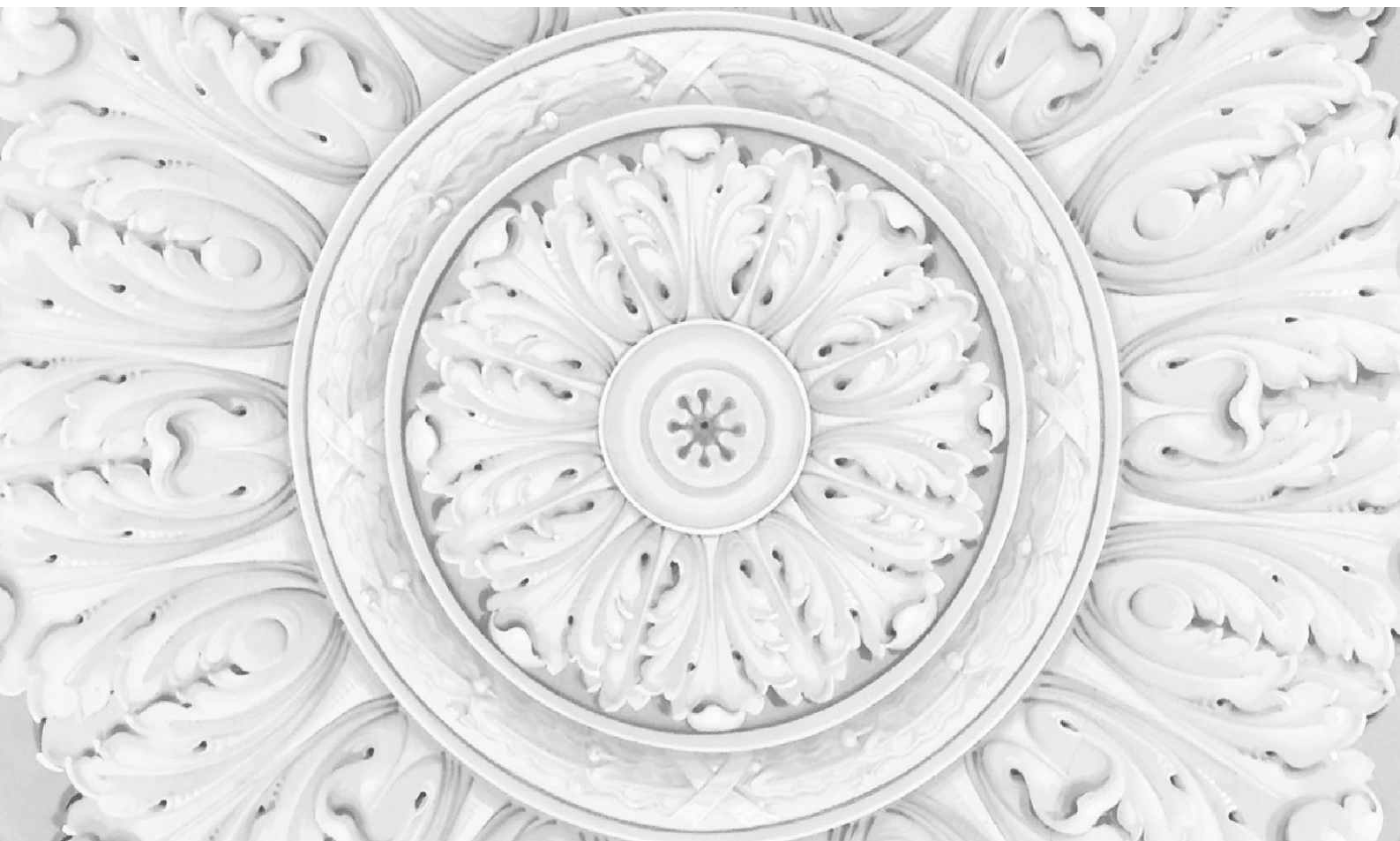




Банк России

Центральный банк Российской Федерации



СЕРИЯ ДОКЛАДОВ ОБ ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

Андрей Андреев

**Прогнозирование инфляции ме-
тодом комбинирования прогнозов
в Банке России**

№14 / Август 2016 г.

Андрей Андреев

Банк России. Email andreevav@cbr.ru

Автор выражает благодарность Ю.Л. Плушевской за полезные комментарии и замечания. Все ошибки, которые могут содержаться в данной работе, являются сферой ответственности автора.

© Банк России, 2016

Адрес ул. Неглинная, 12, Москва, 107016
Телефоны +7 495 771-91-00, +7 495 621-64-65 (факс)
Сайт www.cbr.ru

Все права защищены. Содержание настоящего Доклада (настоящих докладов) выражает личную позицию автора (авторов) и может не совпадать с официальной позицией Банка России. Банк России не несет ответственности за содержание Доклада (докладов). Любое воспроизводство представленных материалов допускается только с разрешения авторов.

Резюме

В статье описывается применение метода комбинирования прогнозов для прогнозирования инфляции в Банке России. В заключении приведена таблица, сопоставляющая точности прогнозов инфляции, полученных с использованием различных методов. Метод комбинирования прогнозов в широком круге случаев дает наилучший результат прогноза на короткий срок.

Ключевые слова: краткосрочное прогнозирование инфляции, комбинирование прогнозов.

JEL классификация: С43, С55, Е37.

ВВЕДЕНИЕ

С 2015 года Банк России перешел к политике таргетирования инфляции. В рамках реализации этого режима денежно-кредитной политики решения принимаются исходя из среднесрочного прогноза, начальным условием формирования которого являются краткосрочные (на горизонт двух кварталов) прогнозы основных макроэкономических показателей, включая инфляцию. Необходимость обеспечения высокого качества прогнозов требует применения комплексных подходов. Важным элементом системы краткосрочного прогнозирования инфляции в Банке России является метод комбинирования прогнозов.

Вышеупомянутый подход к прогнозированию динамики цен широко используется в центральных банках стран, денежно-кредитная политика которых направлена на поддержание ценовой стабильности. Например, в центральных банках Англии, Норвегии, Турции, Швейцарии и в Европейском центральном банке. Например, в исследовании, проведенном в Центральном Банке Турции [1], авторы пришли к выводу, что метод комбинирования прогнозов превосходит по точности все рассматриваемые безусловные прогнозы и прогнозы, использующие экзогенные переменные.

Одним из преимуществ обсуждаемого подхода является возможность использовать при прогнозировании большой объем данных [4]. Вместо включения всех имеющихся переменных одновременно в модель для получения единственного прогноза, которое может отрицательно сказаться на качестве получаемых результатов, метод комбинирования предполагает получение различных прогнозов исследуемой переменной, в каждом из которых можно учесть часть из имеющихся данных. Таким образом, в процессе прогнозирования появляется возможность использовать большое количество переменных без включения всех их одновременно в одну модель.

Второй причиной, по которой может быть полезно использовать метод комбинирования прогнозов, является его гибкость при наличии структурных сдвигов. При построении прогноза с использованием единственной модели для обнаружения структурного сдвига потребуются большое количество наблюдений и переоценка всей взаимосвязи. При этом до того, как структурный сдвиг будет обнаружен, модель будет давать ошибочный результат. В методе комбинирования, в свою очередь, происходит постоянный пересмотр весов различных прогнозов для достижения максимально быстрой подстройки метода к произошедшим структурным сдвигам.

И наконец, комбинирование прогнозов является способом диверсифицировать случайную ошибку. При нахождении средневзвешенного прогноза ошибки, полученные при использовании различных моделей, усредняются [6].

С другой стороны, недостатком метода комбинирования является неопределенность весов отдельных моделей в итоговом прогнозе и необходимость их оценки. Таким образом, оцениваемые веса вносят дополнительную ошибку в итоговый прогноз.

В разделе 1 представлены используемые данные. В разделе 2 описаны модели, применяемые для прогнозирования. В разделе 3 описан общий подход, на котором основывается метод комбинирования. В разделе 4 приводится метод, с помощью которого были выбраны параметры, обеспечивающие наилучшее качество итогового прогноза. Заключение содержит краткую презентацию полученных результатов.

1. ОПИСАНИЕ ДАННЫХ

Для повышения качества прогнозов индекс потребительских цен по Российской Федерации (ИПЦ) разделяется на его субиндексы. Такой подход обладает своими достоинствами и недостатками [3]. Во-первых, использование дезагрегированных данных дает возможность построить модель для более точного прогнозирования инфляции. Например, если динамика цен на различные компоненты потребительской корзины описывается различными процессами, то модель для агрегированного индекса будет иметь неправильную спецификацию. Во-вторых, важным достоинством подхода является усреднение несистематических ошибок прогноза каждого отдельного субиндекса инфляции. При условии независимости ошибок дезагрегирование инфляции снижает волатильность итогового прогноза индекса цен.

Однако, с другой стороны, прогнозирование каждого субиндекса инфляции создает дополнительные трудности с прогнозом весов отдельных компонент в ИПЦ [5].

В рамках применяемого подхода строятся модели для достаточно большого количества субиндексов потребительских цен. В каждой из трех групп товаров и услуг выделяется по шесть субиндексов цен, всего 18 субиндексов ИПЦ. В группе продовольственных товаров выделены субиндексы цен на мясо- и рыбопродукты; яйца кур и молокопродукты; кондитерские изделия, чай, кофе; мучные продукты; плодоовощную продукцию; прочие продовольственные товары. Среди непродовольственных товаров моделируется динамика цен на одежду, медицинские товары и бытовую химию, товары с регулируруемыми ценами, мебель и строительные материалы, технику и автомобили, прочие непродовольственные товары. В группе цен на услуги выделены подгруппы цен на бытовые, образовательные, туристские услуги, услуги пассажирского транспорта, жилищно-коммунального хозяйства и прочие услуги.

В качестве экзогенных используются переменные, отражающие шоки совокупного спроса (реальная зарплата), предложения (продовольственные цены ФАО, производство сельхозпродукции, промышленное производство, индекс цен производителей промышленных товаров, предназначенных для реализации на внутреннем рынке) и реального эффективного обменного курса. Выбор этих показателей учитывает результаты предварительного анализа значимости разнообразных переменных при моделировании ценовых индикаторов. Их будущая динамика оценивается с учетом сценарных условий Банка России.

Оценка моделей происходит на месячных данных. Чтобы исключить влияние сезонных изменений на динамику цен, проведена коррекция сезонности с использованием алгоритма X-12 ARIMA на выборке с начала 2002 года.

2. ИСПОЛЬЗУЕМЫЕ МОДЕЛИ

Метод прогнозирования инфляции с помощью комбинированных прогнозов основывается на использовании различных моделей, которые можно разделить на две группы.

В первую группу можно объединить модели, в которых предполагается, что будущую инфляцию определяет только ее прошлая динамика. К этой группе относятся модели случайного блуждания, авторегрессии с линейным трендом и ненаблюдаемой компоненты.

Модель случайного блуждания (RW) предполагает, что логарифм уровня цен следует процессу случайного блуждания с дрейфом:

$$\ln cpi_t = \beta_0 + \ln cpi_{t-1} + \varepsilon_t,$$

где cpi_t – индекс цен, приведенный к базе, в момент времени t ,

ε_t – случайная ошибка, удовлетворяющая условиям Гаусса–Маркова.

Модель авторегрессии с линейным трендом (LTAR) основана на предположении о том, что динамика индекса цен задается некоторым трендом. При этом отклонения от этого тренда объясняются прошлой динамикой вокруг тренда:

$$\ln cpi_t = \beta_0 + \beta_1 \ln cpi_{t-1} + \beta_2 \ln cpi_{t-2} + \beta_3 trend + \varepsilon_t .$$

Модель ненаблюдаемой компоненты (UC) основана на разложении динамики цен (с исключением сезонности) на две составляющие: тренд и случайную компоненту. При этом темп дрейфа тренда может меняться со временем, отражая структурные изменения в экономике. Разложение происходит с применением фильтра Калмана методом максимального правдоподобия по следующим уравнениям:

$$\begin{aligned} \ln cpi_t &= \ln cpi_{t-1} + drift_t + \varepsilon_t , \\ drift_t &= drift_{t-1} + u_t , \end{aligned}$$

где $drift_t$ отражает скорость месячного роста цен;

u_t – случайная ошибка, удовлетворяющая условиям Гаусса–Маркова.

Вторую группу составляют модели, в которых используются различные экзогенные переменные: модели векторной авторегрессии (стандартной (VAR) и байесовской (BVAR)) и линейной регрессии (OLS). В качестве экзогенных используются переменные, отражающие как ситуацию на внутреннем рынке (индикаторы спроса и предложения), так и внешние факторы, воздействующие на ценовую динамику. Будущая динамика этих переменных учитывается в соответствии со сценарными условиями прогнозов Банка России. В модели включаются все возможные линейные комбинации экзогенных переменных с лагами.

Модели векторной авторегрессии (стандартной и байесовской) строятся отдельно для каждой из трех основных ценовых групп: для группы субиндексов цен на продовольственные товары, непродовольственные товары и услуги. В этих трех моделях в качестве эндогенных переменных выступают субиндексы цен на товары и услуги, входящие в состав соответствующих трех групп. Включение в модель векторной авторегрессии субиндексов цен на товары и услуги, входящие в одну товарную группу, учитывает эффекты замещаемости и комплементарности товаров и услуг. В модели для каждой группы субиндексов цен используется свой набор экзогенных переменных, которые могут включаться в итоговую модель (подробнее о включении отдельных экзогенных переменных см. в разделе 3). Как показали оценки, максимальная точность прогноза достигается при использовании в моделях двухмесячного лага.

Таким образом, формально модель можно записать как

$$Y_t = X_t \beta + Z_t \alpha + \varepsilon_t ,$$

где Y_t – вектор логарифмов всех субиндексов инфляции, составляющих один из трех указанных основных субиндексов ИПЦ:

$$Y_t = \begin{pmatrix} \ln cpi_t^1 \\ \dots \\ \ln cpi_t^k \end{pmatrix} ;$$

X_t – матрица лаговых значений эндогенных переменных:

$$X_t = I_k \otimes (Y_{t-1}, Y_{t-2}) ;$$

Z_t – вектор экзогенных переменных с лагами, включая константу;

α и β – матрицы коэффициентов перед регрессорами;

k – количество субиндексов цен, включенных в каждую модель (по 6 в применяемом подходе).

Модель линейной регрессии, оцениваемая методом наименьших квадратов, строится отдельно для каждого из 18 субиндексов инфляции. Предполагается, что индекс цен в каж-

дом периоде зависит от прошлой динамики цен и набора экзогенных переменных:

$$\ln cpi_t = \beta_0 + \beta_1 \ln cpi_{t-1} + \beta_2 \ln cpi_{t-2} + \alpha Z_t + \varepsilon_t,$$

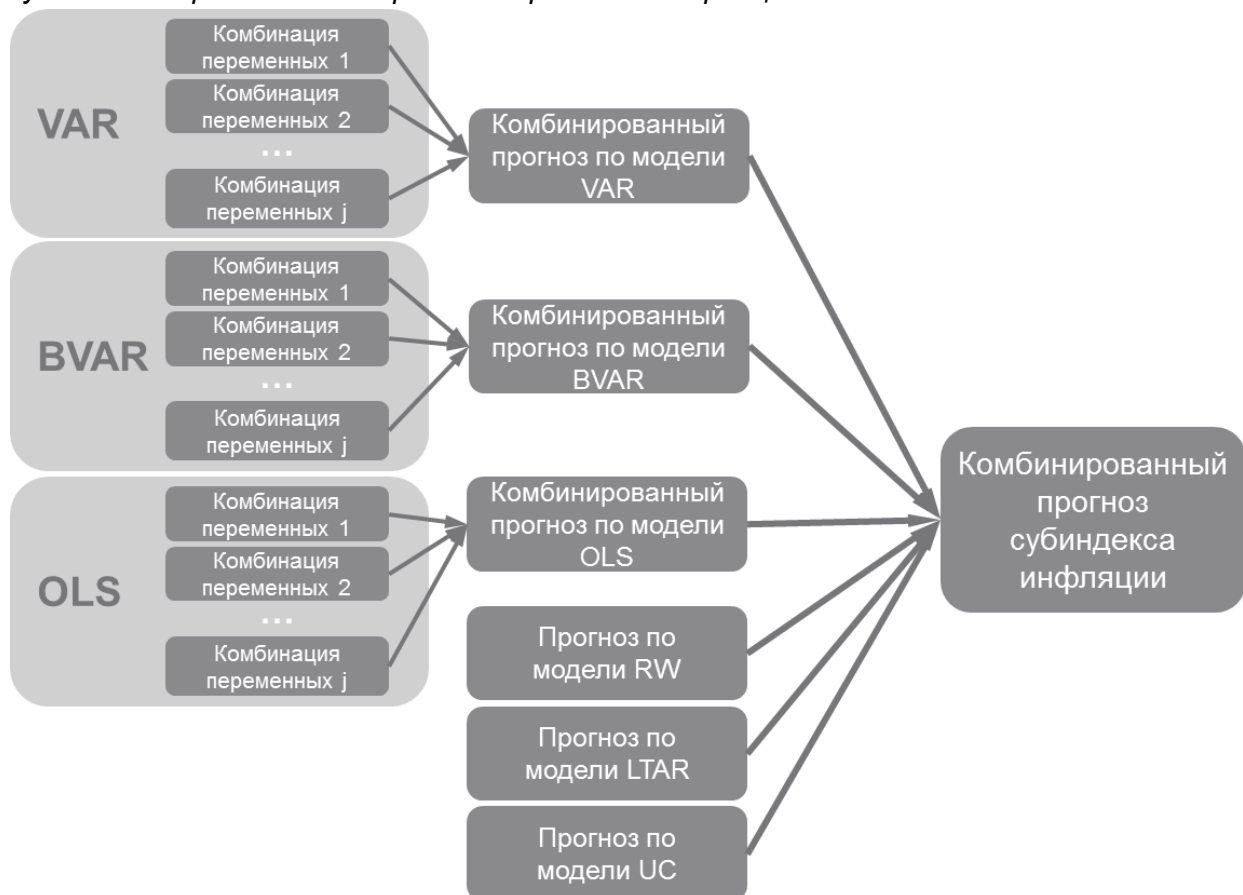
Все описанные модели оцениваются на скользящей выборке длиной 120 месяцев как для исторической оценки их точности, так и для формирования итогового прогноза. Уровень цен является стационарным в разностях, поэтому все описанные модели оцениваются в разностях, чтобы избежать проблемы мнимых регрессий.

3. МЕТОД КОМБИНИРОВАНИЯ

Метод комбинирования прогнозов [2] представляет собой следующий алгоритм (рис. 1).

На первом этапе для каждого субиндекса ИПЦ строятся прогнозы с использованием всех описанных выше моделей. Каждая из безусловных моделей дает по одному прогнозу для каждого субиндекса. В моделях векторной авторегрессии (стандартной, байесовской) и линейной регрессии в качестве объясняющих используются все возможные линейные комбинации выбранных экзогенных переменных, поэтому каждой такой комбинации соответствует свой прогноз динамики цен, их количество для каждого субиндекса ИПЦ определяется количеством модификаций моделей.

Рисунок 1. Алгоритм комбинирования прогнозов инфляции в Банке России



На следующем этапе происходит промежуточное усреднение прогнозов, сделанных с помощью каждой из моделей, использующих различные комбинации экзогенных переменных (векторные авторегрессии и линейная регрессия). Для этого оценивается вневыборочная точность каждого из полученных прогнозов. В качестве меры точности используется обратное от среднеквадратичного отклонения фактической динамики цен от прогнозируемой.

$$v_{i,j} = \frac{T}{\sum_t (cpi_t - cpi_{i,j,t}^f)^2},$$

где $cpi_{i,j,t}^f$ – прогнозируемый уровень цен в периоде t с использованием модели i и комбинации экзогенных переменных j ;

T – количество прогнозных периодов, на котором определяется точность прогнозов.

Прогнозы, полученные для одного субиндекса ИПЦ с помощью одной и той же модели и различных комбинаций экзогенных переменных, комбинируются с весами, пропорциональными их точности; веса нормируются для каждой модели. Таким образом рассчитывается один итоговый прогноз для каждого субиндекса ИПЦ по каждой из моделей, использующих экзогенные переменные. Другими словами, происходит комбинирование между различными наборами экзогенных переменных.

$$cpi_{i,t}^f = \frac{\sum_j v_{i,j} \times cpi_{i,j,t}^f}{\sum_j v_{i,j}},$$

где $cpi_{i,t}^f$ – комбинированный прогноз уровня цен в периоде t с использованием модели i .

Затем аналогичным образом комбинируются прогнозы каждого субиндекса ИПЦ, полученные с использованием различных моделей: для безусловных моделей – обычный прогноз; для моделей векторной авторегрессии и линейной регрессии – комбинированный прогноз с использованием различных наборов экзогенных переменных. Оценивается вневыборочная точность каждого прогноза:

$$w_i = \frac{T}{\sum_t (cpi_t - cpi_{i,t}^f)^2},$$

где $cpi_{i,t}^f$ – прогнозируемый уровень цен в периоде t с использованием модели i ;

T – количество прогнозных периодов, на котором определяется точность прогнозов.

Наконец, для получения финального прогноза динамики каждого субиндекса ИПЦ прогнозы, полученные с использованием различных моделей, агрегируются с весами, пропорциональными их точности:

$$cpi_t^f = \frac{\sum_i w_i \times cpi_{i,t}^f}{\sum_i w_i},$$

где cpi_t^f – комбинированный прогноз уровня цен в периоде t .

4. ОПРЕДЕЛЕНИЕ ВЕСОВ И ВЫБОР ЭКЗОГЕННЫХ ПЕРЕМЕННЫХ

Длина временного интервала, на котором оценивается вневыборочная точность всех прогнозов, влияет на задание весов и, соответственно, на результаты рассматриваемого метода. Максимально короткий интервал предполагает использование самых последних наблюдений, предшествующих прогнозу, для оценки весов различных моделей в итоговом прогнозе. Таким образом, наличие структурных сдвигов и изменения в характере ценовых процессов могут быть учтены через максимально короткий промежуток времени. Однако применение короткого интервала вневыборочного прогноза для оценки точности делает метод чувствительным к случайным ошибкам индивидуальных моделей. Использование более длинного временного интервала для оценки вневыборочной точности прогнозов позволяет снизить роль случайной ошибки в отдельно взятом наблюдении, однако может не выявить значимых изменений в динамике ценовых процессов. На практике вариант подбирается экспертным путем. В нашем случае длина интервала для вневыборочного прогноза составила 1 месяц.

Другая проблема – выбор экзогенных переменных для прогноза. Метод комбинирования прогнозов – это технический подход, который ограничивает возможность экспертных суждений в силу очень большого количества оцениваемых моделей. В этой ситуации для выбора экзогенных переменных может быть использован описанный ранее подход, основанный на взвешивании прогнозов на основе среднеквадратичных отклонений от фактической инфляции. Для повышения точности метода из всех используемых комбинаций экзогенных переменных можно исключить те, которые обладают наименьшей точностью. Таким образом, худшие модели будут отброшены и не окажут возможного отрицательного влияния на результат. С другой стороны, исключение моделей снижает выигрыш, связанный с усреднением ошибок индивидуальных прогнозов.

Сопоставление прогнозов, полученных для различных варьирующихся параметров модели, показало следующее:

1. Независимо от количества исключаемых комбинаций экзогенных переменных большая точность прогнозов достигается при использовании максимально короткого интервала (1 месяц) для определения точности прогноза каждой модели. Таким образом, в российских условиях доминирует важность гибкости пересмотра весов различных моделей, используемых в методе комбинирования.
2. При использовании коротких интервалов для определения весов различных моделей наиболее точным оказывается прогноз, в котором исключается 75% комбинаций экзогенных переменных, которые дают наименьшую точность прогноза. Таким образом, можно сделать вывод о том, что исключение комбинаций экзогенных переменных, которые дают наименее точный прогноз, повышает итоговое качество комбинированного прогноза, несмотря на снижения выигрыша от усреднения случайной ошибки разных комбинаций экзогенных переменных.

Приведенные результаты (рис. 2) были получены при прогнозировании инфляции с января 2013 года по декабрь 2015 года. Оценка вневыборочной точности прогнозов была проведена на временных интервалах длиной от 1 до 24 месяцев. Для проверки обоснованности исключения комбинаций экзогенных переменных, которые дают наименее точные прогнозы, оценены результаты, полученные со всеми комбинациями экзогенных переменных и с исключением 25, 50 и 75% из них. В качестве меры точности была использована U-статистика, которая сравнивает точность прогноза с использованием каждой модели с наивным прогнозом, предполагающим, что годовая инфляция не изменится. Важным достоинством этой величины является отношение ошибки прогноза к фактической инфляции, что дает ошибке в периоды с высокой инфляцией меньший вес, чем в периоды с низкой инфляцией.

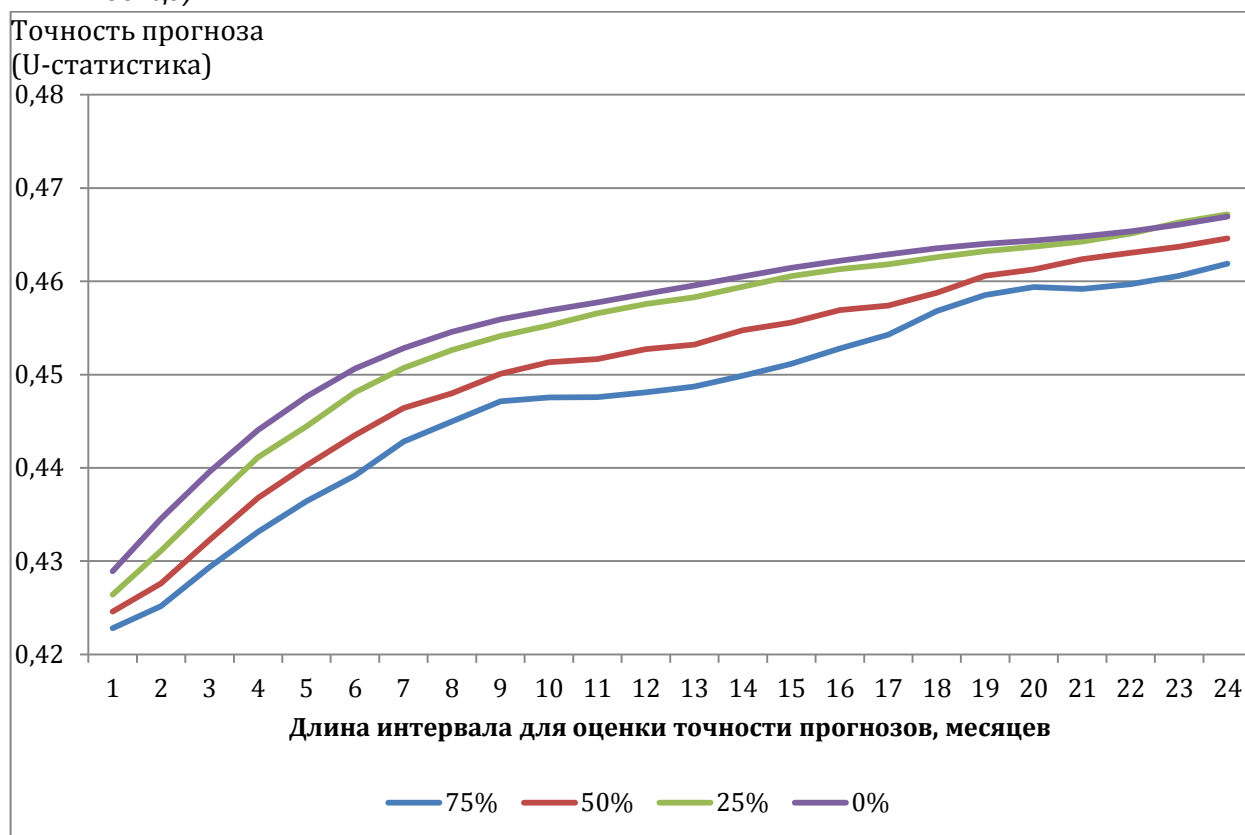
$$U = \frac{\sqrt{\sum_{t=1}^T \left(\frac{\pi_{t+h}^f - \pi_{t+h}}{\pi_t} \right)^2}}{\sqrt{\sum_{t=1}^T \left(\frac{\pi_t - \pi_{t+h}}{\pi_t} \right)^2}},$$

где $\pi_t = \frac{cpi_t - cpi_{t-12}}{cpi_{t-12}} * 100$ – уровень годовой инфляции,

π_t^f – прогноз инфляции,

h – горизонт прогнозирования.

Рисунок 2. Зависимость качества прогноза от доли исключенных комбинаций экзогенных переменных (0, 25, 50, 75%) и длины интервала для определения точности прогноза (1, 2, ... 24 месяца)



РЕЗУЛЬТАТЫ

По описанной выше методологии была проведена оценка точности различных подходов (включая метод комбинирования) к прогнозированию инфляции. С использованием U-статистики оценена точность различных моделей, которые можно использовать для прогнозирования инфляции, и проведено сравнение с методом комбинирования. Для сравнения рассмотрены две группы моделей. В первую группу объединены модели, используемые в методе комбинирования, которые прогнозируют каждый субиндекс потребительских цен с последующей их агрегацией; во вторую группу – классические модели для индекса потребительских цен, которые используют экзогенные переменные и хорошо описывают динамику инфляции: векторная авторегрессия (обычная и байесовская) и линейная регрессия.

Полученные результаты сведены в таблицу 1, которая показывает зависимость U-статистики от горизонта прогнозирования для различных моделей. На большинстве горизонтов прогнозирования метод комбинирования опережает по качеству другие рассмотренные модели. На тех горизонтах прогнозирования, когда метод комбинирования уступает по качеству другим моделям, разница в точности незначительная.

Таблица 1. Качество прогнозов различных моделей на различных горизонтах прогнозирования (U-статистика)

Горизонт прогнозирования, мес.	Прогнозы субиндексов цен с последующей их агрегацией						Классические модели			Комб. прогноз
	BVAR	LTAR	OLS	RW	UC	VAR	BVAR	OLS	VAR	
1	0,431	0,529	0,461	0,728	0,575	0,430	0,506	0,411	0,469	0,396
2	0,417	0,606	0,449	0,744	0,677	0,437	0,601	0,466	0,542	0,405
3	0,406	0,666	0,450	0,769	0,765	0,408	0,682	0,491	0,587	0,411
4	0,426	0,697	0,455	0,787	0,829	0,414	0,755	0,516	0,619	0,413
5	0,483	0,716	0,459	0,802	0,873	0,445	0,838	0,531	0,656	0,441
6	0,527	0,734	0,481	0,818	0,920	0,463	0,930	0,573	0,729	0,470

Синим шрифтом выделена лучшая модель на каждом горизонте прогнозирования.

Описанный метод зарекомендовал себя в центральных банках многих стран, таргетирующих инфляцию. Его широкое применение обусловлено тем, что он помогает разрешить неопределенность в выборе наиболее точного метода прогнозирования, в выборе объясняющих переменных для построения прогноза и тем самым дает возможность избежать больших ошибок, вызванных принятием субъективных решений. Метод комбинирования показал высокое качество прогнозов в период высокой волатильности в России и зарекомендовал себя как важная часть прогнозного инструментария Банка России.

ЛИТЕРАТУРА

1. Akdogan K., Baser S., Chadwick M.G. et al. Short term inflation forecasting models for Turkey and a forecast combination analysis // Central Bank of the Republic of Turkey. 2012. Working paper N12/09.
2. Bates J.M., Granger C.W.J. The combination of forecasts // Operational Research Quarterly. 1969. Vol. 20. P. 451–468.
3. Bermingham C., D'Agostino A. Understanding and Forecasting Aggregate and Disaggregate Price Dynamics // Working Paper 1365, European Central Bank. 2011.
4. Bjornland H.C., Jore A.S., Smith C., Thorsrud L.A. Improving and evaluating short term forecasts at the Norges Bank // Norges Bank Staff Memo. 2008. N 4.
5. Hendry D.F., Hubrich K. Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate // Journal of Business & Economic Statistics. 2011. 29 (2). P. 216–227.
6. Timmermann A. Forecast combinations. in: G. Elliot, C.W.J. Granger and, A. Timmermann. Handbook of Economic Forecasting. 2006. Vol. 1.