МОДЕЛИ ПО СМЕШАННЫМ ДАННЫМ И ИХ ПРИМЕНЕНИЕ В КРАТКОСРОЧНОМ ПРОГНОЗИРОВАНИИ И НАУКАСТИНГЕ

В. И. Малюгин

*Белорусский государственный университет, пр. Независимости, 4,
2203030, Минск, Беларусь, email: Malugin@bsu.by*

Представляются результаты построения и оценки эффективности моделей MIDAS по смешанным данным в задачах краткосрочного прогнозирования и наукастинга месячных и квартальных темпов роста индекса потребительских цен в белорусской экономике на основе ежедневных обменных курсов белорусского рубля по отношению к российскому рублю и доллару США. Показано, что модели MIDAS в целом превосходит по точности прогнозов модели по агрегированным данным.

Ключевые слова: смешанные данные; модель MIDAS; краткосрочное прогнозирование; наукастинг; индекс потребительских цен, обменные курсы валют, эффект переноса обменного курса на инфляцию; точность прогнозов.

MIXED DATA MODELS AND THEIR APPLICATION IN SHORT-TERM FORECASTING AND NOWCASTING

V. I. Malugin

*Belarusian State University, Independence Av., 4,
220030, Minsk, Belarus, Malugin@bsu.by*

The results of the construction and evaluating the effectiveness of MIDAS models using mixed data in the tasks of nowcasting and short-term forecasting of monthly and quarterly growth rates of the consumer price index in the Belarusian economy based on daily exchange rates of the Belarusian ruble against the Russian ruble and the US dollar are presented. It is shown that MIDAS models are superior in forecast accuracy to models based on aggregated data.

Keywords: mixed data; MIDAS model; short-term forecasting; nowcasting; consumer price index; currency exchange rates; exchange rate pass-through to inflation; forecast accuracy.

Введение. Значительная часть ключевых экономических показателей формируется статистическими органами в квартальном представлении, что обуславливает широкое использование квартальных эконометрических моделей среднесрочного прогнозирования [1]. В то же время, многие потенциально полезные предикторы являются доступными из

различных источников с более высокой частотой, например, ежемесячно, ежедневно или еще чаще на финансовых рынках и в виде новостной информации из интернета. Таким образом, на практике для построения моделей могут быть доступны «смешанные данные» с различной частотой наблюдения. В тоже время, традиционный подход в эконометрическом моделировании основан на приведения всех используемых в моделях переменных к одному интервалу с самой низкой частотой наблюдения. Недостатки подхода на основе агрегированных данных наиболее существенно проявляются в задачах краткосрочного прогнозирования и наукастинга [2], т.е. при прогнозировании текущего, ближайшего прошлого и ближайшего будущего состояния процесса по данным, поступающим в реальном времени. Эти недостатки заключаются в следующем.

Рассчитываемые статистическими органами показатели, поступают с задержкой на один и более интервалов наблюдения, т.е. на месяц для месячных и квартал для квартальных показателей. Это, безусловно, ограничивает прогностические возможности моделей, сужая будущий горизонт прогнозирования за счет необходимости построения прогнозов для текущего, а иногда и уже прошедшего, временного интервала. Агрегирование данных влечет за собой потерю информации о динамике высокочастотных переменных внутри низкочастотного интервала наблюдения целевого показателя, что негативно отражается на точности краткосрочных и текущих прогнозов. Это вынуждает пользоваться экспертными оценками. В то же время, использование ежемесячных или ежедневных значений дает возможность осуществлять оперативную корректировку прогнозов для текущего квартала по данным, поступающим в реальном времени [3].

Возможности и типы моделей по смешанным данным. Модели по смешанным данным обладают следующими возможностями:

1) могут включать несколько высокочастотных переменных с разной частотой наблюдения, являясь при этом достаточно «экономными» по числу параметров;

2) предполагают оптимизацию лаговой структуры для различных высокочастотных переменных с помощью методов машинного обучения по малому числу «гиперпараметров»;

3) могут использовать динамические факторы, получаемые с помощью методов динамического факторного анализа и машинного обучения на основе большого массива высокочастотных переменных-предикторов разной частоты;

4) допускают многомерные представления моделей для совместного моделирования и прогнозирования взаимосвязанных низкочастотных переменных и заданного набора высокочастотных предикторов;

5) позволяют строить краткосрочные прогнозы, а также осуществлять корректировку прогнозов для текущего низкочастотного интервала (наукастинга) на основе поступающих в реальном времени данных с более высокой частотой;

б) расширяют возможности векторных моделей с марковскими переключениями состояний за счет использования высокочастотных опережающих переменных с циклическими изменениями.

К настоящему времени получили широкую известность следующие основные типы эконометрических моделей регрессионного и авторегрессионного типа по смешанным данным, а также их модификации [3]:

– регрессионные модели по смешанным данным с ограничениями (*Mixed Data Sampling regression models* – MIDAS) и без ограничений (*Unrestricted MIDAS* – U-MIDAS);

– векторные авторегрессионные модели (*Vector autoregressive models* – VAR) по данным смешанной частоты (*Mixed-Frequency VAR* – MF-VAR) и байесовские векторные авторегрессии (*Bayesian VAR*) по смешанным данным (MF-BVAR);

– динамические факторные модели по смешанным данным (*Dynamic Factors MF-VAR*);

– модели векторной авторегрессии с марковскими переключениями состояний по смешанным данным (*Markov Switching MF-VARX*).

Задачи и результаты проведенного исследования. Цель исследования состояла в построении и оценке эффективности моделей MIDAS для наукастинга и краткосрочного прогнозирования темпов роста месячной и квартальной инфляции в белорусской экономике в зависимости от ежедневных изменений обменных курсов белорусского рубля по отношению к основным валютам. Методология построения и применения моделей MIDAS описывается в работах [4, 5]. Модель MIDAS допускает включение одной и более высокочастотных экзогенных переменных с разными частотами. Предполагается стационарность используемых временных рядов, что требует преобразования временных рядов со стохастическими трендами к стационарному виду.

Модели по смешанным данным ранее не применялись белорусскими специалистами. Первые результаты по их использованию представлены в [6], где решаются следующие задачи:

- наукастинг сезонно скорректированных темпов роста индекса потребительских цен (ИПЦ) «квартал к кварталу» CPI_Q_t , и «месяц к месяцу» CPI_M_t с использованием изменений «день ко дню» обменных курсов белорусского рубля по отношению к основным валютам;

- наукастинг квартальных темпов роста ИПЦ с использованием ежедневных изменений обменных курсов белорусского рубля и месячных темпов роста ИПЦ;

- оценка устойчивости моделей к шоковым воздействиям на разнородных по динамике обменных курсов временных интервалах;

- сравнительный анализ точности краткосрочных прогнозов на основе моделей MIDAS, использующих высокочастотные переменные, и лучших моделей по агрегированным данным.

Экономическим основанием для построенных моделей MIDAS является теоретическое предположение о существовании «эффекта переноса обменного курса на инфляцию» (*exchange rate pass-through effect to inflation*), имеющего место в рамках гипотезы о паритете покупательской способности. Применительно к белорусской экономике «эффект переноса» исследуется в ряде работ, в частности [7]. В отличие от традиционных задач тестирования эффекта переноса для одной валюты, в проведенных исследованиях данный эффект рассматривается по отношению к корзине из двух валют. Приведем краткое описание моделей и результатов.

Введем обозначения:

$Y_t^L (t=1, \dots, T)$ – низкочастотная (*Low frequency*) эндогенная переменная, соответствующая прогнозируемому показателю; переменная Y_t^L может быть квартальной (*Quarterly*) Y_t^Q или месячной (*Monthly*) Y_t^M ;

$X_{1,t}^H, \dots, X_{N_H,t}^H$ – значения высокочастотной экзогенной переменной (*High frequency*), где N_H – число значений экзогенной переменной в интервале наблюдения эндогенной переменной, равно N_M и N_Q для месячной и квартальной переменных.

Для упрощения записи опишем модель MIDAS с одной высокочастотной экзогенной переменной. В предположении, что используются ежедневные значения экзогенной переменной, модель имеет вид [5]:

$$Y_{t+1}^L = \mu + \sum_{i=0}^{p_Y^L-1} \alpha_{i+1} Y_{t-i}^L + \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{N_D-1} w_{i+j \times N_D} (\theta^D) X_{N_D-i,t-j}^D + \sum_{l=1}^r \gamma_l d_{l,t}^L + \eta_{t+1}, \quad t = 1, \dots, T, \quad (1)$$

где: q_X^D – количество лагов экзогенной переменной, включаемых в модель; N_D – число ежедневных наблюдений в рассматриваемом низкоча-

стотном интервале; $X_{N_D-i,t}^D$ ($i = 0, \dots, N_D - 1$) – i -е значение экзогенной переменной для лага $j = 1, \dots, q_X^D$ на временном интервале $t = 1, \dots, T$.

Ключевую роль в модели MIDAS вида (1) играет функция, определяющая способ вычисления лаговых коэффициентов, называемая весовой функцией лаговых коэффициентов для экзогенной переменной

$$w_{i+j \times N_D}(\theta^D) \quad (i = 0, \dots, N_{D-1}, j = 1, \dots, q_X^D), \quad (2)$$

зависящая от вектора параметров $\theta^D = (\theta_1^D, \dots, \theta_m^D)'$ малой размерности $m N_D$, где «'» – знак транспонирования вектора.

Специальные методы вычисления весовой функции (2) позволяют сократить число оцениваемых параметров, по которым осуществляется оптимизация лаговой структуры модели, а также предотвратить эффекты мультиколлинеарности факторов и переобучения модели. Наиболее часто для этих целей применяются модели распределенных лагов Алмона [8] и их модификации, включая модель полиномиально распределенных лагов (*polynomial distributed lags – PDL/Almon weighting models*) и модель экспоненциально распределенных лагов (*exponential Almon lags models*). Построенные в [9] модели MIDAS используют лаги для высокочастотных переменных, определяемые моделью PDL/Almon порядка 3.

Для оценивания моделей использовались доступные на момент проведения исследования данные в следующих временных интервалах: квартальные данные 2015q1–2023q2; месячные данные 2015m1–2023m10; ежедневные данные с 1 января 2015 г. по 30 ноября 2023 г.

Модели MIDAS для квартальных и месячных временных рядов CPI_Q_t и CPI_M_t включают: соответствующие лаговые переменные CPI_Q_{t-1} и CPI_M_{t-1} , а также импульсные фиктивные переменные для учета аддитивных аномальных наблюдений во временных рядах эндогенных переменных в отдельные месяцы или кварталы. Все итоговые модели MIDAS являются статистически адекватными по тестам значимости оценок параметров на уровне значимости 0,05 (для отдельных случаев до 0,07), а также по тестам автокорреляции, гетероскедастичности и нормальности распределения остатков.

В таблице для иллюстрации приводятся результаты сравнительного анализа модели для квартальных темпов роста инфляции CPI_Q по ежедневным изменениям обменных курсов белорусского рубля по отношению к российскому рублю RUR_BYN_D и доллару США USD_BYN_D , а также ежемесячным изменениям инфляции CPI_M .

В таблице приводятся характеристики моделей: сумма квадратов остатков *RSS* (*Residual Sum of Squares*) – целевая функция для машинного обучения моделей MIDAS и информационная статистика Акаике AIC (*Akaike Information Criterion*). Для оценивания точности прогнозов на всем периоде оценивания и вневыборочных (*out of sample*) прогнозных интервалах используются среднеквадратическая ошибка *RMSE* (*Root Mean Squared Error*) и средняя абсолютная ошибка в процентах *MAPE* (*Mean Absolute Percentage Error*).

В качестве альтернативы моделям MIDAS рассматриваются лучшие из построенных модели по агрегированным данным из семейства моделей авторегрессии с распределенными лагами (*autoregressive distributed lag* – ARDL) и авторегрессионных моделей с экзогенными переменными ARX.

Оценки качества моделей и точности прогнозов темпов роста ИПЦ

Периоды оценивания и прогнозирования					Вневыборочный прогноз CPI_Q, %
Период оценивания модели	2015q1 – 2023q2	2015q1 – 2022q4	2015q1 – 2023q1	2015q1 – 2023q2	2023q3
Прогнозный период	2015q1 – 2023q2	2023q1 – 2023q2	2023q2 – 2023q3	2023q3 – 2023q3	
Характеристики моделей и прогнозов	Модель MIDAS по ежедневным обменным курсам и месячным значениям CPI_M				
RMSE	0,4255	0,6750	0,0273	–	100,2
MAPE	0,3209	0,6488	0,0271	–	
RSS	5,1209	4,9742	5,1204	5,1209	
AIQ	1,7020	1,7823	1,7554	1,7020	
	Модель ARX(1) по агрегированным обменным курсам				
RMSE	0,6693	0,8165	1,1948	–	101,7
MAPE	0,5180	0,8011	1,1892	–	
RSS	12,8482	11,1536	11,5159	–	
AIQ	2,2582	2,2027	2,1909	–	

Приведенные в таблице результаты демонстрируют существенное преимущество модели MIDAS над моделью ARX(1) по RMSE и MAPE для двухшаговых прогнозов на различных периодах оценивания моделей и всем другим статистикам. Вневыборочный прогноз для третьего квартала 2023 г. также является значительно более точным у модели MIDAS. Так при истинном значении CPI_Q в этом квартале, равном 100,6% прогнозы на основе моделей MIDAS и ARX соответственно равны 100,2% и

101,7%. Такое же преимущество имеет место и для месячной модели CPI_M по смешанным данным.

Актуальные направления исследований. Особый интерес в макроэкономическом анализе вызывает моделирование квартальных темпов роста реального ВВП на основе моделей по смешанным данным, использующих доступные в месячном выражении компоненты ВВП, а также опережающие статистические индикаторы по опросным и новостным данным [9]. В условиях шоковых воздействий, обуславливающих циклический характер изменения экономической активности («спад – подъем»), представляют особый интерес модели по смешанным данным с переключениями состояний, обобщающие традиционные модели MS-VARX с марковскими переключениями состояний [10].

Библиографические ссылки

1. Система эконометрических моделей для прогнозирования и оценки вариантов денежно-кредитной политики / Ю. С. Харин, В. И. Малюгин, М. В. Пранович, Д. Л. Мурин // Белорусский экономический журнал. 2003. № 3. С. 89–100.
2. *BañBura M., Giannone D., Reichlin L.* Nowcasting // The Oxford Handbook of Economic Forecasting. Oxford University Press, 2012. P. 193–224.
3. *Forni C., Marcellino M.* A survey of econometric methods for mixed frequency data // Working Paper 2013/06, Norges Bank. 2013. – 45 p.
4. *Ghysels E., Sinko A., Valkanov R.* Midas regressions: Further results and new directions // Econometric Reviews. 2006. № 26. P. 53–90.
5. *Andreou E., Ghysels E., Kourtellis A.* Regression models with mixed sampling frequencies / E. Andreou, // Journal of Econometrics. 2010. № 58. P. 246–261.
6. *Малюгин В. И.* Краткосрочное прогнозирование и наукастинг темпов роста инфляции на основе моделей по смешанным данным. Банковский вестник. 2024. № 1/726. С. 23–36.
7. *Мирончик Н. Л., Профатилов С. А.* О влиянии обменного курса на инфляцию // Банковский вестник. 2015. № 10 (627). С. 25–34.
8. *Almon S.* The distributed lag between capital appropriations and expenditures // Econometrica. 1965. № 33. P. 178–196.
9. *Малюгин В. И., Крук Д. Э., Милевский П. С.* Индекс экономических настроений белорусской экономики: методические, модельные и инструментальные средства построения и применения. Банковский вестник. Тематический выпуск «Исследования банка». 2019. №16. 30 с.
10. *Малюгин В. И., Макаревич А. С.* Модель MS-VARX и ее применение для анализа бизнес-цикла белорусской экономики // Банковский вестник. 2020. № 4. С. 22–31.