INTERNATIONAL JOUNAL OF: SCIENCE RESEARCH AND TECHNOLOGY

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
INTERNATIONAL JOURNAL, AUGUST, 2023 EDITIONS
VOL. 14 NO. 9 ISSN: 2623-7861

MODELING MIXED DATA SAMPLING FREQUENCY: A SYSTEMATIC LITERATURE REVIEW AND FURTHER RESEARCH

UMARU HASSAN[1,2] AND MOHD TAHIR ISMAIL[1]

[1]School of Mathematical Science, Universiti Sains Malaysia, 11800 Penang, Malaysia. [2]Department of Statistics, Federal Polytechnic Damaturu, Yobe, Nigeria.

ABSTRACT

Mixed data sampling analysis (MIDAS) is a popular technique for modeling and forecasting economic time series data that have mixed frequencies. This systematic literature review provides an overview of the key concepts, methods, and applications of MIDAS analysis in economics and finance literature. The review covers the period from 2004 to 2023 and

INTRODUCTION

Mixed sampling frequency (MIDAS) is a technique used to estimate regression models with variables sampled at various frequencies. The method involves converting the variables into a common frequency using different weighting schemes. In other words, time series data must be collected using the same frequency when using the conventional regression approach; otherwise, higher frequency data are averaged or aggregated into a lower frequency. MIDAS regression was created to solve these problems[1]. Over the years, MIDAS has gained popularity in econometrics, finance, and other fields that involve analyzing data sampled at different frequencies. This systematic literature review aims to provide an overview of the current state of knowledge on MIDAS, including its applications, strengths, and limitations.

The systematic analysis of MIDAS journals using Preferred Reporting Items for Systematic Reviews and

includes 74 articles published in peer-reviewed journals extracted from Google scholar, Scopus, and Web of Science databases. We employed the method of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), and bibliometric analysis was conducted. The method has been used to model a wide range of economic variables, including inflation, GDP, stock prices, exchange rates, and tourism. The review also discusses the different types of MIDAS models, including the regression-based, time series-based, and state-space models. It has several advantages over traditional methods, including its ability to handle different frequencies and capture short-term dynamics in the data. However, it also has limitations, such as the need for careful selection of the appropriate MIDAS model and the potential for overfitting when using complex models. The review provides guidance on best practices and future directions for MIDAS analysis research. In order to compare MIDAS' performance with those of other approaches to processing mixed-frequency data, more investigation is required on the robustness of MIDAS in various contexts.

Keywords: Mixed data sampling, modeling, forecasting, Systematic literature review

Meta-Analyses (PRISMA) by Moher et al., 2009 [2] protocol is presented in this work. Systematic reviews are particularly valuable because of their potential to lower biases, boost reliability, and enhance communication [3]. Drawing on[4] note that "for a field to progress, it must be conscious of its historical patterns to obtain insights into possible future developments and implications that contribute to the accumulation of knowledge." We think that individuals who 'utilise' the results of systematic reviews will greatly benefit from the methods in which they have been operationalized. The ability to capture the dynamics of economic activity at a more detailed level and the potential to increase the precision of economic projections are some of the possible advantages of MIDAS.

In Narrative literature reviews (NLR), authors summarize, analyze, evaluate, and clarify the issues that have been carried out by other researchers[5]. A systematic literature review (SLR), in contrast to a narrative literature review, is a technique with clearly specified search criteria, study questions, data extractions, and data displays [6]. An SLR's key benefits include being objective, thorough, extensive, and reproducible, to name a few. Additionally, it included a thorough evaluation of top-notch works on the subject and a discussion of them in its stated selection criteria.

In our search, no study has been done that provides an exhaustive analysis of review articles using mixed data sample frequency regression. This work seeks to get beyond that restriction with a stronger focus on MIDAS regression publications. Moreover, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, or PRISMA, approach was selected for this investigation[7].

Author name, title, publication year, keyword, DOI, citation count, abstract, affiliation, journal name, funding information, and references were all exported into a Microsoft Excel file. Records' titles and abstracts were evaluated independently, and blatantly irrelevant publications were removed. After that, we independently and in-depth go through the remaining papers' contents to conduct an eligibility assessment. Any papers that at least partially showed how their methods for locating and selecting the appropriate models were explicit and repeatable were included. This is meant to answer the question, "How did MIDAS regression help in solving mixed frequency problems”?

Significance of the study

The significance of the study lies in its comprehensive review of the Mixed Data Sampling Analysis (MIDAS) technique, highlighting its versatile application in modeling various economic variables with mixed frequencies. By synthesizing 74 articles published over almost two decades and employing a robust methodology, the study provides valuable insights into MIDAS' advantages, such as handling diverse data frequencies, and its limitations, like model selection challenges. Furthermore, the study's call for continued investigation into MIDAS' robustness in different contexts underscores its contribution to advancing research and refining the technique's application.

Related works

Concept of Mixed Data Sampling Frequency (MIDAS) regression.

Mixed data sampling frequency (MIDAS) regression models are a type of regression analysis that allows for the combination of data with different sampling frequencies. In traditional regression analysis, all of the data used in the analysis is typically collected at the same frequency. However, in many real-world applications, data may be collected at different frequencies, such as monthly, quarterly, or yearly. MIDAS models are designed to handle this type of mixed-frequency data. The framework is introduced into the literature by Ghysels et al .[23,38,11]. They do so by allowing for different functional forms for the relationship between the dependent variable and the independent variables at different frequencies[8]. This allows the model to account for the different dynamics that may be present in the data at different frequencies. Some areas of applications include macroeconomics, finance, and marketing[9].

[10] uses mixed data sampling (MIDAS) to link variables sampled at various frequencies without missing high-frequency information. To achieve a parsimonious specification, MIDAS regressions are often based on distributed lag polynomials such as the exponential Almon lag [11]. In general, MIDAS regression seeks to minimise the number of parameters that must be estimated while yet keeping the individual temporal information of the high-frequency data[12][13].

MIDAS regression provides more accurate forecasts when the variables are not aggregated to the same frequencies, as reported in (Franses, 2016). MIDAS regression is generally made to fill the gap between keeping the high-frequency data's individual timing information and minimising the number of parameters that must be estimated[13](Utari et al, 2018). In MIDAS models, the weight function is written as a nonlinear parametric function with only a few parameters due to the parametric restrictions [16] [17][18]. Another virtue of MIDAS regression models, according to [19] and [20], is its capacity to do nowcasting of variables.

MIDAS models have been used in a number of studies to forecast quarterly time series utilising data collected on a monthly, weekly, or daily basis. Researchers like [21] -[22], applied the Almon constraint to test the impact of COVID-19 pandemic on some countries GDP like the U.S., China, Indonesia, and so on. The model accurately captures the drop in the Gross Domestic Product during these periods. MIDAS has provided researchers with new opportunities/possibilities to use any accessible data from various frequencies most efficiently in forecasting/nowcasting without having to deal with the issue of various lags of several macroeconomic time series variables [23].

Other researchers who conducted a study on financial and macroeconomic problems such as stock market returns are [24], GDP[25]-[26]], inflation[27]. Determining how to investigate the diverse nonlinear relationships among variables sampled at different frequencies can be challenging. However, scholars have made an effort to offer some established approaches in an effort to address this difficulty. The Mixed Data Sampling (MIDAS) regression put forward by Ghysels et al.[12] among them directly handles the unprocessed mixed data sampling frequency by adding the distributed lag polynomial weights.

Mixed data sampling frequency and other forecasting models

Combining MIDAS with other models can be very effective when making predictions. The precise data and research topics at hand will determine the specific technique that is used. But, by including MIDAS into the modelling strategy, we may be able to increase the precision of forecasts and learn more about the connections between variables. Guérin and Marcellino [28], introduced a "Markov-switching mixed data sampling (MS-MIDAS) regression" that employs parameter regime shifts to forecast economic activity in the United States. The model's usefulness is supported by simulation and empirical results. Using the penalized least-squares estimator and MIDAS to enforce smoothness via lag distribution, Kapetanios [29] employed daily indicators to predict monthly inflation rates using the penalized least-squares estimate. According to the findings, the commodity price index (CPI) can be used to predict inflation rates [30]. To anticipate actual US GDP growth using crude oil prices, a time-varying parameter called TVP-MIDAS was adopted, where the predictability of GDP growth varies depending on forecasting horizons. The authors claim that TVP-MIDAS beats the other models used in the study. (Examples include OLS regressions with quarterly oil prices and MRS-MIDAS regressions).

Bilgin et al.[31] examined the performance of export, unemployment rate, and stock-exchange index in forecasting Thailand's quarterly GDP growth using MIDAS and U-MIDAS regressions. According to the authors, the unemployment rate is the most accurate predictor of quarterly GDP growth. Their empirical results demonstrated that, independent of the indicators, U-MIDAS outperformed MIDAS. Similar results are documented in Foroni & Schumacher[32]. Le [33] recommends estimating Value at Risk (VaR) and Expected Shortfall (ES) using the Mixed Data Sampling (MIDAS) approach (ES). Out-of-sample VaR and ES projections are compared to known models for a variety of financial assets and backtests using the new approaches. MIDAS-based models outperform traditional GARCH-based forecasts and alternative conditional quantile specifications, notably over multi-day forecast horizons.

GARCH–MIDAS model was used to explores the impact of hot money (Hot money refers to currency that travels swiftly and frequently between financial markets, allowing investors to lock in the best short-term interest rates) on the return and volatility of the Chinese stock market. The empirical findings reveal that there is no linear or nonlinear correlation between the growth rate of hot money and the return on the Chinese stock market, meaning that hot money does not drive the Chinese stock market and vice versa [34].

Barsoum & Stankiewicz [35] looked into the utility of MIDAS models with unconstrained lag polynomials and a Markov-switching component for modelling huge datasets in this research. (For example, quarterly–monthly data.) They use Monte Carlo simulations to compare the MS-U-MIDAS(-AR), MS-MIDAS(-AR), and MS-ADL-MIDAS models for various DGPs. The MS-U-MIDAS class of models has performance that is comparable to or better than that of its constrained counterparts. For many macroeconomic applications, the unconstrained Markov-switching MIDAS model is an excellent alternative to the restricted MS-MIDAS model, especially when the frequency difference is negligible. A new model Group Penalized Unrestricted MIDAS (GP-U-MIDAS) was introduced to extract important variables in high dimensional MIDAS regression, the results from monte Carlo simulation shows that the GP-U-MIDAS model has superior prediction accuracy and variable selection over the P-U-MIDAS, U-MIDAS and FC-U-MIDAS models the same superiority was illustrated using empirical data on US quarterly GDP growth forecasting[36].

In the proposed QRNN-MIDAS model, frequency alignment is conducted on each high frequency variable based on the maximum lag order as determined by information criteria. Then, parameter weighting function is imposed on the frequency alignment to obtain low frequency variable. This allows QRNN model to handle the raw MIDAS data directly and enables QRNN-MIDAS model to extract important information and help explore heterogeneous nonlinear patterns between variables from the raw mixed sampling frequency data. The output results from simulation and real time data show that the proposed QRNN-MIDAS model, compared to several competing models, has better goodness-of-fit and predictive power. More specifically, the QRNN-MIDAS model directly outputs conditional quantiles, enhancing the capacity of the Neural Network in revealing the heterogeneous effect of variables on the whole conditional distribution of a response variable and therefore offering more meaningful information for decision-making.[37]

Ghysels, at el. [38] recently expanded QR to the environment of mixed sampling frequency data. To effectively conduct quantile regression on mixed sample frequency data, a QR-MIDAS model is designed. The QR-MIDAS model expertly portrays heterogeneous effects between mixed sampling frequency variables, allowing for a more comprehensive perspective of the data. Beyond simple linear mean regressions, machine learning methods can also be extended to quantile regression. Typically, [39] introduce the quantile regression neural network (QRNN) approach to explore the heterogeneous nonlinear relationships between variables see [37].

A new MIDAS model is proposed that combines MIDAS with the seasonal autoregressive integrated moving average (MIDAS-SARIMA) process. The generalised dynamic factor model is used to anticipate monthly visitor arrivals in Hong Kong from mainland China using daily composite indices derived from a large number of search queries. The results of the forecasts demonstrate how considerably better this unique model performs than the benchmark model. Furthermore, when comparing forecasts with nowcasts, the latter consistently exceeds [40].

Unrestricted MIDAS regression approaches

Unrestricted Mixed Data Sampling (UMIDAS) is a statistical method that permits blending data with various frequencies without putting any constraints on the regression model's parameters. This demonstrates the adaptability and ability of UMIDAS models to capture intricate connections between variables of various frequencies. Alternatively, Almon's constraint may be excessively harsh on the underlying Data Generating Process (DGP). As a result, Foroni & Schumacher [32] presented the unrestricted MIDAS model, which has no restrictions on the weights of the lag polynomial (U-MIDAS). Foroni et al. [41] compared the standard MIDAS model to the unrestricted MIDAS (U-MIDAS) model. When the frequency mismatch is small, like in macroeconomic and financial applications, the U-MIDAS model can increase prediction accuracy with a reduced model specification without the concern of parameter proliferation.

When the frequency differences between variables in the model are large, the performance of U-MIDAS regression suffers owing to parameter proliferation, which is a disadvantage of this type of analysis. The authors argue that the use of the U-MIDAS model instead of the functional lag polynomial in MIDAS when some macroeconomic variables are used may be beneficial [35]. As the difference in frequency increases, the U-MIDAS becomes unappealing because of parameter proliferation associated with high frequency lag growth [42]. In summary, when the aggregated frequency is small, U-MIDAS regression outperforms the MIDAS model.

The UMIDAS approach enables a more precise and adaptable assessment of the association between variables of various frequencies. However, a major drawback of UMIDAS models is that they could require a lot of processing, particularly if the data sets are sizable. Furthermore, the interpretation of UMIDAS models may be more challenging than interpretation of more constrained models due to the potential complexity of the interactions between the variables.

Limitations on the existing knowledge

The mixed data sampling (MIDAS) frequency model is an econometric model that analyzes time series data with different sampling frequencies. While MIDAS models offer some advantages over traditional time series models, their use has several limitations. Here are some of the main limitations of MIDAS frequency models:

1. Limited data availability: MIDAS models require data at different frequencies, which can be difficult to obtain for some variables. This can limit the range of applications for which MIDAS models are suitable.
2. Complex modeling: MIDAS models are more complex than traditional time series models, which can make them more difficult to estimate and interpret. This complexity can also lead to overfitting of the data and reduce the reliability of the model.
3. Limited ability to capture long-term dynamics: Because MIDAS models are based on short-term data at higher frequencies, they may not capture the long-term dynamics of the variables being analyzed. This can make it difficult to make accurate forecasts for longer time horizons.

ISSN: 2623-7861

4. Sensitivity to parameter choice: The results of MIDAS models can be sensitive to the choice of parameters used in the model. This can make it difficult to choose the optimal model and can lead to uncertainty in the results.

This study offers a helpful overview and in-depth details regarding this topic for upcoming researchers. In particular, the citation analysis provides a good view of the literature by identifying the top publications, authors, and journals. Our review suggests that MIDAS is a useful method for analyzing MIDAS, particularly in forecasting and dynamic modeling. However, researchers should be cautious in selecting appropriate weighting schemes and be aware of the potential for parameter instability. Further research is needed to investigate the robustness of MIDAS in different settings and to compare its performance with other methods for managing mixed-frequency data.

Methodology

The technique employed is a multi-step, systematic literature review. Topic formulation, the first stage, entails defining the keywords that are pertinent to the topic. The second study design is the one where the database sources are chosen. Sampling and data gathering, including searching, saving, and combining processes, are the next steps. Following this data analysis, the appropriate tools for data analysis are identified, and reporting follows. The PRISMA protocol was used to conduct this systematic review [2]. We steadily searched for published articles containing “Mixed data sampling” in their titles, the following steps were considered:

Search strategy: For this systematic search, we developed a search strategy to identify relevant literature. This strategy was customized to three databases, i.e., Scopus, Web of Science (WoS), and Google Scholar, and the search terms used were "mixed data sampling" in two stages. The first search was for all fields, the second was in the title only, and all searches spanned from 2004 to 2023 and included journal articles and conferences. No review paper was identified and published in English only**.**

Selection criteria: The selection criteria were based on the PRISMA [2]. The search mainly focused on mapping existing literature on "mixed data sampling" in the fields of economics, engineering, social sciences, etc. The search was limited to records with the keywords in their titles and dates ranging from 2004 to 2023. The first article on MIDAS in literature was published in 2004, hence the choice for starting date. All articles that did not capture the keywords in the title were excluded from the search. The search focused mainly on Asian and European countries; other continents were excluded. A total of 470 search records were identified at the first stage, out of which 319 search articles were excluded at these stages.

Quality assessment: The only sources used as the basis for this investigation were renowned relevant publications and conference papers. For the purpose of screening, the identified records were exported to Excel sheets and then all records were combined. All duplicates were carefully examined to maintain the review's high standard. To assure the calibre and applicability of the academic literature used in the review process, by taking into account only the publications from a known reputable journal, the abstracts of the articles were thoroughly examined. Each research report underwent a thorough evaluation at a later time. The second set of exclusion criteria was a restriction to publications published in English exclusively. Two non-English items were omitted from the study. After evaluating each item according to the aforementioned inclusion and exclusion criteria, we are left with 74 records to work with. The inclusion and exclusion at each level are depicted in Figure 1 (PRISMA statement).

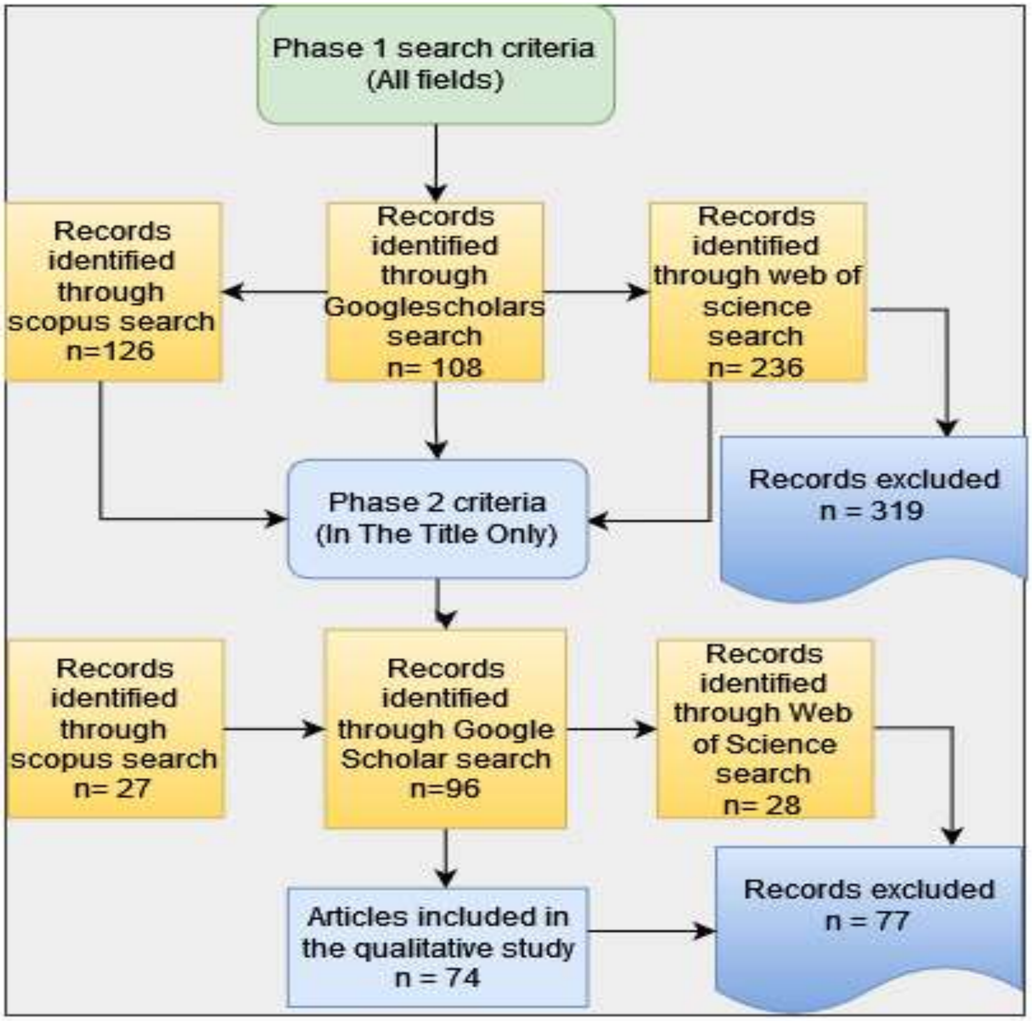
Data extraction: In the data extraction phase1 “all field”, 126 articles were selected from Scopus, 236 articles from WoS, and 108 from Google Scholar and the extraction phase2 “in the title only”, 27 articles were selected from Scopus, 28 articles from WoS and 96 from Google Scholar and the following characteristics were used:

(a) Articles mostly were original papers, and conference papers, in-press papers were excluded.
(b) The article has to be in English language and from the field of social science, economics, managements and statistics.
(c) Extracted records were published between 2004 to 2023.
(d) Extracted records were from Asian and European countries.

The steps involved in data analysis include a summary of publications, citation analysis, authorship analysis, publisher, and term analysis of the title and abstract. The bibliometrix package was utilized along with the R statistical program [43].

The articles that are retrieved from the database and put into an excel sheet were subjected to a thorough bibliometrix examination using this software. Bibliometrix mapping is another method for visualizing and analyzing trends.

Fig.1 PRISMA model chart for systematic literature process



Results and interpretation

A filtered database with 74 possible records served as the basis for the bibliometric study. Biblioshiny, an online interface for the bibliometrix R tool, was used to evaluate the data. Tables 1-4, and, Figures 1–11, with an explanation of the specifics, accompanied the presentation of the analyses' findings. Our search yielded 74 articles that met our inclusion criteria. We found that MIDAS has been applied in a number of fields involving finance, macroeconomics, and energy. The most common application of MIDAS was in forecasting (as it can be confirmed from the keyword cloud in Figure 5), where it has been shown to outperform traditional methods when dealing with mixed-frequency data. Other applications include estimating dynamic models and analyzing the effects of monetary policy. Most of the papers published were between 2015 and 2022. Fig. 2 presents an overview of the number of papers published from 2004 to 2023. A significant number of papers were published in recent years. The papers considered in this review were mostly from reputable journals, as can be seen in figure 3 below, where most of the papers came from Elsevier.

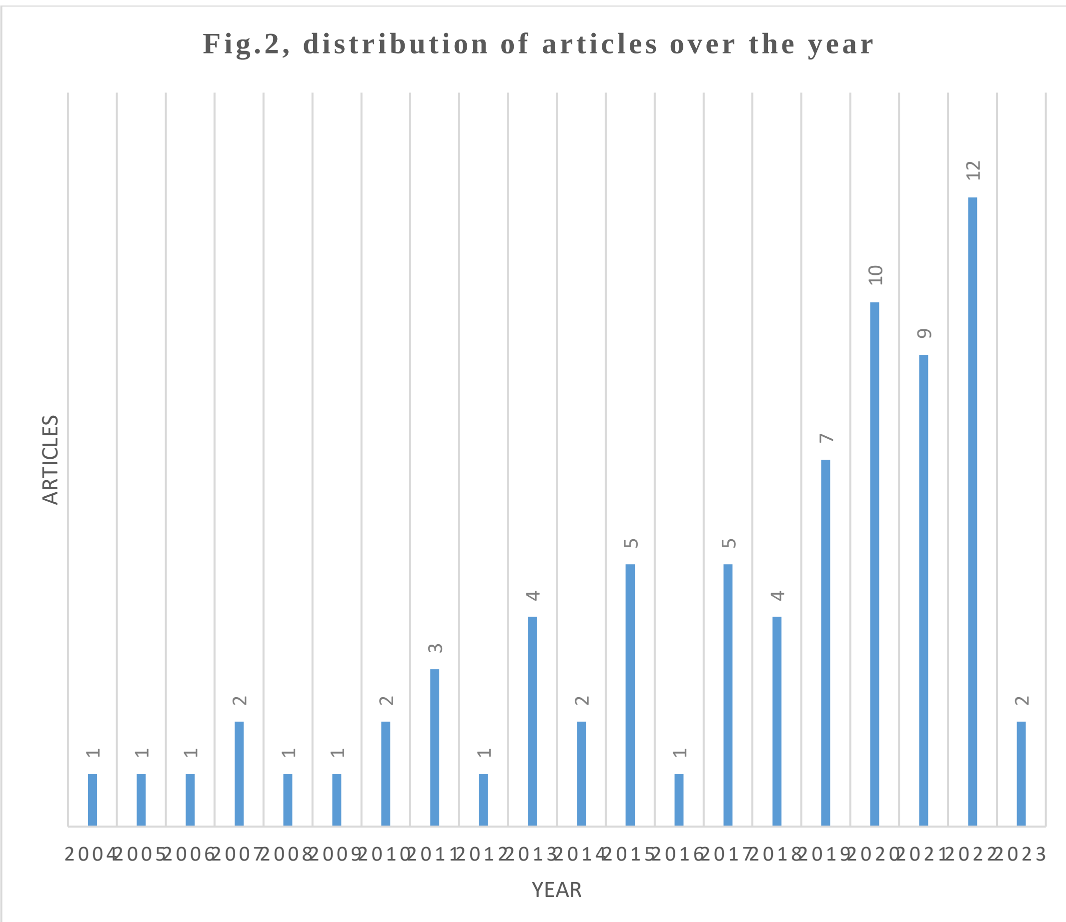


Fig. 2 presents the chart of articles visited across the three databases, 2022 had the highest number of articles considered and most of the records are between 2019 to 2022.

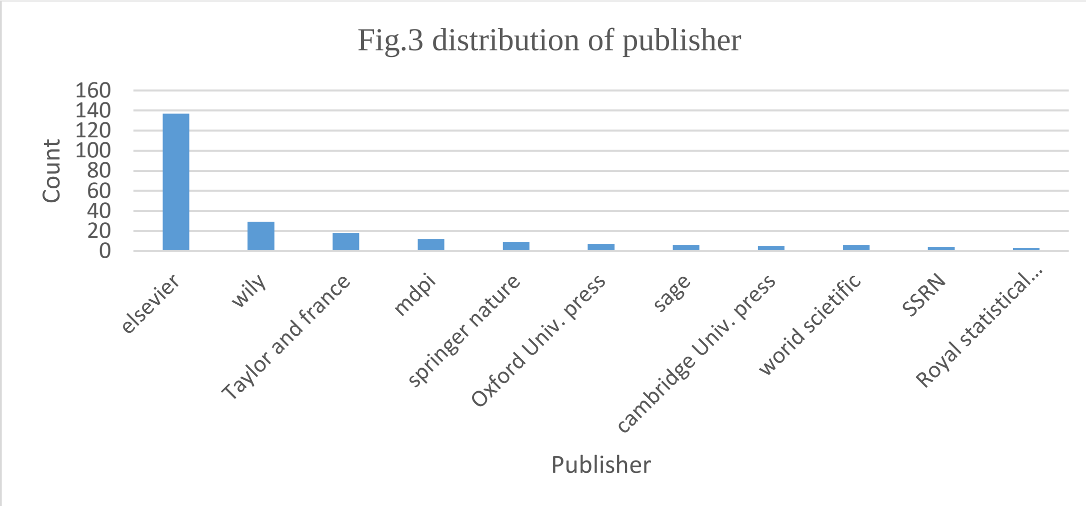


Fig. 3 presents the number of articles with respect to publishers. It can be seen that most of the papers consulted were published by Elsevier, followed by wily at a distance far below.

Researchers have made tremendous progress in the development of a mixed data sampling regression model in terms of forecasting and are nowcasting of various variables across different disciplines. Consequently, various model combinations were built. Figure 4 presents a pie chart of the most prominent of the models considered in the literature, with MIDAS taking the larger share of the sector of the circle, followed by GARCH-MIDAS.

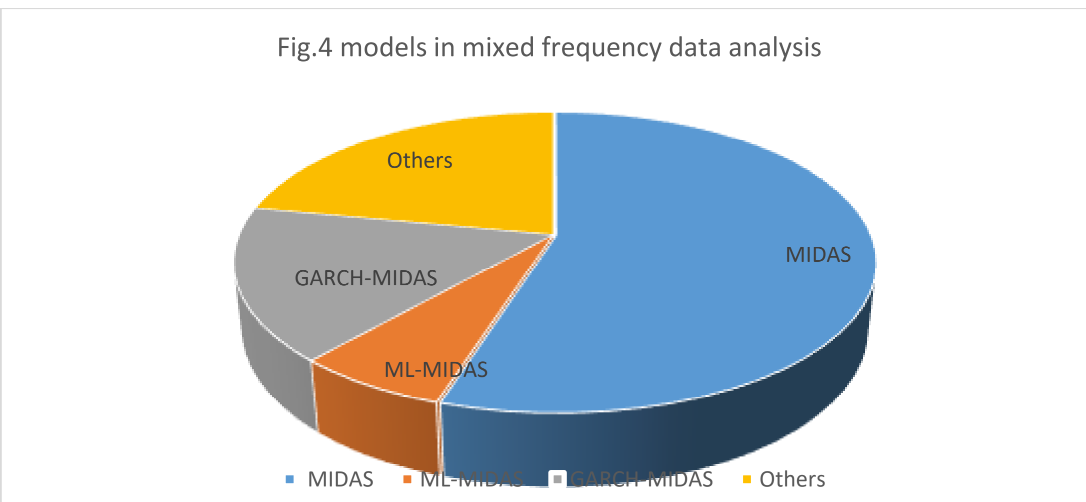


Fig. 4 the pie chart presents the sector distributions of various models in mixed frequency data analysis.

Table 1. Reporting the h, g, and m-indexes

Element	h_index	g_index	m_index	TC	NP	PY_start
JOURNAL OF FORECASTING	6	10	0.5455	285	10	2013
INTERNATIONAL JOURNAL OF FORECASTING	5	5	0.3846	258	5	2011
JOURNAL OF ECONOMETRICS	4	5	0.2222	687	5	2006
ENERGY	3	3	0.6000	104	3	2019
ECONOMIC MODELLING	2	3	0.3333	24	3	2018
ENERGIES	2	2	0.3333	8	2	2018
ENERGY ECONOMICS	2	2	0.2857	224	2	2017
EXPERT SYSTEMS WITH APPLICATIONS	2	2	0.4000	32	2	2019
KNOWLEDGE-BASED SYSTEMS	2	2	0.2857	77	2	2017
RESOURCES POLICY	2	3	1.0000	13	3	2022
TOURISM ECONOMICS	2	2	0.4000	34	2	2019
TOURISM MANAGEMENT	2	2	0.2222	211	2	2015
ANTHROPOLOGIST	1	1	0.1000	1	1	2014
APPLIED ECONOMICS LETTERS	1	1	0.0909	15	1	2013
APPLIED ENERGY	1	1	0.5000	2	1	2022
ASIA PACIFIC JOURNAL OF TOURISM RESEARCH	1	1	0.3333	4	1	2021

Table 1. Reporting the h, g, and m-indexes of journals provides important information about the quality and impact of academic journals. It can help readers identify the most influential and reputable journals in a specific field or research area, and the Journal of Forecasting had the highest index in this report.

Table.2 The first ten (10) most global cited documents

Paper	Total Citations	TC per Year	Normalized TC
GHYSELS E, 2005, J FINANC ECON	501	26.37	1
GHYSELS E, 2006, J ECONOM	434	24.11	1
BANGWAYO-SKEETE PF, 2015, TOUR MANAGE	202	22.44	2.98
COLACITO R, 2011, J ECONOM	152	11.69	1.8
WEI Y, 2017, ENERGY ECON	147	21	4.22
KUZIN V, 2011, INT J FORECAST	135	10.38	1.6
FORONI C, 2015, J R STAT SOC SER A STAT SOC	125	13.89	1.84
ASGHARIAN H, 2013, J FORECAST	121	11	3.38
MA Y-R, 2019, J FORECAST	92	18.4	2.97

ISSN: 2623-7861

MEI D, 2020, ENERGY ECON	77	19.25	4.67

Table 2 reports the most globally cited documents along with their total citations (TC) per year and normalized TC, which provide valuable information about the most influential and impactful publications in a particular field or research area. It can assist readers in identifying key themes and trends in literature.

Fig. 5 world cloud of 40 keywords of mixed data sampling model.

Fig. 5. A word cloud of 40 words in cycle form is a tool for visualization and displays a group of words in different sizes according to their frequency in our review topic.

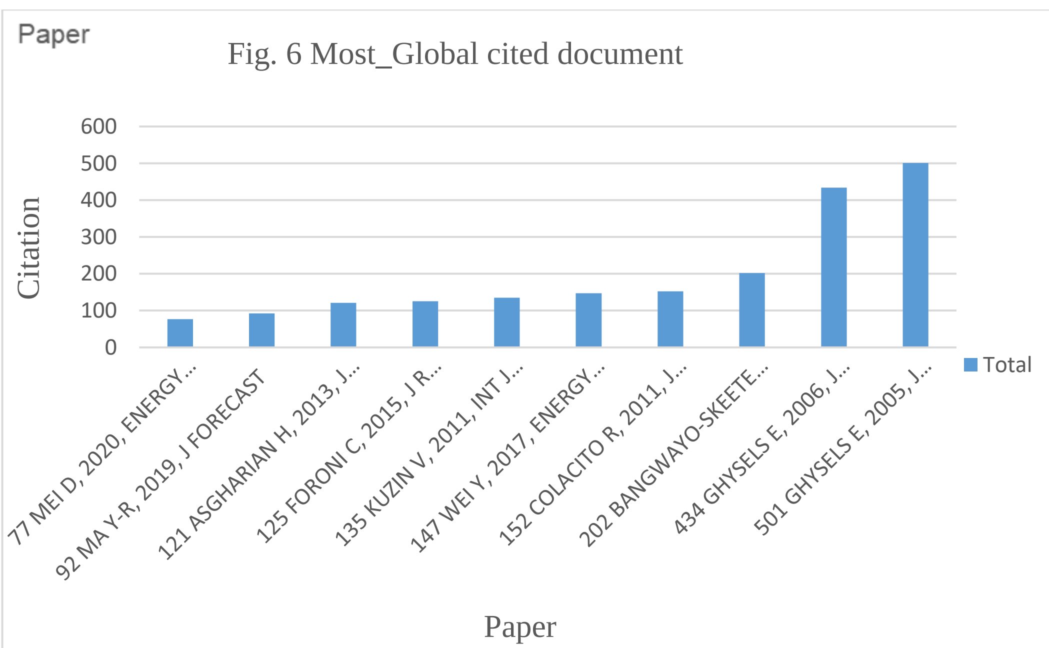


Fig. 6 Identifying the most globally cited documents can be a useful way to identify the most influential papers in a particular field, and to provide context and background for a systematic literature review. It can help to guide the review process and to ensure that the review is grounded in the most important and relevant literature.

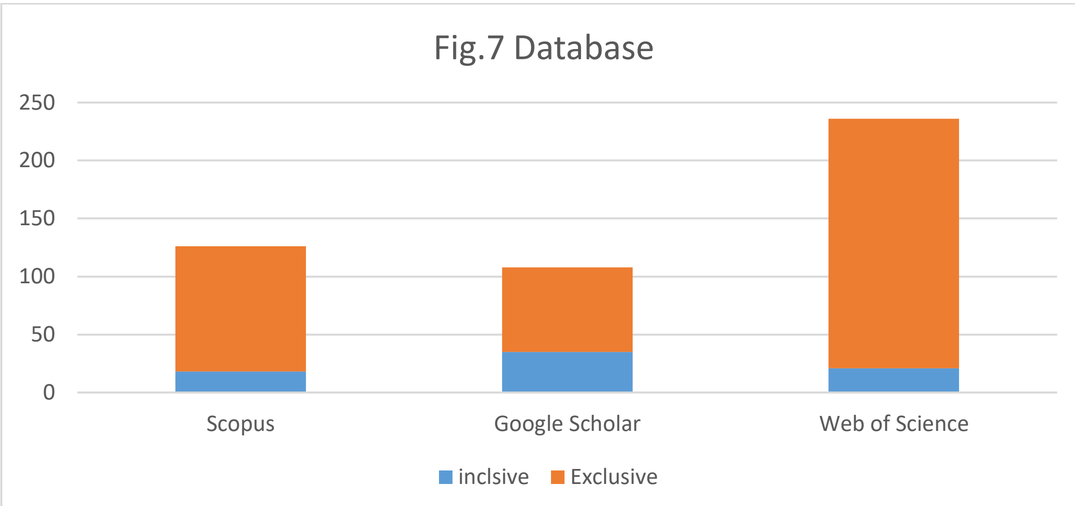


Fig. 7 presents the database chart used to research the records, indicating the number of articles included and excluded during the research process.

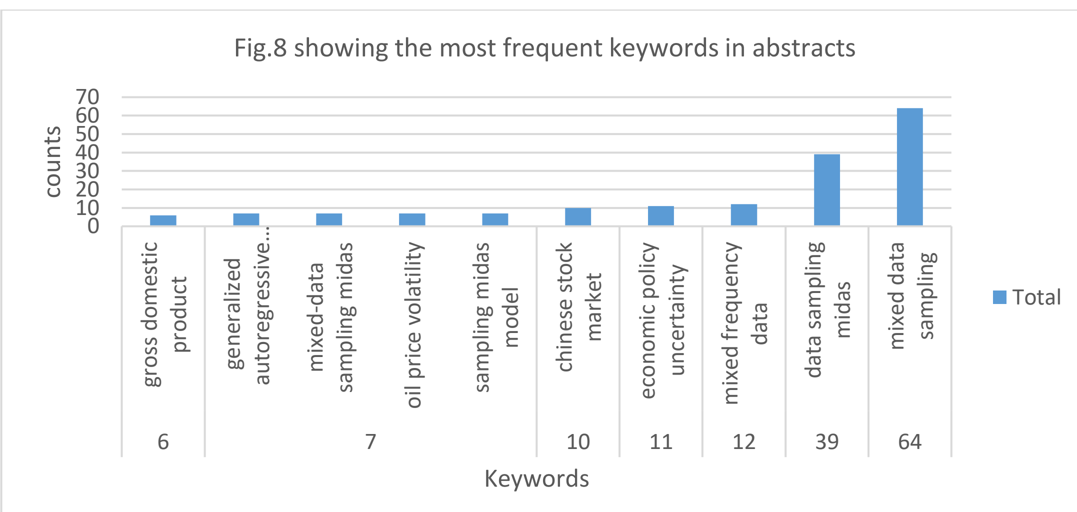


Fig. 8 Reports the most frequent words in abstracts, which is a useful way to summarize the main themes and findings of a systematic literature review. It can help to provide an overview of the literature.

Fig. 9, showing a three-field plot.

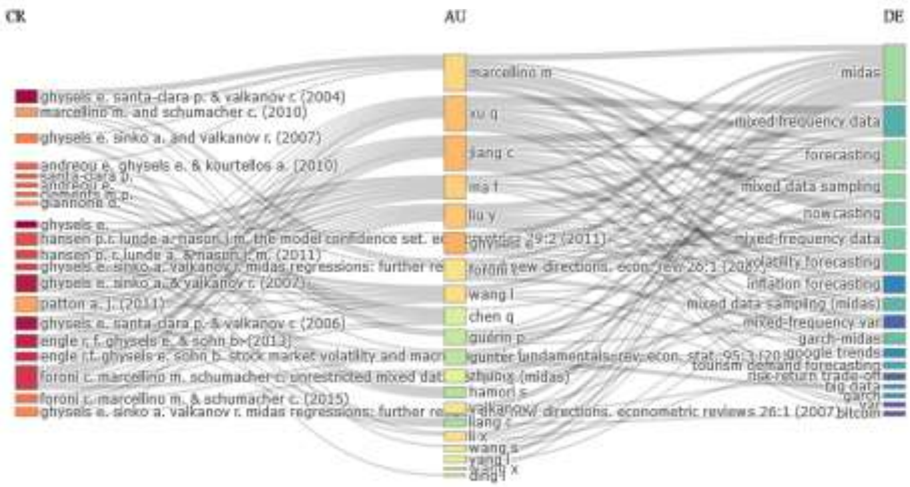
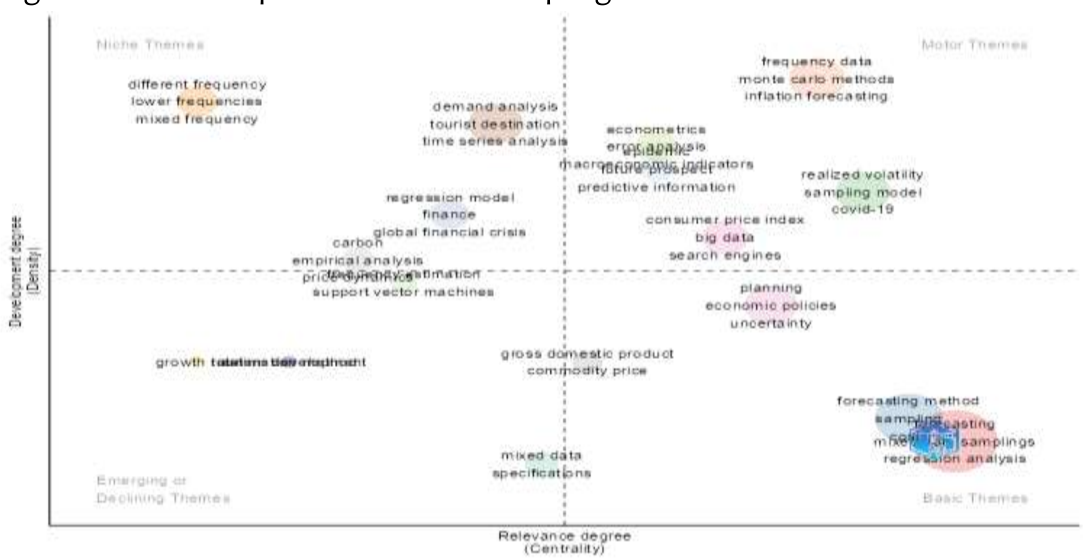


Fig. 9 showing a three-field plot, is a graphical representation that displays the frequency of occurrence of three different variables or categories in relation to each other. The left field is "reference", the middle field is "author", and the right field is "keyword" for the first 20 items.

Fig.10 Thematic maps of mixed data sampling



Thematic maps are used to display the spatial distribution of themes or keywords across a geographic region, while keyword plus is a technique used to identify and analyze the key terms or phrases that appear most frequently in the studies.

TIJSRAT
ISSN: 2623-7861

Fig.11 the co-occurrence network of abstract terms.

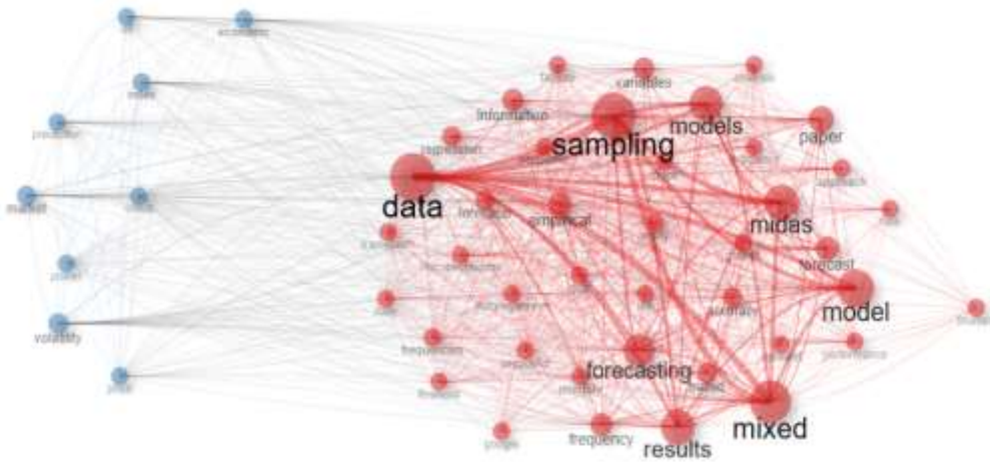


Fig.11 presents the co-occurrence network of abstract terms which can be a useful way to visualize the relationships between different concepts or ideas and identify patterns or trends in the literature. It can help to reveal important insights and inform further analysis and interpretation of the data. The network of co-occurrence terms in a systematic review abstract represents the relationships between terms, with the size of the nodes representing frequency and the colors representing importance or centrality.

Fig. 12 Countries collaboration world map.

A countries collaboration world map can be a useful tool in a systematic literature review to visually represent the collaboration patterns between countries in a particular research area. This type of map provide insights into the geographic distribution of research collaborations and the strength of those collaborations. By visualizing collaboration patterns in this way, researchers can gain insights into the strengths and weaknesses of the collaborative networks in their field and identify opportunities for future collaboration.

Table3 showing the distribution of articles by order of citations.

Title	No. of citation
There is a risk-return trade-off after all	501
Predicting volatility: Getting the most out of return data sampled at different frequencies	434
Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach	202
A component model for dynamic correlations	152
Which determinant is the most informative in forecasting crude oil market volatility: Fundamental, speculation, or uncertainty?	147
MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area	135
Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials	125
The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach	121
Oil financialization and volatility forecast: Evidence from multidimensional predictors	92
Geopolitical risk uncertainty and oil future volatility: Evidence from MIDAS models	77
Exogenous drivers of Bitcoin and Cryptocurrency volatility – A mixed data sampling approach to forecasting	73
Forecasting carbon prices in the Shenzhen market, China: The role of mixed-frequency factors	71

TIJSRAT
ISSN: 2623-7861

A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates	57
Testing for Granger causality with mixed frequency data	52
Volatility forecasting and microstructure noise	47
A multiple support vector machine approach to stock index forecasting with mixed frequency sampling	44
Markov-switching MIDAS models	41
Forecasting performance of global economic policy uncertainty for volatility of Chinese stock market	40
A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data	36
Does Google search index really help predicting stock market volatility? Evidence from a modified mixed data sampling model on volatility	33
Examining the predictive information of CBOE OVX on China's oil futures volatility: Evidence from MS-MIDAS models	28
Cryptocurrency volatility forecasting: A Markov regime-switching MIDAS approach	28

The few papers considered here are based on the number of citations of each paper.

Table 4. A literature table summarizing some key papers on progress in MIDAS regression:

Year	Authors	Title	Key Contributions
2004	Ghysels, Santa-Clara, and Valkanov	"The MIDAS Touch: Mixed Data Sampling Regression Models"	Introduced the MIDAS regression model, which allows for the combination of data with different sampling frequencies.
2008	Ghysels, Sinko, and Valkanov	"MIDAS regressions: Further results and new directions"	Extended the original MIDAS regression model to include time-varying coefficients and to allow for the inclusion of lagged dependent variables.

TIJSRAT
ISSN: 2623-7861

2010	Hillebrand and Medeiros	"The Benefits of Multivariate MIDAS Regression for Business Cycle Analysis"	Showed that a multivariate MIDAS regression model can improve the accuracy of business cycle forecasts compared to a univariate model.
2011	Koop and Potter	"Dynamic Factor Models with Mixed Frequency Data"	Developed a dynamic factor model that allows for the incorporation of mixed-frequency data, which can be used for forecasting macroeconomic variables.
2013	Granziera and Huber	"MIDAS covariates for mixed frequency regression models"	Proposed a method for selecting the optimal lag structure of the covariates in a MIDAS regression model.
2016	Dossou and Razafindrabe	"A mixed data sampling Bayesian quantile regression model"	Extended the MIDAS framework to include a Bayesian quantile regression model, which can be used for estimating conditional quantiles of the dependent variable.
2019	Martens, Doornik, and Hendry	"The pitfalls of MIDAS regression"	Discussed some of the potential pitfalls of using MIDAS regression models, including the possibility of misspecification, data leakage, and overfitting.

This is not an exhaustive list, but these papers provide a good overview of some of the key developments in the field of MIDAS regression over the past two decades.

Recommendation

We recommend this study to researchers and practitioners in economics and finance as an insightful overview of the Mixed Data Sampling Analysis (MIDAS) technique's applications, strengths, and limitations. Its systematic approach, encompassing a wide range of economic variables and employing the PRISMA method, ensures a reliable synthesis of relevant literature. This review provides a solid foundation for informed decision-making when applying MIDAS and points to important avenues for further research in mixed-frequency data analysis.

Conclusion

This systematic literature review provides a mapping of the current state of knowledge on mixed data sampling frequency (MIDAS). We found that MIDAS has been widely applied since 2004 through 2023 in several fields, particularly in forecasting and dynamic modeling. The method has several strengths, including its ability to handle mixed-frequency data, its flexibility in choosing weighting schemes, and its ability to estimate models that are not feasible with traditional methods. However, researchers should be cautious in selecting appropriate weighting schemes and be aware of the potential for parameter instability. Further research is needed to investigate the robustness of MIDAS in different settings and to compare its performance with other methods for handling mixed-frequency data.

Reference

[1] H. Groenvik and Y. Rho, "A self-normalizing approach to the specification test of mixed-frequency models," *Commun Stat Theory Methods*, vol. 47, no. 8, pp. 1913–1922, Apr. 2018, doi: 10.1080/03610926.2017.1332222.

[2] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement David Moher and colleagues introduce PRISMA, an update of the QUOROM guidelines for reporting systematic reviews and meta-analyses", doi: 10.1136/bmj.b2535.

[3] S. Pahlevan-Sharif, P. Mura, and S. N. R. Wijesinghe, "A systematic review of systematic reviews in tourism," *Journal of Hospitality and Tourism Management*, vol. 39, pp. 158–165, Jun. 2019, doi: 10.1016/j.jhtm.2019.04.001.

[4] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *International Journal of Information Management*, vol. 49. Elsevier Ltd, pp. 424–438, Dec. 01, 2019. doi: 10.1016/j.ijinfomgt.2019.07.008.

[5] A. Eykelbosh, "Conducting a Literature Search & Evidence Review: The NCCEH Approach." [Online]. Available: www.ncceh.ca

[6] J. Paul and A. R. Criado, "The art of writing literature review: What do we know and what do we need to know?," *International Business Review*, vol. 29, no. 4, Aug. 2020, doi: 10.1016/j.ibusrev.2020.101717.

[7] S. Pahlevan Sharif, P. Mura, and S. N. R. Wijesinghe, "Systematic Reviews in Asia: Introducing the 'PRISMA' Protocol to Tourism and Hospitality Scholars," 2019, pp. 13–33. doi: 10.1007/978-981-13-2463-5_2.

[8] B. ŞEN DOĞAN Murat MİDİLİÇ, "Forecasting Turkish Real GDP Growth in a Data Rich Environment," 2016.
[9] Y. Liu and Y. Rho, "On the choice of instruments in mixed frequency specification tests," *Commun Stat Theory Methods*, vol. 48, no. 24, pp. 6098–6118, Dec. 2019, doi: 10.1080/03610926.2018.1528367.
[10] E. Andreou, E. Ghysels, and A. Kourtellos, "Regression models with mixed sampling frequencies," *J Econom*, vol. 158, no. 2, pp. 246–261, Oct. 2010, doi: 10.1016/j.jeconom.2010.01.004.
[11] E. Ghysels, A. Sinko, and R. Valkanov, "MIDAS regressions: Further results and new directions," *Econom Rev*, vol. 26, no. 1, pp. 53–90, Jan. 2007, doi: 10.1080/07474930600972467.
[12] A. Ghysels, E. Santa-Clara, P. Valkanov, E. Ghysels, P. Santa-Clara, and R. Valkanov, "UCLA Recent Work Title The MIDAS Touch: Mixed Data Sampling Regression Models Permalink https://escholarship.org/uc/item/9mf223rs Publication Date The MIDAS Touch: Mixed Data Sampling Regression Models *," 2004, [Online]. Available: https://escholarship.org/uc/item/9mf223rs
[13] D. Utari, H. I.-A. C. Proceedings, and undefined 2018, "Comparison of methods for mixed data sampling (MIDAS) regression models to forecast Indonesian GDP using agricultural exports," *aip.scitation.org*, vol. 2021, p. 60005, Oct. 2021, doi: 10.1063/1.5062780.
[14] P. H. Franses, "Yet another look at MIDAS regression," pp. 1–19, 2016.
[15] D. T. Utari and H. Ilma, "Comparison of methods for mixed data sampling (MIDAS) regression models to forecast Indonesian GDP using agricultural exports," *AIP Conf Proc*, vol. 2021, no. 2018, 2018, doi: 10.1063/1.5062780.
[16] Q. Xu, L. Wang, C. Jiang, and X. Zhang, "A novel UMIDAS-SVQR model with mixed frequency investor sentiment for predicting stock market volatility," vol. 132, pp. 12–27, 2019, doi: 10.1016/j.eswa.2019.04.066.
[17] Y. Liu, "Digital Commons @ Michigan Tech STATISTICAL METHODS FOR MIXED FREQUENCY DATA," 2019.
[18] M. Damane, "Forecasting the Government of Lesotho's budget: an AR-MIDAS approach," *African J. of Economic and Sustainable Development*, vol. 7, no. 3, p. 256, 2020, doi: 10.1504/ajesd.2020.105688.
[19] S. Asimakopoulos, J. Paredes, and T. Warmedinger, "Forecasting fiscal time series using mixed frequency data," *ECB Working Paper*, no. 1550, 2013.
[20] C. Baumeister, P. Guérin, and L. Kilian, "Do high-frequency financial data help forecast oil prices ? The MIDAS touch at work," *Int J Forecast*, vol. 31, no. 2, pp. 238–252, 2015, doi: 10.1016/j.ijforecast.2014.06.005.
[21] S. Gunay, G. Can, and M. Ocak, "Forecast of China's economic growth during the COVID-19 pandemic: a MIDAS regression analysis," *Journal of Chinese Economic and Foreign Trade Studies*, vol. 14, no. 1, pp. 3–17, 2021, doi: 10.1108/JCEFTS-08-2020-0053.
[22] T. B. Götz and K. Hauzenberger, "Large mixed-frequency VARs with a parsimonious time-varying parameter structure," *Econom J*, vol. 24, no. 3, pp. 442–461, 2021, doi: 10.1093/ectj/utab001.
[23] E. Ghysels, "The MIDAS Touch : Mixed Data Sampling Regression Models ∗ Pedro Santa-Clara Rossen Valkanov," 2004.
[24] S. Bhaghoe, G. Ooft, and P. H. Franses, "Estimates of Quarterly GDP Growth using MIDAS regressions," no. August, pp. 1–26, 2019.
[25] Z. Pan, Q. Wang, Y. Wang, and L. Yang, "Forecasting U.S. real GDP using oil prices: A time-varying parameter MIDAS model," *Energy Econ*, vol. 72, pp. 177–187, 2018, doi: 10.1016/j.eneco.2018.04.008.

[26] H. Hassani, A. Rua, E. S. Silva, and D. Thomakos, "Monthly forecasting of GDP with mixed-frequency multivariate singular spectrum analysis," *Int J Forecast*, vol. 35, no. 4, pp. 1263–1272, 2019, doi: 10.1016/j.ijforecast.2019.03.021.
[27] D. Science *et al.*, "Predictive Performance of Mixed-Frequency Nowcasting and Forecasting Models Roberto S . Mariano & Suleyman Ozmucur Predictive Performance of Mixed-Frequency Nowcasting and Forecasting Models Roberto S . Mariano & Suleyman Ozmucur," pp. 1–43, 2020.
[28] P. Guérin and M. Marcellino, "Markov-switching MIDAS models," *Journal of Business and Economic Statistics*, vol. 31, no. 1, pp. 45–56, 2013, doi: 10.1080/07350015.2012.727721.
[29] G. Kapetanios, "Analysis of the most recent modelling techniques for big data with particular attention to Bayesian ones," 2018.
[30] L. Tang, H. Pan, and Y. Yao, "K-nearest neighbor regression with principal component analysis for financial time series prediction," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Mar. 2018, pp. 127–131. doi: 10.1145/3194452.3194467.
[31] D. Bilgin, D. Jankovic, and A. Lam, "MIDAS regression using inflation and unemployment to predict GDP," 2018.
[32] C. Foroni, M. Marcellino, and C. Schumacher, "U-MIDAS : MIDAS regressions with unrestricted lag polynomials," vol. 178, no. 1, pp. 57–82, 2011, doi: 10.1111/rssa.12043/abstract.
[33] T. H. Le, "Forecasting value at risk and expected shortfall with mixed," *Int J Forecast*, vol. 36, no. 4, pp. 1362–1379, 2022, doi: 10.1016/j.ijforecast.2020.01.008.
[34] Y. Wei, Q. Yu, J. Liu, and Y. Cao, "Hot money and China's stock market volatility: Further evidence using the GARCH–MIDAS model," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 923–930, 2018, doi: 10.1016/j.physa.2017.11.022.
[35] F. Barsoum and S. Stankiewicz, "Forecasting GDP growth using mixed-frequency models with switching regimes," *Int J Forecast*, vol. 31, no. 1, pp. 33–50, 2015, doi: 10.1016/j.ijforecast.2014.04.002.
[36] Q. Xu, X. Zhuo, C. Jiang, X. Liu, and Y. Liu, "Group penalized unrestricted mixed data sampling model with application to forecasting US GDP growth," *Econ Model*, vol. 75, no. June, pp. 221–236, 2018, doi: 10.1016/j.econmod.2018.06.021.
[37] Q. Xu, S. Liu, C. Jiang, and X. Zhuo, "QRNN-MIDAS: A novel quantile regression neural network for mixed sampling frequency data," *Neurocomputing*, 2021, doi: 10.1016/j.neucom.2021.06.006.
[38] E. Ghysels, P. Santa-Clara, and R. Valkanov, "There is a risk-return trade-off after all," *J financ econ*, vol. 76, no. 3, pp. 509–548, Jun. 2005, doi: 10.1016/j.jfineco.2004.03.008.
[39] A. J. Cannon, "Neural networks for probabilistic environmental prediction: Conditional Density Estimation Network Creation and Evaluation (CaDENCE) in R," *Comput Geosci*, vol. 41, pp. 126–135, 2012, doi: 10.1016/j.cageo.2011.08.023.
[40] L. Wen, "Forecasting Tourism Demand with an Improved Mixed Data Sampling Model," 2020, doi: 10.1177/0047287520906220.
[41] C. Foroni, M. Marcellino, C. S.-J. of the R. Statistical, and undefined 2015, "Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials," *JSTOR*, Accessed: Feb. 15, 2023. [Online]. Available: https://www.jstor.org/stable/43965717
[42] E. Ghysels, V. Kvedaras, and V. Zemlys-Balevičius, "Mixed data sampling (MIDAS) regression models," *Handbook of Statistics*, vol. 42, pp. 117–153, 2020, doi: 10.1016/bs.host.2019.01.005.
[43] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J Informetr*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.