

Прикладная эконометрика, 2020, т. 59, с. 113–127.

Applied Econometrics, 2020, v. 59, pp. 113–127.

DOI: 10.22394/1993-7601-2020-59-113-127

И. П. Станкевич<sup>1</sup>

## Сравнение методов наукастинга макроэкономических индикаторов на примере российского ВВП

*Работа посвящена изучению точности оценок текущих темпов прироста ВВП (наукастов) на основе более оперативных данных. Сравняется качество наукастов для большого числа моделей: MIDAS (модели со смешанными данными) разных модификаций, в том числе с регуляризацией и снижением размерности матрицы объясняющих переменных при помощи метода главных компонент, и MFVAR (байесовских векторных авторегрессий смешанной частоты) с априорным распределением Миннесоты. В качестве объясняющих переменных используется набор индексов, характеризующих компоненты ВВП по производству. Показывается превосходство моделей, основанных на векторных авторегрессиях, над другими типами моделей, анализируется динамика ошибок наукастов, приводятся оценки месячных темпов прироста ВВП, полученные при помощи MFVAR моделей.*

**Ключевые слова:** наукастинг; ВВП; MIDAS модели; модели со смешанной частотой.

**JEL classification:** C53; E37.

### 1. Введение

Прогнозирование текущих значений публикуемых со значительной задержкой макроэкономических показателей (nowcasting, в литературе на русском языке часто оставляемый без перевода) на основе значений более оперативных индикаторов является важной задачей с точки зрения оперативной подстройки действий экономических агентов под изменяющиеся макроэкономические условия. Оно может позволить производить регулярный мониторинг экономической активности, оперативно регулировать денежно-кредитную и фискальную политику в зависимости от состояния экономики, помочь компаниям более точно планировать инвестиции, продажи и производство. Из всех макроэкономических показателей одним из самых важных, но при этом не самым оперативным, является ВВП, что обуславливает актуальность задачи наукастинга ВВП.

Кроме того, некоторые методы позволяют получать оценки месячных рядов ВВП (а статистика ВВП в месячном разрезе отсутствует), что само по себе представляет значительный интерес с точки зрения создания моделей и систем для оперативного прогнозирования

<sup>1</sup> Станкевич Иван Павлович — Национальный исследовательский университет «Высшая школа экономики»; ФИАН им. П. Н. Лебедева РАН; [istankevich@hse.ru](mailto:istankevich@hse.ru).

и оценки состояния экономики. Интересным может также быть анализ расхождений в наукастах («прогнозах» на текущий период) и публикуемых данных: они могут свидетельствовать о наличии проблем в статистике, что достаточно распространено для оперативных данных по ВВП.

С технической точки зрения можно выделить несколько основных подходов к наукастингу. Во-первых, это связующие уравнения, основанные на прогнозировании стандартными методами высокочастотного ряда с последующим его агрегированием в данные более низкой частоты и использованием полученного индикатора в качестве объясняющей переменной в низкочастотном уравнении для интересующего показателя. Этот подход был представлен, например, в работе (Ingenito, Trehan, 1996), и сохраняет значительную популярность до сих пор, см. (Schumacher, 2016; Soybilgen, Yazgan, 2018 и др.). Другой очень распространенный класс эконометрических моделей — это так называемые MIDAS (Mixed Data Sampling) модели со смешанными данными (подразумевается смешанная частота данных). Этот подход основан на использовании в уравнении для ряда более низкой частоты (к примеру, квартального) в качестве объясняющих переменных нескольких относящихся к текущему периоду значений ряда более высокой частоты (к примеру, трех месячных значений другого показателя, относящихся к данному кварталу). В зависимости от спецификации может вводиться ряд ограничений на значения коэффициентов модели: модели без ограничений обычно называются неограниченными MIDAS моделями (U-MIDAS). Представленные в (Ghysels et al., 2006, 2007), модели этого класса широко применяются для наукастинга макроэкономических показателей, см., например, (Chernis, Sekkel, 2017; Ferrara, Marsilli, 2019; Marcellino, Schumacher, 2010). Из технически более сложных подходов можно выделить векторные авторегрессии смешанной частоты (Mixed Frequency Vector Autoregression, MFVAR), основанные на описании совместной динамики высокочастотных показателей и ненаблюдаемого разложения низкочастотных показателей на данные более высокой частоты, см., к примеру, (Kuzin et al., 2011) или (Schorfheide, Song, 2015) в приложении к наукастингу. Существуют и байесовские обобщения этого подхода (McCracken et al., 2015).

Попытки применения методов машинного обучения к наукастингу макроэкономических показателей достаточно ограничены и сводятся обычно к методам снижения размерности или автоматического отбора переменных. Tiffin (2016) сравнивает методы отбора переменных на основе L1 и L2 регуляризации (Lasso и Ridge регрессии и Elastic Net регрессии как комбинации этих двух подходов) и использование методов случайного леса для наукастинга ВВП. Serpi et al. (2019) сравнивают различные методы отбора переменных (кроме уже упомянутых регуляризаций, они используют метод главных компонент и LARS) также применительно к наукастингу ВВП. В целом, ограниченное использование методов машинного обучения для наукастинга представляется оправданным ввиду того, что большая часть этих методов требует размера обучающей выборки большего порядка, чем обычно доступно в макроэкономике.

Вопрос наукастинга российского ВВП исследован довольно слабо, можно отметить только несколько работ. В исследовании (Микош, Соланко, 2019) изучаются вопросы как наукастинга, так и прогнозирования ВВП России при помощи данных со смешанной периодичностью и стандартного набора моделей (связующие уравнения, MIDAS, U-MIDAS). Однако эта работа основное внимание уделяет выявлению факторов, наиболее ценных для наукастинга и краткосрочного прогнозирования ВВП, вопросы же сравнения моделей и даже подбора параметров использованных моделей не затрагиваются (и, как отмечают сами авторы, работы,

сравнивающие широкий спектр моделей на российских данных, отсутствуют). Подобным образом выполнена и более ранняя работа (Поршаков и др., 2016), где основной упор делается на сравнение точности групп факторов (более 100 показателей, объединенных в группы «опережающие показатели», «показатели реального сектора» и «финансовые показатели») в рамках одной модели. В качестве основного инструмента выбраны динамические факторные модели (ДФМ), приводится сравнение результатов с моделью случайного блуждания и связующими уравнениями. Наукастинг ВВП России также затрагивается в работе (Яковлева, 2018), но не в плане подбора и сравнения моделей, а с точки зрения выделения переменных, в данном случае — анализа большого массива новостных текстов.

Таким образом, настоящая работа в определенном смысле заполняет существующий пробел в литературе по наукастингу российского ВВП, представляя сравнение качества работы достаточно широкого круга моделей.

## 2. Методология

Данная работа предлагает сравнение вневыборочной точности наукастов (прогноз величины показателя на текущий период, полученный на основе текущих значений других индикаторов), полученных различными моделями как чисто эконометрического характера, так и использующими отдельные приемы, свойственные машинному обучению. Из стандартных подходов в работе не рассматриваются связующие уравнения, продемонстрировавшие достаточно низкую точность в большинстве предшествующих исследований, и динамические факторные модели. Они заменяются на идейно близкие методы, основанные на методе главных компонент. Напомним, что динамические факторные модели, берущие свое начало в работе (Geweke, 1977), сводятся к описанию большого числа переменных набором (ненаблюдаемых) величин меньшей размерности, что очень близко перекликается со снижением размерности, для которого используются главные компоненты. Другой аргумент против использования динамических факторных моделей — достаточно небольшой набор объясняющих переменных, что влечет за собой отсутствие необходимости искать комбинации этих факторов. Вместо использования очень большого набора переменных, зачастую слабо связанных с ВВП, в данной работе используется небольшой набор показателей, характеризующих отдельные компоненты ВВП по производству (тем самым прямо связанных с ВВП). Метод главных компонент применяется здесь на высокочастотных данных (на вход принимается по три месячных значения каждой из объясняющих переменных), выполняя тем самым двойную роль снижения размерности и смены частотности данных.

### 2.1. MIDAS

В качестве базовой эконометрической модели в работе используется модель с данными смешанной частоты MIDAS, которая в наиболее общей форме может быть записана в виде:

$$y_t = \sum_{j=1}^p \alpha_j y_{t-j} + \sum_{i=0}^k \sum_{j=0}^{m_i} \beta_j^{(i)} x_{tm_i-j}^{(i)} + u_t,$$

где  $y_t$  — ряд данных более низкой частоты (в нашем случае — квартальный ряд темпов прироста ВВП);  $x_t^{(i)}$  — ряды объясняющих факторов более высокой частоты;  $m_i$  — число наблюдений объясняющей переменной  $x_t^{(i)}$ , соответствующих одному значению (одному периоду) в зависимой переменной. При этом в большинстве формулировок MIDAS моделей предполагается наличие ограничений на значения параметров  $\beta_j^{(i)}$  общего вида:

$$\beta_j^{(i)} = g^{(i)}(j, \lambda),$$

где конкретный вид функции  $g^{(i)}$  зависит от типа используемой MIDAS модели. По сути, эти ограничения определяют способ получения данных более низкой частоты из данных более высокой частоты.

Если ограничения на значения параметров не вводятся, полученная модель называется неограниченной MIDAS моделью (U-MIDAS).

Из MIDAS моделей с ограничениями, сравниваются следующие версии:

- неэкспоненциальные полиномы Алмона (Clements, Galvão, 2009)

$$\beta_j = \frac{\exp(\lambda_1(j+1) + \dots + \lambda_r(j+1)^r)}{\sum_{s=0}^d \exp(\lambda_1(s+1) + \dots + \lambda_r(s+1)^r)};$$

- экспоненциальные полиномы Алмона

$$\beta_j = \sum_{s=0}^d \lambda_s j^s;$$

- полиномы на основе распределения Гомпертца

$$\beta_j = \frac{\exp(\lambda_2 j - \lambda_1 e^{\lambda_2 j})}{\sum_{s=0}^d \exp(\lambda_2 s - \lambda_1 e^{\lambda_2 s})};$$

- полиномы на основе бета-распределения

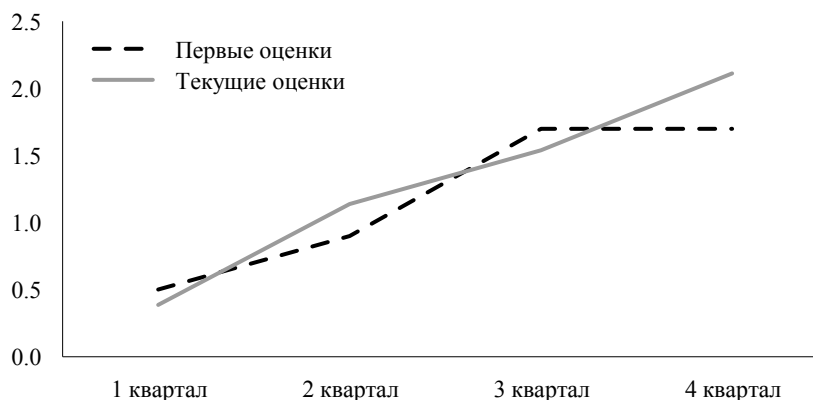
$$\beta_j = \frac{\left(\xi + (1-\xi) \frac{j-1}{m-1}\right)^{\lambda_1-1} \left(1-\xi - (1-\xi) \frac{j-1}{m-1}\right)^{\lambda_2-1}}{\sum_{s=0}^d \left(\xi + (1-\xi) \frac{s-1}{m-1}\right)^{\lambda_1-1} \left(1-\xi - (1-\xi) \frac{s-1}{m-1}\right)^{\lambda_2-1}},$$

где  $\lambda_s, \xi$  — параметры модели, которые при оценке модели рассчитываются вместе с коэффициентами. Разные версии MIDAS модели, хотя и обладают несколько различающимися свойствами, имеют общую цель ввести ограничения на параметры для снижения размерности задачи и не допустить, в терминах из машинного обучения, переобучение модели, сохранив при этом достаточную гибкость для правильной подстройки под данные.

Данная работа использует стандартную версию MIDAS модели без авторегрессионной части и с «технической» точки зрения основана на реализации моделей в пакете *midasr* языка *R* (Ghysels et al., 2016). Отказ от авторегрессионной части обусловлен тем, что данные по ВВП за предыдущий квартал (наиболее оперативные) могут внести в модель значительную

неточность в силу того, что такие данные, как правило, несколько раз корректируются. С учетом того, что при оценке моделей используются исторические данные, уже пересмотренные раньше, сравнивать наукасты по историческим данным и по оперативным в условиях запаздывания зависимой переменной было бы некорректно.

К сожалению, официальные ряды первых оценок темпов прироста ВВП не публикуются Росстатом, поэтому нет возможности измерить точность моделей на исторических данных в приближенных к реальности условиях. На основе оперативных публикаций Росстата («Срочные информации и справки по актуальным вопросам», раздел «О производстве и использовании ВВП (ежеквартальная)», доступны с начала 2019 г.)<sup>2</sup> можно восстановить первые оценки только за 2019 г. Их сравнение с текущими официальными данными приведено на рис. 1.



**Рис. 1.** Первые и текущие оценки квартальных темпов прироста ВВП России за 2019 г.

Расхождение в 4-м квартале 2019 г. превышает 0.4%, что примерно соответствует средней абсолютной ошибке наукаста лучших моделей (см. раздел 4). Масштабы пересмотров нивелируют ценность таких данных. Помимо этого, существуют свидетельства, что оценка одних и тех же моделей на данных до и после пересмотра статистики может давать качественно разные результаты (Chang, Li, 2018).

## 2.2. MFBVAR

Байесовские векторные авторегрессии смешанной частоты (Mixed-Frequency Bayesian VAR, MFBVAR) представляют собой версию стандартных векторных авторегрессий, модифицированную для использования с данными разной частоты. Предполагается, что при исходном высокочастотном (в данном случае месячном) процессе, описываемом стандартной VAR моделью, часть переменных наблюдается только на более низкой частоте. При этом, например, наблюдаемые квартальные значения являются средними из ненаблюдаемых месячных значений.

В данной работе используется версия байесовской векторной авторегрессии с априорным распределением Миннесоты, модифицированная для данных со смешанной частотой (McCracken et al., 2015). Общая идея априорного распределения Миннесоты, предложенная в работах (Litterman, 1979, 1986), заключается в том, что все рассматриваемые ряды

<sup>2</sup> См. <https://rosstat.gov.ru/compendium/document/50798>.

предполагаются процессами случайного блуждания с независимыми ошибками. Такой подход позволяет оценивать модели с большим числом переменных и лагов на данных сравнительно небольшого объема за счет исключения «лишних» переменных. Использование большинства других априорных распределений подразумевает необходимость вводить какие-то дополнительные предположения о свойствах данных и взаимосвязях показателей, а в данной работе этого хотелось бы избежать.

Расчеты проведены с использованием пакета *mfvar* языка *R* (Ankargren, Yang, 2019).

### 2.3. Линейные модели с регуляризацией

В качестве подхода, сохраняющего идею исключения «лишних» переменных, но более свойственного машинному обучению, рассматриваются линейные модели с регуляризацией. В работе представлены модели с L1 регуляризацией (известные также в эконометрике как LASSO регрессии), построенные на базе U-MIDAS модели. На практике при применении MIDAS моделей, как правило, используется только одна объясняющая переменная, что обуславливается небольшим объемом данных и необходимостью целого набора дополнительных регрессоров при введении в модель одной новой объясняющей переменной. Так, при использовании месячных данных для объяснения квартальных — это три дополнительных регрессора, при использовании данных более высокой частоты число регрессоров возрастает. Регуляризация позволяет снизить остроту этой проблемы и использовать наборы из нескольких объясняющих переменных в рамках одной модели.

LASSO-U-MIDAS модель может быть записана как

$$y_t = \beta^0 + \sum_{i=1}^k \sum_{j=0}^{m_i} \beta_j^{(i)} x_{tm_i-j}^{(i)} + u_t$$

с целевой функцией

$$\sum_{t=1}^T (y_t - \hat{y}_t)^2 + \lambda \sum_{i=1}^k \sum_{j=0}^{m_i} |\beta_j^{(i)}| \rightarrow \min .$$

Параметр регуляризации  $\lambda$  подбирается кросс-валидацией.

Одним из самых серьезных ограничений стандартной LASSO регрессии является отсутствие у нее *oracle property* (способность модели корректно отбирать и состоятельно оценивать ненулевые коэффициенты). В общем случае отбор переменных при помощи LASSO регрессии может быть несостоятельным (см. (Zou, 2006)). Для параметров, истинное значение которых равно нулю, вероятность получения нулевой оценки может не стремиться к 1, таким образом, в модели могут остаться «лишние» переменные. В качестве решения этой проблемы предлагается использовать адаптивное LASSO, предложенное в (Zou, 2006), которое обладает *oracle property*, в том числе для моделей большой размерности (Huang et al., 2008) и авторегрессионных моделей (Kock, 2012). Адаптивное LASSO подразумевает использование весов в целевой функции

$$\sum_{t=1}^T (y_t - \hat{y}_t)^2 + \lambda \sum_{i=1}^k \sum_{j=0}^{m_i} w_{ij} |\beta_j^{(i)}| \rightarrow \min ,$$

где  $w_{ij}$  — веса, полученные при помощи состоятельных оценок коэффициентов модели (в качестве таковых в (Zou, 2006) предлагается использовать МНК-оценки коэффициентов).

Модели, полученные на основе адаптивного LASSO, в таблицах отмечены как ALASSO-U-MIDAS.

Последней рассмотренной в данной работе моделью является модификация LASSO-U-MIDAS, в которой объясняющие переменные перед использованием преобразованы в главные компоненты. Такой подход позволяет увеличить объем информации, содержащийся в одной переменной, что может положительно сказаться на качестве результирующих моделей в условиях очень небольшого числа наблюдений, свойственного для макроэкономических задач.

### 3. Данные и оценка качества

Все модели оцениваются и сравниваются на официальных квартальных данных по ВВП России, в базовых ценах 2016 г., временной период — с 1-го квартала 2002 г. по 4-й квартал 2018 г. Все показатели рассчитываются как темпы прироста к соответствующему периоду предыдущего года для обеспечения сопоставимости и снятия необходимости сезонно корректировать данные. В качестве объясняющих переменных используется следующий набор показателей:

- $I_{Prom}$  — индекс промышленного производства;
- $I_{mining}$  — индекс производства по добыче полезных ископаемых;
- $I_{proc}$  — индекс обрабатывающего производства;
- $I_{en}$  — индекс производства в секторе «Обеспечение электрической энергией, газом и паром; кондиционирование воздуха»;
- $I_{agr}$  — индекс реального объема сельскохозяйственного производства;
- $I_{cons}$  — индекс реального объема работ, выполненных по виду деятельности «Строительство»;
- $I_{hous}$  — ввод в действие жилых домов;
- $I_{trans}$  — коммерческий грузооборот транспорта;
- $I_{trade}$  — индекс реального оборота розничной торговли.

Зависимая переменная — темп прироста ВВП в данном квартале к соответствующему кварталу предыдущего года.

Все объясняющие переменные также пересчитаны в темпы прироста к соответствующему кварталу предыдущего года, что обеспечивает сопоставимость показателей.

В качестве меры качества полученной модели используется средняя абсолютная ошибка:

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

где  $Y_t$  — фактически наблюдаемый темп прироста ВВП,  $\hat{Y}_t$  — предсказанный данной моделью темп прироста ВВП,  $T$  — число периодов времени (кварталов), на котором тестируется модель. Средняя абсолютная ошибка является оптимальной метрикой в ситуации, когда предсказываемая величина имеет типичные значения около нуля, потому что не предполагает возведения в квадрат величин, близких к нулю (как MSE и RMSE), что может исказить масштаб.

Для получения прогноза из данных удаляется информация о текущем квартале (как о темпе прироста ВВП, так и о значениях объясняющих переменных), оцененная на такой выборке модель затем применяется к объясняющим переменным, соответствующим данному

кварталу, для получения наукаста ВВП. Таким образом, все использованные прогнозы — вневыборочные.

Для тестирования моделей сравниваются результаты по последним 10 точкам: вторая половина 2016, весь 2017 и 2018 гг. (сравнительно однородный и стабильный с точки зрения макроэкономических условий период), и результаты по последним 20 точкам (2014–2018 гг. — период, значительно более разнородный).

#### 4. Результаты

Для начала рассмотрим результаты, полученные на последних 10 точках, а затем сравним их с результатами на последних 20 точках.

MIDAS модели, как правило, строятся с использованием только одной месячной объясняющей переменной с последующим выбором показателя, обеспечивающего наименьшую ошибку прогнозирования. Сравнение разных спецификаций модели и объясняющих переменных дает результаты, приведенные в табл. 1 (жирным шрифтом выделены лучшие результаты для каждой спецификации модели).

**Таблица 1.** Средняя абсолютная ошибка MIDAS моделей по последним 10 точкам, %

	IProm	Imining	Iproc	Ien	Iagr	Icons	Ihous	Itrans	Itrade
U-MIDAS	1.24	2.40	1.14	2.20	1.58	2.70	1.41	3.32	<b>1.08</b>
MIDAS nealmon	1.24	2.40	<b>1.10</b>	2.17	1.57	2.53	1.35	3.27	1.29
MIDAS almonp	1.23	2.41	1.10	2.17	1.58	2.51	1.41	3.30	<b>0.96</b>
MIDAS gompertz	1.25	2.40	<b>1.10</b>	2.18	1.57	2.48	1.34	3.27	1.29
MIDAS beta	1.24	2.39	<b>1.08</b>	2.19	1.57	2.59	1.32	3.28	1.29

Наибольшую точность наукастинга обеспечивают индекс обрабатывающего производства и индекс реального оборота розничной торговли, из всех спецификаций модели лучше оказывается MIDAS с экспоненциальными полиномами Алмона. Этот факт немного расходится с утверждением из (Микош, Соланко, 2019) о том, что неэкспоненциальные полиномы Алмона выигрывают в точности у экспоненциальных (хотя в упомянутой работе сравнение и не проводится), но согласуется с результатом о преимуществе полиномов Алмона перед другими спецификациями MIDAS моделей (Clements, Galvão, 2009; Marcellino, Schumacher, 2010).

Одним из достоинств байесовских векторных авторегрессий с априорным распределением Миннесоты является достаточно низкая склонность к переобучению, что позволяет при оценке модели использовать все имеющиеся переменные, варьируя только число лагов (на данных высокой частоты, в нашем случае — месячных). Оценки точности приведены в таблице 2.

**Таблица 2.** Средняя абсолютная ошибка MFBVAR моделей по последним 10 точкам, %

	Число лагов					
	3	4	5	6	7	8
MAE	0.45	0.44	0.41	0.41	0.23	0.39



Заметен рост точности по мере увеличения количества использованных лагов, вплоть до модели с 7 лагами. Заметим также, что ошибка для оптимальной MFBVAR модели более чем в 4 раза ниже, чем для оптимальной MIDAS модели.

Среди линейных моделей с регуляризацией сравнивались LASSO-U-MIDAS, ALASSO-U-MIDAS, PC-LASSO-U-MIDAS и PC-ALASSO-U-MIDAS (последние две — с расчетом главных компонент) в двух версиях: с использованием всего набора объясняющих переменных и с использованием лишь переменных, продемонстрировавших наиболее высокую точность в стандартных MIDAS моделях. Последнее мотивировано тем, что даже для моделей с регуляризацией избыточное число переменных с низкой объясняющей силой может привести к переобучению и снижению качества прогноза, хотя этот эффект будет менее ярко выражен, чем для моделей без регуляризации. Это особенно актуально для метода, основанного на главных компонентах, поскольку использование «лишних» переменных при расчете главных компонент может снизить объем полезной информации, приходящейся на одну компоненту.

При расчете моделей с отбором переменных были использованы следующие показатели: индекс промышленного производства, индекс обрабатывающего производства, индекс реального объема сельскохозяйственного производства, индекс реального оборота розничной торговли. Два из них показали самую высокую точность для стандартных MIDAS моделей, другие два также показали высокую точность, помимо этого, они характеризуют другие сектора экономики, что может оказаться полезным с точки зрения описания функционирования экономики. Оценки точности приведены в таблице 3.

**Таблица 3.** Средняя абсолютная ошибка MIDAS моделей с регуляризацией по последним 10 точкам, %

Модель	Все переменные	С отбором переменных
LASSO-U-MIDAS	1.29	0.71
ALASSO-U-MIDAS	1.16	0.61
PC-LASSO-U-MIDAS	1.45	0.67
PC-ALASSO-U-MIDAS	1.39	0.71

В соответствии с исходными предположениями, предварительный отбор переменных значительно повышает точность моделей. Интересно, что точность моделей с регуляризацией даже без отбора переменных практически не уступает точности простых MIDAS моделей, основанных на одной переменной. Использование же предварительного отбора переменных в сочетании с регуляризацией приводит к точности, значительно превосходящей результаты стандартных моделей.

Отдельно стоит отметить превосходство адаптивного LASSO над стандартным LASSO в большинстве случаев: соблюдение oracle property оказывается важным не только в теоретическом плане, но и с точки зрения качества прогнозов результирующих моделей. Самой точной среди всех рассмотренных модификаций MIDAS моделей оказалась основанная именно на адаптивном LASSO.

Полученные результаты (см. таблицы 4–6) оказываются устойчивыми и при рассмотрении более длительного периода построения прогнозов, включающего в себя периоды значительной макроэкономической нестабильности.

Среди всех MIDAS моделей самыми результативными оказываются основанные на индексе обрабатывающего производства (при этом модели с полиномами на основе бета-распределения немного выигрывают у моделей с полиномами Алмона).

**Таблица 4.** Средняя абсолютная ошибка MIDAS моделей по последним 20 точкам, %

	IProm	Imining	Iproc	Ien	Iagr	Icons	Ihous	Itrans	Itrade
U-MIDAS	1.96	2.83	<b>1.64</b>	2.93	3.12	2.04	3.45	3.39	2.10
MIDAS nealmon	1.94	2.82	<b>1.62</b>	2.94	3.09	1.98	3.28	3.36	2.02
MIDAS almonp	1.93	2.83	<b>1.61</b>	2.98	3.13	1.94	3.46	3.38	2.04
MIDAS gompertz	1.94	2.82	<b>1.62</b>	2.95	3.09	1.95	3.28	3.36	2.02
MIDAS beta	1.94	2.82	<b>1.60</b>	2.96	3.09	1.99	3.30	3.36	2.01

MFBVAR по-прежнему демонстрируют самую высокую точность среди всех рассмотренных моделей, при этом, в отличие от предыдущего случая, различия в точности моделей минимальны, но с небольшим преимуществом модели с 3 лагами перед моделями с большим числом лагов.

**Таблица 5.** Средняя абсолютная ошибка MFBVAR моделей по последним 20 точкам, %

	Число лагов					
	3	4	5	6	7	8
MAE	0.55	0.58	0.59	0.56	0.56	0.57

Линейные модели с регуляризацией и отбором переменных выигрывают у MIDAS, но отстают от MFBVAR. По-прежнему, модели, основанные на адаптивном LASSO, демонстрируют самую высокую точность.

**Таблица 6.** Средняя абсолютная ошибка MIDAS-моделей с регуляризацией по последним 20 точкам, %

Модель	Все переменные	С отбором переменных
LASSO-U-MIDAS	2.20	1.54
ALASSO-U-MIDAS	2.07	1.28
PC-LASSO-U-MIDAS	2.32	1.38
PC-ALASSO-U-MIDAS	2.40	1.59

Отдельный интерес представляет изучение ошибок во времени. Для этого рассмотрим на последних 20 точках динамику абсолютной ошибки лучшей из модификаций MIDAS (ALASSO-U-MIDAS) и лучшей MFBVAR модели (с 3 лагами) — рис. 2.

Обращает на себя внимание качественно разное поведение графиков ошибок: для ALASSO-U-MIDAS (как, кстати, и для других версий MIDAS моделей) не заметно выраженной сезонности, однако ошибка имеет явную тенденцию к возрастанию с пиком в третьем квартале 2017 г. Ошибка же для MFBVAR модели (с любым числом лагов поведение качественно одинаково) имеет ярко выраженную сезонную структуру с пиком в 4-м квартале, что заметно для всех годов, кроме 2018. В 2018 г. Росстатом был произведен существенный пересмотр внутригодовых данных, что могло привести к более равномерному распределению ошибки внутри года.

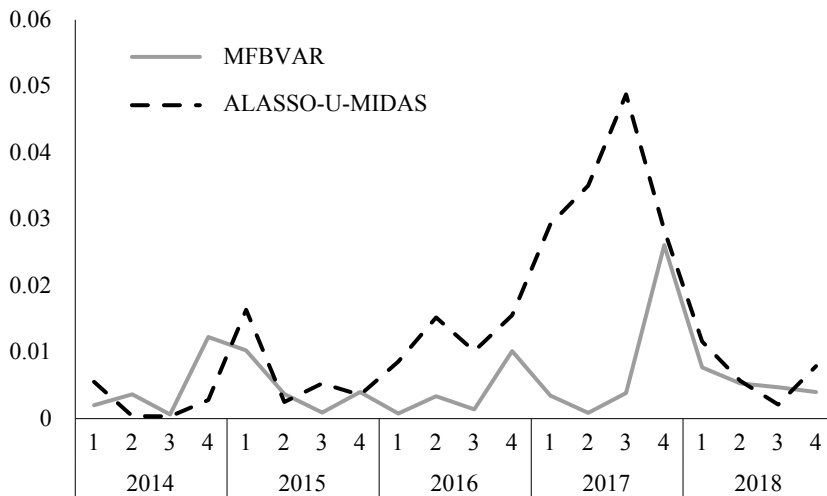


Рис. 2. Абсолютная ошибка наукастов MFBVAR и ALASSO-U-MIDAS моделей

В целом можно сказать, что MFBVAR по сравнению с модификациями MIDAS модели более корректно описывает ВВП, но более чувствителен к сезонным эффектам и «досчетам статистики» в конце года.

К достоинствам MFBVAR, помимо более высокой точности, можно отнести возможность получения оценки месячного ряда ВВП (в нашем случае — темпов прироста в месячном разрезе). Оценка месячных темпов прироста ВВП в сравнении с квартальными данными приведена на рис. 3.

Возможность такого разложения позволяет использовать модели этого класса не только как инструмент наукастинга ВВП, но и как инструмент оперативного мониторинга состояния экономики.

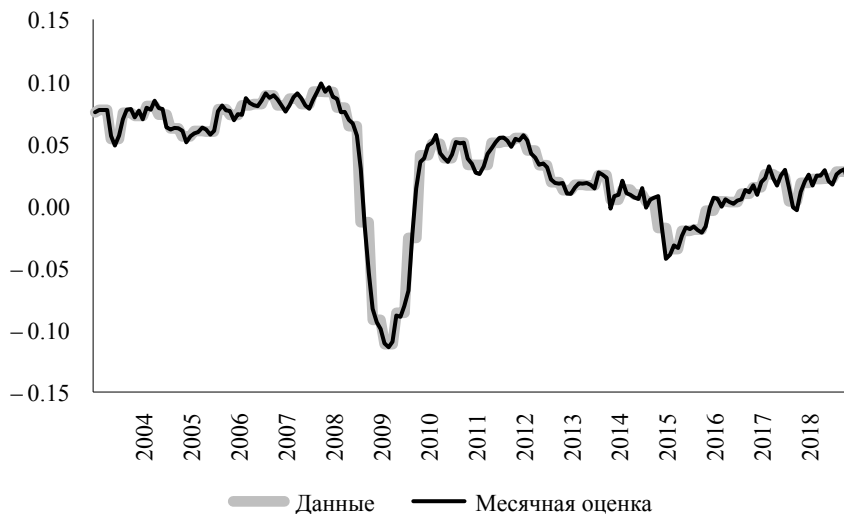


Рис. 3. Оценка месячных темпов прироста ВВП

## 5. Заключение

В работе сравнивается качество наукастов для различных моделей: MIDAS нескольких модификаций, в том числе с регуляризацией и снижением размерности объясняющих переменных при помощи метода главных компонент, и байесовских векторных авторегрессий смешанной частоты с априорным распределением Миннесоты (MFBVAR).

Показано, что для разных тестовых периодов (как более стабильных с точки зрения макроэкономических условий, так и более разнородных) самую высокую точность демонстрируют MFBVAR модели, далее идут модификации MIDAS моделей с регуляризацией и предварительным отбором переменных. Доминирование моделей с несколькими переменными может свидетельствовать о недостаточности стандартного в литературе о наукастинге подхода, основанного на использовании MIDAS переменных с одной объясняющей переменной. Помимо этого, ошибки MFBVAR моделей имеют ярко выраженную сезонную структуру с пиками в 4-м квартале, что может быть связано с особенностями расчета статистики и отнесению расхождений в годовых и квартальных оценках ВВП на 4-й квартал. Кроме того, MFBVAR модели позволяют получить оценки месячных темпов прироста ВВП, что делает их полезным инструментом для оперативного мониторинга состояния экономики.

## Список литературы

Микош Х., Соланко Л. (2019). Прогнозирование роста российского ВВП с использованием данных со смешанной периодичностью. *Деньги и кредит*, 78 (1), 19–35.

Поршаков А. С., Пономаренко А. А., Синяков А. А. (2016). Оценка и прогнозирование ВВП России с помощью динамической факторной модели. *Журнал Новой экономической ассоциации*, 2 (30), 60–76.

Яковлева К. (2018). Оценка экономической активности на основе текстового анализа. *Деньги и кредит*, 77 (4), 26–41.

Ankargren S., Yang Y. (2019). Mixed-frequency Bayesian VAR models in R: The mfbvar Package. [https://cran.r-project.org/web/packages/mfbvar/vignettes/mfbvar\\_jss.pdf](https://cran.r-project.org/web/packages/mfbvar/vignettes/mfbvar_jss.pdf).

Сепни О., Güney I. E., Swanson N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, 35 (2), 555–572.

Chang A. C., Li P. (2018). Measurement error in macroeconomic data and economics research: Data revisions, gross domestic product, and gross domestic income. *Economic Inquiry*, 56 (3), 1846–1869.

Chernis T., Sekkel R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics*, 53 (1), 217–234.

Clements M. P., Galvão A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24 (7), 1187–1206.

Ferrara L., Marsilli C. (2019). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *The World Economy*, 42 (3), 846–875.

Geweke J. (1977). The dynamic factor analysis of economic time series. In: *Latent Variables in Socio-Economic Models*, 365–383. Amsterdam: North-Holland.

Ghysels E., Kvedaras V., Zemlys V. (2016). Mixed frequency data sampling regression models: The R package midasr. *Journal of Statistical Software*, 72 (4), 1–35.

Ghysels E., Santa-Clara P., Valkanov R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131 (1–2), 59–95.

Ghysels E., Sinko A., Valkanov R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26 (1), 53–90.

Huang J., Ma S., Zhang C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 1603–1618.

Ingenito R., Trehan B. (1996). Using monthly data to predict quarterly output. *Economic Review (Federal Reserve Bank of San Francisco)*, 3, 3–11.

Kock A. B. (2012). On the oracle property of the adaptive LASSO in stationary and nonstationary autoregressions. *CREATES Research Papers*, 2012–05. Aarhus University.

Kuzin V., Marcellino M., Schumacher C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27 (2), 529–542.

Litterman R. B. (1979). Techniques of forecasting using vector autoregressions. *Working Paper No. 115*. Federal Reserve Bank of Minneapolis.

Litterman R. B. (1986). Forecasting with Bayesian vector autoregressions — five years of experience. *Journal of Business & Economic Statistics*, 4 (1), 25–38.

Marcellino M., Schumacher C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72 (4), 518–550.

McCracken M. W., Owyang M., Sekhposyan T. (2015). Real-time forecasting with a large, mixed frequency, Bayesian VAR. *Working Paper 2015–30*. Federal Reserve Bank of St. Louis.

Schorfheide F., Song D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33 (3), 366–380.

Schumacher C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, 32 (2), 257–270.

Soybilgen B., Yazgan E. (2018). Evaluating nowcasts of bridge equations with advanced combination schemes for the Turkish unemployment rate. *Economic Modelling*, 72, 99–108.

Tiffin A. (2016). Seeing in the dark: A machine-learning approach to nowcasting in Lebanon. *Working Paper 16/56*. International Monetary Fund.

Zou H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101 (476), 1418–1429.

Поступила в редакцию 11.06.2020;  
принята в печать 30.08.2020.

Stankevich I. Comparison of macroeconomic indicators nowcasting methods: Russian GDP case. *Applied Econometrics*, 2020, v. 59, pp. 113–127.

DOI: 10.22394/1993-7601-2020-59-113-127

### Ivan Stankevich

National Research University Higher School of Economics (NRU HSE), Moscow, Russian Federation; istankevich@hse.ru

## Comparison of macroeconomic indicators nowcasting methods: Russian GDP case

The paper compares the nowcasting quality of a range of models of Russian GDP using high-frequency data. The models compared are MIDAS in different specifications, including models with regularization and dimensionality reduction using principal components and Mixed-Frequency Bayesian VAR with Minnesota prior. Indices corresponding with GDP by production components are used as explanatory variables. Nowcasts of MFBVAR models are shown to have higher accuracy than obtained by any type of MIDAS models on different test time periods. We also analyze dynamics of nowcasting errors of models and calculate monthly estimate of GDP growth rate that can be obtained with MFBVAR models.

**Keywords:** nowcasting; GDP; MIDAS models; mixed frequency models.

**JEL classification:** C53; E37.

## References

Mikosch H., Solanko L. (2019). Forecasting quarterly Russian GDP growth with mixed-frequency data. *Russian Journal of Money and Finance*, 78 (1), 19–35 (in Russian).

Porshakov A. S., Ponomarenko A. A., Sinyakov A. A. (2016). Nowcasting and short-term forecasting of Russian GDP with a dynamic factor model. *Journal of the New Economic Association*, 2 (30), 60–76 (in Russian).

Yakovleva K. (2018). Text mining-based economic activity estimation. *Russian Journal of Money and Finance*, 77 (4), 26–41 (in Russian).

Ankargren S., Yang Y. (2019). Mixed-frequency Bayesian VAR models in R: The mfbvar Package. [https://cran.r-project.org/web/packages/mfbvar/vignettes/mfbvar\\_jss.pdf](https://cran.r-project.org/web/packages/mfbvar/vignettes/mfbvar_jss.pdf).

Cepni O., Güney I. E., Swanson N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, 35 (2), 555–572.

Chang A. C., Li P. (2018). Measurement error in macroeconomic data and economics research: Data revisions, gross domestic product, and gross domestic income. *Economic Inquiry*, 56 (3), 1846–1869.

Chernis T., Sekkel R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics*, 53 (1), 217–234.

Clements M. P., Galvão A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24 (7), 1187–1206.

Ferrara L., Marsilli C. (2019). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *The World Economy*, 42 (3), 846–875.

Geweke J. (1977). The dynamic factor analysis of economic time series. In: *Latent Variables in Socio-Economic Models*, 365–383. Amsterdam: North-Holland.

Ghysels E., Kvedaras V., Zemlys V. (2016). Mixed frequency data sampling regression models: The R package midasr. *Journal of Statistical Software*, 72 (4), 1–35.

Ghysels E., Santa-Clara P., Valkanov R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131 (1–2), 59–95.

Ghysels E., Sinko A., Valkanov R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26 (1), 53–90.

Huang J., Ma S., Zhang C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 1603–1618.

Ingenito R., Trehan B. (1996). Using monthly data to predict quarterly output. *Economic Review (Federal Reserve Bank of San Francisco)*, 3, 3–11.

Kock A. B. (2012). On the oracle property of the adaptive LASSO in stationary and nonstationary autoregressions. *CREATES Research Papers*, 2012–05. Aarhus University.

Kuzin V., Marcellino M., Schumacher C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27 (2), 529–542.

Litterman R. B. (1979). Techniques of forecasting using vector autoregressions. *Working Paper No. 115*. Federal Reserve Bank of Minneapolis.

Litterman R. B. (1986). Forecasting with Bayesian vector autoregressions — five years of experience. *Journal of Business & Economic Statistics*, 4 (1), 25–38.

Marcellino M., Schumacher C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72 (4), 518–550.

McCracken M. W., Owyang M., Sekhposyan T. (2015). Real-time forecasting with a large, mixed frequency, Bayesian VAR. *Working Paper 2015–30*. Federal Reserve Bank of St. Louis.

Schorfheide F., Song D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33 (3), 366–380.

Schumacher C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, 32 (2), 257–270.

Soybilgen B., Yazgan E. (2018). Evaluating nowcasts of bridge equations with advanced combination schemes for the Turkish unemployment rate. *Economic Modelling*, 72, 99–108.

Tiffin A. (2016). Seeing in the dark: A machine-learning approach to nowcasting in Lebanon. *Working Paper 16/56*. International Monetary Fund.

Zou H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101 (476), 1418–1429.

Received 11.06.2020; accepted 30.08.2020.