.

In the fully additive case, the model becomes:

$$y_i = \mu_i + \varepsilon_i$$
$$\mu_i = level_i + slope_i + season_i$$

Each of the states ($level_i$, $slope_i$, and $season_i$) is structured similarly. For example, the additive $level_i$ is:

$$level_i = level_{i-1} + \alpha e_{i-1},$$

where α is a smoothing parameter between 0 and 1, and e_{i-1} is the previous period error. Intuitively, this equation suggests that the current level estimate is updated by α times the last error. Given that the error is the difference between the actuals (y_i) and the forecast ($\hat{y}_i$) for the case of exponential smoothing that has only an additive level, the model can be written in two alternative forms to help explain its function:

$$y_i = \mu_i + \varepsilon_i$$
$$\mu_i = level_i$$
$$level_i = level_{i-1} + \alpha e_{i-1}$$

or equivalently:

$$y_i = \mu_i + \varepsilon_i$$
$$\mu_i = level_i$$
$$level_i = \alpha \cdot actuals_{i-1} + (1-\alpha) level_{i-1}$$ [3]

[3] the level is a simple distributed lag of realized values.

The second set of equations suggest that the smoothing parameter α decides by how much to update the previous level with the last observed actuals. Noting that $0 < \alpha < 1$, a percentage contribution interpretation becomes possible. For example, if $\alpha = 0.2$, the last estimated level is updated by 20 percent of the last observation. All other states operate similarly, requiring an additional parameter for each additional state, and a model may have any of these states on their own or together. This provides 30 possible ETS models. However, typically we restrict the selection to a subset of those (see Table 2), as some models can be unstable. For instance, combining an additive trend and multiplicative seasonality can result in calculation issues when the additive trend pushes the observations to zero or negative values.

Table 2. Subset of Permitted ETS Models

Model component	Specification									
	1	2	3/4	5/6	7	8	9/10	11/12	13	14/15
Error										
Additive	✓		✓		✓		✓			
Multiplicative		✓		✓		✓		✓	✓	✓
Trend										
None	✓	✓			✓	✓			✓	
Additive/damped			✓	✓			✓	✓		✓
Season										
None	✓	✓	✓	✓						
Additive					✓	✓	✓	✓		
Multiplicative									✓	✓

Source: Kourentzes and Athanasopoulos (2019).

The estimation of the model parameters is done using maximum likelihood. This also facilitates the calculation of information criteria that enable automatic model selection, such as the Akaike Information Criterion (AIC). A detailed exposition of ETS is provided by Ord et al., (2017) and the theoretical underpinnings of the state space model are discussed by Hyndman et al., (2008). ETS models can be further modified to include exogenous regressors (Kourentzes and Petropoulos, 2016), which can help model additional factors that are relevant for liquidity forecasting, such as calendar events.

4.1.3. Autoregressive Integrated Moving Average Models

The Autoregressive Integrated Moving Average (ARIMA) family of models is a flexible class of models used for time series forecasting in a wide range of settings. In general, the ARIMA model is defined as:

$$(1 - \phi(B))(1 - B)^d y_t = (1 + \theta(B))\varepsilon_t$$

Here B is the backshift operator that lags a variable, i.e., $By_t = y_{t-1}$, $B^2 y_t = y_{t-2}$, etc. The order of differencing d is typically equal to 1 (or in rare cases 2) for non-stationary series and 0 for stationary series. The term $\left(1 - \phi(B)\right) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is known as the autoregressive polynomial (or order p) and the term $\left(1 + \theta(B)\right) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^p$ is known as the moving polynomial (or order q). The term ε_t is a random innovation term. The nomenclature ARIMA(p,d,q) is used to describe an ARIMA model. For example, an ARIMA model with $p = 2$, $d = 1$ and $q = 2$ would be referred to as an ARIMA(2,1,2) model.

An important extension to ARIMA models is the seasonal ARIMA (SARIMA), which allows the modeling of patterns that repeat themselves every m observations. In general, SARIMA take the form:

$$(1-\phi(B))(1-\Phi(B^m))(1-B)^d(1-B^m)^D y_t = (1+\theta(B))(1+\theta(B^m))\varepsilon_t$$

where P, D, and Q are the orders of the seasonal autoregressive component, seasonal differencing, and seasonal moving component. The nomenclature ARIMA(p,d,q)(P,D,Q)[m] is used to describe such models. Seasonal ARIMA models of this form are only capable of explicitly capturing one seasonality of periodicity m. It is possible to extend them by introducing additional seasonal differences and polynomials, although this can quickly become very expensive in terms of data and complicate the identification of model orders substantially. An alternative approach to include additional seasonal cycles is by including regressors (ARIMAX). ARIMAX models are also useful to incorporate other information in the models, such as holidays, paydays, etc., that are relevant for liquidity forecasting. Details about the ARIMA model family can be found in Ord et al. (2017). Note that RegARIMA and ARIMAX are closely connected, often being identical or simply implying estimation methodology differences.

For the specification of ARIMA we follow the methodology proposed by Hyndman and Khandakar (2008). In brief, they propose to first make the time series stationary by testing for the appropriate order of differencing. Then, start with a simple model and calculate a relevant information criterion, such as AIC. Iterate the autoregressive and moving average orders by ±1 and calculate the resulting AIC. At each step, choose the model with the lowest AIC and repeat the local search until the information criterion cannot be improved further.

Finally, it is easy to show algebraically that there are equivalences between the purely additive ETS models and ARIMA (Hyndman et al., 2008), while it is simple to obtain multiplicative forms by applying ARIMA on log-transformed data. Mixed-form ETS models are not encompassed by ARIMA. Furthermore, given the constrained modeling alternatives of ETS, it is often easier to identify a well-performing ETS model when compared to ARIMA. Therefore, we use both modeling families in the liquidity forecasting framework.

4.1.4. Trigonometric Seasonality Box-Cox, ARMA errors, Trend, and Seasonal Components

The Trigonometric Seasonality BOX-Cox, ARMA errors, Trend, and Seasonal Components (TBATS) model incorporates many of the features of the models already introduced. With TBATS, seasonality and trend are handled via exponential smoothing (using trigonometric terms for the former, see Section 4.1.5), a Box-Cox transformation is used, and ARIMA errors are incorporated. A particularly attractive feature of the TBATS model is its ability to handle multiple seasonalities, without any special treatment. For additional details, the reader is referred to De Livera et al. (2011).

4.1.5. Modeling Multiple Seasonalities with Trigonometric Encoding

In a regression context, seasonality can also be introduced in models using binary indicator variables. For example, day-of-week effects can be modeled using only four indicator variables (five-day week) of the form:

$$D_t^{(Sun)} = \begin{cases} 1 \text{ if day t is a Sunday,} \\ 0 \text{ otherwise.} \end{cases}$$

Similar indicators (or dummies) can be defined for Mon, Tue, Wed, and Thur. These indicators are then included in a vector of covariates x'_t and the ARIMA model has the same specification as before, but with y_t replaced by $y_t - x'_t\beta$. A similar modification can be made for ETS. Of course, both ARIMA and ETS can also model seasonality directly, without the inclusion of regressors.

Daily time series can exhibit multiple seasonal cycles that must be accounted for in the modeling. These include day in the week, day in the month, and day in the year, corresponding to different cyclicities in the data. This substantially complicates the creation of forecasts, as many models typically incorporate a single seasonal periodicity. Three elements are of interest in modeling multiple seasonalities: the length of the seasonal cycles, their encoding, and the efficiency of the latter, as we aim for parsimonious models. To resolve questions raised by the first element, one counts how many days are in each periodicity. For example, there are five days in the week (without weekends). However, day-of-the-month seasonality is more challenging as months have a different number of days. To overcome this, quarterly seasonality is used, as a quarter contains a fixed number of weeks, and, by extension, days.

The multiple seasonal cycles are encoded using trigonometric indicator variables. Given the length of a season of s periods, s/2 pairs of trigonometric variables are constructed, with $i = 1, \dots, s/2$:

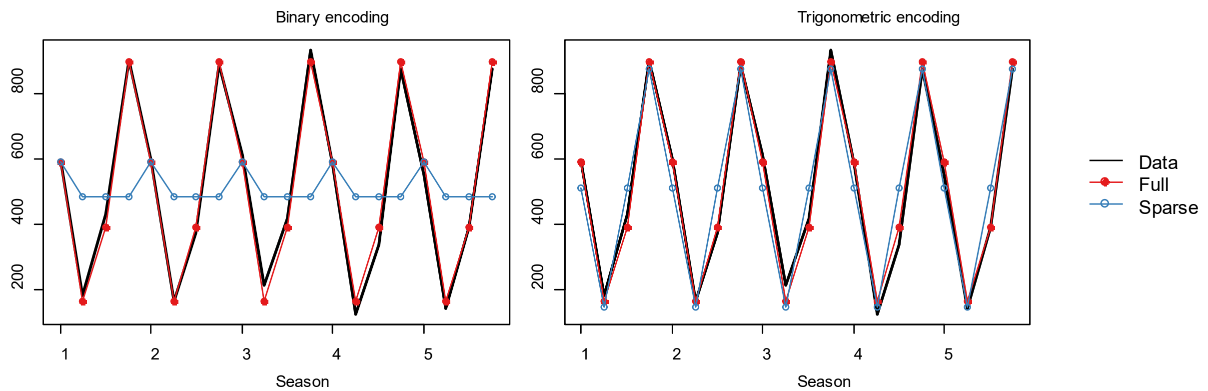
$$d_i = cos\left(\frac{2i\pi t}{s}\right),$$

$$d_{i+s/2} = sin\left(\frac{2i\pi t}{s}\right),$$

where $t = 1, \dots, n$ (with n being the sample size). When s is an odd number, $s/2$ is rounded up to the closest integer. This encoding is mathematically equivalent to using s binary indicator variables, in which case, each binary indicator would encode the level of a particular day in the season (Ghysels and Osborn, 2001). Note that one of the indicators will correspond to a constant, resulting in $s - 1$ informative indicator variables.

A major advantage of trigonometric encoding over its binary counterpart is how it behaves when indicators are removed to achieve sparsity. Consider a time series with a single seasonality and $m = 4$, i.e., a quarterly seasonality. This would require 3 binary indicators, or 3 trigonometric indicators (with the 4^{th} being a constant and therefore excluded, as similarly, the 3 binary indicators assume the presence of a constant in the model). A sparse encoding would use less than 3 indicators, and in our example, we retain only the first indicator. Figure 5 visualizes the outcome. With full formulation, the resulting model fit is identical from both binary and trigonometric encodings. With sparse formulation, the binary encoding correctly captures the quarter for which the indicator is provided, while for the rest, the mean value of the time series is estimated (from the constant in the model). In the case of trigonometric encoding, the model is still able to approximate a simplified version of the seasonality. Eliminating terms with trigonometric seasonality lowers the quality of the approximation but is still able to model the whole length of the seasonal cycle. Therefore, with trigonometric encoding we can produce more parsimonious models that are easier to estimate.

Figure 5. An Example of Sparse Binary and Trigonometric Seasonality Encoding



Source: IMF staff calculations.

A disadvantage of the trigonometric seasonality is that the corresponding coefficients are not interpretable in the same way as with the binary encoding. For the latter, the coefficient signifies the vertical shift from the coefficient of the constant term in the model to obtain the seasonal value. With the trigonometric encoding, the coefficients should be interpreted as a Fourier decomposition of the time series.

For the liquidity forecasting framework, we propose to use binary encoding for the day-of-the-week seasonality, as it requires minimal additional model terms and allows direct interpretation of the resulting coefficients. For the remaining seasonalities we recommend using a trigonometric representation, which is then made sparse. This is done using a regression, which can either rely on stepwise selection with AIC, or, preferably, lasso regression. The lasso regression is tasked to find a good compromise between how well the model fits the data and its complexity as measured by the number of parameters it has. Models with more parameters (and therefore input variables) are better at modeling the observations, but can potentially overfit, capturing the randomness in the time series instead of the underlying structure. Lasso offers a better search of the model space than stepwise regression. More details about the lasso regression can be found in Ord et al., (2017) and Kourentzes and Sagaert (2018).

To keep the complexity of the regression modeling low, first the trend of the time series is removed by subtracting a centered moving average from the time series (Ord et al., 2017). The centered moving average simply calculates the average of all values within a season that effectively models the trend in the time series. This is subtracted from the data, and the residuals are then modeled with different trigonometric indicator variables as explanatory variables.

Finally, it should be stressed that the seasonality modeled using binary or trigonometric indicators will be deterministic. This means that the shape of the seasonality will not evolve over time. In contrast, the seasonality modeled with autoregressive and moving average terms, in ETS, ARIMA, and TBATS, is stochastic, which means that its shape can evolve over time. A detailed discussion of the differences between the two and ways to identify them is provided by Ghysels and Osborn (2001). For the objectives of the liquidity forecasting framework, we assume that longer period seasonalities evolve slower, such as the annual seasonality, and therefore there is little loss of accuracy by modeling it as a deterministic seasonality. On the other hand, for the day-of-the-week seasonality this can be a strong assumption. Therefore, this seasonality is preferably treated by the models directly to retain its stochastic nature.

4.1.6. Level Shifts

Various effects can introduce level shifts to the autonomous factors. In recent years, in many countries, the effect of Covid-19 resulted in very strong level shifts in liquidity. A level shift can be permanent or transient, with the time series of interest returning to its original level. For models that can use regressors (ETS and ARIMA) a level shift can be added with a binary indicator variable:

$$\mathrm{D}_t = \begin{cases} 1 \text{ if t occurs after the level shift,} \\ 0 \text{ otherwise.} \end{cases}$$

If the analysts identify a level shift as transient, one can model this with two binary indicators, one modeling the first level shift, and another modeling the termination of the shift. More economically, if the resulting level is the same as the one before the shift, one can use a single binary indicator which takes values of 1 only for the duration of the transient shift.

Note that models with autoregressive terms will tend to smooth the level shift over multiple periods, depending on the autoregressive structure. This can help us model shifts that are happening over a number of periods, without the need to include more complex encoding.

4.1.7. Calendar and Special Events

Both ETS and ARIMA models can be augmented by adding regressors in the same fashion as with conventional regression modeling. Taking the series of currency in circulation (CIC) as an example, Cabrero et al. (2009) identified the deterministic structure of the RegARIMA model as follows:

i. Fixed and Moving Festivals: These include holidays that have either fixed dates or change annually, impacting consumer behavior and cash demand.

ii. Intramonthly Effects Related to Payrolls: These are regular fluctuations associated with payroll cycles that influence cash withdrawals and deposits.

iii. Trading Day (Intraweek) Effects: This accounts for variations in economic activity based on different days of the week, reflecting trading patterns and financial transactions.

These effects can be incorporated into the ARIMA and ETS models using binary indicators. We distinguish a number of options in the modeling of these. First, these effects may be restricted solely to the period of the special event. Second, if the model includes autoregressive terms, their effect can be spread to the following periods. Third, we may want to introduce the leading and training effects of a special event. This can be done with the introduction of multiple binary dummies, giving us full control of the modeled response. Alternatively, a more economical approach is to use a parabolic indicator. Some days before the holiday or event, the indicators start to increase quadratically and reach the maximum of the parabolic curve with a value of 1. Then the indicators begin to decay quadratically and become 0. Such encoding can best capture the real movement of liquidity caused by CIC. The public will start withdrawing cash to prepare, for example, for a celebration right ahead of the holiday, while after the holiday the demand for money will dwindle. These indicators are calculated as

$$D_t = \max\left(0, 1 - \left(\frac{t - t_{holiday}}{7}\right)^2\right),$$

where the effect is spread across seven periods. In some cases, it may be advisable to use two parabolic indicators, centered across two consecutive periods, to economically capture complex responses.

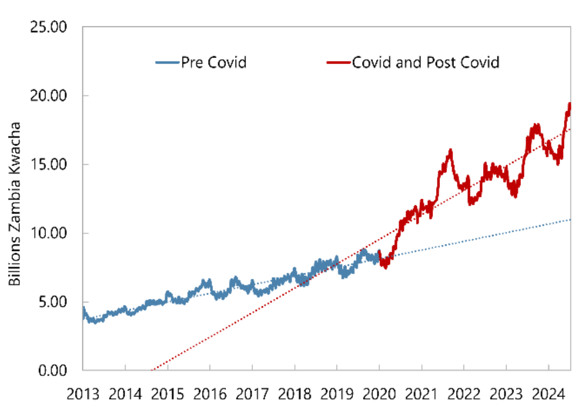
4.1.8. Augmented ETS and ARIMA Models

The various modeling options described in Sections 4.1.5–4.1.7 can be incorporated into ETS and ARIMA models, allowing them to model the multiple complexities of the autonomous factors. Figure 6 provides examples of using these options from cases in Guatemala and the United Arab Emirates (UAE). A potential risk is that analysts oversaturate the models with additional regressors, which may or may not be informative. This can make the model cumbersome to use and difficult to estimate. To avoid this, we recommend evaluating the inclusion of the various indicators using lasso or stepwise based on AIC regression. In the liquidity forecasting framework all regressors, apart from those modeling level shifts, are evaluated in this way before being introduced to the models. Seasonality indicators and special event regressors are evaluated separately, in two consecutive steps.

Finally, for the case of ARIMA, we consider alternatives with and without seasonality when considering the additional regressors. This corresponds to modeling the day-of-the-week seasonality as stochastic, deterministic, or mixed. The choice is resolved by comparing the AIC of the resulting ARIMA.

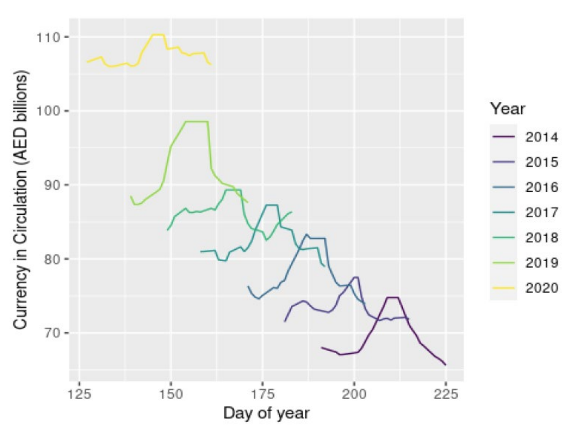
Figure 6. Modeling Deterministic Structure as Regressors: Example of CIC

Permanent or transitory structural breaks may exist due to economic, banking, or exchange rate crises. Continuous indicators can be used to codify structural breaks.



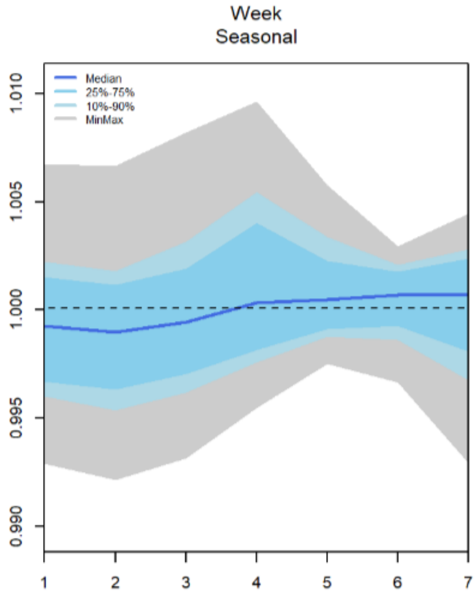
Source: Haver (2024)

Increasing demand for currency may start right before a holiday and gradually decrease right after, forming a parabolic arch. The figure below presents the case of Eid-al-Fitr in the UAE, and the holiday may be moving.



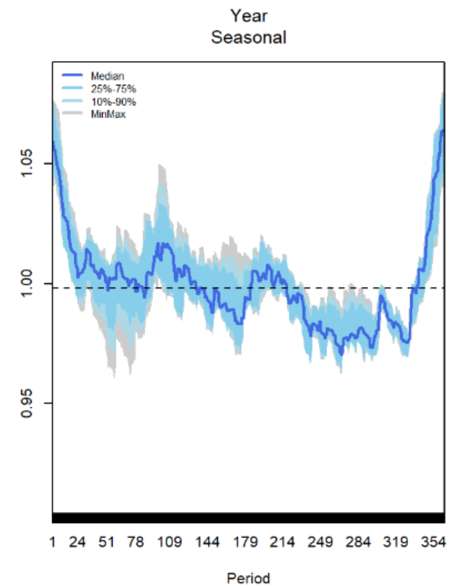
Source: El Gemayel (2022)

An example of Guatemalan CIC demand, which tends to increase close to weekends. This may be because wages are paid on Fridays, or the public are withdrawing cash for weekends. Indicator variables can be used to model such effects.



Source: Gallardo (2024)

Guatemalan CIC also demonstrate Intra-yearly seasonality, which reflects a huge increase of CIC due to the year-end holiday season. Trigonometric terms will be the more parsimonious way to model the intra-yearly effect.



Source: Gallardo (2024)

4.2 Volatility Models

Volatility models are appropriate for forecasting series with high volatility. Normally, these models will be applied to forecast NFA. Unlike CIC and GAB, the NFA series is usually volatile with little structural information to be modeled. While the mean of NFA seems to follow a random walk, the first difference does exhibit conditional heteroskedasticity, indicating that generalized autoregressive conditional heteroskedasticity (GARCH)-type models probably will be more appropriate for NFA forecasts. Several conditional volatility models are fitted to obtain the probabilistic forecasts of NFA, which will be later transformed into point forecasting using bootstrapping simulation.

The most popular family of conditional volatility models is the GARCH model. The variance is modeled as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \, \sigma_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \, e_{t-i}^2$$

The proposed framework also makes available two variations of GARCH, which allows for asymmetric effects of shocks on volatility: eGARCH and gjrGARCH. The specification of the exponential GARCH model (eGARCH) is given by:

$$\log\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \log\sigma_{t-i}^2 + \sum_{i=1}^{q} \alpha_i \, g(\epsilon_{t-i}).$$

where $g(\epsilon_t) = \theta\epsilon_t + \lambda(|\epsilon_t| - E(|\epsilon_t|))$. An advantage of this specification is its asymmetry, since the sign and magnitude of innovations have different effects on the variance.

The gjr (Glosten-Jagannathan-Runkle)-GARCH specification is given by:

$$\sigma_t^2 = \omega + \delta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \phi\epsilon_{t-1}^2 I_{t-1}$$

where $I_{t-1} = 0$ if $\epsilon_{t-1} \geq 0$ and $I_{t-1} = 1$ if $\epsilon_{t-1} < 0$. Like eGARCH, this specification allows for asymmetric effects. For more details on the volatility models the reader is referred to Tsay (2005).

Similar to the extrapolative time series models, we recommend using the random walk as a benchmark for the volatility models.

4.3 Hierarchical Reconciliation Methods

To support OMO, the main quantity of interest is the net liquidity injection, AGG = NFA – CIC - GAB. One approach would be to forecast each autonomous factor separately and aggregate them, which is common practice in many central banks. An alternative would be to directly forecast AGG. As forecasts are imperfect, we anticipate that for a set of forecasts $\widehat{AGG}_t$, $\widehat{CIC}_t$, $\widehat{GAB}_t$, and $\widehat{NFA}_t$ the following will be true:

$$\widehat{AGG}_t = AGG_t + e_{AGG,t}, \quad (3)$$

$$\widehat{CIC}_t = CIC_t + e_{CIC,t}, \quad (4)$$

$$\widehat{GAB}_t = GAB_t + e_{GAB,t}, \quad (5)$$

$$\widehat{NFA}_t = NFA_t + e_{NFA,t}, \quad (6)$$

and

$$\widehat{AGG}_t = \widehat{NFA}_t - \widehat{CIC}_t - \widehat{GAB}_t + \tilde{e}_t, \quad (7)$$

where $\tilde{e}_t$ is the reconciliation error, i.e., how much forecasting the disaggregate autonomous factors disagrees with forecasting directly the AGG. By replacing in Equation (7) the expressions above that include the forecast errors (3) – (6), we can easily see that the reconciliation error is connected with the forecast errors. Therefore, if we were to eliminate the reconciliation error, we can anticipate some reduction of the forecast errors. Intuitively, forecast reconciliation suggests that each forecast contains an incomplete set of information, greater than what is available to the whole problem. By reconciling the forecasts, this information is blended and therefore has the potential to improve the quality of the forecasts. In fact, Panagiotelis et al. (2021) provide a proof that the total forecast errors across the hierarchy, after reconciliation, will be reduced, irrespective of how the reconciliation is done. Note, that this does not necessarily prove that the forecasts errors of AGG, the quantity of interest, will always be better, and therefore empirical evaluation remains important.

Forecast reconciliation is mathematically achieved using a restricted forecast combination. First, we define a vector that contains all the individual bottom-level (disaggregate) observations, $\boldsymbol{b}_t =$

$(CIC, GAB, NFA)'$ and a vector that contains all individual observations across the hierarchy $\boldsymbol{y}_t = (AGG_t, \boldsymbol{b}_t)'$. We can now write a summing matrix $\boldsymbol{S}$, so that:

$$\boldsymbol{S} = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t.$$

Observe that the first row of S matches the aggregation equation (2) while the rest of the rows simply point to one of the bottom-level series. We can rewrite the expression for forecasts:

$$\widehat{\boldsymbol{y}}_t = \boldsymbol{S}\widehat{\boldsymbol{b}}_t + \tilde{\boldsymbol{e}}_t$$

that is a compact expression of Equation (7). Additionally, we can see that it resembles a regression formulation. Wickramasuriya et al. (2019) showed that the reconciliation errors can be minimized by

$$\widetilde{\boldsymbol{y}}_\mathrm{t} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}\,\widehat{\boldsymbol{y}}_t \qquad (8)$$

where $\boldsymbol{W}$ is the variance-covariance matrix of the forecast errors of $\widehat{\boldsymbol{y}}$. In practice, the estimation of $\boldsymbol{W}$ can be challenging and, instead, we opt for various approximations: (i) the OLS approximation, where $\boldsymbol{W} = \boldsymbol{I}$, the identity matrix; (ii) the structural (STR), where $\boldsymbol{W}$ is a diagonal with each element being the row-wise sum of $\boldsymbol{S}$; (iii) the weighted least squares (WLS), where $\boldsymbol{W}$ is a diagonal with each element being the MSE of the corresponding forecasting model for that row; and (iv) the MinT, which estimates the complete variance-covariance matrix shrinking the off-diagonal elements toward zero. A detailed discussion of the approximations for the variance-covariance matrix and their implications is provided by Pritularga et al. (2021), while a detailed overview of hierarchical reconciliation is given by Athanasopoulos et al. (2023).

Observe that Equation (8) can be written as $\widetilde{\boldsymbol{y}}_\mathrm{t} = \boldsymbol{S}\mathbf{G}\widehat{\boldsymbol{y}}_t$ where $\boldsymbol{G} = (\boldsymbol{S}'\boldsymbol{W}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}^{-1}$. This makes the forecast combination underlying forecast reconciliation apparent, where the forecasts across the whole hierarchy are combined with the weights in $\mathbf{G}$ to construct reconciled bottom-level forecasts, i.e., the autonomous factors, that are multiplied by $\boldsymbol{S}$ to provide reconciled forecasts for the complete hierarchy, as per $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$. Moreover, note that the forecasts in $\widehat{\boldsymbol{y}}_t$ are model independent, meaning that they can originate from any of the models outlined in Sections 4.1. and 4.2. Or otherwise, and they may be further adjusted, based on expert judgment by the analysts, before the reconciliation takes place.

As the forecasting performance of the reconciled AGG forecasts remains an empirical question, we recommend the use of two benchmarks. First, a direct forecast of the AGG without any reconciliation, and second, a direct aggregation from the three autonomous factors to the AGG (bottom-up).

V. Forecast Evaluation and Selection

The selection of the best forecast is an empirical question, for which we need to collect evidence of the predictive performance of the various models on test data. To this end, we need to define appropriate metrics and evaluation schemes. In this section, we introduce the various metrics that we use, the evaluation scheme, and then provide heuristics to aid the users of the liquidity forecasting framework.

5.1 Evaluation Metrics

The most common practice to calibrate OMO is to rely on point forecasts of the mean. The proper score for evaluating these forecasts is the Mean Squared Error (MSE) due to its quadratic loss (Gneiting and Raftery, 2007). For scaling reasons, in practice, the Root Mean Squared Error (RMSE) is preferred over the MSE for reporting. Liquidity can be subject to shocks, which are both difficult to forecast and not representative of normal conditions. Therefore, it can be desirable to consider the Mean Absolute Error (MAE), which is the proper score for the median of the predictive distribution, and is thus more robust to extreme errors, which may occur due to shocks. Both metrics measure the magnitude of forecast errors (accuracy). It is useful to measure the bias of forecasts, which captures their tendency to over or under forecast. This can be done with the Mean Error (ME). Formally these metrics are defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i),$$

where y_i is the actual value, $\hat{y}_i$ is the forecast for the same period, and n is the number of periods considered. The errors should be measured across forecasts of the same horizon, e.g., one period ahead, and then, can be further averaged to provide an overall performance measurement. All RMSE, MAE, and ME provide errors in the scale and units of the time series and cannot be used to summarize across different time series. If a scale-independent version of the metrics is required, then we recommend dividing the metrics by the in-sample RMSE or MAE (for ME as well) of the random walk. These metrics are known in the literature as scaled metrics and avoid many of the pitfalls of other scale-independent metrics. The reader is referred to Athanasopoulos and Kourentzes (2021) for more details.

Forecasts with very wide predictive distributions can imply large forecast uncertainty, and vice versa. Note that this statement assumes that the models are well calibrated, as models can provide unreasonably narrow prediction intervals, particularly when they are overfit to the training data. Forecasts are not anticipated to be equal with future observations, but, instead, to follow their trajectory closely. This is due to the randomness in all stochastic time series. Therefore, it is helpful to evaluate the forecasts probabilistically. Intuitively, this evaluation investigates whether the expected percentage of observations falls within the respective prediction intervals.

We argue that the central bank should forecast the complete predictive distribution to calibrate OMO and publish the forecast. The predictive distribution can be interpreted as a representation of the risk that a forecast carries. This risk can be due to the modeling risk, i.e., the ability of the forecasting models to approximate well the underlying data-generating process, and the inherent stochasticity of the modeled series. When using the point forecast, which corresponds to the mean of the predictive distribution, the implicit assumption is that the consequences of under-allotting the operation are the same as those of over-alloting. Note that this corresponds to the 50th percentile, the median, for symmetric distributions. More generally, if this assumption is relaxed, which is more realistic, and the under- and over-allotting have different costs, then a different percentile should be targeted, reflecting the risk preference of the bank. Communicating the predictive distribution, rather than the point prediction, provides more information to users, particularly for the associated forecast uncertainty. Users can thus factor in the forecasting risk in their decisions. Additionally, providing the predictive distribution can mitigate potential anchoring to the point forecast. We therefore recommend publishing the predictive distribution.

To assess the predictive distribution, we focus on the width of the prediction intervals. For this purpose, one can use the Mean Interval Score (MIS). The MIS penalizes forecasts for the width of the prediction intervals, as arbitrarily large intervals would include all observations, and for the number of observations that fall outside the intervals. Therefore, it prefers forecasts with as narrow intervals as possible, while containing the desired number of observations within these. For the evaluation, the 95 percent intervals are considered, i.e., covering 95 percent of the observations. The metric is calculated as:

$$MIS = \frac{1}{n}\sum_{i=1}^{n}\left((U-L)+\frac{2}{\alpha}(L-Y_i)1\{Y_j<L\}+\frac{2}{\alpha}(Y_i-U)1\{Y_i>U\}\right)$$

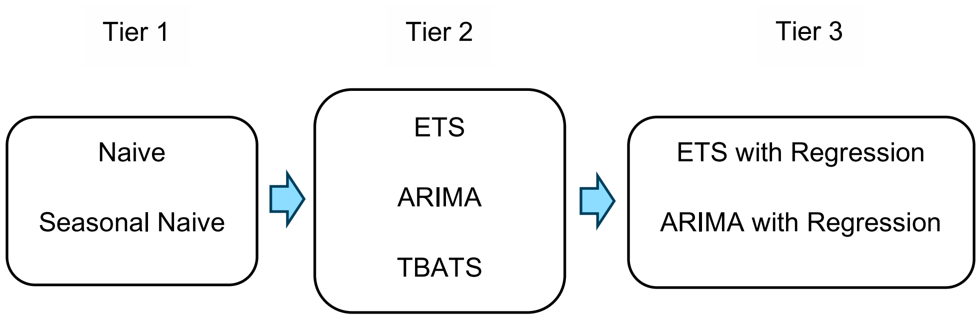
where α corresponds to the desired confidence level, e.g., 95 percent, U and L are the upper and lower prediction intervals matching α, and 1(·) is an indicator function that takes the value of 1 when its condition is true and 0 otherwise.

If a specific quantile is of interest, for example, due to an asymmetry of costs, the Pinball loss can be used instead. The function of the Pinball is similar to MIS but only for one side of the interval. Other metrics cater to different evaluations of the predictive distribution. A detailed treatment is provided by Gneiting and Raftery (2007).

5.2 Evaluation Scheme

We collect errors on a test set to identify the appropriate forecast for a time series. Naturally, there is a requirement to collect several error measurements to make a confident assessment of the performance of a forecasting model. Likewise, the optimal performance remains unknown, as the data-generating process of a time series is unknown. Nonetheless, benchmark model forecasts can help bound the acceptable performance of forecasts. To this end, we organize forecasting models in different tiers of complexity, as illustrated in Figure *7*. For similar forecasting errors, simpler models are preferable, as they require fewer modeling choices and they have less potential to overfit the data. Tier 1 includes the level and seasonal random walks that provide an upper bound on the acceptable forecasting performance. Tier 2 relies on ETS, and ARIMA in level, seasonal, or fully automatic specifications, and TBATS. All of these models, together with their specification methodologies, have been extensively researched and tested in the literature. In practice, they require minimal intervention from the analyst, and they therefore provide a set of benchmark forecasts that can capture key aspects of the liquidity time series. The last tier includes ETS and ARIMA models augmented with the various regressors discussed in Section 4. Although these models are the only ones that can capture various effects present in the liquidity time series, which are well documented in the literature, they also have the largest potential to overfit, or to be overly complex with potential estimation issues. Therefore, to provide reliable forecasts, we argue that these models should be preferred only when there is strong evidence in their favor. Similarly, for volatility forecasts we consider the Naive as a benchmark over the more complex GARCH models.

Figure 7. Forecasting Models Tiered in Groups of Increasing Complexity

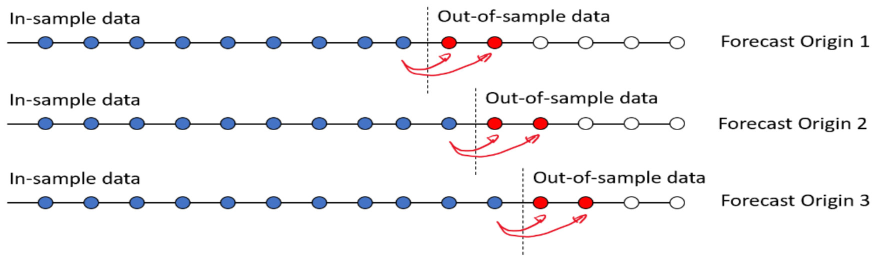


Source: IMF staff.

To collect a sufficient sample of test errors we rely on a rolling origin evaluation scheme. For this, some observations are retained aside as a test set, which is not used for model specification, and simulates unseen data. Unlike in-sample evaluation, which measures the goodness of a model, out-of-sample evaluation can indicate the predictive power, which aligns with our goal of forecasting. Out-of-sample evaluation can help mitigate the potential overfitting of forecasting models. Relying on a single out-of-sample measurement of the forecast accuracy of competing models is not adequate. Consider that each observation of the time series contains structure and noise. Relying on a single measurement makes it impossible to know whether the potentially high or low accuracy is due to modeling the structure, or due to the randomness present in the time series. However, as the noise is randomly distributed around the structure of the time series, averaging across multiple error measurements will tend to cancel out some of this randomness and evaluate the performance of the forecasts against the underlying structure.

The rolling origin out-of-sample evaluation helps to achieve this. Figure 8 provides a representation of how it operates. Using q observations as in-sample to specify the forecasting models (blue dots), forecasts for h-steps ahead are generated for the out-of-sample period (red dots). The errors from this first forecast origin are collected. In general, in the next step, the in-sample is increased by one period. Using the new $q + 1$ in-sample, the forecasting models are tuned again, and new h-step forecasts are generated from the second forecast origin. This process is repeated until all the available out-of-sample have been exhausted. If u is the size of the out-of-sample set, then $u - h + 1$ rolling forecasts and respective errors are generated. The reader can find more details about the advantages of rolling origin evaluation in Tashman (2000). Note that the rolling origin scheme is also referred to as cross-validation over time.

Figure 8. The Out-of-sample Cross-validation Scheme



Source: IMF staff.

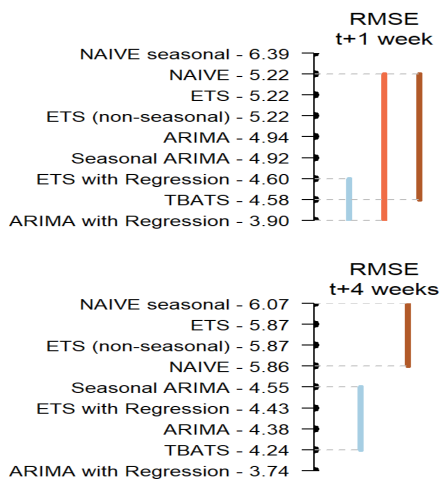
5.3 Testing the Forecasting Performance

Having collected forecast errors, the next question that needs to be resolved is what constitutes a sufficient difference between the performance of forecasts, to prefer a model in a higher tier. We rely on non-parametric statistical testing, namely, the Friedman and the post-hoc Nemenyi tests, which can facilitate multiple comparisons without the need for repeated pairwise testing. The non-parametric nature of the tests is convenient as it does not place any distributional assumptions on the forecast errors. However, as such, non-parametric tests are typically weaker than their parametric counterparts, requiring more samples. However, this is often not an issue in liquidity forecasting, as there is an abundance of data.

The Friedman test first evaluates whether at least one of the methods is statistically different from the rest. The Nemenyi test then groups the methods into subgroups. Briefly, the Nemenyi test calculates the mean rank for each forecasting model and a critical distance. Any forecasting model that falls within the mean rank ± critical distance is grouped together. The simpler model within the group with the lowest mean rank is the preferred forecast. Additional details for these non-parametric tests can be found in Hollander et al. (2003), and an example of how to apply them in forecasting comparisons in Kourentzes and Athanasopoulos (2019). Derivations of the critical values for multiple models and various sample sizes are provided by Kourentzes (2023).

Figure 9 provides an example of the use of the non-parametric tests. The vertical axis represents the mean rank of the various forecasts. At each origin of the rolling origin evaluation the errors of the different models are collected and ranked, with the lowest rank being best. This is done over all origins in the test set and the mean is calculated. Forecasts that are connected with vertical lines belong to the same group. From a group, the simplest method is preferred. In the example, we see evidence that ARIMA with Regression is performing better than the rest of the forecasting models across multiple forecast horizons of interest. If we focus on the t+1 week horizon, ARIMA with Regression, ETS with Regression, and TBATS are all grouped together as best performing.

Figure 9. An Example Nemenyi Test Comparison for Three Different Target Forecast Horizons



Source: IMF staff calculations.

Balancing across multiple error metrics and comparisons can become cumbersome. Here we provide some recommendations to help analysts. Note that these recommendations are heuristics and are based on our experience with OMO and liquidity forecasting, and alternative rules can be formulated.

- If the focus is on the point forecast of the mean, we recommend that the analyst first rely on RMSE comparisons. If a clear winning forecast does not emerge, then we recommend consulting MAE results, and eventually ME. MIS can provide additional evidence, as it helps gauge the associated uncertainty of the forecasts.

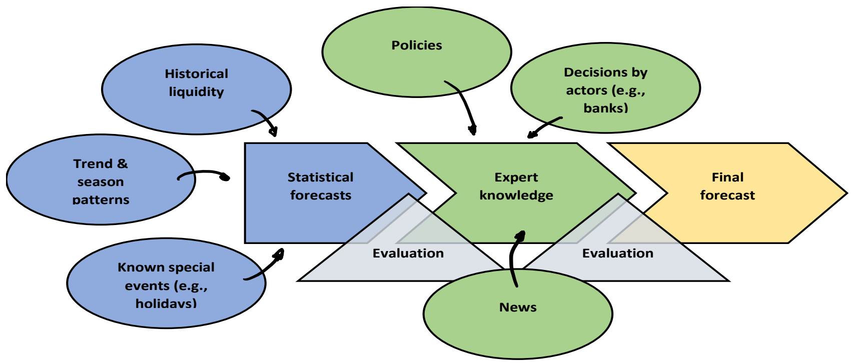
- If the focus is on the predictive distribution, MIS becomes necessary and should be consulted first. Similarly, MIS is important for assessing volatility forecasts.

Note that the ranking of the forecasts may change across metrics, hence our recommendation is to weight more the RMSE, and then the MAE, with the rest following. Moreover, although the forecasting framework provides summary forecast errors, we recommend relying on the comparisons offered by the statistical tests to support the choice of the appropriate forecast.

5.4 The Forecasting Process: Statistical Forecasts and Expert Judgment

A complete short-term liquidity forecasting framework should be a process of combining statistical modeling and expert adjustments originating from communications with related stakeholders. The expectation is that statistical forecasts can automatically make use of well-structured information to provide baseline forecasts. The structured information includes historical observations and indicator variables for holidays and special events. Additionally, through better communications with the counterparties of the central bank, additional information such as budget revenues, gold purchase plans, etc., can be obtained, which can be used to reduce the uncertainty of forecasts. This information can be simply added to the forecasts (see Figure 10). More importantly, at each layer of forecasting, it is critical to evaluate performance to guide any improvements.

Figure 10. A Schematic of the Process for Liquidity Forecasting

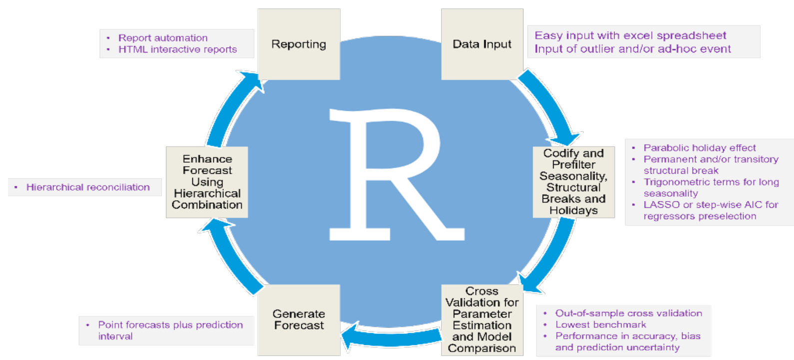


Source: IMF staff.

The literature has investigated in detail the efficacy of expert-judged adjustments in forecasts (Lawrence et al., 2006; Arvan et al., 2019). The main findings are that experts are influenced by various behavioral biases that can result in inconsistent performance of adjustments. This motivates our proposed separate evaluation of the performance of expert adjustments, to provide feedback to analysts and guide them to improve future adjustments. Sroginis et al., (2023) find that experts can be overwhelmed by contextual information, which can lead to erroneous adjustments. To mitigate this, we recommend analysts use the augmented ETS and ARIMA to incorporate all structured information in the models and limit their adjustment to unstructured information that cannot be easily incorporated into a regression model. Another helpful practice has been to decompose the expert-judged adjustment into its constituents, where the experts have to account for all of the different factors that make up their total adjustment. This has been found to help mitigate overly optimistic adjustments. The reader is referred to Sroginis et al. (2023) and Arvan et al. (2019) for a detailed overview of the findings on behavioral adjustment in the literature.

The proposed framework for liquidity forecasting is implemented in R, an open-source mathematical programming language. Figure 11 provides an overview of the forecasting workflow. Once the analysts have prepared the data of the autonomous factors, and codified the various holidays, level shifts, and other events, the implementation provides various automations to obtain rolling origin errors for the various described forecasting models. Aided by the reports produced by the framework, the analyst is asked to choose the best forecasts for each of the time series. Forecasts are then generated and reconciled. The analyst is provided with the relevant reports to aid in the selection of which hierarchical reconciliation is the most appropriate, leading to the final forecast output.

Figure 11. The Workflow for Statistical Modeling of Short-term Liquidity



Source: IMF staff.

The implementation offers various automations to enable the analysts to conduct the analysis easily. It is not required to reconfigure the models at every forecasting cycle; however, it is strongly advised that this be done periodically. The frequency at which reconfiguration is done should be a balance between the workload of the analysts and computational resources.

5.5 Limitations and Extensions

The framework incorporates static and dynamic forecast combinations of pools of models. Forecast pooling is considered an advantageous step in the forecasting process, aiding in both mitigating modeling uncertainty and improving forecast accuracy (Elliot and Timmerman, 2013; Kourentzes et al., 2019). Nonetheless, these are not currently part of the main forecasting workflow. Our experience with applying the liquidity forecasting framework in various countries is that the forecast combinations offer marginal improvements for the additional modeling complexity. Moreover, as the hierarchical reconciliation is a part of the main workflow, the final forecasts incorporate a form of forecast combination. Our experiments have indicated that there are diminishing returns to using two "layers" of forecast combination, which has guided our choice to retain the forecast combination and pooling options as optional. During the COVID-19 pandemic we found evidence that the additional dynamic forecast combination was helpful, as the various forecasting models were failing in complimentary ways.

The liquidity forecasting framework does not automate any machine learning or artificial intelligence methods. We have trialed the use of shallow neural network and random forest forecasts. We did not find evidence of forecasting performance benefits, particularly given the additional computational cost. In recent years, global forecasting methods have demonstrated good performance. A potential implementation for the liquidity forecasting framework would be to train models across liquidity data from multiple countries, to leverage the benefits of global learning. Although this entails data collection complications, it is a fruitful avenue for future research.

Likewise, another modeling methodology that has the potential to further improve the forecasting performance of the models is the use of Temporal Hierarchies (Athanasopoulos et al., 2017). Much like the forecast reconciliation across the autonomous factors, Temporal Hierarchies enable reconciliation over time scales. Beyond the forecasting accuracy gains, this modeling approach can help bridge the gap between short- and long-term liquidity forecasting. At more aggregate time scales, it is simpler to generate long-term forecasts, but also to incorporate additional macroeconomic variables. The contained information can supplement the short-term forecasts in a mechanism similar to the one described in for the reconciliation of forecasts across autonomous factors. Kourentzes and Athanasopoulos (2019) propose a cross-temporal formulation that incorporates the reconciliation used here with temporal hierarchies in a single modeling step and could be a useful extension of the liquidity forecasting framework.

VI. Technical Assistance Approach

Below we provide a recommended process for using the liquidity forecasting framework in Technical Assistance (TA) missions:

i. At least six weeks before the mission (T-6), request the time series. Ideally, the TA recipient should provide at least three years of central bank daily balance sheet data and convert them into a liquidity table, showing autonomous factors, banks' reserves at the central bank, and monetary operations. In the absence of this extensive set of data, use the standardized data request in Appendix I. Share with the authorities the mission requirements in terms of software (R and RStudio) and staffing (staff with coding capacity).[4]

    - At T-6, prepare the data for modeling. Follow up with the authorities on any problems encountered in preparing the data. Run the models. Prepare a (PowerPoint) presentation including: (i) time series descriptive analysis and model configuration; (ii) model selection statistics; (iii) point forecasts output with predictive intervals; (iv) possible forecast enhancement with reconciliation; and (v) alignment with existing liquidity table. The standardized output template is presented in Appendix II.

ii. At T-2, the standardized results are presented to the TA recipient virtually. A Q&A session is organized. The team will explain that the in-person mission will mainly consist of a workshop with the staff of the TA recipient running the model on their own devices. The TA recipient is reminded of the mission requirement, including staffing for the workshop and software.

iii. From T-2 to T-0, a draft report is prepared and circulated internally in MCMCO for preliminary comments.

iv. At T-0, the draft report is shared with the TA recipient for its comments (due before the end of the mission). The workshop session starts. The workshop consists of: (i) set-up of infrastructure and working environment with relevant packages; (ii) R coding essentials; (iii) customization of model configurations; (iv) execution of forecasting codes; (v) interpretation of the results; (vi) alignment with existing liquidity table and OMO calibration; and (vii) discussion of qualitative input and other

[4] Liquidity forecasting should be in the policy implementation department. However, adequate quantitative and coding skills may have to be drawn from the research or policy areas, creating the need for inter-departmental cooperation.

liquidity impacting scenarios. At T+1 (end of the mission), the authorities provide comments on the draft reports and the draft is circulated for formal review at IMF Headquarters.

v. At T+3, the comments are included with the aide-mémoire and the report is finalized.

The TA should advocate in favor of publishing the forecast if forecast uncertainty is also disclosed. The forecast would inform the central bank counterparties' bidding at the central bank monetary operations, which should help the central bank to achieve its operational target regardless of its operational framework. However, the central bank should also publish the distribution forecast to allow the user of the forecast to factor in forecasting risk.

Due to the complexity of the proposed models, the TA should prioritize skill development within central banks. The TA experience is that coding skills are more broadly available inside central banks than organically expected. That said, the TA should dedicate significant time to hand-on training in the context of workshops using the central bank data and the code developed by the mission. The TA should ensure that central bank staff could prepare the forecast independently before the end of mission. One approach is to request the central bank colleagues to prepare and present the forecast at their senior management at the end of mission meeting.

Except for the training session mentioned above, some other challenges encountered during the previous TAs and their solutions include:

(1) Difficulties in inter- and/or intra-departmental data sharing. The mission team should assess the data availability and make sure the data are sufficient for the modeling before the mission. In addition, the mission team should emphasize the importance of data communication channels with relevant counterparties and elaborate on how expert judgment is integral to the whole forecasting process. If there is delayed delivery of data, the mission team should advise on rescheduling the forecasting date or adopting flexible forecasting horizons.

(2) IT or hardware restrictions preventing the installation of RStudio and relevant packages. The mission team should instruct the central bank counterparts on the installation of the infrastructure and make sure there are no IT restrictions on running such open-source software and packages. Central bank colleagues are usually advised to consult with their IT department in advance to clear existing restrictions.

(3) Open-source packages version update causing bugs in the codes. Because the toolkit is designed in a way to reduce the users' learning curve, we have hidden underlying codes from the users. Some package updates, such as the removal of a certain function, will undermine the functionality of the toolbox. It requires our effort to modify the underlying code to make the updated package become compatible again. The mission team will advise the forecaster not to update any package or RStudio related to the forecasting exercise.

VII. Conclusions

In this document, we have proposed a comprehensive forecasting framework designed to tackle the intricacies of predicting both autonomous factors and aggregate liquidity. Recognizing the limitations of existing forecasting approaches, which often necessitate complex specifications and maintenance, our framework is designed to be generic and adaptable. Moreover, it offers a high degree of automation, thereby aiding teams of analysts with different degrees of expertise.

The framework is a holistic process consisting of not only statistical modeling, but also a modeling process, helping to select and generate resilient forecasts and incorporate expert judgment. The statistical component of the framework is detailed with two distinct families of models tailored to accommodate the unique characteristics of autonomous factors. These are rigorously tested and compared using cross-validation techniques so that the best-performing model can be identified. Furthermore, we leverage forecast reconciliation methods to further reduce modeling uncertainty and improve forecast accuracy.

In conclusion, our proposed forecasting framework represents a significant advancement in central bank liquidity management, offering a generic yet efficient pipeline to predict short-term liquidity supply. The application of this framework is expected to facilitate decision-making processes, optimize liquidity interventions, and ultimately contribute to a more stable financial environment.

Appendix I. Data Request Template for Liquidity Forecasting

IMF: Monetary and Capital Markets Department

Central Bank Operations Division

Technical Assistance Mission: Liquidity Forecasting

INTRODUCTION

MCM will soon field a mission to assist the name of Central Bank (CB) with the conduct of its monetary policy focusing on liquidity forecasting. In preparation for this mission, we ask for the following information and data to be provided by date. Please provide all data in Excel format.

Institutional Arrangements

1. Other than commercial banks, what entities (financial and non-financial) hold accounts at CB?
    a. Is there a Memorandum of Understanding with these entities that dictate how these accounts can be operated, for example, imposing limits on maximum and minimum balances, or notification periods for transfer of funds?
2. Is there a Memorandum of Understanding governing the cash and debt management relationship between CB and the Ministry of Finance? If so, please provide a copy.
3. What information does the Ministry of Finance provide to CB regarding its cash management activities?

Liquidity Forecasting

4. Please describe the current liquidity monitoring and forecasting process.
    a. Which divisions within CB are involved?
    b. What is the interval (e.g., daily) and horizon (e.g., one week) of the forecasts?
    c. Please provide the liquidity template.
5. How are the autonomous factors forecasted (i.e., net foreign assets, government account, and currency in circulation)? Please provide the specifications of all models used for forecasting the autonomous factors.
6. Is there ex-post evaluation of forecasting accuracy undertaken? If so, please describe this process.
7. Are liquidity forecasts published?

8. Is there a model for estimating the (precautionary) demand for reserves and, if so, over what intervals (e.g., daily, weekly)? Please provide the specification of any model used.

Data Requirements

Time Series or Textual Information	***Frequency***	***Period***
A sample of the Liquidity Table.		Any Day
List of seasonal factors you think are relevant for liquidity in your country, including not only the list of dates but also the interval (e.g., +/- 6 days around a religious holiday, etc.). Please differentiate between: – Calendar seasonality: end of the week, end of month, end of quarter, etc. – Religious holidays. – National non-religious holidays. – Others (summer months, etc.).		
Time series of the above list (e.g., the past 3 years), but also for the current and next years.		
List of any ad-hoc item that could be relevant as a structural break, or to better understand the specificities of your country (for instance, “after 2011, the dynamic has changed due to a new policy,” etc.).		
Liquidity Forecasting Details Done by CB or Counterparties		
Past forecasts of currency in circulation, overall net balance of government account at the central bank (by the Treasury or CB), net foreign assets, and/or other items that have an impact on liquidity. Please include: • Forecasted values. • Dates when the forecasts were made (e.g., forecast for April 5th based on March 30th information). • Forecast horizon. • If expert adjustments of the forecasts are made, and these are stored separately, or somehow indicated, please provide this information as well. • Any analysis done on the forecasting errors, charts, etc. • The main parameters of the model, as well as the training sample. • If possible, please share with us the codes you use.		

Currency in Circulation		
Withdrawals of banknotes and coins ("increase of currency in circulation") by commercial banks from the central bank.	Daily	2017–July 2023
Deposits of banknotes and coins ("decrease of currency in circulation") by commercial banks to the central bank.	Daily	2017–July 2023
If the breakdown withdrawals/deposits are not available, please provide time series of currency in circulation.	Daily	2017–July 2023
Time series of exogeneous factor(s) relevant for forecasting currency in circulation in your country: for instance, interest rate, exchange rate, high frequency trade data, tourism data, etc.	Daily	2017–July 2023
Position of the Government at Central Bank		
Overall net balance of the government account at the central bank.	Daily	2017–July 2023
Daily credit and debit of government deposits at the central bank.[1] If possible, please provide the decomposition of the credit and debit of government deposits between: – Revenues (taxes and tariffs). – Current expenditures (public servant salaries, pensions, etc.). – Debt service. – Maturing government debt. – New issues of government debt. – Other items.	Daily	2017–July 2023
Daily central bank claims on the government split between: – Securities: please provide the amount, average duration, maturity, and interest rate. – Loans and advances: please provide the amount, average maturity, and interest rate. – Others: please provide the amount, average maturity, and interest rate.	Daily	2017–July 2023
Time series of exogeneous factors relevant for the estimation: subsidies, grants, etc. (if relevant).	Daily	2017–July 2023

[1] The Treasury can simultaneously withdraw and deposit at the Central Bank on the same day, which is why we ask for the data separately.

Net Foreign Assets		
Time series of net foreign assets at CB.	Daily	2017–July 2023
Central bank purchase of foreign assets from commercial banks.	Daily	2017–July 2023
Central bank sales of foreign assets to commercial banks.	Daily	2017–July 2023
Time series of exogeneous factors relevant for the estimation: exchange rate, interest rate spread, data on sovereign fund/public pension fund, SWIFT payments data, capital flows data (if relevant).	Daily	2017–July 2023
Other Items on Balance Sheet		
Time series of other relevant items explaining the systemic liquidity in your country.	Daily	2017–July 2023
Time series of any exogeneous factors relevant for the estimation of these other items.	Daily	2017–July 2023
Details on the auctions of the last 3 years for liquidity-injecting operations per maturity (if any): – Instrument type (CB bill or repo). – Number of participants. – Amount announced. – Amount submitted. – Allocated amount. – Minimum rate required. – Minimum rate submitted. – Maximum rate submitted. – Minimum rate allocated (marginal rate). – Weighted average rate. – Percentage of allocation at minimum rate. – Eligible collateral.	Daily	2017–July 2023

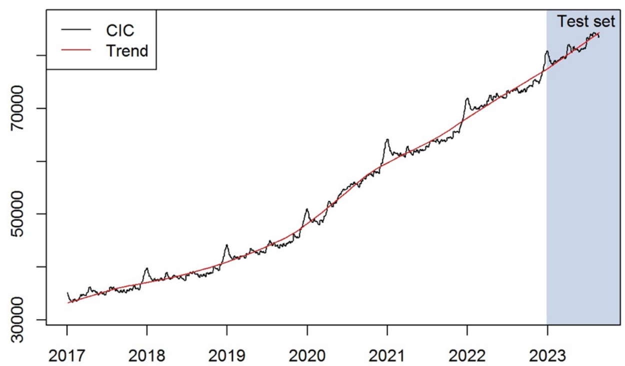
Appendix II. Standardized Output

The contents of this appendix are based on Gallardo (2024),

A. Time Series Descriptive Statistics Report: An Example of CIC

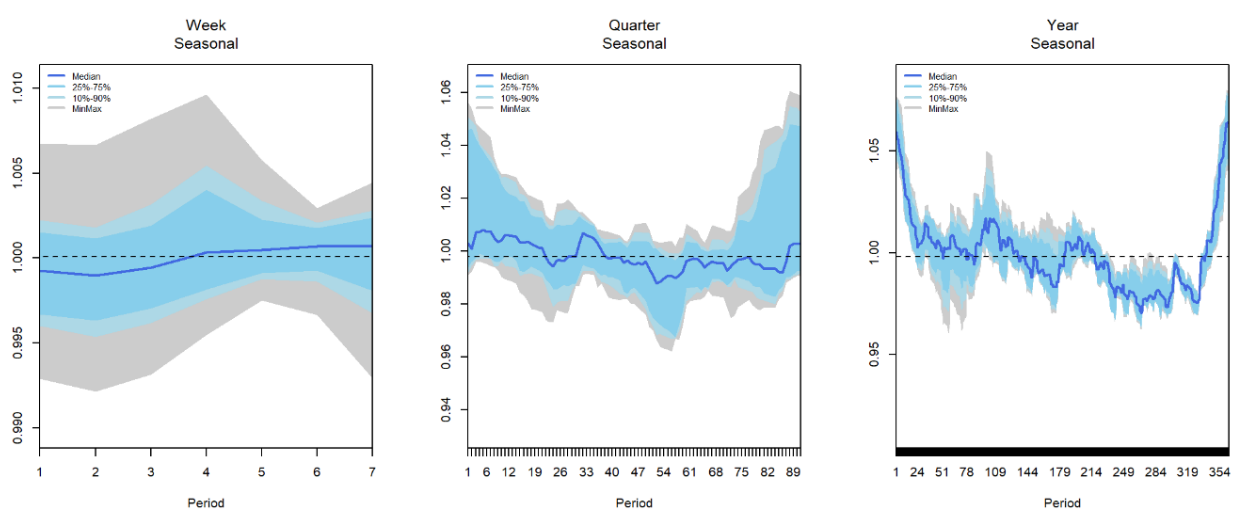
Models fit for Currency in Circulation over the period 2022-12-29 to 2023-07-20.

The following reports on time series pattern exploration for Currency in Circulation. Data are measured in millions of Guatemalan quetzals (GTQ). The period used for testing the forecasting models is highlighted. The long-term trend is captured using a centered moving average and is plotted in red.



Daily time series can exhibit multiple seasonal patterns: (i) day of the week; (ii) day of the month; and (iii) day of the year. As the number of days in a month are not fixed, the seasonal profile within a quarter is investigated.

The plots below are generated by de-trending the time series and then plotting all seasons as a distribution. For instance, the "Week" seasonal plot shows the distribution of values for days of the week, more specifically, the minimum, 10 percent, 25 percent, the median, 75 percent, 90 percent, and the maximum values. Seasonality is identified as a clear pattern across the periods of the season.



B. Forecasting Models Cross-Validation and Comparison Report: An Example of CIC

Models fit for Currency in Circulation over the period 2022-12-29 to 2023-07-20.

The following reports on models fit to data from 2022-12-29 to 2023-07-20. Data are measured in GTQ millions. Results are summarized for Currency in Circulation.

ACCURACY RESULTS

We provide Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for forecasting up to one, two, and four weeks ahead. RMSE is more sensitive to extreme values than the MAE. The lower the value is, the more accurate the results are. Sorted by one-week accuracy value.

RMSE

	Method	Week1	Week2	Week4
1	ARIMA with Regression	241.8	336.5	430.9
2	ETS with Regression	258.5	397.4	591.0
3	ARIMA	360.6	608.8	840.6
4	TBATS	368.4	610.8	860.6
5	ETS	370.5	638.3	903.7
6	NAIVE	378.1	609.4	869.4
7	Seasonal ARIMA	379.4	642.2	891.2
8	ETS (non-seasonal)	390.4	658.8	917.1
9	NAIVE seasonal	652.8	801.5	1,007.8

MAE

	Method	Week1	Week2	Week4
1	ARIMA with Regression	191.7	269.7	348.0
2	ETS with Regression	208.7	321.6	475.4
3	NAIVE	263.9	443.0	644.8
4	ARIMA	280.2	469.9	632.8
5	TBATS	281.6	483.4	698.6
6	ETS	283.3	490.0	692.8
7	Seasonal ARIMA	294.4	483.4	667.5
8	ETS (non-seasonal)	302.3	506.6	707.3
9	NAIVE seasonal	468.4	587.6	763.8

BIAS RESULTS

Unbiased forecasts should have a Mean Error (ME) close to zero. Sorted by the one-week bias.

ME

	Method	Week1	Week2	Week4
1	ETS (non-seasonal)	-30.2	-70.5	-205.7
2	Seasonal ARIMA	34.6	47.6	43.4
3	ARIMA	37.2	51.0	42.3
4	ARIMA with Regression	-37.3	-71.7	-148.3
5	TBATS	-41.9	-88.2	-224.5
6	ETS	-43.2	-93.0	-234.9
7	NAIVE	-91.5	-151.7	-301.5
8	ETS with Regression	-106.4	-174.2	-319.8
9	NAIVE seasonal	-110.4	-170.6	-320.4

PREDICTIVE DISTRIBUTION EVALUATION

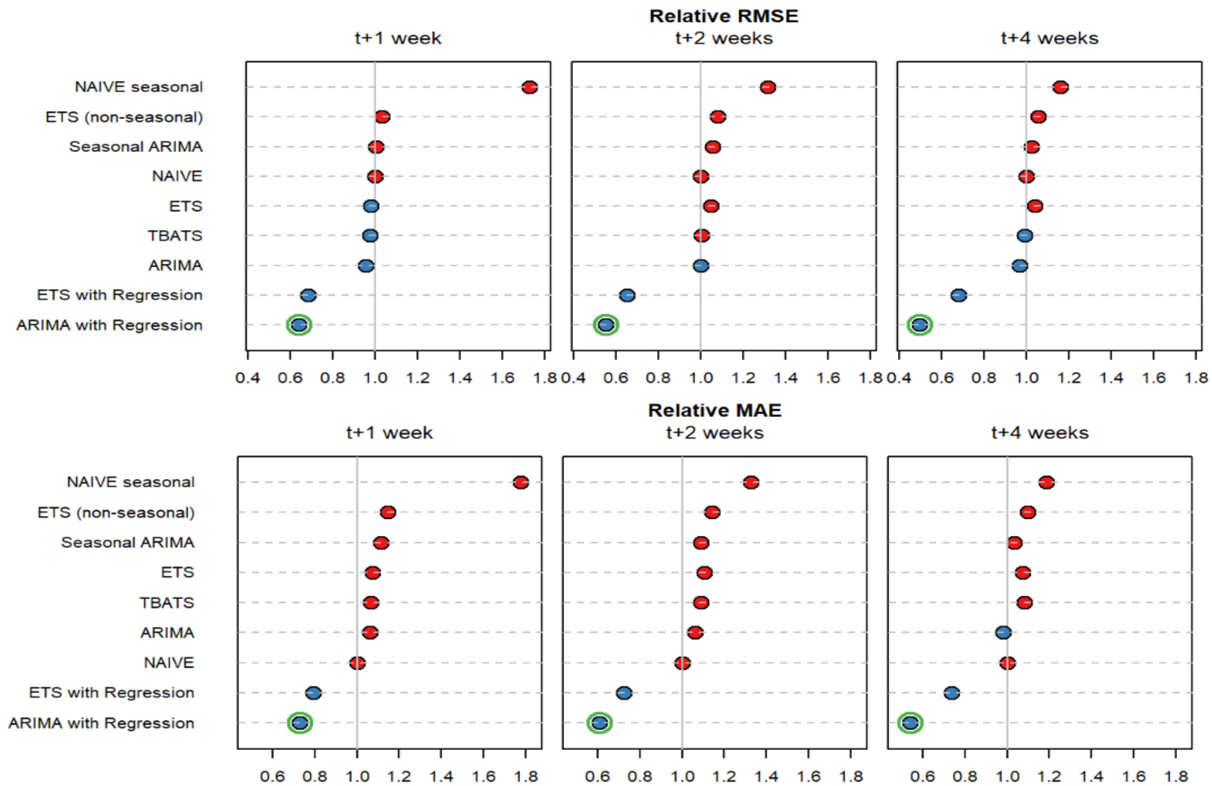
The Mean Interval Score (MIS) assesses how well the 95 percent prediction intervals capture the real distribution of the observed data. The lower the value is, the better the performance is, sorted by the 1-week horizon.

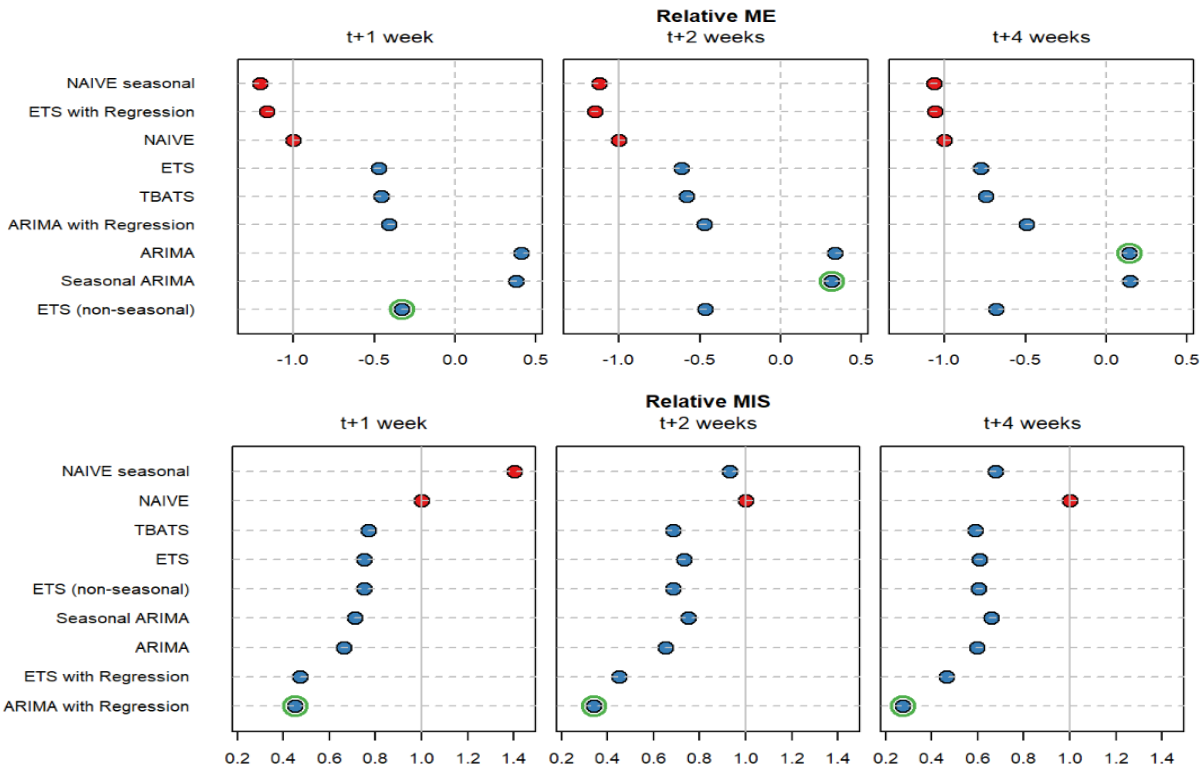
MIS

	Method	Week1	Week2	Week4
1	ARIMA with Regression	1,243.2	1,740.0	2,123.4
2	ETS with Regression	1,308.6	2,301.0	3,592.9
3	ARIMA	1,838.5	3,348.5	4,620.1
4	Seasonal ARIMA	1,969.8	3,843.2	5,100.2
5	ETS (non-seasonal)	2,075.3	3,504.6	4,659.2
6	ETS	2,083.6	3,758.9	4,689.3
7	TBATS	2,133.7	3,508.4	4,553.7
8	NAIVE	2,769.7	5,123.0	7,728.2
9	NAIVE seasonal	3,881.8	4,757.1	5,228.5

VISUAL SUMMARY

Performance of the models across the various metrics is plotted below. The forecasts are ordered from worst to best, according to each criterion, for producing forecasts for one-week ahead. The best forecast for each horizon is highlighted with a green circle. When a Naive forecast is available, the errors are provided relative to it. Any forecast less accurate than the Naive (performance equal to 1) should not be considered.



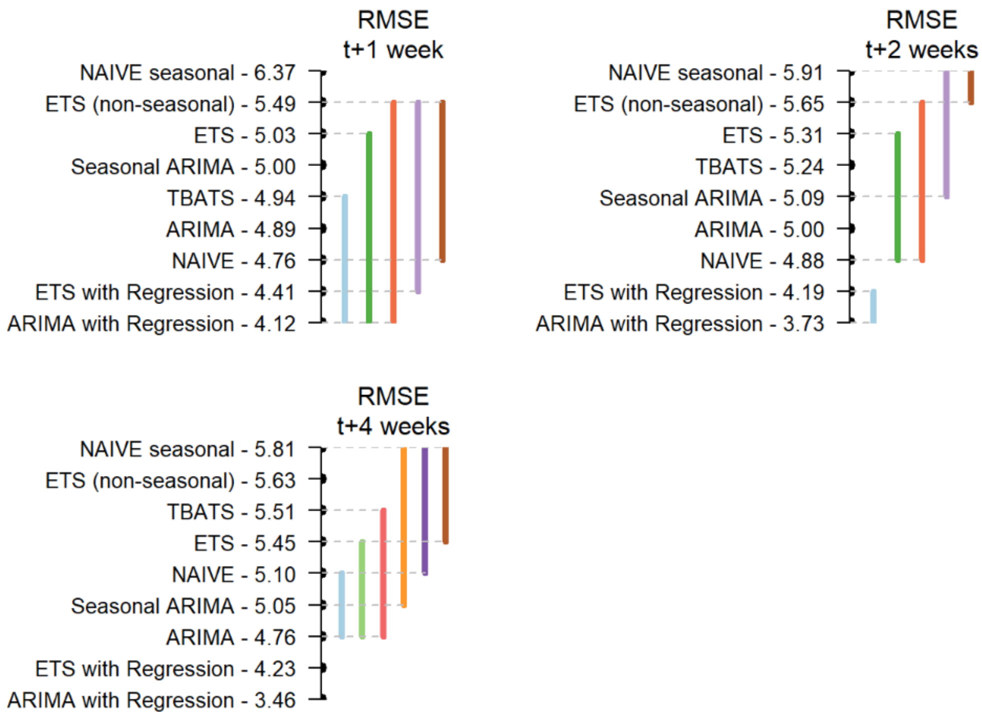


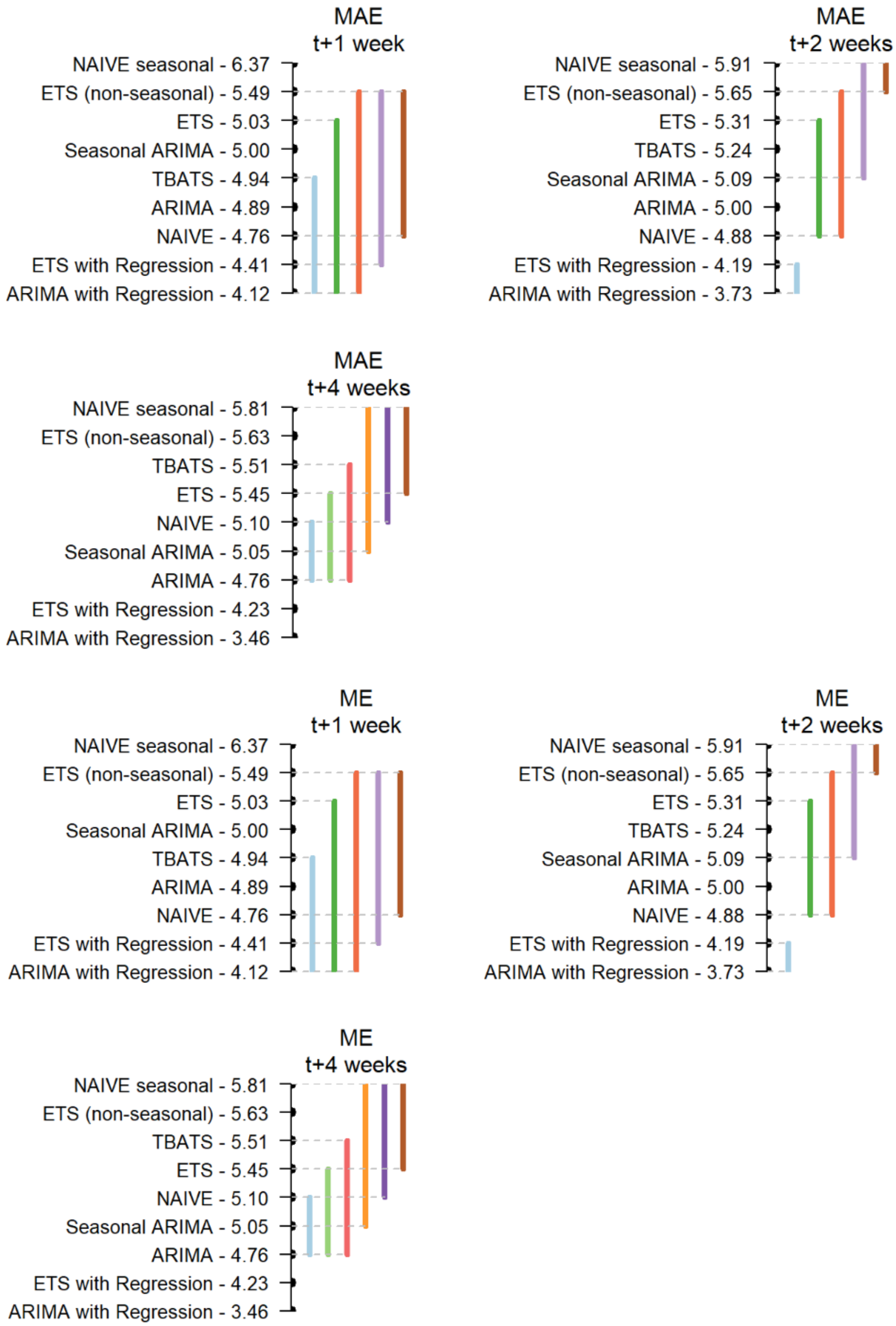
STATISTICAL TESTING

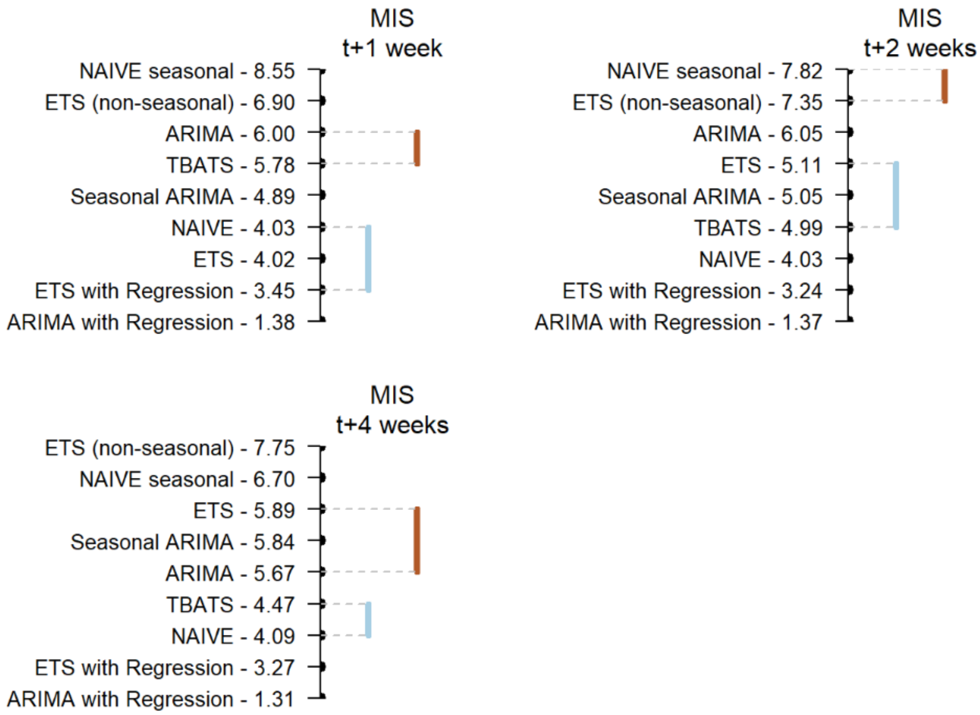
We test whether the reported differences between forecasts are statistically significant. As the errors are tracked over different periods in the test set, the results will vary as more evidence is collected. Statistical tests attempt to account for this in comparing the models.

We use the Friedman and the Nemenyi non-parametric tests and report their results at a 95 percent significance level. In the visualizations below, the models are ranked according to their mean rank (the lower, the better), which is shown next to the model's name. Models that are connected by a vertical line are grouped, and statistically speaking, there is no evidence of statistical differences. From a group, we are statistically indifferent about which model to prefer. Nonetheless, simpler models are more resilient and simpler to maintain and understand.

Note that as more test periods are accumulated, the results will become clearer, and for a limited number of periods these tests have limited power to distinguish between the competing forecasts.


RMSE
t+1 week
NAIVE seasonal - 6.37
ETS (non-seasonal) - 5.49
ETS - 5.03
Seasonal ARIMA - 5.00
TBATS - 4.94
ARIMA - 4.89
NAIVE - 4.76
ETS with Regression - 4.41
ARIMA with Regression - 4.12
RMSE
t+2 weeks
NAIVE seasonal - 5.91
ETS (non-seasonal) - 5.65
ETS - 5.31
TBATS - 5.24
Seasonal ARIMA - 5.09
ARIMA - 5.00
NAIVE - 4.88
ETS with Regression - 4.19
ARIMA with Regression - 3.73
RMSE
t+4 weeks
NAIVE seasonal - 5.81
ETS (non-seasonal) - 5.63
TBATS - 5.51
ETS - 5.45
NAIVE - 5.10
Seasonal ARIMA - 5.05
ARIMA - 4.76
ETS with Regression - 4.23
ARIMA with Regression - 3.46


MAE
t+1 week
NAIVE seasonal - 6.37
ETS (non-seasonal) - 5.49
ETS - 5.03
Seasonal ARIMA - 5.00
TBATS - 4.94
ARIMA - 4.89
NAIVE - 4.76
ETS with Regression - 4.41
ARIMA with Regression - 4.12
MAE
t+2 weeks
NAIVE seasonal - 5.91
ETS (non-seasonal) - 5.65
ETS - 5.31
TBATS - 5.24
Seasonal ARIMA - 5.09
ARIMA - 5.00
NAIVE - 4.88
ETS with Regression - 4.19
ARIMA with Regression - 3.73
MAE
t+4 weeks
NAIVE seasonal - 5.81
ETS (non-seasonal) - 5.63
TBATS - 5.51
ETS - 5.45
NAIVE - 5.10
Seasonal ARIMA - 5.05
ARIMA - 4.76
ETS with Regression - 4.23
ARIMA with Regression - 3.46
ME
t+1 week
NAIVE seasonal - 6.37
ETS (non-seasonal) - 5.49
ETS - 5.03
Seasonal ARIMA - 5.00
TBATS - 4.94
ARIMA - 4.89
NAIVE - 4.76
ETS with Regression - 4.41
ARIMA with Regression - 4.12
ME
t+2 weeks
NAIVE seasonal - 5.91
ETS (non-seasonal) - 5.65
ETS - 5.31
TBATS - 5.24
Seasonal ARIMA - 5.09
ARIMA - 5.00
NAIVE - 4.88
ETS with Regression - 4.19
ARIMA with Regression - 3.73
ME
t+4 weeks
NAIVE seasonal - 5.81
ETS (non-seasonal) - 5.63
TBATS - 5.51
ETS - 5.45
NAIVE - 5.10
Seasonal ARIMA - 5.05
ARIMA - 4.76
ETS with Regression - 4.23
ARIMA with Regression - 3.46

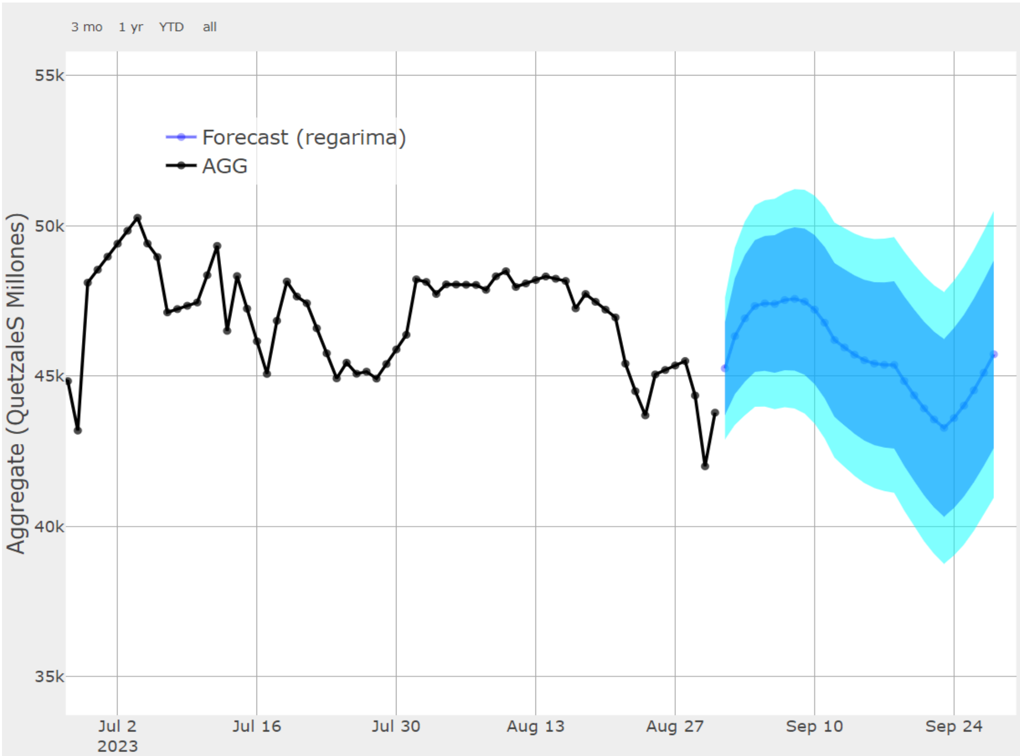


C. Forecasting Results: An Example of Aggregated Liquidity

The following contains details for forecasting of AGG. All models were forecasted on 2023-08-31.

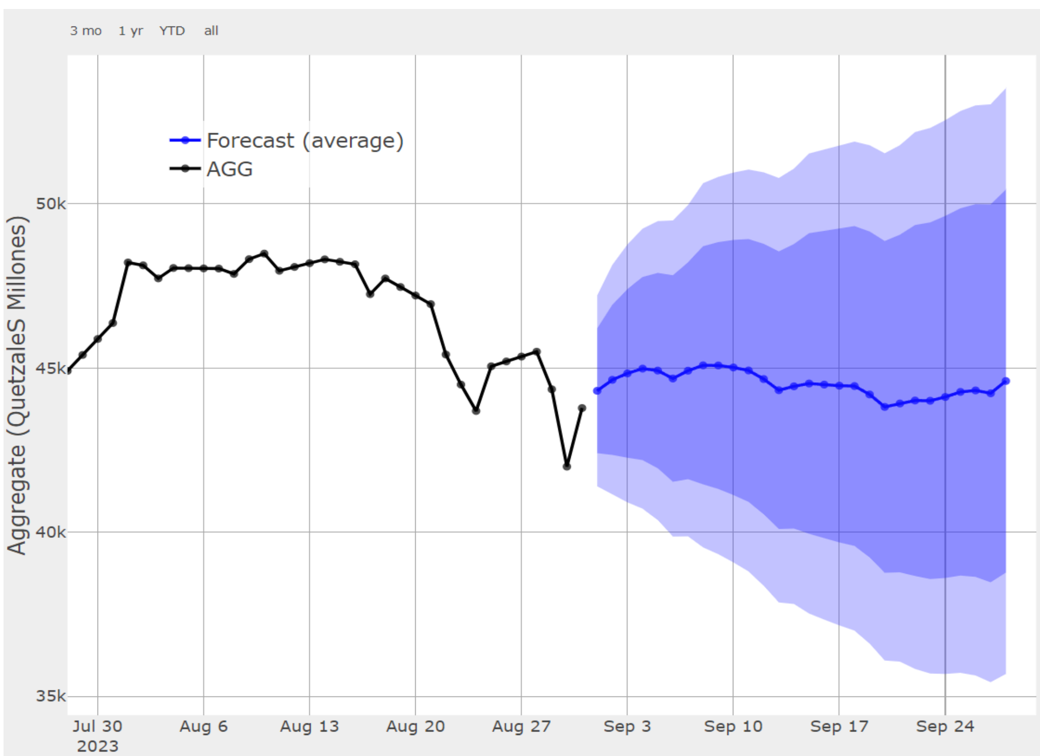
LATEST FORECASTS (SELECTED MODEL)

The model selected for AGG is ARIMA with Regression. This selection can be changed in the configuration file. The plot below shows forecasts with prediction intervals. The darker bands indicate an 80 percent prediction interval, while the lighter bands indicate a 95 percent prediction interval. The observed data is shown in black.



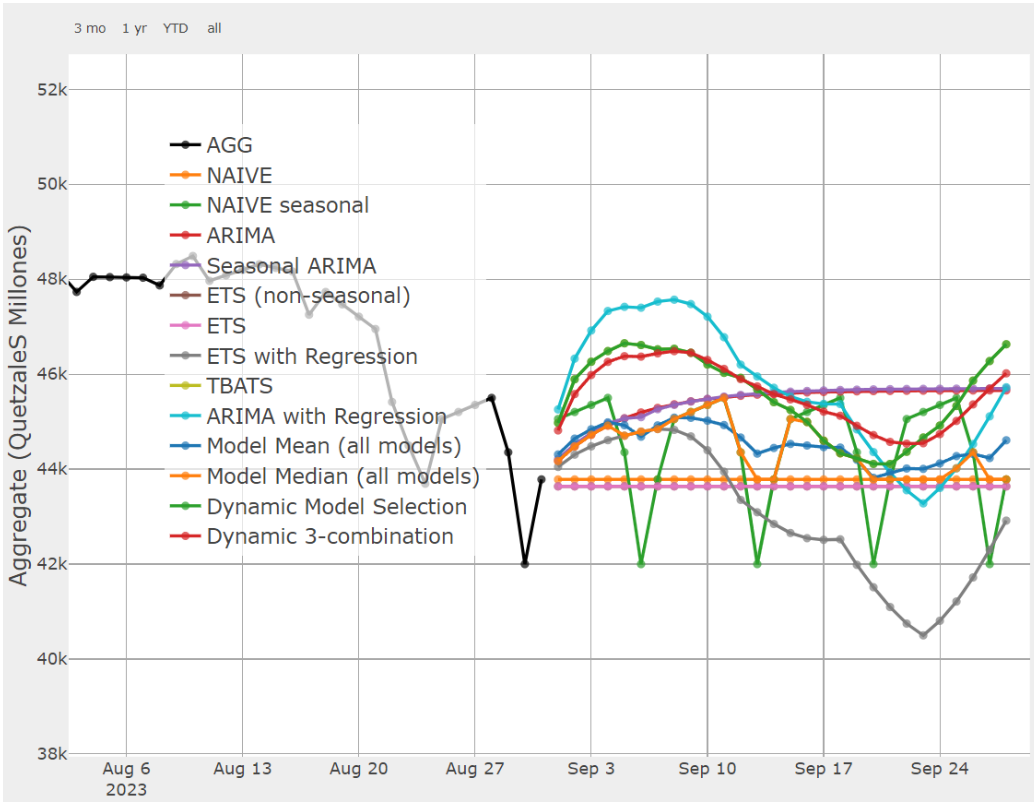
LATEST FORECASTS (MODEL AVERAGES)

The forecasts below correspond to the **equally weighted average of all models.**



FORECASTS FOR ALL MODELS

The chart below shows the forecasts for all models.



D. Enhancement of Aggregated Liquidity Forecast through Reconciliation Report

Models fit over the period 2023-01-05 to 2023-07-27

The following contains details of reconciling liquidity forecasts. The models chosen for each autonomous factor are as follows:

- State Account Balance: ETS with Regression.
- Net Foreign Accounts: ETS with Regression.
- Net Other Assets: ARIMA with Regression.
- Currency in Circulation: ARIMA with Regression.
- Aggregate: ARIMA with Regression.

These selections can be changed in the configuration file.

Most Recent Forecast for 2023-08-31

One-step-ahead forecasts for the most recent origin date (2023-08-31) are tabulated below for both reconciled and unreconciled forecasts. The method selected for reconciliation is the STR method (shaded), while the best method based on past performance (RMSE) is the OLS method (yellow).

Forecast Evaluation

Validation period used for the evaluation: 2023-01-05 to 2023-07-27.

RMSE (ACCURACY)

The table below summarizes RMSE for forecasts using different reconciliation methods. The chosen method (STR) is shown shaded while the best method based on average RMSE across horizons (OLS) is shown in yellow.

	Method	Week1	Week2	Week4
1	OLS	1,530.987	1,576.952	1,637.343
2	Base (Unreconciled)	1,539.931	1,576.947	1,631.999
3	MinT	1,560.999	1,595.187	1,657.947
4	Bottom Up	1,624.146	1,740.319	1,829.821

MAE (ACCURACY)

The table below summarizes MAE for forecasts using different reconciliation methods. The chosen method (STR) is shown shaded while the best method based on average MAE across horizons (OLS) is shown in yellow.

	Method	Week1	Week2	Week4
1	OLS	1,128.307	1,193.661	1,264.856
2	Base (Unreconciled)	1,140.167	1,194.323	1,267.480
3	MinT	1,160.670	1,207.308	1,297.477
4	Bottom Up	1,245.110	1,340.951	1,415.599

ME (BIAS)

The table below summarizes ME for forecasts using different reconciliation methods. The chosen method (STR) is shown shaded while the best method based on average ME across horizons (MinT) is shown in yellow.

	Method	Week1	Week2	Week4
1	Base (Unreconciled)	-89.492	-146.328	-267.574
2	OLS	-44.230	-105.004	-241.105
3	MinT	2.202	-117.799	-260.571
4	Bottom Up	136.816	60.289	-135.230

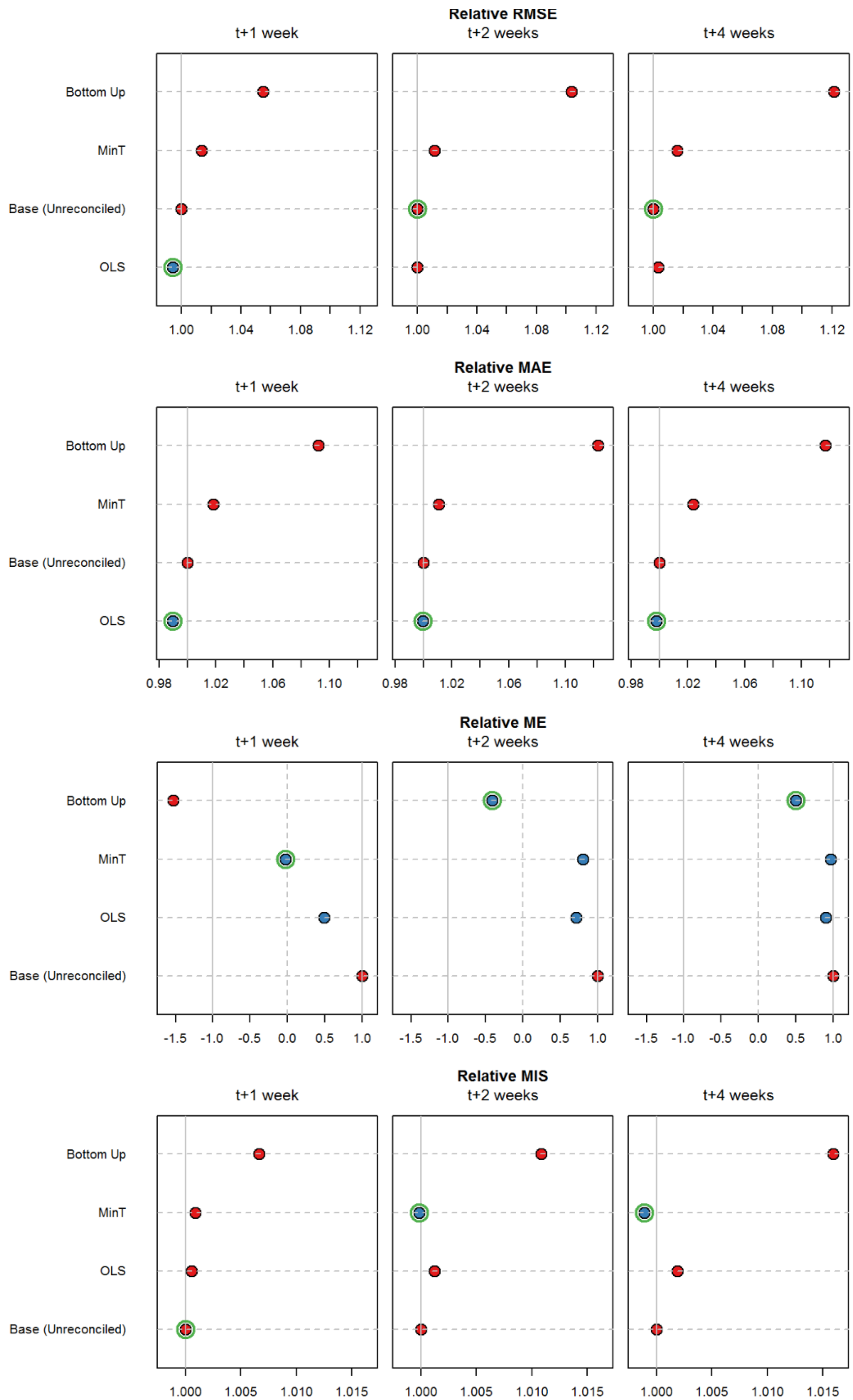
MIS (PREDICTIVE DISTRIBUTION)

The table below summarizes MIS for forecasts using different reconciliation methods. The chosen method (STR) is shown shaded while the best method based on average MIS across horizons (Base (Unreconciled)) is shown in yellow.

	Method	Week1	Week2	Week4
1	Base (Unreconciled)	99,108.458	99,710.535	100,577.040
2	OLS	99,162.843	99,829.813	100,767.040
3	MinT	99,199.067	99,697.048	100,466.502
4	Bottom Up	99,768.098	100,793.964	102,178.093

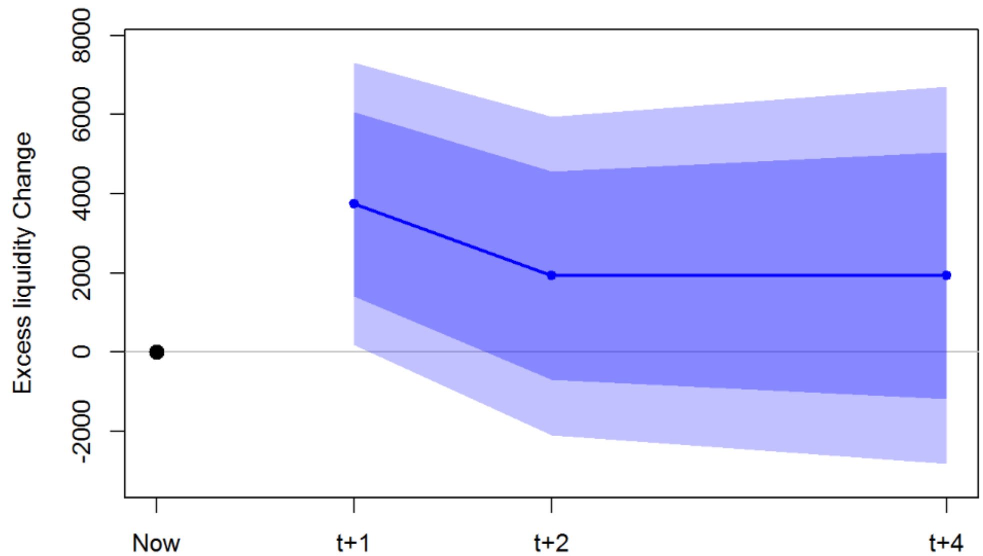
VISUAL SUMMARY

Performance of the models across the various metrics is plotted below. The forecasts are ordered from worst to best, according to each criterion, for producing forecasts for one week ahead. The best forecast for each horizon is highlighted with a green circle. All errors are presented as relative to the unreconciled (base) forecast.

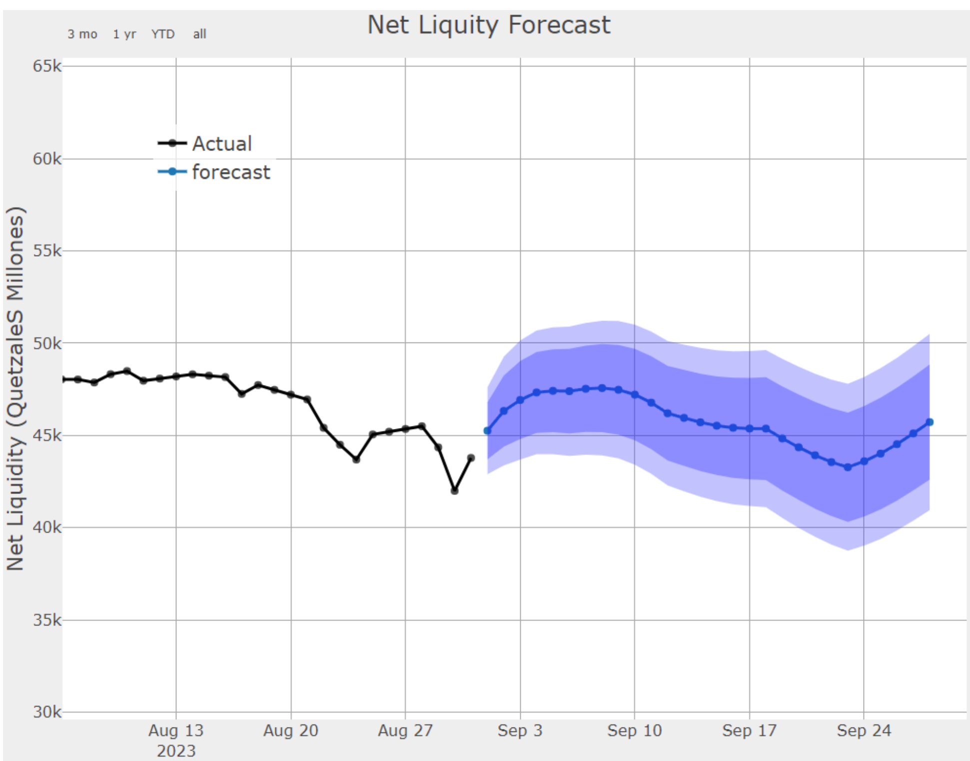

Relative RMSE
t+1 week
t+2 weeks
t+4 weeks
Bottom Up
MinT
Base (Unreconciled)
OLS
1.00
1.04
1.08
1.12
Relative MAE
0.98
1.02
1.06
1.10
Relative ME
-1.5
-1.0
-0.5
0.0
0.5
1.0
Relative MIS
1.000
1.005
1.010
1.015

E. Enhanced Forecasts for Aggregated Liquidity Report

The changes in Net Liquidity for one, two, and four weeks ahead are provided below.



Forecasts or the net liquidity position by the chosen method (STR) are plotted below.



References

Arvan, M., Fahimnia, B., Reisi, M., and Siemsen, E. “Integrating human judgement into quantitative forecasting methods: A review,” *Omega*, vol. 86, 2019, pp. 237–252.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Petropoulos, F. “Forecasting with temporal hierarchies,” *European Journal of Operational Research*, vol. 262, no. 1, 2017, pp. 60–74.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Panagiotelis, A. “Forecast reconciliation: A review,” *International Journal of Forecasting*, vol. 40, no. 2, 2023, pp. 430-456.

Athanasopoulos, G., and Kourentzes, N. “On the evaluation of hierarchical forecasts.” International Journal of Forecasting, vol. 39, no. 4, 2021, pp. 1502-1511.

Bell, W. R., & Hillmer, S. C. "Modeling time series with calendar variation," *Journal of the American Statistical Association*, vol. 78, no. 383, 1983, pp. 526–534.

Cabrero, A., Camba-Mendez, G., Hirsch, A., and Nieto, F. “Modelling the daily banknotes in circulation in the context of the liquidity management of the European Central Bank”. *Journal of Forecasting*, vol. 28, no.3, 2009, pp. 194-217.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. “Forecasting time series with complex seasonal patterns using exponential smoothing,” *Journal of the American Statistical Association*, vol. *106, no.* 496, 2011, pp. 1513–1527.

El Gemayel, J., Lafarguette, R., MCM, J. F., ITD, K. M., del Castillo Penna, R. A., Merrouche, O., and Panagiotelis, A. N. *United Arab Emirates: Technical Assistance Report – Liquidity Management and Forecasting*, International Monetary Fund, 2022.

El Hamiani Khatat, M. *Monetary policy and models of currency demand*, International Monetary Fund, 2018.

Elliott, G., and Timmermann, A., eds. *Handbook of economic forecasting*, Newnes, 2013.

Gallardo, D., Chen, Z., and Veyrune, R. *Guatemala: Technical Assistance Report—The Statistical Component of Liquidity Forecasting*, International Monetary Fund, 2024.

Ghysels, E., and Osborn, D. R. *The econometric analysis of seasonal time series*, Cambridge University Press, 2001.

Gneiting, T., and Raftery, A. E. “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. *102, no.* 477, 2007, pp. 359–378.

Gray, S. *Liquidity forecasting*, Bank of England, 2008.

Harvey, A., Koopman, S. J., and Riani, M. "The modeling and seasonal adjustment of weekly observations," *Journal of Business & Economic Statistics*, vol. 15, no. 3, 1997, pp. 354–368.

Hlaváček, M., Čada J., and Hakl F. "The application of structured feedforward neural networks to the modelling of the daily series of currency in circulation," *International Conference on Natural Computation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

Hollander, M., Wolfe, D. A., & Chicken, E. Nonparametric statistical methods. John Wiley & Sons, 2003.

Hyndman, R.J., and Athanasopoulos, G. "Forecasting: principles and practice," 3rd edition, OTexts: Melbourne, Australia, 2021. OTexts.com/fpp3.

Hyndman, R. J., and Khandakar, Y. "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 27, 2008, pp. 1–22.

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

Ikoku, A. "Modeling and forecasting currency in circulation for liquidity management in Nigeria," *CBN Journal of Applied Statistics*, vol. 5, no. 1, 2014, pp. 79–104.

Iskandar, I., Willett, Roger., and Xu, S. "The development of a government cash forecasting model," *Journal of Public Budgeting, Accounting & Financial Management*, vol. 30, no. 4, 2018, pp. 368–383.

Kourentzes, N. "tsutils: Time Series Exploration, Modelling and Forecasting," R package version 0.9.4, https://CRAN.R-project.org/package=tsutils, 2013.

Kourentzes, N., and Athanasopoulos, G. "Cross-temporal coherent forecasts for Australian tourism," *Annals of Tourism Research*, vol. *75*, 2019, pp. 393–409.

Kourentzes, N., Barrow, D., and Petropoulos, F. "Another look at forecast selection and combination: Evidence from forecast pooling," *International Journal of Production Economics*, vol. 209, 2019, pp. 226–235.

Kourentzes, N., and Petropoulos, F. "Forecasting with multivariate temporal aggregation: The case of promotional modelling" *International Journal of Production Economics*, vol. 181, 2016, pp. 145–153.

Kourentzes, N., and Sagaert ,Y. "Incorporating Leading Indicators into Sales Forecasts," *Foresight: The International Journal of Applied Forecasting*, vol. 48, 2018, pp. 24–40.

Koziński, W., and Świst, T. "Short-term currency in circulation forecasting for monetary policy purposes– the case of Poland," *Financial Internet Quarterly*, vol. 11, no. 1, 2015, pp. 65–75.

Lawrence, M., Goodwin, P., O'Connor, M., and Önkal, D. "Judgmental forecasting: A review of progress over the last 25 years," *International Journal of Forecasting*, vol. 22, no. 3, 2006, pp. 493–518.

Nasiru, S., Luguterah, A., and Anzagra, L. "The efficacy of ARIMAX and SARIMA models in predicting monthly currency in circulation in Ghana," *Mathematical Theory and Modeling*, vol. 3, no. 5, 2013, pp. 73–81.

Ord, K., Fildes, R., and Kourentzes, N. *Principles of Business Forecasting*. 2nd ed. Wessex Press Publishing Co, 2017.

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., and Hyndman, R. J. "Forecast reconciliation: A geometric view with new insights on bias correction," *International Journal of Forecasting*, vol. 37, no. 1, 2021, pp. 343–359.

Pritularga, K. F., Svetunkov, I., and Kourentzes, N. "Stochastic coherency in forecast reconciliation," *International Journal of Production Economics*, vol. 240, 2021, article 108221.

Sroginis, A., Fildes, R., and Kourentzes, N. "Use of contextual and model-based information in adjusting promotional forecasts," *European Journal of Operational Research*, vol. 307, no. 3, 2023, pp. 1177–1191.

Tashman, L. J. "Out-of-sample tests of forecasting accuracy: An analysis and review," *International Journal of Forecasting*, vol. *16, no.* 4, 2000, pp. 437–450.

Tsay, R. S. *Analysis of Financial Time Series*. John Wiley & Sons, 2005.

Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman. "Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series through Trace Minimization," *Journal of the American Statistical Association*, vol. 114, no. 526, 2019, pp. 804–819.

Williams, M. *Government cash management: Its interaction with other financial policies,* International Monetary Fund, 2010.