



МОДЕЛИ ПО ДАННЫМ РАЗНОЙ ЧАСТОТЫ И ИХ ПРИМЕНЕНИЯ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Бовт Тимофей Анатольевич

Научный руководитель: В.И. Малюгин

Цели работы и постановка решаемой задачи по данным разной частоты

Цели работы:

- подготовка аналитического обзора моделей по смешанным данным;
- построение моделей по смешанным данным на реальных данных белорусской экономики;
- сравнительный анализ точности прогнозирования альтернативных типов моделей по смешанным данным.

Постановка задачи.

В качестве примера приложения моделей по данным разной частоты решается задача исследования зависимости показателя ВВП Беларуси от показателя ИПЦ (инфляции) Беларуси и обменных курсов валют относительно белорусского рубля. Данные имеют следующую частоту:

- показатель ВВП – квартальный временной ряд;
- показатель ИПЦ – месячный временной ряд;
- обменные курсы валют – дневные временные ряды.

Проблема прогнозирования по данным разной частоты

Обычно все часто применяемые регрессионные модели машинного обучения работают с данными, заданными в одной частоте. Но некоторые данные из сферы экономики, как правило, формируются в квартальных представлениях. Параллельно с этим какие-либо объясняющие факторы могут быть собраны с более высокой частотой, будь то ежемесячные, еженедельные или ежедневные представления.

Популярные способы решения этой проблемы:

- наивное приведение данных более высокой частоты к нужной нам более низкой частоте, иначе говоря, агрегация данных более высокой частоты;
- специальные подходы для заполнения пропущенных значений.

Модели по агрегированным данным

Модели MIDAS регрессии строятся на основе моделей с распределенным запаздыванием (DL). Пусть

- y_t – эндогенная переменная;
- x_t – экзогенная переменная;
- ε_t – белый шум;
- β_0 – свободный член.

Используя эти обозначения, DL-модель формулируется как

$$y_t = \beta_0 + \sum_{i=0}^p b_i x_{t-i} + \varepsilon_t, \quad (1)$$

Более продвинутой модификацией является авторегрессионная модель с распределенным запаздыванием (ARDL), которая формулируется в виде

$$\sum_{i=0}^p \beta_i y_{t-i} = \sum_{j=0}^q \alpha_j x_{t-j} + \varepsilon_t. \quad (2)$$

Модели по данным разной частоты

Чтобы ввести модель Mixed Data Sampling (MIDAS) регрессии, введем обозначения:

- $y_t^{(q)}$ — эндогенная квартальная переменная;
- $x_t^{(m)}$ — экзогенная месячная переменная;
- $\varepsilon_t^{(m)}$ — белый шум;
- $\beta_0, \beta_1 \in \mathbb{R}$ — свободные переменные;
- $b(L^{1/m}, \Theta) = \sum_{j=0}^p b(j, \Theta) L^{j/m}$, где $L^{j/m} x_t^{(m)} = x_{(t-j)/m}^{(m)}$ — лаговый оператор.

Тогда базовая MIDAS модель может быть сформулирована в виде

$$y_t^{(q)} = \beta_0 + \beta_1 b(L^{1/m}, \Theta) x_t^{(m)} + \varepsilon_t^{(m)}. \quad (3)$$

Также можем записать это уравнение в виде

$$y_t^{(q)} = \beta_0 + \beta_1 \sum_{j=0}^p b(j, \Theta) x_{(t-j)/m}^{(m)} + \varepsilon_t^{(m)}. \quad (4)$$

Лаговые многочлены

Базовые MIDAS модели отличаются между собой в зависимости от выбора лагового оператора $b(L^{1/m}, \Theta) = \sum_{j=0}^p b(j, \Theta) L^{j/m}$. Фактически задание этого оператора определяет способ агрегации данных высокой частоты в ряд более низкой частоты. Наиболее распространенными являются следующие виды функции лаговых коэффициентов:

- экспоненциальные лаги Алмона

$$B(j, \Theta) = \frac{e^{\Theta_1 j + \dots \Theta_n j^n}}{\sum_{s=0}^p e^{\Theta_1 s + \dots \Theta_n s^n}}; \quad (5)$$

- бета лаги

$$B(j, \Theta_1, \Theta_2) = \frac{f(\frac{j}{K}, \Theta_1; \Theta_2)}{\sum_{s=0}^p f(\frac{s}{p}, \Theta_1; \Theta_2)}, \quad (6)$$

где

$$f(x, \Theta_1, \Theta_2) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(\Theta_1 + \Theta_2)}{\Gamma(\Theta_1)\Gamma(\Theta_2)}. \quad (7)$$

Альтернативные модели по данным разной частоты

В курсовой работе также рассматриваются математические определения и других моделей предназначенных для работы по данным разной частоты:

- MIDAS-модели многих экзогенных переменных;
- нелинейные MIDAS-модели;
- многомерные MIDAS-модели;
- линейные MIDAS-модели с регуляризацией;
- U-MIDAS-модели, или неограниченные MIDAS-модели;
- MF-VAR, или векторная авторегрессия смешанной частоты;
- MF-BVAR, или байесовские векторные авторегрессии смешанной частоты;
- MS-MFVAR, или векторная авторегрессия по смешанным данным с марковскими переключениями состояний;
- DFM, или динамические факторные модели по смешанным данным.

Оценка точности моделей по данным смешанной частоты

Для оценки качества прогнозов моделей наиболее популярными являются три следующих критерия:

- средняя абсолютная ошибка (MAE);
- средняя абсолютная ошибка в процентах (MAPE);
- корень из среднеквадратической ошибки (RMSE).

При построении наукастов не учитывается информация о последнем доступном квартале. Рассматриваемые модели сравниваются по последним 12 точкам, в которых построены ретроспективные прогнозы, и проверяются на будущем прогнозе, который построен на невошедшем квартале.

Построение базовых моделей MIDAS для решаемой задачи

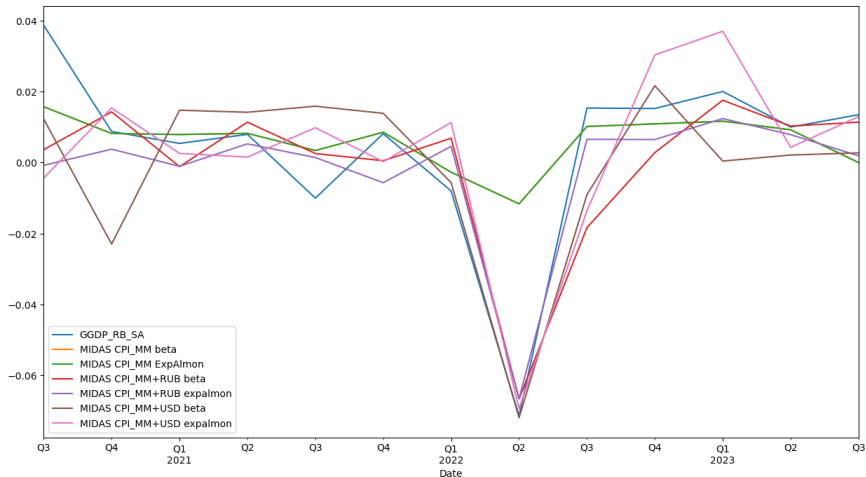


Рис.: Графики ретроспективных прогнозов по 12 последним точкам моделей MIDAS регрессии

Оценки точности построенных моделей

Таблица: Retrospective Evaluation Metrics

Model	MAE	MAPE	RMSE
MIDAS CPI_MM Beta	0.010561	0.473983	0.018825
MIDAS CPI_MM ExpAlmon	0.010561	0.473978	0.018825
MIDAS CPI_MM+RUB Beta	0.010875	0.816500	0.015378
MIDAS CPI_MM+RUB ExpAlmon	0.010394	0.784645	0.013824
MIDAS CPI_MM+USD Beta	0.013635	1.152350	0.016956
MIDAS CPI_MM+USD ExpAlmon	0.013467	0.993322	0.017887

Таблица: Future Evaluation Metrics

Model	MAE	MAPE	RMSE
MIDAS CPI_MM Beta	0.004407	0.787724	0.004407
MIDAS CPI_MM ExpAlmon	0.004407	0.787740	0.004407
MIDAS CPI_MM+RUB Beta	0.001626	0.290612	0.006263
MIDAS CPI_MM+RUB ExpAlmon	0.001587	0.283747	0.006301
MIDAS CPI_MM+USD Beta	0.004614	0.824704	0.012502
MIDAS CPI_MM+USD ExpAlmon	0.002240	0.400436	0.005648

Сравнение лучшей MIDAS модели с моделями DL и ARDL

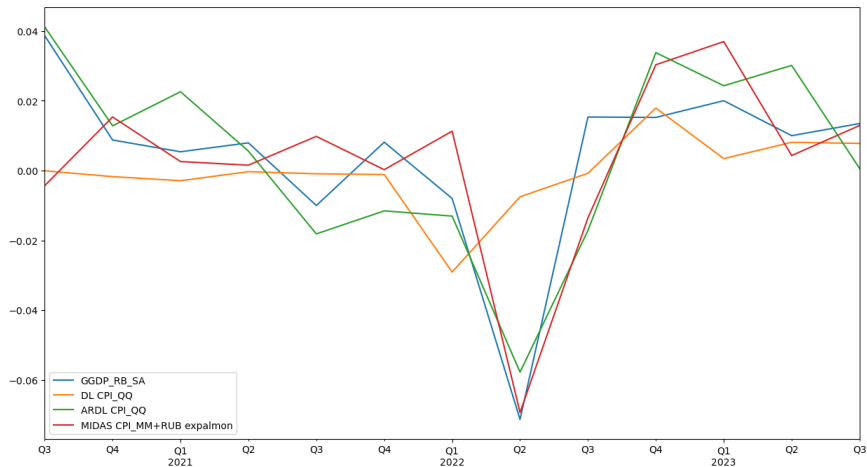


Рис.: Графики ретроспективных прогнозов по 12 последним точкам моделей MIDAS, DL и ARDL

Оценка точности моделей MIDAS, DL и ARDL

Таблица: Evaluation Metrics

Model	MAE	MAPE	RMSE
MIDAS CPI_MM+RUB ExpAlmon	0.013467	0.993322	0.017887
DL CPI_QQ	0.016318	1.001932	0.023231
ARDL CPI_QQ	0.012386	1.122417	0.015147

Заключение

Основные результаты работы:

- были математически определены и рассмотрены модели для работы по данным разной частоты;
- были исследованы статистические свойства временных рядов ВВП, ИПЦ и курсов валют;
- были построены модели MIDAS регрессии и модели с распределенным запаздыванием;
- была проведена оценка точности ретроспективных и будущих прогнозов для моделей MIDAS регрессии.

Используемые источники

- 1 Foroni, C. A survey of econometric methods for mixed frequency data / C. Foroni, M. Marcellino // Working Paper 2013/06, Norges Bank.
- 2 Ghysels, E., Santa-Clara P., Valkanov R. 2002. The MIDAS touch: Mixed data sampling regression models, Working paper, UNC and UCLA.
- 3 Макеева, Н.М., Наукастинг элементов использования ВВП России / Н.М. Макеева, И.П. Станкевич // Статья 2022/10, Экономический журнал ВШЭ.
- 4 Foroni, C. Unrestricted Mixed Data Sampling (U-MIDAS): MIDAS Regressions With Unrestricted Lag Polynomials / C. Foroni, M. Marcellino, C. Schumacher // Discussion paper 2015, Deutsche Bundesbank.
- 5 Станкевич И.П. Сравнение методов наукастинга макроэкономических индикаторов на примере российского ВВП // Прикладная эконометрика 2020. С. 113–127.
- 6 Харин, Ю. С. Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук // Минск : БГУ, 2011.