

CHAPTER II

LITERATURE REVIEW

This chapter presents the concepts, theories, and principles related to the research, which serve as the foundation for time series model development and statistical data analysis. The chapter covers the models used in this study, including MIDAS, Regime-Switching, and Self-Organizing Map (SOM), as well as the criteria for evaluating model performance and related studies. These are used as guidelines for conducting the research and developing models with greater accuracy and reliability.

2.1 Stationary of Time-Series Analysis

In time series analysis, stationarity is a fundamental property that is important for the analysis and modeling of time series data because most statistical models, such as the Autoregressive model, Moving Average model, and ARIMA, are based on the assumption that the data are stationary. In other words, the statistical characteristics of the stochastic process do not change over time. If this assumption is not satisfied, the parameter estimation and forecasting may become unreliable (George E. P. Box and Gwilym M. Jenkins, 1976; James D. Hamilton, 1994).

Definition of Strictly Stationary:

Definition 2.1. According to Brockwell and Davis (2016), a time series $\{Y_t\}$ is said to be *strictly stationary* if

$$(Y_1, Y_2, \dots, Y_n) \stackrel{d}{=} (Y_{1+h}, Y_{2+h}, \dots, Y_{n+h}) \quad \text{for all } n \geq 1,$$

where $\stackrel{d}{=}$ denotes that the joint probability distributions are identical.

Although strict stationarity is a condition that completely characterizes the proba-

bility distribution of a stochastic process, it is a restrictive condition and difficult to verify using real data. Moreover, most time series analysis methods, such as autocorrelation function analysis and parameter estimation of models, rely only on the properties of the mean and covariance. Therefore, weak stationarity is considered a sufficient and more appropriate condition in practice.

Therefore, in the construction and analysis of time series models, the assumption of weak stationarity is commonly adopted as the main framework because it can be verified from the data and is consistent with the statistical methods used for time series analysis and forecasting.

Definition of Weak Stationarity:

Definition 2.2. According to Brockwell and Davis (2016), a time series $\{Y_t\}$ is said to be *weakly stationary* if it satisfies the following conditions:

1. The mean is constant over time:

$$E(Y_t) = \mu.$$

2. The variance is constant and finite:

$$\text{Var}(Y_t) = \sigma^2 < \infty.$$

3. The covariance depends only on the time lag h and not on the time index t :

$$\text{Cov}(Y_t, Y_{t+h}) = \gamma(h).$$

If a time series is non-stationary, it affects time series analysis, particularly in linear models. The parameters of the model may become unstable because the structure of the data changes over time, causing the estimated values to be ineffective for forecasting (Hamilton, 1994). Moreover, non-constant variance may violate the assumption of white noise, leading to errors in statistical testing and confidence intervals, and may result in the

problem of spurious regression. Granger and Newbold (1974) demonstrated that analyzing relationships between non-stationary variables may produce statistically significant results even when no true relationship exists.

The stationarity of data is commonly tested using the Unit Root Test through the Augmented Dickey–Fuller (ADF) test, which was developed by David A. Dickey and Wayne A. Fuller (1979). The ADF test can be expressed in the following form:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^p \delta_i \Delta Y_{t-i} + \varepsilon_t,$$

where

- ΔY_t is the first difference of the data,
- α is a constant term,
- βt is the linear trend term,
- γ is the parameter used to test for the presence of a unit root,
- p is the number of lag terms added to correct the autocorrelation problem,
- ε_t is the white noise error term.

The hypotheses of the test are given by

$$H_0 : \gamma = 0,$$

$$H_1 : \gamma < 0.$$

Rejection of the null hypothesis indicates that the time series does not contain a unit root and is therefore stationary. The decision is based on the p-value associated with the test statistic. If the p-value is less than the chosen significance level (e.g., 0.05), the null hypothesis is rejected. Otherwise, there is insufficient evidence to reject the null hypothesis, suggesting that the series contains a unit root and is non-stationary. In such cases, appropriate transformations, such as differencing, should be applied before proceeding with further time series analysis.

In cases where the data are non-stationary, the data can be transformed into stationary form using several methods, such as:

- Differencing, which is commonly used to eliminate trends that change over time (Box and Jenkins, 1976).
- Log Transformation, which is used to reduce non-constant variance, especially in financial data (Tsay, 2010).
- Seasonal Differencing, which is used to remove seasonal patterns in the data (Hyndman and Athanasopoulos, 2018).

Stationarity is an important condition in time series analysis because it affects the accuracy of parameter estimation and forecasting. If the data are non-stationary, it may lead to misleading conclusions. Therefore, testing stationarity using the Augmented Dickey–Fuller (ADF) test and transforming the data into stationary form are important steps before constructing time series models.

2.2 White Noise

White noise is a fundamental stochastic process in time series analysis and is commonly used as the error assumption in time series models (Box and Jenkins, 1976).

Definition of White Noise

Definition 2.3. Let $\{\varepsilon_t\}$ be a sequence of random variables. The process $\{\varepsilon_t\}$ is called *white noise* if it satisfies the following conditions:

1. The mean is zero:

$$E(\varepsilon_t) = 0.$$

2. The variance is constant:

$$\text{Var}(\varepsilon_t) = \sigma^2.$$

3. The observations are uncorrelated:

$$\text{Cov}(\varepsilon_t, \varepsilon_s) = 0, \quad t \neq s.$$

This process is denoted by $\varepsilon_t \sim WN(0, \sigma^2)$, which means that ε_t has zero mean, constant variance, and no autocorrelation between observations at different time periods. White noise is considered a special case of a weakly stationary process and is commonly used as the error term in time series models.

Furthermore, testing whether the residuals are white noise can be performed using the Ljung–Box test, which is used to examine the independence of data across multiple lags simultaneously (Ljung and Box, 1978).

2.3 ACF and PACF

Time series analysis not only considers the stationarity of the data, but also examines the linear relationship between observations at different time periods. This relationship can be measured using the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF), which are fundamental tools for investigating the structure of time series data and for identifying the appropriate form of time series models.

2.3.1 Autocorrelation Function: ACF

Definition 2.4. The Autocorrelation Function (ACF) at lag k , denoted by ρ_k , is a measure of the linear relationship between the current value of a time series Y_t and its lagged value at k periods. The autocorrelation coefficient takes values between -1 and 1 . In statistical time series analysis for stationary time series, the ACF at lag k (Hamilton, 1994; Cryer and Chan, 2008) is defined as the ratio between the autocovariance at lag k and the variance of the time series as follows:

$$\rho_k = \frac{\gamma_k}{\gamma_0},$$

where

- ρ_k is the autocorrelation coefficient at lag k ,
- γ_k is the autocovariance at lag k ,
- γ_0 is the variance of the time series.

In practice, the ACF is commonly used as a guideline for identifying the order of a Moving Average model, denoted by $MA(q)$, by considering the pattern of the autocorrelation plot. If the ACF exhibits a cut-off after lag q , it may indicate an $MA(q)$ model. On the other hand, if the ACF gradually decays over time, it is generally characteristic of an Autoregressive model (Box et al., 2015).

For white noise processes, which have no temporal correlation, the Autocorrelation Function (ACF) is equal to zero for all non-zero lags (Brockwell and Davis, 2016), that is,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}(Y_t, Y_{t-k})}{\gamma_0} = \frac{0}{\gamma_0} = 0,$$

for all $k \neq 0$ and $\rho_0 = 1$ since it represents the correlation of the series with itself.

Demonstrating that white noise has an ACF equal to zero for all non-zero lags is important because it serves as a fundamental criterion for comparing the structure of time series data. If the data behave as white noise, no temporal relationship will be observed. Conversely, if the ACF values differ significantly from zero at certain lags, it indicates that the data contain temporal dependence, which can be used in constructing time series models. In addition, this property is commonly used to assess model adequacy. The residuals of a good model should behave as white noise, meaning that the ACF of the residuals should be close to zero at all lags. If temporal correlation remains in the residuals, it indicates that the model has not fully explained the structure of the data.

2.3.2 Partial Autocorrelation Function: PACF

Definition 2.5. The Partial Autocorrelation Function (PACF) at lag k is the linear correlation between the current value of a time series Y_t and its lagged value Y_{t-k} after controlling for the effects of the intermediate lagged variables. Mathematically, the PACF at lag k can be defined as the conditional correlation as follows (Hamilton, 1994):

$$\phi_{kk} = \text{Corr}(Y_t, Y_{t-k} \mid Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}).$$

In addition, the PACF can also be defined through the linear regression representation of an autoregressive process as follows:

$$Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \dots + \phi_{kk}Y_{t-k} + \varepsilon_t,$$

where ϕ_{kk} is the coefficient representing the direct relationship between Y_t and Y_{t-k} after controlling for the effects of the other lagged variables.

The PACF plays an important role in identifying the order of an Autoregressive model. For an $AR(p)$ model, the PACF typically exhibits a cut-off after lag p , or approaches zero when the lag is greater than p (Box et al., 2015; Tsay, 2010).

2.4 Linear Process

A Linear Process is an important concept in time series analysis and is used to describe the characteristics of a stochastic process in which the current value of the series is influenced by past shocks or disturbances through a systematic weighting structure. This concept allows the fundamental properties of the data, such as the mean, variance, and correlation structure across different time periods, to be studied more clearly. Moreover, many time series models, such as AR, MA, and ARMA models, can be represented within the framework of a linear process under certain conditions. Therefore, this concept serves as an important foundation for studying the weak stationarity properties of time series data.

Definition of Causal Linear Process

Definition 2.6. According to Brockwell and Davis (2016), the time series $\{Y_t\}$ is called a linear process if it can be represented as

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$

for all t , where $\{\varepsilon_t\}$ is a white noise process and $\{\psi_j\}$ is a sequence of constants satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

The process is said to be causal if $\psi_j = 0$ for all $j < 0$ in which case the representation simplifies to $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$.

Theorem: Stationarity of a Causal Linear Process

Theorem 2.1. *According to Brockwell and Davis (2016), let $\{Y_t\}$ be a linear process defined by*

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\} \sim WN(0, \sigma^2)$. If the sequence of coefficients $\{\psi_j\}$ is absolutely summable, that is

$$\sum_{j=0}^{\infty} |\psi_j| < \infty,$$

then the process $\{Y_t\}$ is weakly stationary.

Proof

To prove that $\{Y_t\}$ is weakly stationary, three conditions must be verified: constant mean, finite and constant variance, and an autocovariance function that depends only on the lag h .

1. Constant Mean

We examine the expectation of Y_t :

$$E[Y_t] = E \left[\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \right].$$

Since $\sum |\psi_j| < \infty$, the Fubini's Theorem (or dominated convergence) allows us to interchange the expectation and the summation:

$$E[Y_t] = \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}],$$

Given that for all (White Noise property):

$$E[Y_t] = \sum_{j=0}^{\infty} \psi_j(0) = 0.$$

The mean is constant at 0, which is independent of t .

2. Finite and Constant Variance

The variance of Y_t is defined as:

$$\text{Var}(Y_t) = E[Y_t^2] - (E[Y_t])^2.$$

Since $E[Y_t] = 0$:

$$\text{Var}(Y_t) = E \left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \right)^2 \right] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k E[\varepsilon_{t-j}, \varepsilon_{t-k}].$$

By the property of white noise, $E[\varepsilon_{t-j}, \varepsilon_{t-k}] = \sigma^2$ if $j = k$, and 0 otherwise:

$$\text{Var}(Y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2.$$

Since absolute summability ($\sum |\psi_j| < \infty$) implies square summability ($\sum \psi_j^2 < \infty$), the variance is finite and independent of t .

3. Autocovariance depends only lag k

Let $\text{Cov}(Y_t, Y_{t-k}) = \gamma(k)$ for any lag $k \geq 0$:

$$\text{Cov}(Y_t, Y_{t-k}) = E[Y_t, Y_{t-k}] = E \left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \right) \left(\sum_{m=0}^{\infty} \psi_m \varepsilon_{t-k-m} \right) \right].$$

Expanding the product and applying the expectation:

$$\text{Cov}(Y_t, Y_{t-k}) = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} \psi_j \psi_m E[\varepsilon_{t-j}, \varepsilon_{t-k-m}].$$

The expectation $E[\varepsilon_{t-j}, \varepsilon_{t-k-m}]$ is non-zero (equal to σ^2) only when the time indices

match: $t - j = t - k - m$, which simplifies to $j = k + m$. Substituting j :

$$\gamma(k) = \sigma^2 \sum_{m=0}^{\infty} \psi_{k+m} \psi_m.$$

The result $\gamma(k)$ is independent of the time step and depends only on the lag k .

Conclusion; Since all three conditions are satisfied, the linear process $\{Y_t\}$ is weakly stationary.

2.5 Development of Time-Series Models

The development of Time Series Models has continuously evolved in accordance with theoretical advances and practical applications. In the early stages, the models mainly focused on describing the linear relationships within the data. Later, more advanced models were developed to accommodate more complex data structures, including nonlinearity, volatility, and changes in data regimes. The models developed in each period reflect the concepts and limitations of the methodologies available at that time. In this study, the development of important models will be presented in chronological order to illustrate the trend of model development, with emphasis only on the models related to the context of this research.

2.5.1 Autoregressive Model: (AR)

The Autoregressive Model (AR) is a fundamental model for time series analysis. The concept of the model is to explain the current value of a variable using its own past values under a linear relationship. The concept of the Autoregressive Model was systematically developed during the twentieth century and was originally based on the work of George Udny Yule in 1927. An Autoregressive Model of order p , denoted by $AR(p)$, can be written as follows (Yule, 1927; Box et al., 2015):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t,$$

where

- Y_t represents the time series value at time t ,
- ϕ_0 is a constant term,
- $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ are the model parameters,
- ε_t is the random error term at time t .

In specifying the structure of the model, ε_t is commonly assumed to be a white noise process with zero mean, constant variance, and no temporal correlation, consistent with the definition of white noise presented in Section 2.2.

Selection of the Order of the $AR(p)$ Model Determining the order of the Autoregressive model, denoted by $AR(p)$, is an important step in constructing a time series model because the value of p represents the number of lagged observations of the dependent variable used to explain the current value. If the order is specified too low, the model may not adequately capture the structure of the data. On the other hand, if the order is specified too high, the model may become unnecessarily complex and may reduce forecasting performance (Box and Jenkins, 1976).

The order of the model is commonly determined by considering the Partial Autocorrelation Function (PACF), since the PACF of an $AR(p)$ process remains significant up to lag p and exhibits a cut-off behavior afterward. Therefore, it can be used as a guideline for determining the order of the model. However, in practice, the selection is usually considered together with information criteria such as AIC or BIC in order to confirm the appropriateness of the selected model order.

Stationary of the $AR(p)$ process In this section, we establish the conditions under which an Autoregressive process of order p , denoted as $AR(p)$, satisfies the requirements for weak stationarity. The derivation follows the logic of transforming an $AR(p)$ model into a causal linear process representation.

1. The AR(p) model with backshift operator

Consider the AR(p) process defined by

$$\Phi(B)Y_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

where $\Phi(B)$ denotes the autoregressive polynomial in the backshift operator B , and

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

Provided that the stationarity condition is satisfied, the process can be expressed in its infinite moving-average representation as

$$Y_t = \Phi(B)^{-1} \varepsilon_t = \Psi(B) \varepsilon_t,$$

where

$$\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j.$$

2. Partial Fraction Decomposition and The Root Condition

To determine the explicit form of the coefficients $\{\psi_j\}$, we decompose the reciprocal of the autoregressive polynomial. Let $r_1, r_2, \dots, r_p$ be the distinct roots of the characteristic equation $\Phi(B) = 0$. The operator $\Psi(B)$ can be factored as:

$$\Psi(B) = \frac{1}{\Phi(B)} = \frac{1}{\left(1 - \frac{1}{r_1} B\right) \left(1 - \frac{1}{r_2} B\right) \dots \left(1 - \frac{1}{r_p} B\right)}.$$

By applying partial fraction decomposition, we obtain:

$$\Psi(B) = \frac{A_1}{1 - \frac{1}{r_1} B} + \frac{A_2}{1 - \frac{1}{r_2} B} + \dots + \frac{A_p}{1 - \frac{1}{r_p} B}.$$

Using the geometric series expansion for each term, the operator becomes:

$$\Psi(B) = \sum_{i=1}^p A_i \left[1 + \frac{1}{r_i} B + \left(\frac{1}{r_i} B \right)^2 + \dots \right] = \sum_{j=0}^{\infty} \psi_j B^j.$$

This expansion is convergent if and only if $\left| \frac{1}{r_i} \right| < 1$ for all $i = 1, \dots, p$, which leads to

the fundamental root condition:

$$|r_i| > 1, \quad \forall i.$$

This implies that all roots of the characteristic polynomial must lie outside the unit circle.

3. Proof of Absolute Summability and Stationarity

From equation $Y_t = \Phi(B)^{-1} \varepsilon_t = \Psi(B) \varepsilon_t$, the process can be written as a causal linear process:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where the coefficients are given by $\psi_j = \sum_{i=1}^p A_i \left(\frac{1}{r_i}\right)^j$. Given the root condition $|r_i| > 1$, there exists a constant C and the smallest absolute root r_{min} such that:

$$\sum_{j=0}^{\infty} |\psi_j| \leq \sum_{j=0}^{\infty} C \left| \frac{1}{r_{min}} \right|^j < \infty.$$

Since the sequence $\{\psi_j\}$ is shown to be absolutely summable, we can directly apply the Theorem of Stationarity for Linear Processes (as detailed in Section 2.4). This confirms that the $AR(p)$ process is weakly stationary, as it satisfies the requirement that its first and second moments remain independent of time t .

Parameter Estimation of the Model In this study, the parameters of the AR model are estimated using the Least Squares Estimation (LSE) method. The concept of this method is to determine the parameter values that minimize the sum of squared residuals (Gujarati and Porter, 2009; Hamilton, 1994).

Consider the model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \quad t = p+1, \dots, n.$$

Let the residual at time t be

$$\varepsilon_t = Y_t - \phi_0 - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p}.$$

The objective function of the Least Squares method is

$$S(\phi_0, \phi_1, \dots, \phi_p) = \sum_{t=1}^n \varepsilon_t^2$$

or

$$S(\phi_0, \phi_1, \dots, \phi_p) = \sum_{t=1}^n (Y_t - \phi_0 - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p})^2.$$

The estimation of parameters using the Least Squares method is the process of finding the values of $\phi_0, \phi_1, \dots, \phi_p$ that minimize S (Draper and Smith, 1998).

In addition, the $AR(p)$ model can be expressed in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\phi} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & Y_0 & \cdots & Y_{1-p} \\ 1 & Y_1 & \cdots & Y_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Y_{n-1} & \cdots & Y_{n-p} \end{bmatrix}, \quad \boldsymbol{\phi} = \begin{bmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}.$$

When written in matrix form, the Least Squares estimator of the parameter vector $\boldsymbol{\phi}$ is

$$\hat{\boldsymbol{\phi}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

provided that $(\mathbf{X}^T \mathbf{X})^{-1}$ exists, where $\mathbf{X}^T$ denotes the transpose of $\mathbf{X}$.

The $AR(p)$ model has the advantage of having a simple structure and being able to explain the relationship between the current value and past values. In addition, the Autoregressive model is an important foundation of the ARIMA model (Box et al., 2015) and modern hybrid time series models. However, the limitation of the model is that it is suitable only for linear relationships and requires the assumption that the time series data are stationary. In some cases, the structure of time series data may not depend solely on past values, but may also be influenced by past error terms. This leads to the discussion of the Moving Average (MA) model in the next section.

2.5.2 Moving Average Model (MA)

The Moving Average Model (MA) is another fundamental model in time series analysis. The concept of the model is that the current value of a variable can be explained by the current random error term and past random error terms under a linear relationship. This concept was originally developed from the work of Slutsky in 1927 (Slutsky, 1927). An MA model of order q can be written as follows:

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

where

- Y_t represents the time series value at time t ,
- θ_0 is the constant term of the model,
- $\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the model,
- ε_t is the random error term at time t .

In the Moving Average (MA) model, ε_t is commonly assumed to be a white noise process with zero mean, constant variance, and no temporal correlation. Under the concept of the MA model, the current value of Y_t is influenced by the current and past random error terms over only a limited number of q time periods. Therefore, the MA model is considered suitable for describing the short-term relationship in time series data.

Selection of the Order of the MA(q) Model Determining the order of the Moving Average model, denoted by MA(q), is an important step in constructing a time series model because the value of q represents the number of past random error terms that still influence the current value. If the order is specified too low, the model may not adequately capture the structure of the data. On the other hand, if the order is specified too high, the model may become unnecessarily complex and reduce forecasting performance.

In general, the order of the MA model is commonly determined by considering the Autocorrelation Function (ACF), since the ACF of an MA(q) process remains significant

up to lag q (Box et al., 2015; Hamilton, 1994) and exhibits a cut-off behavior afterward. That is, for $k > q$, we have $\rho_k = 0$. Therefore, in theory, the ACF plot can be used as a preliminary guideline for determining the order of the model. However, in practice, the order of the model is usually considered together with information criteria such as AIC and BIC in order to confirm the appropriateness of the model.

Stationary of the MA(q) process The Moving Average process of order q , or MA(q), is defined as:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} = \Theta(B) \varepsilon_t,$$

where $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$. Connection to Linear Processes (Section 2.4):

The MA(q) model is inherently a special case of the linear process defined in Section 2.4. By setting the coefficients $\psi_j = \theta_j$ for $j \leq q$ and $\psi_j = 0$ for $j > q$, we can observe the following:

1) Absolute Summability: Since the MA(q) process consists of a finite number of terms, the condition $\sum_{j=0}^{\infty} |\psi_j| < \infty$ is always satisfied for any finite values of θ .

2) Stationarity Guaranteed: Following the theorem established in Section 2.4, any absolutely summable linear process is weakly stationary. Consequently, every MA(q) process is inherently weakly stationary, with a constant mean $E[Y_t] = 0$ and a time-invariant variance $\text{Var}(Y_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$.

Conclusion: Unlike the AR(p) process, the stationarity of an MA(q) model does not depend on the roots of its characteristic polynomial. It is stationary by definition as a finite and absolutely summable linear process. (Brockwell and Davis, 2016)

Parameter Estimation of the Model For parameter estimation in the Moving Average model, denoted by MA(q), the commonly used method is Conditional Least Squares (CLS). This method operates through a recursive calculation process to determine the parameter values that minimize the Sum of Squared Errors (SSE), as follows:

1. Initial Condition

Since the MA(q) model at time $t = 1$ requires past error terms prior to the observation period, namely $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)}$, which cannot be observed, it is common in practice to specify the initial condition by setting these initial error terms equal to zero, that is, $\varepsilon_t = 0$ for $t \leq 0$ in order to begin the calculation process.

2. Error Calculation

After specifying the initial values, the model calculates the error terms ε_t sequentially from $t = 1$ to n using the previously calculated error terms according to the following equation:

$$\varepsilon_t = Y_t - \theta_0 - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}.$$

3. Optimization

The process is iterated to determine the set of parameters $(\theta_0, \theta_1, \dots, \theta_q)$ that minimizes the objective function or the value of

$$S(\theta_0, \theta_1, \theta_2, \dots, \theta_q) = \sum_{t=1}^n \varepsilon_t^2.$$

In general, this process relies on numerical methods implemented through statistical software in order to obtain the most accurate parameter estimates (Brockwell and Davis, 2016).

The Moving Average model, denoted by MA(q), has the advantage of appropriately explaining the effects of past random error terms on the current value. The model structure reflects the short-term relationship of time series data and naturally satisfies the stationarity property for finite-order cases. However, the limitation of the model is that parameter estimation is more difficult than that of the AR model, and the selection of the model order may not always be clear for some real datasets.

In addition, the Moving Average model is an important component of the combined AR and MA model, known as the Autoregressive Moving Average (ARMA) model.

When applied to non-stationary data through differencing, it leads to the Autoregressive Integrated Moving Average (ARIMA) model, which will be discussed in the next section.

2.5.3 ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used time series models for the analysis and forecasting of non-stationary time series data. The model was systematically developed from the concepts of George Box and Gwilym Jenkins, which were presented in the influential work *Time Series Analysis: Forecasting and Control*, first published in 1970, and later became the foundation of the approach known as the Box–Jenkins methodology (Box and Jenkins, 1976).

The main concept of the ARIMA model is to combine the structure of the Autoregressive (AR) model and the Moving Average (MA) model with the differencing process in order to transform non-stationary data into stationary form before constructing the model. Therefore, the ARIMA model is suitable for time series data that exhibit trends or changing mean levels over time.

The general form of the ARIMA model is denoted by $\text{ARIMA}(p, d, q)$, where

- p is the order of the Autoregressive component,
- q is the order of the Moving Average component,
- d is the number of differencing operations required to make the data stationary.

The $\text{ARIMA}(p, d, q)$ model can be written in the general form as follows:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t,$$

where

- B is the backshift operator, where $BY_t = Y_{t-1}$,
- $(1 - B)^d$ is the differencing operator of order d ,
- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p$,

- $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$,
- ε_t is the random error term at time t .

In the structure of the model, ε_t is assumed to be a white noise process with the properties described in Section 2.2.

The ARIMA model consists of three components: the Autoregressive component $AR(p)$, the differencing component used for non-stationary data (Integrated: I), and the Moving Average component $MA(q)$. Each component can explain the structure of the model as follows:

1. The Autoregressive component $AR(p)$ explains the relationship between the current value of the data and its past values over p previous time periods. The model can be written as follows:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

2. The Integrated component (I) begins with testing the stationarity of the time series data using the Augmented Dickey-Fuller (ADF) Test (as described in Section 2.1). If the data are found to be non-stationary, the differencing process is applied once ($d = 1$), and the ADF Test is performed again. If the test indicates that the data are stationary, then the parameter d is equal to 1. However, if the data remain non-stationary, higher-order differencing is applied. The differencing operator of order d can be written in terms of the Backshift Operator as follows:

Case $d = 1$:

$$(1 - B)Y_t = Y_t - Y_{t-1}$$

Case $d = 2$:

$$(1 - B)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

The value of d represents the number of differencing operations required to remove the trend or non-stationary characteristics of the time series data in order to prepare the data for model construction in the next step.

3. The Moving Average component $MA(q)$ explains the current value of the data using past random error terms over q previous time periods. The model can be written as follows:

$$Y_t = \theta_0 + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} + \varepsilon_t$$

Combining these three components results in the $ARIMA(p, d, q)$ model, which can be written in the following form: $\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t$

Selection of the Order of the $ARIMA(p, d, q)$ Model In determining the order of the $ARIMA(p, d, q)$ model, the process begins by transforming the data into stationary form through statistical testing using the ADF Test (Dickey and Fuller, 1979; Said and Dickey, 1984). After that, the orders of the parameters p and q are considered. The order of $AR(p)$ is identified using the Partial Autocorrelation Function (PACF), as described in the section *Selection of the Order of the $AR(p)$ Model*, while the order of $MA(q)$ is identified using the Autocorrelation Function (ACF), as described in the section *Selection of the Order of the $MA(q)$ Model*.

In addition, the identification of the model order is considered together with information criteria such as AIC or BIC (Akaike, 1974; Schwarz, 1978) in order to select the model with the highest efficiency and the least complexity. Finally, the model adequacy is examined to ensure that the residual terms behave as white noise.

Stationary of the $ARIMA(p, d, q)$ process The stationarity of an $ARIMA$ process is fundamentally evaluated through the properties of its transformed series (Hamilton, 1994; Brockwell and Davis, 2016). The conditions for stationarity are categorized based on the order of differencing (d) as follows:

1. Transformation and Differencing Requirements

The requirement for achieving a stationary series depends on the initial state of the data:

- **Case $d = 0$:** When d is equal to zero, the original series $\{Y_t\}$ is assumed to be inherently stationary. In this scenario, no differencing transformation is

required to stabilize the statistical properties of the series, such as its mean and variance.

- **Case $d > 0$ (The ARIMA Process):** For any $d > 0$, the original series $\{Y_t\}$ is inherently non-stationary, typically characterized by the presence of one or more unit roots (stochastic trends). The application of the differencing operator $(1 - B)^d$ is necessary to eliminate these trends (Box et al., 2015; Hyndman and Athanasopoulos, 2018). While the raw series $\{Y_t\}$ lacks weak stationarity, the transformed series

$$W_t = (1 - B)^d Y_t$$

is rendered stationary.

2. Autoregressive Stability (The Unit Root Condition)

For the process to maintain weak stationarity (either the original series when $d = 0$, or the differenced series when $d > 0$), the characteristic equation associated with the autoregressive polynomial:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$$

must satisfy the stability condition. Specifically, all complex roots of $\Phi(B) = 0$ must lie outside the unit circle in the complex plane (Hamilton, 1994; Brockwell and Davis, 2016). This ensures that the influence of historical observations dissipates over time rather than persisting indefinitely.

3. Invertibility of the Moving Average Component

Although the moving average polynomial $\theta(B)$ does not dictate stationarity, it governs the invertibility of the model. For the model to be uniquely identifiable and representable as an infinite AR process, the roots of the characteristic equation:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q = 0$$

are conventionally required to lie outside the unit circle (Box et al., 2015; Tsay, 2010). This condition ensures that the current observation can be expressed as a convergent linear combination of past observations.

Parameter Estimation of the Model After determining the appropriate order of the ARIMA(p, d, q) model, the next step is to estimate the parameters of the autoregressive and moving average components, namely $\phi_1, \phi_2, \dots, \phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$, in order for the model to appropriately describe the characteristics of the data. In this study, the parameter estimation method used is the Least Squares Estimation method (Brockwell and Davis, 2016), in which the principle is to select the parameter values that minimize the sum of squared residuals.

For parameter estimation, the residual of the model at time t is defined as

$$e_t(\phi, \theta) = \phi(B)(1 - B)^d Y_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q},$$

which is the residual calculated from the observed data under the specified parameter values. In general, the residuals can be recursively computed from the ARIMA model equation. The objective function is then defined as the sum of squared residuals as follows:

$$S(\phi, \theta) = \sum_{t=1}^n e_t^2(\phi, \theta),$$

where $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ and $\theta = (\theta_1, \theta_2, \dots, \theta_q)$

Therefore, parameter estimation using the Least Squares method is the process of finding the estimates $\hat{\phi}$ and $\hat{\theta}$ that minimize $S(\phi, \theta)$, that is,

$$(\hat{\phi}, \hat{\theta}) = \arg \min_{\phi, \theta} S(\phi, \theta).$$

Since the parameters in the Moving Average (MA) component appear in a nonlinear form through the lagged residuals, the parameter estimates cannot be directly obtained in closed form as in ordinary linear regression. Therefore, iterative numerical optimization methods (Box et al., 2015; Nocedal and Wright, 2006) are required to search for the

optimal values, such as the Gauss–Newton method or the Levenberg–Marquardt method. The calculation process begins with initial parameter values and repeatedly updates them until the change in the objective function becomes very small or converges to a minimum point.

The ARIMA model is one of the most widely used tools in time series analysis because it has a clear structure and is supported by a strong statistical theoretical foundation. One of the main strengths of the model is its ability to handle non-stationary data through the differencing process, making it suitable for data with trends or changing mean levels over time. In addition, the model can effectively explain temporal relationships through the autoregressive (AR) and moving average (MA) components, together with a systematic estimation and diagnostic procedure based on the Box–Jenkins methodology. As a result, ARIMA is considered an effective tool for short- to medium-term forecasting of univariate time series data. However, the ARIMA model also has several important limitations. First, the model was developed under the framework of single-frequency data, which requires all variables to share the same time frequency context, such as monthly or quarterly data. Consequently, the model cannot directly integrate variables with different frequencies. Second, the model focuses on explaining linear relationships and therefore may not adequately capture complex dynamics or nonlinear relationships. Finally, increasing the lag order may lead to the problem of overparameterization, which can reduce out-of-sample forecasting performance. Particularly in economic and financial research, the data often consist of mixed-frequency variables, such as a monthly dependent variable combined with daily or quarterly explanatory variables. The ARIMA model cannot efficiently accommodate this type of data structure. For this reason, models capable of directly integrating mixed-frequency data were developed, namely the Mixed Data Sampling (MIDAS) model, which has become increasingly popular in recent years.

2.5.4 Mixed Data Sampling Model: MIDAS Model

In general econometric time series analysis, such as the ARIMA (Autoregressive Integrated Moving Average) model, there is usually a basic requirement that all variables in the system must have the same data frequency. For example, if the objective is to

forecast Gross Domestic Product (GDP), which is quarterly data, the explanatory variables must also be in quarterly frequency. However, in practice, economic and market data are often collected at different frequencies, such as monthly interest rates or daily stock price indices. If these variables are directly analyzed together with quarterly variables, a frequency mismatch problem will occur. In the past, this problem was commonly addressed through temporal aggregation, which involves converting high-frequency data into lower-frequency data (for example, averaging daily data to obtain monthly values). However, this approach often leads to the loss of important information, particularly short-term volatility and leading economic signals, thereby reducing forecasting performance (Marcellino and Schumacher, 2010). To overcome these limitations, Eric Ghysels and colleagues (2004) introduced the MIDAS (Mixed Data Sampling) model, which is a framework that allows explanatory variables with different frequencies to be incorporated into the same equation without converting them to lower-frequency data beforehand. As a result, the characteristics and information contained in the original data can be preserved more completely (Ghysels, Santa-Clara, and Valkanov, 2004).

The MIDAS model is a model that explains the relationship between a low-frequency dependent variable (Y_t) and a high-frequency independent variable (X_t) through a distributed lag weighting structure, as shown in the following equation:

$$Y_t = \beta_0 + \beta_1 \sum_{k=1}^K w_k(\nu) X_{t-k/m} + \varepsilon_t,$$

where

- Y_t is the dependent variable observed at low frequency at time t ,
- $X_{t-k/m}$ is the independent variable observed at high frequency,
- K is the maximum number of lags considered,
- $w_k(\nu)$ is the lag weight at lag k , which depends on the parameter ν ,
- β_0 is the constant term of the model,
- β_1 is the coefficient of the weighted sum,

- ε_t is the random error term, and
- m is the frequency ratio between the explanatory variable X and the dependent variable Y .

An important advantage of the MIDAS model is its ability to directly utilize the potential of high-frequency data to explain or forecast low-frequency variables, such as using daily raw data to estimate monthly economic indicators. For this reason, MIDAS has become widely used in macroeconomic, financial, and asset price forecasting research (Foroni, Marcellino, and Schumacher, 2015), and is considered an important tool in modern time series analysis that effectively reduces the limitations of single-frequency models.

Multivariate MIDAS: In real economic and social conditions, the dependent variable is often influenced simultaneously by various surrounding factors. An analysis using only a single explanatory variable may not comprehensively explain the complex relationship structure. For example, forecasting asset prices may require the simultaneous consideration of financial market indices, exchange rates, and commodity prices. For this reason, the MIDAS model was extended to the Multivariate MIDAS Regression framework in order to support the analysis involving p explanatory variables, as shown in the following equation:

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j \sum_{k=1}^K w_{j,k}(\nu_j) X_{j,t-k/m} + \varepsilon_t,$$

where

- Y_t is the dependent variable observed at low frequency at time t ,
- p is the total number of explanatory variables,
- $X_{j,t-k/m}$ is the j th explanatory variable,
- K is the maximum lag considered for the j th variable,
- $w_{j,k}(\nu_j)$ is the lag weight at lag k for the j th variable, which depends on the parameter ν_j ,

- β_0 is the constant term of the model,
- β_j is the coefficient of the j th explanatory variable,
- ε_t is the random error term.

From the above equation, it can be seen that each explanatory variable can have a different number of lags and weighting structure, providing the model with high flexibility in reflecting temporal effects. For example, some variables may affect Y_t in the short term, while others may exert cumulative long-term effects. MIDAS can appropriately represent these effects through the weighting function assigned to each variable. An important advantage of the multivariate MIDAS model is its ability to combine data from multiple sources with different frequencies into a single equation without the need to transform all variables into the same frequency beforehand. Consequently, this approach helps reduce information loss and improve forecasting performance, particularly when the dependent variable is influenced simultaneously by multiple factors (Foroni et al., 2015). However, as the number of explanatory variables increases, problems such as multicollinearity may arise, parameter estimation becomes more complex, and the model may be more susceptible to overfitting. Therefore, many studies apply variable selection methods, dimensionality reduction techniques, or combine MIDAS with machine learning methods in order to improve model performance. At present, the multivariate MIDAS model has become highly popular in macroeconomics, financial market forecasting, asset pricing, energy analysis, and public health data analysis because it can efficiently utilize information from multiple sources and multiple frequencies simultaneously.

MIDAS-AR: The MIDAS-AR model is an extension of the Mixed Data Sampling (MIDAS) model by incorporating autoregressive terms of the dependent variable into the equation. This allows the model to simultaneously reflect both the influence of high-frequency explanatory variables and the temporal persistence of the dependent variable. This type of model is suitable for time series data in which the current value is related to its past values, such as gross domestic product, inflation rates, asset prices, and other economic and financial variables. Clements and Galvão (2008) applied the MIDAS framework to

forecast economic growth in the United States by incorporating lagged values of the dependent variable into the forecasting equation in order to capture the persistence characteristics of macroeconomic data. This approach is considered one of the important foundations of the MIDAS-AR model and has become widely used in econometric studies. The general form of the MIDAS-AR model can be written as follows:

$$Y_t = \beta_0 + \sum_{r=1}^P \phi_r Y_{t-r} + \sum_{j=1}^p \beta_j \sum_{k=1}^K w_{j,k}(\nu_j) X_{j,t-k/m} + \varepsilon_t$$

where

- Y_t is the dependent variable observed at low frequency at time t ,
- p is the total number of explanatory variables,
- P is the autoregressive order,
- ϕ_r is the coefficient of the r th lagged dependent variable,
- $X_{j,t-k/m}$ is the j th explanatory variable,
- K is the maximum lag considered for the j th variable,
- $w_{j,k}(\nu_j)$ is the lag weight at lag k for the j th variable, which depends on the parameter ν_j ,
- β_0 is the constant term of the model,
- β_j is the coefficient of the j th explanatory variable,
- ε_t is the random error term.

An important advantage of the MIDAS-AR model is its ability to efficiently utilize mixed-frequency data while simultaneously capturing the temporal relationship of the dependent variable through its past values. Therefore, the model is suitable for data with high persistence. In addition, the model reduces the information loss that commonly occurs when high-frequency data are aggregated into lower-frequency data before analysis (Andreou, Ghysels, and Kourtellis, 2010). However, the application of the model still has important limitations, particularly the complexity of parameter estimation because the

weighting functions are generally nonlinear. Consequently, estimation typically requires the use of Nonlinear Least Squares (NLS). Furthermore, caution must be taken regarding overfitting problems resulting from specifying too many lag terms, as well as the common factor restriction assumption, which may affect forecasting accuracy when the dynamic structures of the dependent and explanatory variables differ substantially. Therefore, the MIDAS-AR model is considered an appropriate approach for forecasting time series data that require both mixed-frequency explanatory variables and lagged dependent variables. The model effectively combines the advantages of autoregressive models and MIDAS models into a unified framework.

Weight Functions: An important core component of the MIDAS model is the weighting function, which determines the level of influence of high-frequency explanatory variables at different lag values on the low-frequency dependent variable. The use of weighting functions allows a large number of parameters to be estimated using a small parameter vector (ν), which helps reduce the dimensionality curse problem and prevents overfitting when the number of lag terms (K) is large (Ghysels, Sinko, and Valkanov, 2007).

In general, the weighted component of the high-frequency variable can be written in the following general form:

$$Z_t = \sum_{k=1}^K w_k(\nu) X_{t-k/m},$$

where

- Z_t is the weighted sum of the high-frequency explanatory variable,
- $X_{t-k/m}$ is the value of the explanatory variable at lag k (where m is the frequency ratio between the high-frequency and low-frequency data),
- $w_k(\nu)$ is the weighting function at lag k ,
- ν is the parameter vector controlling the characteristics and shape of the weighting distribution.

In this study, the Exponential Almon Lag weighting function is selected because it has high flexibility and requires only two parameters for estimation, namely ν_1 and ν_2 . The

weighting function can be written as follows:

$$w_k(\nu_1, \nu_2) = \frac{\exp(\nu_1 k + \nu_2 k^2)}{\sum_{j=1}^K \exp(\nu_1 j + \nu_2 j^2)}, \quad k = 1, \dots, K.$$

An important characteristic of this function is its ability to reflect various temporal effect structures, such as geometric decay or hump-shaped effects in which the influence initially increases before gradually decreasing. These characteristics are consistent with the behavior of economic and financial market variables.

In order for the estimation to have statistical meaning and to be interpreted as relative weights, the weighting function in the MIDAS model must satisfy the following fundamental theoretical properties.

Theorem: Normalization

Theorem 2.2. *Let the weighting function be defined as*

$$w_k(\nu) = \frac{g_k(\nu)}{\sum_{j=1}^K g_j(\nu)},$$

where $g_k(\nu) > 0$ for all values of k . Then, the sum of all weights is always equal to one:

$$\sum_{k=1}^K w_k(\nu) = 1,$$

where $g_k(\nu) = \exp(\nu_1 k + \nu_2 k^2)$ and $g_j(\nu) = \exp(\nu_1 j + \nu_2 j^2)$

Proof

Consider the sum of the weighting functions:

$$\sum_{k=1}^K w_k(\nu) = \sum_{k=1}^K \left[\frac{g_k(\nu)}{\sum_{j=1}^K g_j(\nu)} \right].$$

Since the denominator

$$\sum_{j=1}^K g_j(\nu)$$

is constant with respect to the summation index k , it can be factored out of the summation:

$$\sum_{k=1}^K w_k(\nu) = \frac{1}{\sum_{j=1}^K g_j(\nu)} \sum_{k=1}^K g_k(\nu) = 1$$

Theorem: Positivity

Theorem 2.3. *For the Exponential Almon lag weighting function, the weights at all lag periods are always positive:*

$$w_k(\nu_1, \nu_2) > 0, \quad \forall k \in \{1, 2, \dots, K\}.$$

Proof

From the property of the exponential function, whose range is always positive real numbers ($\exp(x) > 0$ for all $x \in \mathbb{R}$), both the numerator $\exp(\nu_1 k + \nu_2 k^2)$ and every term in the denominator are positive. Therefore, the ratio of positive real numbers is always positive.

Theorem: Boundedness

Theorem 2.4. *The weighted sum Z_t always lies within the bounds of the original data if $w_k(\nu) \geq 0$ and $\sum w_k(\nu) = 1$. That is, $\min_{k=1, \dots, K} (X_{t-k/m}) \leq Z_t \leq \max_{k=1, \dots, K} (X_{t-k/m})$.*

Proof

Let $m = \min(X_{t-k/m})$ and $M = \max(X_{t-k/m})$. Then, $m \leq X_{t-k/m} \leq M$. Multiplying by $w_k(\nu) \geq 0$ and summing over k gives

$$m \sum_{k=1}^K w_k(\nu) \leq \sum_{k=1}^K w_k(\nu) X_{t-k/m} \leq M \sum_{k=1}^K w_k(\nu).$$

Since $\sum_{k=1}^K w_k(\nu) = 1$, it follows that, $m \leq Z_t \leq M$.

From the mathematical properties of the weighting functions presented above, it can be seen that the specification of weights in the MIDAS model plays an important role in the performance of the model in several aspects.

First, regarding model stability, specifying that each weight is nonnegative and that the sum of all weights is equal to one causes the weighted sum Z_t to behave as a weighted average of past observations. As a result, the obtained values do not deviate

excessively from the range of the original data and help reduce fluctuations that may arise from inappropriate weighting schemes. This property is important for parameter estimation and forecasting because it contributes to greater model stability.

Second, regarding the interpretation of results, expressing Z_t as a convex combination allows a clear interpretation that it is obtained from combining past high-frequency observations with different weights assigned to different time periods. In other words, observations from some periods may exert greater influence on the dependent variable than others, depending on the form of the weighting function selected. Consequently, this helps explain the temporal impact structure of past information in a systematic manner.

In addition, the use of weighting functions helps reduce the number of parameters that need to be estimated because it is unnecessary to estimate the coefficient of each lag separately. Instead, the effects of many lag terms can be represented using only a small number of parameters. As a result, the model becomes more parsimonious and more suitable for practical applications with real-world data.

Therefore, the weighting function is not merely a technical component of the MIDAS model, but also an important mechanism that provides mathematical appropriateness, estimation stability, and clear interpretability of the model results.

Parameter Estimation of the Model The parameter estimation of the MIDAS model aims to obtain the most appropriate parameter estimates in order for the model to explain the data and provide accurate forecasting. In general, the parameters to be estimated consist of the constant term of the model, the coefficients of the explanatory variables, and the parameters of the weighting function (Ghysels et al., 2007; Andreou et al., 2010), and in some cases may also include the coefficients of lagged dependent variables. Since the parameters of the weighting function appear in a nonlinear form, they cannot be estimated directly using the Ordinary Least Squares (OLS) method. Therefore, this study uses the Nonlinear Least Squares (NLS) method (Ghysels et al., 2007; Hamilton, 1994) for parameter estimation by finding the set of parameters that minimizes the sum of squared residuals, as follows:

$$\hat{\psi} = \arg \min_{\psi} \sum_{t=1}^T \left(Y_t - \hat{Y}_t(\psi) \right)^2,$$

where ψ represents the vector of all model parameters, and $\hat{Y}_t(\psi)$ is the estimated value obtained from the model under that parameter set. In the numerical computation, the researcher uses the L-BFGS-B algorithm (Byrd et al., 1995; Zhu et al., 1997), which is an optimization method suitable for nonlinear objective function minimization problems and allows parameter bounds to be specified. The estimation process is performed iteratively until the objective function converges to the minimum value according to the specified criterion.

Since the MIDAS model can be classified into several forms according to the number of explanatory variables and dynamic components, this study divides the presentation of parameter estimation into three cases: the univariate MIDAS model, the multivariate MIDAS model, and the MIDAS model with autoregressive terms, as described in the following sections.

Case 1: BASIC MIDAS

The model contains only one explanatory variable (Ghysels et al., 2004). The equation can be written as follows:

$$Y_t = \beta_0 + \beta_1 \sum_{k=1}^K w_k(\nu) X_{t-k/m} + \varepsilon_t,$$

where

- β_0 is the constant term of the model,
- β_1 is the coefficient of the explanatory variable,
- ν is the parameter of the weighting function.

The estimation in this case is to determine the values of β_0 , β_1 , and ν that minimize the sum of squared residuals. This model is suitable for studying the effect of only one main factor on the change of the dependent variable.

Case 2: Multivariate MIDAS

In the case where there is more than one explanatory variable (Andreou et al., 2010; Foroni et al., 2015), the model can be extended as follows:

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j \sum_{k=1}^K w_{j,k}(\nu_j) X_{j,t-k/m} + \varepsilon_t,$$

where $j = 1, 2, \dots, p$ represents the index of each explanatory variable, in which each variable can have different numbers of lag terms, data frequencies, and weighting function structures.

The parameter estimation in this case requires the estimation of β_0 , β_j , and ν_j for all variables simultaneously under the principle of minimizing the sum of squared residuals. This type of model is suitable for cases where the dependent variable is influenced by multiple factors simultaneously, which helps improve the ability to explain the data more effectively.

Case 3: MIDAS-AR

In some cases, the dependent variable may exhibit persistence from its own past values, meaning that the current value is also influenced by lagged values of the dependent variable. Therefore, the MIDAS model is extended by including autoregressive terms (Clements and Galvão, 2008; Ghysels et al., 2007), resulting in the MIDAS-AR model as follows:

$$Y_t = \beta_0 + \sum_{r=1}^P \phi_r Y_{t-r} + \sum_{j=1}^p \beta_j \sum_{k=1}^K w_{j,k}(\nu_j) X_{j,t-k/m} + \varepsilon_t,$$

where ϕ_r is the coefficient of the r th lagged dependent variable, and P is the order of the autoregressive process. The parameter estimation in this case requires the simultaneous estimation of β_0 , ϕ_r , β_j , and ν_j , while still applying the Nonlinear Least Squares method in the same manner as the general MIDAS model.

Each type of MIDAS model has a different structure depending on the number of explanatory variables and temporal components. However, the parameter estimation process still relies on the same principle, namely finding the set of parameters that allows

the model to explain empirical data and provide forecasts with the highest efficiency.

In summary, the MIDAS family of models represents an important development in time series analysis and mixed-frequency forecasting because it allows high-frequency data to be directly incorporated with low-frequency dependent variables without the need for prior data aggregation. This helps reduce information loss and improve forecasting performance. These models have been continuously extended from the basic MIDAS framework to more complex forms, such as multivariate MIDAS and MIDAS-AR, which can effectively capture the influence of multiple factors together with the dynamics of past dependent variables. However, although MIDAS models have advantages in handling mixed-frequency data, they are still fundamentally based on the framework of linear relationships and assume that the model parameters remain constant throughout the entire period. This assumption may not be consistent with real-world data, which often exhibit structural changes, economic fluctuations, or market behaviors that vary over time. For this reason, researchers have developed approaches that can better accommodate changes in data conditions by integrating the MIDAS framework with models that allow parameters to vary according to different regimes or system states. This has led to the development of the Regime-Switching MIDAS model, which will be discussed in the next section.

2.6 Regime-Switching Model

2.6.1 Regime Concept in Financial Markets

In the context of financial markets, asset prices and returns often exhibit behavior that changes according to economic conditions, monetary policy, or geopolitical events. As a result, the statistical characteristics of time series data do not remain constant under a single structural regime, but instead display structural changes over different periods, referred to as “regimes” or states (Hamilton, 1989). In each regime, the parameters of the system, such as the mean (μ) or variance (σ^2), may differ significantly. Mathematically, a

state variable S_t can be defined to represent the regime at time t as follows:

$$S_t \in \{1, 2, \dots, \mathcal{M}\},$$

where $\mathcal{M}$ is the number of specified regimes. For example, when $\mathcal{M} = 2$, the regimes may represent a normal state and a crisis state. Understanding this concept helps overcome the limitations of traditional linear models, which generally assume time-invariant parameters that remain constant throughout the entire period, an assumption that is often inconsistent with the reality of financial markets.

2.6.2 Threshold Regime-Switching Model

This is a model that determines the regime from the value of a directly observable variable. When the value of such variable is greater than or less than a threshold value, the system switches into another regime. This concept was systematically developed by Tong (1978) and Tong and Lim (1980).

An example of the basic form of the Threshold Autoregressive Model (TAR) can be written as follows:

$$Y_t = \begin{cases} \alpha_1 + \phi_1 Y_{t-1} + \epsilon_t, & q_t \leq c, \\ \alpha_2 + \phi_2 Y_{t-1} + \epsilon_t, & q_t > c, \end{cases}$$

where q_t is the threshold variable, and c is the threshold value.

For this study, the Threshold Regime-Switching Model can be divided into two forms as follows:

1. Threshold MIDAS

$$Y_t = \begin{cases} \beta_{0,1} + \sum_{j=1}^p \beta_{j,1} \sum_{k=1}^K w_{j,k}(\nu_{j,1}) X_{j,t-k/m} + \epsilon_t, & Y_{t-1} \leq c, \\ \beta_{0,2} + \sum_{j=1}^p \beta_{j,2} \sum_{k=1}^K w_{j,k}(\nu_{j,2}) X_{j,t-k/m} + \epsilon_t, & Y_{t-1} > c. \end{cases}$$

2. Threshold MIDAS-AR

$$Y_t = \begin{cases} \beta_{0,1} + \sum_{r=1}^P \phi_{r,1} Y_{t-r} + \sum_{j=1}^p \beta_{j,1} \sum_{k=1}^K w_{j,k}(\nu_{j,1}) X_{j,t-k/m} + \varepsilon_t, & Y_{t-1} \leq c, \\ \beta_{0,2} + \sum_{r=1}^P \phi_{r,2} Y_{t-r} + \sum_{j=1}^p \beta_{j,2} \sum_{k=1}^K w_{j,k}(\nu_{j,2}) X_{j,t-k/m} + \varepsilon_t, & Y_{t-1} > c. \end{cases}$$

where

Y_{t-1} is the threshold variable,

c is the threshold value, and

d is the delay parameter.

The advantage of this type of model is that it is easy to interpret and suitable for data with clear turning points, such as when prices break through support or resistance levels. However, an important limitation is that an appropriate threshold value must be specified, and the regime switching occurs discontinuously.

2.6.3 Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) is a regime-switching model in which the regime or system state is defined as a hidden state that cannot be directly observed, but can be inferred from the behavior of the observed data. The fundamental concept of this model was developed by Baum and Petrie (1966), and was later widely applied in econometrics and finance, especially in the work of Hamilton (1989), which applied the Markov switching concept to explain business cycles.

Unlike the Threshold Regime-Switching Model, which determines the regime from directly observable variables, the HMM assumes that regimes evolve according to a Markov process. That is, the current state depends only on the state from the previous period, which can be written as follows:

$$P(S_t = j \mid S_{t-1} = i) = p_{ij},$$

where

- S_t is the state of the system at time t ,
- i, j are the regime indices,
- p_{ij} is the probability that the system changes from state i to state j .

The transition probability matrix for the case of two regimes can be written as follows:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

where

- p_{11} is the probability that the system remains in regime 1,
- p_{12} is the probability that the system changes from regime 1 to regime 2,
- p_{21} is the probability that the system changes from regime 2 to regime 1,
- p_{22} is the probability that the system remains in regime 2.

The sum of probabilities in each row must be equal to 1.

For this study, two regimes are specified by using gold returns and gold price volatility as important information for estimating the hidden states of the gold market in order to classify normal market periods and high-volatility market periods before integrating the results with the MIDAS model. The Hidden Markov Model combined with the MIDAS framework can be divided into two forms as follows:

1. HMM-MIDAS

$$Y_t = \begin{cases} \beta_{0,1} + \sum_{j=1}^p \beta_{j,1} \sum_{k=1}^K w_{j,k}(\nu_{j,1}) X_{j,t-k/m} + \varepsilon_t, & S_t = 1, \\ \beta_{0,2} + \sum_{j=1}^p \beta_{j,2} \sum_{k=1}^K w_{j,k}(\nu_{j,2}) X_{j,t-k/m} + \varepsilon_t, & S_t = 2. \end{cases}$$

2. HMM-MIDAS-AR

$$Y_t = \begin{cases} \beta_{0,1} + \sum_{r=1}^P \phi_{r,1} Y_{t-r} + \sum_{j=1}^p \beta_{j,1} \sum_{k=1}^K w_{j,k}(\nu_{j,1}) X_{j,t-k/m} + \varepsilon_t, & S_t = 1, \\ \beta_{0,2} + \sum_{r=1}^P \phi_{r,2} Y_{t-r} + \sum_{j=1}^p \beta_{j,2} \sum_{k=1}^K w_{j,k}(\nu_{j,2}) X_{j,t-k/m} + \varepsilon_t, & S_t = 2, \end{cases}$$

where

- S_t is the hidden state at time t ,
- p_{ij} is the transition probability from state i to state j ,
- $\beta_{0,s}$ is the constant term under regime s ,
- $\beta_{j,s}$ is the coefficient of the explanatory variable under regime s ,
- $\phi_{r,s}$ is the autoregressive coefficient under regime s ,
- $\nu_{j,s}$ is the weighting function parameter under regime s .

An important advantage of the HMM is its flexibility in adapting to market conditions without requiring a predefined threshold value. This allows the model to capture changes in unobserved states through a probabilistic process. However, the model is complex in terms of parameter estimation, and the results are sensitive to the specification of initial parameters, which may affect model stability.

Although the HMM is effective in identifying temporal regime switching, relying on only a small number of variables may not be sufficient to explain highly complex market conditions. Therefore, machine learning techniques such as the Self-Organizing Map (SOM) have recently been applied to cluster multivariate data and identify market patterns through dimensionality reduction before entering the HMM process. This approach helps improve forecasting accuracy and allows the determination of the number of regimes to be more consistent with the actual behavior of the data.

2.7 Self-Organizing Map

Self-Organizing Map (SOM) is an unsupervised learning technique developed by Teuvo Kohonen during the 1980s. Its objective is to cluster highly complex data and transform multidimensional data into a two-dimensional map that can be easily interpreted while preserving the structural relationships of the original data as much as possible (Kohonen, 1982, 1990).

The main concept of SOM is to feed multivariate data into a network of processing units (neurons) arranged on a two-dimensional grid. Each unit contains a weight vector representing the characteristics of the data. When a new observation enters the system, SOM searches for the unit that is most similar to the observation, called the Best Matching Unit (BMU). Then, the weights of the BMU and neighboring units are adjusted to move closer to the sample data. This process is repeatedly performed until systematic clustering of the data is achieved.

Let the input vector at time t be written as

$$x_t = (x_{1t}, x_{2t}, \dots, x_{nt})^T,$$

where

- x_t is the input vector at time t ,
- $x_{1t}, x_{2t}, \dots, x_{nt}$ are the values of variables 1 to n at time t ,
- n is the number of variables used as input data.

The BMU search uses the minimum Euclidean distance principle as follows:

$$c = \arg \min_i \|x_t - w_i\|,$$

where w_i denotes the weight vector associated with unit i . The weight vector of each unit is then updated according to

$$w_i(t+1) = w_i(t) + \eta(t)h_{ci}(t)[x_t - w_i(t)],$$

where $\eta(t)$ is the learning rate and $h_{ci}(t)$ is the neighborhood function satisfying

$$0 \leq h_{ci}(t) \leq 1, \quad \text{with} \quad h_{cc}(t) = 1.$$

The learning rate is a parameter that determines the magnitude of the weight update of the SOM network in each learning iteration. In other words, it controls how much the processing units move toward the new input data. When $\eta(t)$ has a high value, the weight adjustment becomes larger, causing the network to learn faster. In contrast, when $\eta(t)$ has a low value, the weight adjustment becomes smaller, resulting in slower but more stable learning. In financial market research, SOM plays an important role in identifying market behavior patterns, such as low-volatility periods, high-volatility periods, or trend transition periods. These characteristics may not be clearly observable from raw data using only a single variable. In this study, SOM is used as a preprocessing step before estimating the Hidden Markov Model (HMM) in order to help cluster the data and extract structural patterns of the gold market from important variables, namely gold returns and gold price volatility. The SOM results are then used as supplementary information for estimating the hidden states of the HMM in order to improve the clarity of regime classification.

After regime classification, the obtained states are incorporated into the MIDAS–AR model for forecasting the dependent variable. The main model of this study is the SOM–HMM–MIDAS–AR Model, which is a hybrid model combining artificial intelligence techniques, hidden state models, and mixed-frequency data models.

The hybrid SOM–HMM–MIDAS–AR model can be written in the following general form:

$$Y_t = \begin{cases} \beta_{0,1} + \sum_{r=1}^P \phi_{r,1} Y_{t-r} + \sum_{j=1}^p \beta_{j,1} \sum_{k=1}^K w_{j,k}(\nu_{j,1}) X_{j,t-k/m} + \varepsilon_t, & S_t = 1, \\ \beta_{0,2} + \sum_{r=1}^P \phi_{r,2} Y_{t-r} + \sum_{j=1}^p \beta_{j,2} \sum_{k=1}^K w_{j,k}(\nu_{j,2}) X_{j,t-k/m} + \varepsilon_t, & S_t = 2, \end{cases}$$

where S_t is the hidden state of the system at time t , which is estimated from the Hidden Markov Model using data clustered by the Self-Organizing Map (SOM).

2.8 Information Criteria

In statistical and time series modeling, model selection is an important step because a model with too many parameters may lead to overfitting, while a model that is too simple may not be able to adequately explain the structure of the data (underfitting). Therefore, information criteria have been developed to compare multiple models under the same dataset by considering the balance between the ability to explain the data and the complexity of the model (Burnham and Anderson, 2002; Box et al., 2015).

The most commonly used information criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which can be applied to various types of models, such as linear models, ARIMA time series models, regression models, and other probabilistic models (Hamilton, 1994; Brockwell and Davis, 2016).

2.8.1 Akaike Information Criterion (AIC)

The AIC was developed by Hirotugu Akaike (1974) based on Information Theory in order to estimate the information loss when a model is used to represent the true process (Akaike, 1974).

$$AIC = -2\ln(L) + 2k,$$

where

- L is the maximum likelihood of the model,
- k is the number of parameters to be estimated.

The interpretation of the AIC value is that a lower AIC indicates a more appropriate model when compared with other models fitted to the same dataset.

2.8.2 Bayesian Information Criterion (BIC)

The BIC, or Schwarz Information Criterion, was proposed by Gideon Schwarz in 1978 and developed based on the Bayesian approach (Schwarz, 1978).

$$BIC = -2\ln(L) + k\ln(n),$$

where

- L is the maximum likelihood of the model,
- k is the number of parameters to be estimated,
- n is the number of observations.

The interpretation of the BIC value is similar to that of the AIC . A model with a lower BIC value indicates a more appropriate model when compared under the same dataset. In practice, the BIC is suitable for research focusing on identifying an appropriate model structure and avoiding overfitting, including studies that require clear parameter interpretation. However, one limitation is that, in some cases, it may select a model that is too simple, which can reduce forecasting performance.

Both AIC and BIC have the same objective, namely model comparison and model selection, but the two criteria emphasize different aspects. The AIC is more suitable for studies focusing on forecasting accuracy, whereas the BIC is more suitable for studies emphasizing model parsimony and theoretical interpretation (Lütkepohl, 2005). Therefore, in applied mathematics and time series research, both AIC and BIC are commonly considered together. If a model provides low values for both criteria in the same direction, it increases confidence that the model is appropriate in terms of both performance and complexity. Nevertheless, they should be used together with residual diagnostics and actual forecasting performance in order to achieve a more complete model selection process (Box et al., 2015).

2.9 Forecasting Performance Evaluation

Evaluating the performance of forecasting models is essential for assessing the accuracy and reliability of predictions. In this study, three widely used statistical measures are employed: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). These indicators provide complementary perspectives on model performance.

2.9.1 Mean Absolute Error

MAE calculates the average absolute deviation between the predicted values ($\hat{Y}_t$) and the actual values (Y_t), treating all errors equally. It is straightforward to interpret and provides practical insights into forecasting accuracy. The formula is defined as:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|,$$

where Y_t represents the actual value and $\hat{Y}_t$ the forecast value at time t .

2.9.2 Mean Absolute Percentage Error

MAPE expresses forecast accuracy as a percentage error by comparing the predicted values ($\hat{Y}_t$) with the actual values (Y_t). This measure is particularly useful for comparing forecasting performance across datasets with different scales. The formula is defined as:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|.$$

2.9.3 Root Mean Square Error

RMSE measures the deviation between the predicted values ($\hat{Y}_t$) and the actual values (Y_t), placing greater weight on large errors. Consequently, RMSE is highly sensitive to severe deviations. The formula is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}.$$

By employing all three measures together, the evaluation offers a comprehensive assessment of model performance—capturing overall error magnitude (RMSE), relative percentage error (MAPE), and uniformly weighted average error (MAE).

2.10 Diebold–Mariano Test: DM-Test

The comparison of forecasting model performance should not rely only on average error measures such as RMSE, MAE, or MAPE because these values represent only the overall magnitude of forecasting errors and cannot clearly indicate whether the differences between models are statistically significant. Therefore, forecasting research commonly applies the Diebold–Mariano test (DM-Test), developed by Francis X. Diebold and Roberto S. Mariano in 1995, to examine whether two forecasting models have significantly different predictive accuracy (Diebold and Mariano, 1995).

The principle of this test is to compare the forecast loss of two models, such as Model 1 and Model 2. Let the actual value at time t be denoted by y_t , and the forecasts from each model be denoted by $\hat{y}_{1,t}$ and $\hat{y}_{2,t}$. The forecast errors can be calculated as follows:

$$e_{1,t} = y_t - \hat{y}_{1,t} \quad e_{2,t} = y_t - \hat{y}_{2,t}.$$

Next, define a loss function $L(\cdot)$, such as squared error loss or absolute error loss, and define the loss differential as

$$d_t = L(e_{1,t}) - L(e_{2,t}).$$

If the mean of d_t is equal to zero, this indicates that the two models do not differ in forecasting accuracy. However, if the mean significantly differs from zero, it can be concluded that one model outperforms the other (Diebold and Mariano, 1995; Harvey et al., 1997).

The DM-Test statistic can be written as

$$DM = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}},$$

where

- $\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t$ is the mean of the loss differential,
- $\widehat{Var}(\bar{d})$ is the estimated variance of the mean loss differential,

- T is the number of periods used for forecast evaluation.

Under the null hypothesis, the test statistic approximately follows the standard normal distribution.

Hypotheses

$H_0 : E(d_t) = 0$, meaning that the two models have no difference in forecasting ability.

$H_1 : E(d_t) \neq 0$, meaning that the two models have different forecasting ability.

If the p-value is less than the specified significance level, such as 0.05, the null hypothesis is rejected, and it can be concluded that the forecasting accuracy of the two models differs significantly. In the case where the DM statistic is positive, this indicates that Model 2 produces a smaller forecast loss than Model 1. In contrast, if the DM statistic is negative, this indicates that Model 1 performs better. However, this interpretation depends on the ordering of d_t .

The DM-Test has the advantage of directly comparing the forecasting accuracy of models based on forecast errors. In addition, it supports various forms of loss functions, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), making it flexible for application according to the characteristics of the data. Moreover, it is a standard method that has been widely accepted in forecasting and econometric research (Diebold and Mariano, 1995).

2.11 Related Research

Hamilton (1989) developed the Markov-Switching Model by allowing the parameters of an autoregressive process to change according to a discrete-state Markov chain. Maximum Likelihood Estimation combined with a nonlinear filter was employed to estimate the probabilities of being in different regimes. Using post-World War II U.S. Real GNP data, the study revealed a clear alternation between expansion and recession regimes, with recessions associated with a permanent decline in the level of GNP of approximately 3%.

Teräsvirta (1994) advanced the construction and evaluation of Smooth Transition Autoregressive (STAR) models, particularly the Logistic STAR (LSTAR) and Exponential STAR

(ESTAR) forms. The approach incorporates linearity testing to detect nonlinearity, selection of the delay parameter, and determination of the appropriate model type. Estimation is carried out via nonlinear least squares, with residual diagnostics applied for validation. Using both simulated data and empirical series such as Canadian lynx populations and German industrial production, the study demonstrated that STAR models provide a more flexible and continuous representation of regime transitions compared to linear or Threshold Autoregressive (TAR) models.

Hamilton and Susmel (1994) extended the regime-switching concept to financial volatility analysis by developing the Markov-Switching ARCH (SWARCH) model. This framework integrates Engle's (1982) ARCH model, which captures conditional heteroskedasticity, with Hamilton's (1989) Markov Switching mechanism, which models regime changes in economic activity. The SWARCH model allows volatility to shift flexibly across periods, enabling a better representation of high- and low-volatility regimes. Using weekly U.S. interest rate data, the study found that SWARCH outperformed the traditional ARCH model by more accurately capturing the alternation between high- and low-volatility states. These findings highlight the importance of incorporating regime structures into volatility modeling for a deeper understanding of financial market behavior and empirical financial econometrics.

Atsalakis and Valavanis (2009) reviewed intelligent forecasting methods in financial markets and highlighted SOM as an effective preprocessing and clustering tool for identifying hidden market structures before prediction. Their survey concluded that SOM-based hybrid models often outperform purely linear forecasting techniques in volatile financial environments.

Guérin and Marcellino (2013) introduced the Markov-Switching MIDAS (MS-MIDAS) model by integrating regime changes into the Mixed Data Sampling framework. The proposed model allows MIDAS parameters to vary across economic states governed by a latent Markov chain. Using U.S. and U.K. quarterly GDP data together with higher-frequency financial indicators, the study found that MS-MIDAS significantly improved both macroeconomic forecasting accuracy and real-time recession probability detection compared with standard MIDAS models.

Tsui, Xu, and Zhang (2013) conducted a study forecasting Singapore's GDP growth

one quarter ahead using mixed-frequency data. The research compared MIDAS regression with direct regression and time-averaging regression. The results demonstrated that the MIDAS model incorporating high-frequency stock return data significantly outperformed the other two approaches, with weekly stock returns providing the most accurate forecasts.

Nguyen and Le (2014) proposed a two-stage stock forecasting architecture combining Self-Organizing Map (SOM) with fuzzy Support Vector Machines. SOM was first used to partition market states into homogeneous clusters, after which fuzzy-SVM models were trained within each cluster. Experimental results on stock market datasets demonstrated superior forecasting accuracy compared with standalone machine learning models.

Barsoum and Stankiewicz (2015) proposed the Markov-Switching U-MIDAS (MS-U-MIDAS) model, which combines unrestricted MIDAS lag structures with regime-switching dynamics. The model was applied to U.S. GDP growth forecasting using a large mixed-frequency dataset. Empirical results showed that the MS-U-MIDAS specification produced similar or superior nowcasting and forecasting performance relative to conventional MIDAS approaches.

Ghysels and Qian (2019) investigated estimation techniques for the MIDAS regression model. Instead of relying on the commonly used but computationally complex Nonlinear Least Squares (NLS), they proposed using Ordinary Least Squares (OLS) combined with Polynomial Parameter Profiling. This approach incorporated a Beta polynomial weighting scheme to handle the lag parameters, simplifying linear estimation while preserving the efficiency and desirable properties of U-MIDAS. Their method enhanced the flexibility of MIDAS estimation and broadened its applicability to datasets with differing frequencies.

Fang et al. (2020) developed the GARCH-MIDAS model, which integrates short-term and long-term volatility components. The study applied variable selection using an Adaptive-Lasso penalty to identify the most influential macroeconomic and financial variables for long-term stock market volatility. Three key predictors were selected: housing starts, default spread, and realized volatility. Out-of-sample evaluation indicated that incorporating variable selection significantly improved forecasting accuracy, demonstrating the potential of the GARCH-MIDAS framework in predicting long-term stock market

volatility.

Tan et al. (2021) examined volatility modeling of Bitcoin (BTC) returns using a Time-Varying Transition Probability Markov-Switching GARCH (TV-MSGARCH) model with exogenous variables such as daily trading volume and Google search intensity. Comparisons with standard models, including GARCH, GJR-GARCH, Threshold GARCH, and MSGARCH, showed that TV-MSGARCH with skewed and fat-tailed distributions provided the best in-sample fit. Moreover, the specification incorporating daily trading volume and Student-t error distribution achieved the highest out-of-sample forecasting accuracy. The study also identified time-varying transition probabilities and evaluated shortfall forecasts under both long and short trading positions, where TV-MSGARCH consistently outperformed benchmark models.

Cavicchioli (2021) proposed a closed-form matrix solution for computing higher-order moments and kurtosis in univariate Markov-Switching GARCH (MS-GARCH) models, along with asymptotic theory for the estimators of these quantities. These results facilitate normality testing and were validated through Monte Carlo simulations. The analytical framework was further applied to financial market data to identify periods of heightened volatility, such as during financial crises and the COVID-19 pandemic.

Kole and van Dijk (2023) employed the Markov-Switching Vector Autoregressive (MS-VAR) model to analyze the effects of economic and financial shocks under different regimes, such as recessions and bear markets. The MS-VAR framework effectively captured both gradual adjustments and discrete regime shifts. The authors developed a comprehensive methodological approach, including regime conditional first- and second-moment estimation and the construction of impulse response functions for both moments. Application to stock and bond market data revealed that forecasts of mean, volatility, and autocorrelations varied significantly across regimes. Moreover, shocks were found to exhibit strong nonlinearity, with more severe and persistent effects during downturn regimes.

Chan, Poon, and Zhu (2024) developed a Time-Varying Parameter MIDAS model for nowcasting U.S. real GDP. The framework allows regression coefficients to evolve over time rather than remain fixed. The study reported that the time-varying MIDAS model outperformed static MIDAS and benchmark regime-switching MIDAS models during periods

of structural instability.

Wang et al. (2024) investigated insurance demand forecasting in China using the Mixed Data Sampling (MIDAS) framework. The study utilized monthly data including the Consumer Confidence Index (CCI), Economic Policy Uncertainty (EPU) index, and Producer Price Index (PPI), along with quarterly data on Total Insurance Depth (TID). The results indicated that the MIDAS model provides highly accurate short-term forecasts, particularly when appropriate weighting functions and forecasting windows are applied. Moreover, the findings revealed that insurance demand is most sensitive to consumer confidence, especially during the COVID-19 period. This study underscores the importance of MIDAS as a tool for improving forecasting accuracy and supporting policy-making in the Chinese insurance sector.

Ma et al. (2024) proposed a hybrid model that combines Reserve Restricted Mixed Data Sampling (RR-MIDAS) with deep learning architectures, specifically CNN-LSTM, for forecasting stock price volatility using mixed-frequency data. The research emphasized handling macroeconomic variables and technical indicators with different frequencies and employed the Optuna framework to optimize parameter selection. Experimental results demonstrated that the proposed hybrid model significantly reduced forecasting errors, including RMSE, MAE, MSLE, MAPE, SMAPE, and QLIKE, compared to traditional models. On average, accuracy improvements ranged from 12% to 30%. These findings highlight the effectiveness and robustness of integrating RR-MIDAS with CNN-LSTM in forecasting stock price volatility under mixed-frequency conditions.

Although prior studies have separately demonstrated the forecasting capability of MIDAS models for mixed-frequency data and SOM models for nonlinear clustering, empirical evidence directly integrating SOM-based regime classification with MIDAS forecasting remains limited. Existing regime MIDAS studies mainly rely on Markov-switching mechanisms, while SOM offers a data-driven and nonparametric alternative for identifying hidden regimes. Therefore, combining SOM with MIDAS may provide a flexible framework capable of capturing nonlinear market states and improving forecasting performance under structural changes.