



НАЦИОНАЛЬНЫЙ БАНК КАЗАХСТАНА

**СИСТЕМА СЕЛЕКТИВНО-  
КОМБИНИРОВАННОГО ПРОГНОЗА  
ИНФЛЯЦИИ (SSCIF):  
выбор оптимальной техники  
прогнозирования динамики  
потребительских цен в условиях  
структурного шока (на примере  
Казахстана)**

**Департамент исследований и статистики  
Экономическое исследование №2017-9**

Тулесуов Олжас

Экономические исследования и аналитические записки Национального Банка Республики Казахстан (далее – НБРК) предназначены для распространения результатов исследований НБРК, а также других научно-исследовательских работ сотрудников НБРК. Экономические исследования распространяются для стимулирования дискуссий. Мнения, высказанные в документе, выражают личную позицию автора и могут не совпадать с официальной позицией НБРК.

Система селективно-комбинированного прогноза инфляции (SSCIF): выбор оптимальной техники прогнозирования динамики потребительских цен в условиях структурного шока (на примере Казахстана)

Ноябрь 2017 года

**NBRK – WP – 2017 – 9**

**Система селективно-комбинированного прогноза инфляции (SSCIF):  
выбор оптимальной техники прогнозирования динамики  
потребительских цен в условиях структурного шока (на примере  
Казахстана)**

Тулеев Олжас<sup>1</sup>

**Аннотация**

Как показывает практика моделирования, применение традиционного подхода в прогнозировании инфляции, основанного на использовании одной наиболее точной по сравнению с другими модели, не всегда может отвечать всем требованиям получения точных и робастных прогнозов зависимой переменной ввиду наличия структурных сдвигов и шоковых явлений в экономических процессах. Одним из подходов, способных превосходить по качеству и точности предсказаний инфляции другие методы, является подход **комбинирования прогнозов**, впервые предложенный Bates и Granger в 1969 году.

В данной работе автор предлагает модификацию ставшей уже «классической» техники комбинированного прогнозирования путем развития **селективного комбинирования прогнозов** на примере данных инфляции Казахстана. Так, Национальный Банк Казахстана в качестве опытного образца реализации данной техники применяет систему селективно-комбинированного прогноза инфляции или SSCIF – the System of Selective Combined Inflation Forecast.

Таким образом, в данной работе описаны и раскрыты технические аспекты построения системы селективно-комбинированного прогноза на примере данных инфляции в Казахстане, детально рассмотрены классы моделей SSCIF, а также представлены эмпирические оценки и сравнение точности прогнозов инфляции вне выборки (out-of-sample) на периоде до 9 месяцев отдельно по каждому модельному подходу SSCIF и по всей системе.

**Ключевые слова:** прогнозирование инфляции, комбинированный прогноз, селективно-комбинированный прогноз, макроэкономические показатели.

**Классификация JEL:** C32; C53; E31.

---

<sup>1</sup> Тулеев Олжас – главный специалист-аналитик, Управление макроэкономического прогнозирования и мониторинга, Департамент исследований и статистики, Национальный Банк Республики Казахстан. E-mail: Olzhas.Tuleuov@nationalbank.kz

## Содержание

Введение.....	3
Обзор литературы.....	6
Методология исследования и исходные данные.....	11
Обсуждение полученных результатов.....	23
Заключение.....	24
Список литературы.....	26
Приложение	

## 1. Введение

Традиционно построение моделей прогнозирования инфляции и в целом временных рядов макроэкономических показателей в эконометрическом контексте сходится к оценке различных модельных классов и семейств регрессий, что позволяет решить задачи прогнозирования зависимой переменной. При этом согласно традиционным эконометрическим подходам, из всего набора оцененных моделей на основании различных показателей точности их прогнозов, рассчитанных на тестовом участке выборки фактических наблюдений, отбирается так называемая «модель-чемпион». Предполагается, что данная единственная модель является наиболее оптимальной с точки зрения прогнозирования, так как на тестовом участке выборки фактических наблюдений, которые не были учтены при обучении этой модели, ее ошибка прогноза была наименьшей.

Как показывает практика моделирования, применение данного традиционного подхода не всегда может отвечать всем требованиям получения точных и робастных прогнозов зависимой макроэкономической переменной. Это связано с тем, что исторические наблюдения множества макроэкономических показателей, в частности инфляции, могут отражать экономические процессы, не содержащие ярко выраженной и четкой закономерной зависимости лишь от конкретного набора предикторов, которые были включены в отобранную модель-чемпион. Другими словами, наличие структурных сдвигов и шоковых явлений в экономических процессах демонстрирует, что количественная оценка степени зависимости целевого показателя от определенного набора объясняющих показателей, которая может быть максимально точно описана на историческом периоде и соответствовать всем статическим параметрам, не гарантирует сохранения степени выявленных связей на горизонте вне-выборочного прогноза (*out-of-sample forecast*). Дополнительно проблема может заключаться в том, что прогнозирование макроэкономических переменных на основании ограниченного набора предикторов и одной модели-чемпиона будет малоэффективной, так как сам целевой макроэкономический показатель является результатом агрегирования других менее агрегированных компонентов. Так, отдельные прогнозы более дезагрегированных компонентов на основании отдельных индивидуальных функций от такого же набора предикторов, как и в модели общего агрегированного показателя, могут быть намного точнее.

Таким образом, в этом случае возникает необходимость использовать большой набор информации при моделировании, способный максимально, всесторонне и детально описать весь изучаемый и прогнозируемый процесс. Но при этом возникает проблема, связанная с невозможностью сочетать в одной модели-чемпионе одновременно все оценочные техники и более расширенный набор предикторов, способных максимально описать

изучаемый экономический процесс, в частности, инфляцию. С технической точки зрения такая проблема будет выражаться в следующем:

- на фоне увеличения количества предикторов при неизменности количества их наблюдений произойдет снижение степеней свободы оцениваемой модели, а также возникнет существенная мультиколлинеарность между объясняющими переменными;

- вследствие сочетания в одном прогнозном уравнении множества различных модельных подходов будет наблюдаться искажение и значительное отклонение полученных количественных оценок коэффициентов при объясняющих переменных от экономически верных и интуитивно логичных причинно-следственных связей.

Одним из подходов, способных решить данную проблему, является метод комбинирования моделей, или метод комбинирования прогнозов.

Метод комбинирования прогнозов используется для решения задач прогнозирования, как временных рядов, так и перекрестных данных. Суть данного метода заключается в том, чтобы для прогнозирования целевого показателя применять весь максимально широкий набор модельных подходов, оценка которых является целесообразной, а не единственную модель-чемпион, отобранную из общего набора. При этом для получения итогового прогноза целевого показателя прогнозы по каждой из имеющихся моделей взвешиваются на основе весов, рассчитанных в зависимости от точности каждой модели на тестовом периоде фактической выборки. Это значительно диверсифицирует случайную ошибку, а также позволяет с наименьшими потерями и техническими издержками учитывать очень широкий объем информации в процессе моделирования. В свою очередь, в части макроэкономического прогнозирования, реализация такого подхода позволяет переходить от предсказания одномерного временного ряда общего агрегированного макропоказателя к моделированию и прогнозированию многомерных временных рядов, которые являются дезагрегированными компонентами целевого зависимого показателя. При этом чувствительность дезагрегированных компонентов к неизменному набору предикторов будет существеннее и качественнее, чем в случае с единым зависимым агрегированным показателем. Вместе с тем, метод комбинирования прогнозов дает возможность увеличивать набор предикторов, включая их в независимые друг от друга модели, что позволяет избежать проблемы снижения степеней свободы и мультиколлинеарности.

В настоящее время метод комбинирования моделей, называемый в науке данных «ансамблевым обучением» (ensemble learning), также широко используется при машинном обучении больших массивов данных (Big Data), сочетая в себе такие подходы к обучению, как бустинг (boosting), бэггинг (bagging), случайные леса (random forest), деревья многомерных регрессий (multivariate regression trees) и другие более традиционные методы. Развитие алгоритмов комбинирования моделей на Big Data показало, что построение ансамблей этих моделей превосходит по качеству предсказаний другие

методы, став тем самым одним из мощных методов машинного обучения (Шитиков, Мастицкий, 2017).

В случае прогнозирования макроэкономических показателей метод комбинирования моделей по аналогии с наукой данных также реализуется через сочетание различных параметрических и непараметрических подходов, но уже более характерных для классической и байесовской эконометрики, на основании которых анализируются и моделируются в отличие от big data временные ряды с ограниченным числом наблюдений. При этом в качестве основных подходов к моделированию выступают множественные линейные и нелинейные факторные регрессии, авторегрессионные интегрированные модели со скользящими средними, различные виды векторных авторегрессий и моделей коррекции ошибок, авторегрессионные модели с распределенными лагами, модели ненаблюдаемой компоненты и пространства состояний, модели случайного блуждания и др. В частности, применительно к макроэкономическим данным, метод комбинирования моделей в настоящее время широко используется в прогнозировании инфляционных процессов.

В отношении чисто эконометрической реализации метода комбинирования прогнозов элементы машинного обучения используются для создания алгоритмов, позволяющих самостоятельно рассчитывать веса каждой модельной техники в общем влиянии на конечный прогноз целевой переменной, а также с учетом данных весов самостоятельно и автоматически усреднять прогнозные результаты по всем моделям в один общий прогноз.

Стоит отметить, что в ходе исследований автора было выявлено, что метод «классического» комбинирования прогнозов обладает определенным недостатком, связанным со снижением точности и робастности тех итоговых взвешенных прогнозов, которые на горизонте прогноза находятся ближе к точке начального прогнозирования. Так, если по данному методу производить комбинацию прогнозов на  $k$  месяцев (периодов) вперед, то окажется, что более «дальние» прогнозы будут точнее, чем более «близкие». Причины таких различий автор связывает с тем, что для расчета комбинационных весов на тестовом периоде времени учитывается средняя точность каждой модели за весь тестовый период (например, в среднем за последние  $k$  месяцев). В рамках данной работы подобные веса условно называются «стационарными комбинационными весами». Взамен этого автор предлагает **селективно** определять точность каждой модели и рассчитывать их комбинационные веса для отдельных временных периодов в рамках всего тестового периода (например, за последний месяц, за последние 2 месяца, за последние 3 месяца, ..., за последние  $k$  месяцев). Эти веса, в свою очередь, автор называет «рекурсивными комбинационными весами». При этом весь предлагаемый подход краткосрочного предсказания динамики зависимой переменной автором был назван **селективным комбинированием прогнозов**.

Одним из примеров успешной реализации данной техники является система селективно-комбинированного прогноза инфляции, или SSCIF – the System of Selective Combined Inflation Forecast, которая применяется в Национальном Банке Казахстана (далее – НБК). Так, в условиях реализации политики инфляционного таргетирования, к которой НБК перешел в августе 2015 года, одну из первостепенных ролей играет моделирование инфляции, а также повышение точности прогнозов, полученных на основании данных модельных техник. В этой связи, SSCIF используется НБК в качестве основного инструмента краткосрочного прогнозирования инфляционных процессов в Казахстане с октября 2016 года. Данная система стала результатом эволюции множества предыдущих модельных техник предсказания инфляции, которые разрабатывались и применялись НБК с момента перехода на режим инфляционного таргетирования. В настоящее время SSCIF включает четыре класса моделей (авторегрессия с линейным трендом, случайное блуждание, множественная линейная факторная регрессия и байесовская векторная авторегрессия), 78 переменных и более 200 уравнений. Также SSCIF интегрирована в систему анализа и прогнозирования для принятия решения по денежно-кредитной политике<sup>2</sup> НБК и применяется как в период проведения прогнозного раунда в блоке краткосрочного прогнозирования инфляции, так и в межпрогнозные периоды. Стоит отметить, что одним из немаловажных преимуществ SSCIF является возможность построения сценарных прогнозов инфляции, исходя из предположений динамики различных экзогенных факторов, как цена на нефть, номинальный обменный курс тенге, объем денежной массы и др.

Таким образом, в данной работе описаны и раскрыты технические аспекты построения системы селективно-комбинированного прогноза инфляции (SSCIF) в Казахстане, детально рассмотрены классы моделей SSCIF, а также представлены оценки и сравнение точности вне-выборочных (out-of-sample) прогнозов инфляции на периоде до 9 месяцев отдельно по каждому модельному подходу SSCIF и по всей системе.

Дальнейшая структура работы представлена обзором литературы, рассмотрением методологии построения и применения SSCIF, описанием исходных данных, применяемых при оценке моделей SSCIF, обсуждением результатов вне-выборочного прогнозирования инфляции, заключением и списком использованных источников.

## 2. Обзор литературы

Первый научный подход к изучению и анализу комбинирования прогнозов был предложен Bates и Granger в работе 1969 года, где ученые продемонстрировали возможность объединения в один сводный прогноз двух отдельных наборов прогнозов. При этом для прогноза в качестве

---

<sup>2</sup> FPAS – forecast and policy analysis system



целевого показателя использовались данные по пассажирским авиаперевозкам. В качестве исходных методов получения прогнозов применялись методология Бокса-Дженкинса и экспоненциальное сглаживание Брауна. Для определения весов данных подходов прогнозирования авторы первыми предложили использовать прошлые ошибки каждого из исходных прогнозов, после чего рассчитанные веса применялись для составления комбинированного прогноза. Основным выводом Bates и Granger является то, что комбинированный прогноз может иметь более низкую среднеквадратическую ошибку, чем каждый из исходных одиночных прогнозов. Несмотря на свою наивность и простоту, данная работа является ценной с точки зрения научно обоснованной иллюстрации преимущества точности комбинированного прогноза над одиночными исходными прогнозами. Помимо этого, в работе представлено сопоставление двух различных подходов к прогнозированию – параметрического прогнозирования на основании методологии Бокса-Дженкинса и непараметрического прогнозирования согласно экспоненциальному сглаживанию Брауна. Тем самым, авторы демонстрируют еще одно преимущество применения комбинации прогнозов. Данное преимущество можно определить, как возможность учитывать и сочетать в одном итоговом прогнозе различные формы связи переменных и информации, которые не могут быть включены и определены в единственном из исходных прогнозов. Дополнительно авторы продемонстрировали, что верный подбор метода оценки прошлой ошибки исходных прогнозов для расчета весов при комбинировании также может влиять на точность итогового сводного прогноза. Впоследствии подход комбинирования, предложенный Bates и Granger, стал называться **методом оптимального взвешивания**.

Немаловажное место в изучении метода комбинирования прогнозов занимает исследование Granger и Ramanathan (1984). В своей работе авторы предлагают несколько способов улучшения точности и качества прогнозов, полученных методом комбинирования, на примере прогнозирования квартальных цен на свинину. В качестве передовой и наилучшей идеи повышения точности комбинированных прогнозов среди прочих предложенных подходов демонстрируется регрессионный способ определения весов комбинации исходных прогнозных техник, основанный на неограниченном методе наименьших квадратов (unrestricted least square). При этом в данной регрессии в качестве зависимой переменной используются фактические данные прогнозируемого показателя за определенный период времени, в роли же объясняющих факторов выступают вне-выборочные прогнозы этого показателя за тот же период времени, рассчитанный по всем имеющимся методам. В дальнейшем коэффициенты при объясняющих факторах данной регрессии служат весами соответствующих прогнозных техник для комбинирования исходных прогнозов. Примечательным является то, что авторы делают вывод о необходимости включения в данную

регрессию постоянного компонента, что улучшает итоговый результат комбинации прогнозов. Также согласно выводам авторов, **регрессионный подход определения весов** и дальнейшего взвешивания прогнозов для целей их комбинирования дает возможность получать более точные прогнозы по сравнению с **методом оптимального взвешивания** (Bates, Granger, 1969). Стоит отметить, что, по нашему мнению, использование регрессионного метода взвешивания имеет ограничения в виде возможной смещенности оценок весов, так как входящие в регрессию объясняющие переменные в виде исходных прогнозов анализируемого показателя в большинстве случаев могут быть коррелированными между собой. Это, в свою очередь, приводит к возникновению мультиколлинеарности оцененной регрессии.

В настоящее время в теоретических и прикладных эконометрических исследованиях комбинированного прогнозирования широко освещаются и цитируются работы американского экономиста Timmermann. Так, в одном из ранних совместных исследований Timmermann и Aofi (2006) на примере макроэкономических данных стран G7 продемонстрировали методику построения устойчивых (persistence) по точности прогнозов на основе **стратегии условного комбинирования** линейных и нелинейных моделей, которые предварительно кластеризуются, исходя из качества экспериментальных вне-выборочных прогнозов. При этом в работе представлено, что одни и те же модели могут обладать разной степенью точности на разных оцениваемых выборках, а также доказано, что точность условно комбинированных прогнозов кластеризованных моделей всегда превышает точность как отдельных прогнозов каждой из этих моделей, так и результат их обычного усреднения. Данная работа обладает подробным математическим описанием стратегии условного комбинирования прогнозов и анализа полученных результатов, в связи с чем по настоящее время применяется многими учеными-экономистами в качестве отправной точкой тщательного изучения и исследования вопросов комбинирования прогнозов.

Значимая часть прикладных работ по комбинированию прогнозов представлена разработками и исследованиями различных центральных банков. В свою очередь, это связано с тем, что различные методы комбинирования модельных прогнозов дают возможность более точно и качественно предсказывать такие целевые показатели центральных банков, как инфляция и другие макроэкономические переменные.

Так, в совместной работе Akdogan и других авторов (2012) реализованы комбинации прогнозов инфляции в Турции на основании широкого набора моделей:

- одномерные модельные уравнения (UM – unconditional mean, RW – random walk, AR – autoregressive, SARIMA – seasonal autoregressive integrated moving average);

- нелинейные методы моделирования (TAR – transition autoregressive, STAR – smooth transition autoregressive, ESTAR – exponential smooth transition autoregressive);

- непараметрические модели, основанные на фильтрациях по методу Hodrick-Prescott, Baxter-King, Christiano-Fitzgerald, а также модель UC – unobserved component (фильтр Kalman);
- модифицированные структурные модели с изменяющимися во времени коэффициентами на базе уравнения кривой Филлипса;
- модели VAR – vector autoregressive, BVAR – Bayesian vector autoregressive;
- динамические факторные модели.

При этом комбинирование прогнозов данных моделей осуществляется на основании двух подходов: **Performance Based (PB)** и **Recent Best Forecaster (RBF)**. Объединяет данные два подхода то, что оба являются модификацией метода **оптимального взвешивания**, предложенного Bates и Granger (1969). Различия в том, что метод комбинирования PB, предложенный Stock и Watson (2004), предполагает, что для расчета весов каждой модели присваивается «дисконтный» показатель от 0 до 1. Так, модели с более точным псевдо-вне-выборочным прогнозом получают дисконты близкие к 1, в свою очередь, более слабым по точности псевдо-вне-выборочных прогнозов моделям присваиваются дисконты близкие к 0. При этом показателем точности каждой модели выступает усредненное кумулятивное значение RMSE (среднеквадратическое отклонение) за весь период псевдо-вне-выборочного прогнозирования. Вместе с тем, по методу RBF по аналогии с PB также присваиваются разные веса моделям, исходя из их точности на тестовом периоде, но в этом случае наиболее точной модели присваивается вес 1, остальным моделям 0.

В целом, данная работа группы турецких экономистов систематизировала наиболее распространенные эконометрические методы краткосрочного прогнозирования инфляции, комбинируя их прогнозы по ранее предложенным подходам. При этом авторы не предложили модифицированные или иные альтернативные способы комбинирования прогнозов инфляции. Тем не менее, результаты исследования также подтвердили, что комбинация прогнозов имеет ряд преимуществ над одиночными прогнозами отдельных моделей.

Более продвинутые методы комбинирования прогнозов в рамках деятельности центрального банка представлены в работе Assenmacher-Wesche и Pesaran (2008). Так, авторы, применяя векторную модель коррекции ошибок, оценивают структурные долгосрочные коинтеграционные отношения между такими различными макроэкономическими показателями Швейцарии, как выпуск, инфляция, краткосрочная процентная ставка и другие. В работе показано, как различные спецификации данной модели используются для получения нескольких прогнозов основных целевых макропоказателей. Результирующим прогнозом каждого целевого показателя является комбинация индивидуальных прогнозов по отдельным спецификациям исходной модели. При этом для комбинирования прогнозов авторы применяют равные веса; веса, рассчитанные по информационным

критериям Акайке; а также веса, для расчета которых применяются подходы, приближенные к методам машинного обучения (Yang 2004; Sancetta 2006). Преимуществом данной работы является демонстрация возможностей комбинирования прогнозов, исходя только из результатов относительно «строго» структурированных макроэкономических моделей, избегая каких-либо других поведенческих модельных оценок целевых показателей. Но при этом авторы ограничились лишь оценкой одного типа структурной модели с различной спецификацией, оставив без внимания другие подходы строго теоретизированного макроэкономического моделирования.

Работа сотрудников Банка Норвегии (Bjørnland, Jore and etc., 2008) также подчеркивает необходимость применения комбинации модельных прогнозов взамен результатам единственной модели. В частности, в работе описывается система краткосрочного прогнозирования Банка Норвегии (SAM – the System for Averaging Models), включающая такие классы моделей, как:

- AutoRegressive Integrated Moving Average (ARIMA) models;
- Monthly Indicator (MI) models;
- Factor models;
- Bivariate Vector AutoRegressive (2VAR) models;
- VAR with money (VARm) models;
- Vector AutoRegressive (VAR) models;
- Bayesian VAR (BVAR models);
- Term structure models (Tstruc);
- Error correction model (EMOD);
- Dynamic stochastic general equilibrium model (DSGE).

Результаты исследования демонстрируют, что комбинированный прогноз инфляции Норвегии имеет наименьшее отклонение от факта по сравнению с прогнозами отдельных модельных техник. При этом в работе отсутствует детальное и аналитическое описание техники комбинирования прогнозов, что дает возможность полагать, что речь идет о «классическом» подходе к комбинированию.

В работе Андреева (2016), демонстрирующей технику краткосрочного прогнозирования инфляции в Банке России, представлено описание техники комбинирования прогнозов 18 субиндексов российского ИПЦ. В свою очередь, каждый из субиндексов моделируется по 6 модельным подходам (random walk, autoregressive, multiple regression, VAR, Bayesian VAR, unobserved components model). Комбинация прогнозов субиндексов ИПЦ производится по методу оптимального взвешивания (Bates, Granger, 1969). Итоговый прогноз общего ИПЦ является результатом взвешивания комбинированных прогнозов 18 субиндексов ИПЦ в соответствии с их весами в потребительской корзине. К числу преимуществ данного исследования следует отнести то, что прогнозирование инфляции осуществляется на основании моделирования и прогнозирования дезагрегированных компонент ИПЦ, что позволяет качественнее оценивать

динамику в целом всей инфляции. Вместе с тем, оценка весов при комбинировании моделей является стационарной и основывается на учете псевдо-вне-выборочных тестовых прогнозов лишь на один месяц. Применение подобного интервала псевдо-вне-выборочного прогноза для оценки весов делает метод чувствительным к случайным ошибкам индивидуальных моделей при прогнозировании более, чем на 1 месяц. Так, результаты прогнозирования по данной технике на более дальний горизонт времени могут переоценивать фактическую инфляцию.

### 3. Методология исследования и исходные данные

#### «Классическое» комбинирование прогнозов

Так называемое в данном исследовании «классическое» комбинирование прогнозов широко применяется для построения взвешенных прогнозов различных макроэкономических и финансовых показателей во многих центральных банках мира. При этом «классическая» методология комбинирования прогнозов, которая в общем виде сходится к методу оптимального взвешивания (Bates, Granger, 1969) технически состоит из нескольких основных этапов:

1. Первичная оценка имеющихся прогнозных моделей на сокращенной выборке данных (без тестового участка фактической выборки);
2. Расчет тестовых «псевдо-вне-выборочных» прогнозов по данным моделям на выбранном тестовом участке фактической выборки;
3. Оценка точности полученных «псевдо-вне-выборочных» прогнозов по оцененным моделям в сравнении с фактическими наблюдениями;
4. Расчет весов используемых моделей на основании оцененных показателей точности их «псевдо-вне-выборочных» прогнозов;
5. Итоговая переоценка имеющихся прогнозных моделей на полной выборке данных;
6. Вне-выборочный прогноз целевого показателя по всем имеющимся оцененным прогнозным техникам;
7. Получение комбинированного прогноза в результате взвешивания итоговых вне-выборочных прогнозов по весам, рассчитанным на четвертом этапе.

Для того чтобы математически описать данный алгоритм действий по «классическому» комбинированию прогнозов (**применительно к данным типа временных рядов**), необходимо ввести несколько обозначений:

$T_0$  – точка времени, от которого осуществляется вне-выборочное прогнозирование;

$H$  – количество временных периодов вперед, на которые необходимо осуществить вне-выборочное прогнозирование ( $H \geq 0$ );

$T_p = (T_0 - H)$  – точка времени, от которого осуществляется тестирование точности прогнозных моделей;

$t_p \in \{T_p + 1, \dots, T_0\}$  – период времени, на котором осуществляется тестирование точности прогнозных моделей;

$T_k = (T_0 + H)$  – точка времени, до которого осуществляется вневыборочное прогнозирование;

$t_k \in \{T_0 + 1, \dots, T_k\}$  – период времени, на котором осуществляется вневыборочное прогнозирование;

$m$  – количество моделей, используемых при комбинировании прогнозов.

Общий вид прогнозных моделей представим следующим образом:

$$y_t^i = f_i(x_t^i; \theta_t^i) + \varepsilon_t^i, \quad (1.1)$$

$$i \in \overline{1, m}; \quad (1.2)$$

$y_t^i$  –  $i$ -ая модель прогнозирования показателя  $y$ , оцененная на временной выборке  $t$ ;

$x_t^i$  – вектор объясняющих переменных для модели  $i$ , оцененной на временной выборке  $t$ ;

$\theta_t^i$  – вектор неизвестных параметров модели  $i$ , оцененной и учитывающей информацию временной выборки  $t$ ;

$\varepsilon_t^i$  – вектор ошибок для модели  $i$ , оцененной на временной выборке  $t$ .

При этом первичная оценка прогнозных моделей (1.1 – 1.2) согласно первому этапу алгоритма комбинирования прогнозов, выглядит как:

$$y_{T_p}^i = f_i(x_{T_p}^i; \theta_{T_p}^i) + \varepsilon_{T_p}^i, \quad (2.1)$$

$$i \in \overline{1, m}; \quad (2.2)$$

$y_{T_p}^i$  –  $i$ -ая модель прогнозирования показателя  $y$ , оцененная на временной выборке, ограниченной точкой  $T_p$ ;

$x_{T_p}^i$  – вектор объясняющих переменных для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_p$ ;

$\theta_{T_p}^i$  – вектор неизвестных параметров модели  $i$ , оцененной и учитывающей информацию временной выборки, ограниченной точкой  $T_p$ ;

$\varepsilon_{T_p}^i$  – вектор ошибок для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_p$ .

Отметим, что неизвестные параметры  $\theta_{T_p}^i$  оценены лишь на сокращенной выборке данных, следовательно, они не учитывают информацию о процессах, которые произошли на тестовом участке фактической выборки. При этом для случая «классического» комбинирования прогнозов первичную оценку прогнозных моделей вида (2.1 – 2.2) в рамках данной работы условно назовем *стационарной первичной оценкой*.

Далее в соответствии с алгоритмом разрабатываются тестовые «псевдо-вне-выборочные» прогнозы зависимой переменной и производится расчет точности полученных прогнозов:

$$\sigma_{i,H} = F(\hat{y}_{t_p}^i; y_{t_p}), \quad (3.1)$$

$$i \in \overline{1, m}, \quad (3.2)$$

$$t_p \in \{T_p + 1, \dots, T_0\}, \quad (3.3)$$

$$F \in \{RMSE, MAE, MAPE\}; \quad (3.4)$$

$\sigma_{i,H}$  – показатель точности стационарных «псевдо-вне-выборочных» прогнозов модели  $i$  на тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ ;

$F$  – показатель оценки точности «псевдо-вне-выборочных» прогнозов модели  $i$  на стационарном тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ ; включающий показатели RMSE (среднеквадратическое отклонение), MAE (среднее абсолютное отклонение), MAPE (среднее относительное отклонение);

$\hat{y}_{t_p}^i$  – тестовые «псевдо-вне-выборочные» прогнозы зависимой переменной по модели  $i$  на тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ ;

$y_{t_p}$  – фактические значения зависимой переменной на тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ .

В общем виде методы оценки «псевдо-вне-выборочных» прогнозов  $F$  представляют собой способы измерения вариации между тестовыми прогнозами и фактическими данными тестового участка выборки. В этой связи, более точным среди всех моделей будет прогноз, имеющий наименьшее значение показателя  $\sigma_{i,H}$ .

Исходя из оценок точности «псевдо-вне-выборочных прогнозов», производится расчет весов каждой модели:

$$w_{i,H} = \frac{\sigma_{i,H}^{-1}}{\sum_{i=1}^m \sigma_{i,H}^{-1}}, \quad (4.1)$$

$$\sum_{i=1}^m w_{i,H} = 1, \quad (4.2)$$

$$i \in \overline{1, m}; \quad (4.3)$$

$w_{i,H}$  – веса прогнозов модели  $i$ , или вес каждой модели;  
 $\sigma_{i,H}^{-1}$  – обратно пропорциональное значение показателя точности стационарных «псевдо-вне-выборочных» прогнозов модели  $i$  на тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ .

В результате итоговой переоценки моделей, производимой после расчета их весов, необходимых для дальнейшего комбинирования, они принимают следующий вид:

$$y_{T_0}^i = f_i(x_{T_0}^i; \theta_{T_0}^i) + \varepsilon_{T_0}^i, \quad (5.1)$$

$$i \in \overline{1, m}; \quad (5.2)$$

$y_{T_0}^i$  –  $i$ -ая модель прогнозирования показателя  $y$ , оцененная на временной выборке, ограниченной точкой  $T_0$  (итоговые оценки моделей  $i$ );

$x_{T_0}^i$  – вектор объясняющих переменных для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_0$ ;

$\theta_{T_0}^i$  – вектор неизвестных параметров модели  $i$ , оцененной и учитывающей информацию временной выборки, ограниченной точкой  $T_0$ ;

$\varepsilon_{T_0}^i$  – вектор ошибок для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_0$ .

В этом случае оценки неизвестных параметров  $\theta_{T_0}^i$  учитывают больший объем информации, включая тот, который был представлен на тестовом участке фактической выборки данных и не учитывался при оценке неизвестных параметров  $\theta_{T_p}^i$ .

Результаты прогноза зависимой переменной на основании итоговых переоцененных моделей (5.1 – 5.2) взвешиваются по соответствующим стационарным весам  $w_{i,H}$ :

$$\hat{Y}_w = \sum_{i=1}^m \hat{y}_{t_k}^i * w_{i,H}, \quad (6.1)$$

$$\hat{Y}_w = \begin{bmatrix} \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,H} \\ \vdots \\ \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,H} \end{bmatrix}; \quad (6.2)$$



$\hat{Y}_w$  – вектор взвешенных по стационарным весам прогнозов зависимой переменной по моделям  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ ;

$\hat{y}_{t_k}^i$  – итоговые вне-выборочные прогнозы по моделям  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ ;

$w_{i,H}$  – стационарные веса прогнозов модели  $i$ .

При этом итоговый комбинированный прогноз зависимой переменной соответствует взвешенным прогнозам всех имеющихся  $m$  моделей:

$$\hat{Y}_{combined} = \hat{Y}_w = \begin{bmatrix} \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,H} \\ \vdots \\ \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,H} \end{bmatrix}; \quad (7)$$

$\hat{Y}_{combined}$  – вектор, содержащий комбинированный прогноз зависимой переменной, или комбинация итоговых вне-выборочных прогнозов по всем моделям  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ .

Таким образом, результатом «классического» комбинирования прогнозов являются итоговые взвешенные прогнозы всех моделей в виде вектора  $\hat{Y}_{combined}$  (7). Ввиду того, что моделям с наименьшей вариацией точности прогнозов придается меньший вес, а более точным прогнозным техниками присваивается больший вес, итоговый комбинированный прогноз будет иметь точность прогноза выше среднего качества.

При всех преимуществах комбинированного прогноза над одиночными прогнозами моделей-чемпионов также существует один недостаток данного «классического» подхода к комбинированию. Как видно из значений вектора  $\hat{Y}_{combined}$  (7), взвешивание псевдо-вне-выборочных прогнозов по каждой модели  $i$  на каждый момент времени периода  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$  производится по весам  $w_{i,H}$ . Необходимо указать, что рассчитанные веса моделей  $w_{i,H}$  являются **стационарными**, то есть учитывают качество прогнозов каждой модели в среднем за весь тестовый период фактической выборки данных, а не за отдельные его участки. Другими словами, стационарные веса  $w_{i,H}$  отражают **среднюю сглаженную точность** моделей  $i$  за весь тестовый период  $t_p \in \{T_p + 1, \dots, T_0\}$ .

Наряду с этим, стационарные веса  $w_{i,H}$  рассчитаны на основании моделей  $i$  с оценками параметров  $\theta_{T_p}^i$ . В свою очередь, как отмечалось выше, оценка  $\theta_{T_p}^i$  не учитывает информацию, отражающую изменения динамики экономических явлений и структурных сдвигов, которые проявлялись внутри тестового периода фактической выборки  $t_p \in \{T_p + 1, \dots, T_0\}$ .

По этим причинам «классическая» комбинация прогнозов на  $h$ -шагов вперед (или на момент времени  $T_0 + H$ ) при условии, что  $H > Q$ , будет иметь в результате сглаженные (менее точные) прогнозы на любой момент времени  $T_0 + Q$ , так как веса моделей определены, исходя из их точности на стационарном периоде с точки  $T_0 - H$  до  $T_0$ , а не на периоде с точки  $T_0 - Q$  до  $T_0$ . Другими словами, для комбинирования вне-выборочных прогнозов на момент времени  $T_0 + Q$ , который входит в прогнозный период  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ , требуется использовать индивидуальный вес  $w_{i,Q}$ , который рассчитан на оценках неизвестных параметров  $\theta_t^i$  моделей  $i$  с учетом большего объема информации об экономических явлениях и структурных сдвигах, произошедших близко к точке  $T_0$  – моменту начала осуществления вне-выборочных прогнозов.

Автором в рамках данной работы предложен подход для развития «классического» метода комбинирования прогнозов, позволяющий решить вышеназванную проблему стационарных весов и получать более точные взвешенные вне-выборочные прогнозы. Суть предлагаемого подхода заключается в расчете **рекурсивных весов** моделей взамен **стационарным весам** и в соответствующей селекции комбинированных прогнозов из широкого набора всех возможных комбинаций итоговых вне-выборочных прогнозов. Данный подход сочетания прогнозов автором назван селективным комбинированием прогнозов.

### Селективно-комбинированное прогнозирование

Главным отличием селективного комбинирования прогнозов от «классического» является то, что на первом этапе алгоритма комбинирования производятся **первичные рекурсивные оценки** взамен **стационарным** (1.1 – 1.2):

$$y_{T_p+H-h}^i = f_i(x_{T_p+H-h}^i; \theta_{T_p+H-h}^i) + \varepsilon_{T_p+H-h}^i, \quad (8.1)$$

$$i \in \overline{1, m}, \quad (8.2)$$

$$h \in \overline{1, H}; \quad (8.3)$$

$y_{T_p+H-h}^i$  –  $i$ -ая модель прогнозирования показателя  $y$ , оцененная на временной выборке, ограниченной точкой  $T_p + H - h$  (рекурсивные тестовые оценки моделей  $i$ );

$x_{T_p+H-h}^i$  – вектор объясняющих переменных для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_p + H - h$ ;

$\theta_{T_p+H-h}^i$  – вектор неизвестных параметров модели  $i$ , оцененной и учитывающей информацию временной выборки, ограниченной точкой  $T_p + H - h$ ;

$\varepsilon_{T_p+H-h}^i$  – вектор ошибок для модели  $i$ , оцененной на временной выборке, ограниченной точкой  $T_p + H - h$ .

В данном случае рекурсивные первичные оценки прогнозных моделей производятся  $h$  раз от 1 до  $H$  на каждый момент времени в рамках тестового периода  $t_p \in \{T_p + 1, \dots, T_p + H\}$ .

В этой связи, дальнейшая оценка точности тестовых прогнозов моделей производится также **рекурсивно**  $h$  раз от 1 до  $H$  на каждый момент времени в рамках периода  $t_p \in \{T_p + 1, \dots, T_p + H\}$ :

$$\sigma_{i,h} = F(\hat{y}_{t_p}^i; y_{t_p}), \quad (9.1)$$

$$i \in \overline{1, m}, \quad (9.2)$$

$$t_p \in \{T_p + H - h + 1, \dots, T_0\}, \quad (9.3)$$

$$h \in \overline{1, H}, \quad (9.4)$$

$$F \in \{RMSE, MAE, MAPE\}; \quad (9.5)$$

$\sigma_{i,h}$  – показатель точности рекурсивных «псевдо-вне-выборочных» прогнозов модели  $i$  на тестовом периоде  $t_p \in \{T_p + H - h + 1, \dots, T_0\}$  при  $h \in \overline{1, H}$ ;

$F$  – метод оценки точности прогнозов «псевдо-вне-выборочных» прогнозов модели  $i$  на рекурсивном тестовом периоде  $t_p \in \{T_p + H - h + 1, \dots, T_0\}$  при  $h \in \overline{1, H}$ , включающий показатели RMSE (среднеквадратическое отклонение), MAE (среднее абсолютное отклонение), MAPE (среднее относительное отклонение);

$\hat{y}_{t_p}^i$  – тестовые «псевдо-вне-выборочные» прогнозы зависимой переменной по модели  $i$  на тестовом периоде  $t_p \in \{T_p + H - h + 1, \dots, T_0\}$  при  $h \in \overline{1, H}$ ;

$y_{t_p}$  – фактические значения зависимой переменной на тестовом периоде  $t_p \in \{T_p + 1, \dots, T_0\}$ .

В свою очередь  $h$  показателей точности прогнозов для каждой модели  $i$  применяются для расчета соответствующих **рекурсивных весов**:

$$w_{i,h} = \frac{\sigma_{i,h}^{-1}}{\sum_{i=1}^m \sigma_{i,h}^{-1}}, \quad (10.1)$$

$$\sum_{i=1}^m w_{i,h} = 1, \quad (10.2)$$

$$i \in \overline{1, m}, \quad (10.3)$$

$$h \in \overline{1, H}; \quad (10.4)$$

$w_{i,h}$  – рекурсивные веса прогнозов модели  $i$  для  $h \in \overline{1, H}$ ;  
 $\sigma_{i,h}^{-1}$  – обратно пропорциональное значение показателя точности рекурсивных «псевдо-вне-выборочных» прогнозов модели  $i$  на тестовом периоде  $t_p \in \{T_p + H - h + 1, \dots, T_0\}$  при  $h \in \overline{1, H}$ .

Как можно заметить из системы уравнений (10.1 – 10.4), общее число рекурсивных весов  $w_{i,h}$  для каждой модели  $i$  равно максимальному значению  $h$ , то есть  $H$ , тогда как при «классическом» комбинировании для каждой модели  $i$  применяется единственный стационарный вес  $w_{i,H}$  (4.1 – 4.3).

Вследствие того, что количество рекурсивных весов  $w_{i,h}$  (при  $H > 1$ ) превышает единицу, итоговые взвешенные по этим рекурсивным весам прогнозы всех моделей  $i$  будет представлен в виде матрицы, а не вектора, как в случае «классического» комбинирования (6.1 – 6.2):

$$\hat{Y}_w = \sum_{i=1}^m \hat{y}_{t_k}^i * w_{i,h}, \quad (11.1)$$

$$h \in \overline{1, H}, \quad (11.2)$$

$$\hat{Y}_w = \begin{bmatrix} \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,1} & \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,2} & \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,3} & \dots & \sum_{i=1}^m \hat{y}_{T_0+1}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,1} & \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,2} & \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,3} & \dots & \sum_{i=1}^m \hat{y}_{T_0+2}^i * w_{i,H} \\ \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,1} & \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,2} & \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,3} & \dots & \sum_{i=1}^m \hat{y}_{T_0+3}^i * w_{i,H} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,1} & \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,2} & \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,3} & \dots & \sum_{i=1}^m \hat{y}_{T_0+H}^i * w_{i,H} \end{bmatrix}; \quad (11.3)$$

$\hat{Y}_w$  – матрица размерности  $H \times H$ , которая содержит взвешенные по рекурсивным весам прогнозы зависимой переменной по модели  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ ;

$\hat{y}_{t_k}^i$  – итоговые вне-выборочные прогнозы по модели  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ ;

$w_{i,h}$  – рекурсивные веса прогнозов модели  $i$ .

Можно заметить, что результаты «классического» комбинирования, выраженные вектором  $\hat{Y}_{combined}$  (7), соответствуют последнему столбцу  $\hat{Y}_w$  матрицы взвешенных по рекурсивным весам прогнозов моделей  $i$  (11.3). При этом **селективно-комбинированный вне-выборочный прогноз** моделей  $i$  представлен в диагональных значениях матрицы  $\hat{Y}_w$  (11.3):

$$\hat{Y}_{combined}^{selective} = diagonal[\hat{Y}_w] = \begin{bmatrix} \sum_{i=1}^m \hat{Y}_{T_0+1}^i * w_{i,1} \\ \sum_{i=1}^m \hat{Y}_{T_0+2}^i * w_{i,2} \\ \sum_{i=1}^m \hat{Y}_{T_0+3}^i * w_{i,3} \\ \vdots \\ \sum_{i=1}^m \hat{Y}_{T_0+H}^i * w_{i,H} \end{bmatrix}; \quad (12)$$

$\hat{Y}_{combined}^{selective}$  – вектор, содержащий селективно-комбинированный прогноз зависимой переменной, или селективная комбинация итоговых вне-выборочных прогнозов по модели  $i$  на периоде  $t_k \in \{T_0 + 1, \dots, T_0 + H\}$ ;

$diagonal[\hat{Y}_w]$  – диагональные значения матрицы  $\hat{Y}_w$ .

При сравнении векторов  $\hat{Y}_{combined}(7)$  и  $\hat{Y}_{combined}^{selective}$  (12) заметно, что в случае селективного комбинирования для каждого прогноза модели  $i$  на  $T_0 + h$  ( $h \in \overline{1, H}$ ) шагов вперед соответствует индивидуальный рекурсивный вес этой модели  $w_{i,h}$  ( $h \in \overline{1, H}$ ), тогда как для «классического» комбинирования используется единственный стационарный вес  $w_{i,h}$  ( $h = H$ ).

Как показывают эмпирические расчеты, представленные в данной работе, подобное различие между двумя подходами к комбинированию прогнозов дает большее преимущество в пользу селективно-комбинированного прогноза. Данное преимущество, как было сказано выше, отражается в возможности учета при вне-выборочном прогнозировании тех изменений в динамике экономических процессов, в том числе структурных шоков, которые произошли на тестовом участке фактической выборки данных.

### **Система селективно-комбинированного прогноза инфляции для Казахстана.**

Система селективно-комбинированного прогноза инфляции, или SSCIF (the System of Selective Combined Inflation Forecast), является практической и эмпирической реализацией метода селективного комбинирования прогнозов, описанной выше, применительно к данным по инфляции. В настоящее время данная система используется в НБК для целей краткосрочного прогнозирования инфляции в Казахстане в рамках системы FPAS.

SSCIF, применяемая в НБК, моделирует динамику инфляции Казахстана, выраженную показателем индекса потребительских цен (далее - ИПЦ). При этом ИПЦ моделируется и прогнозируется, исходя из динамики своих дезагрегированных показателей согласно Классификатору индивидуального потребления по целям. Другими словами, в SSCIF в качестве зависимых переменных моделей используются 66 субиндексов цен на группы и разделы товаров и услуг, которые суммарно составляют ИПЦ в Казахстане (таблица 1 Приложения). При этом в качестве экзогенных факторов, которые отчасти объясняют динамику данных 66 субиндексов цен

на потребительские товары и услуги, представлены показатели обменных курсов тенге к иностранным валютам, инфляции в России, цен в отрасли промышленности Казахстана, тенговой денежной массы и реальных денежных доходов в Казахстане, а также мировых цен пшеницу и нефть (Таблица 1 Приложения). Подобный подход моделирования общего агрегированного макропоказателя (ИПЦ), основанный на моделировании отдельных его дезагрегированных составляющих частей (субиндексы ИПЦ), позволяет более детализировано и точно определить чувствительность целевого показателя к изменению динамики объясняющих факторов.

Все зависимые и объясняющие переменные моделей в составе SSCIF являются месячными сезонно-сглаженными показателями и используются в виде разницы логарифмов, что позволяет добиться стационарности всех временных рядов при 5%-ом уровне значимости. При этом в моделях применяется выборка наблюдений всех зависимых и объясняющих переменных, начиная с января 2005 года.

Таким образом, в рамках SSCIF производится селективно-комбинированное прогнозирование 66 субиндексов цен потребительских товаров и услуги в Казахстане. По результатам селективного комбинирования прогноза целевых 66 показателей производится их взвешивание согласно собственным весам в потребительской корзине Казахстана. Итоговый взвешенный селективно-комбинированный прогноз является прогнозом общего ИПЦ, отражающим будущую динамику инфляции на горизонте прогноза.

Для прогнозирования каждого из 66 субиндексов ИПЦ Казахстана в рамках SSCIF применяется четыре класса моделей безусловного и условного прогнозирования. К числу моделей безусловного прогнозирования относятся *модель авторегрессии первого порядка с линейным трендом (наивное прогнозирование)* и *модель случайного блуждания (стохастическое прогнозирование)*. К подходам условного прогнозирования относятся *многофакторная регрессионная модель* и *байесовская векторная авторегрессионная модель*. При оценке точности каждого модельного прогноза на тестовом участке фактической выборки с целью последующего расчета весов каждой модели в SSCIF используется показатель среднеквадратического отклонения – RMSE.

## Описание моделей, применяемых в SSCIF

### *Модели авторегрессии первого порядка с линейным трендом (LTAR)*

$$\Delta y_t^k = \beta_0 + \beta_1 \Delta y_{t-1}^k + \beta_1 trend^k + \varepsilon_t^k, \quad (13.1)$$

$$k \in \overline{1, 66} \quad (13.2)$$

$\Delta y_t^k$  – прирост логарифма субиндекса ИПЦ  $k$ ;  
 $\Delta y_{t-1}^k$  – прирост логарифма субиндекса инфляции  $k$  в предыдущем месяце;  
 $trend^k$  – трендовый компонент субиндекса инфляции  $k$ ;  
 $\varepsilon_t^k$  – случайные ошибки, характерные для уравнения субиндекса инфляции  $k$ .

Всего оценивается 66 моделей данного типа (13.1). С экономической точки зрения использование данного класса моделей в SSCIF объясняется тем, что модели линейного тренда с авторегрессией учитывают инерционность цен на товары и услуги, тем самым позволяя оценить величину «эффекта залипания цен» или «эффекта меню» на цены каждой отдельной товарной и сервисной группы и раздела, суммарно составляющие всю корзину ИПЦ. Также представленный модельный класс является классическим подходом оценки инфляционных ожиданий при моделировании динамики ИПЦ. Следует отметить, что в данном случае речь идет лишь об адаптивных инфляционных ожиданиях.

#### *Модели случайного блуждания (RW)*

$$\Delta(\Delta y_t^k)^j = \beta_0 + u_t^{kj}, \quad (14.1)$$

$$u_t^{kj} \in N(0, 1), \quad (14.2)$$

$$k \in \overline{1, 66}, \quad (14.3)$$

$$j \in \overline{1, 500}; \quad (14.4)$$

$\Delta(\Delta y_t^k)^j$  – первая разность прироста логарифма субиндекса ИПЦ  $k$  симуляции случайного блуждания порядка  $j$ ;

$u_t^{kj}$  - случайная ошибка со стандартным нормальным распределением для субиндекса инфляции  $k$  при симуляции случайного блуждания порядка  $j$ .

Как видно из записи (14.1 – 14.4) в SSCIF для моделирования динамики каждого из 66 субиндексов ИПЦ оцениваются 500 моделей случайного блуждания (всего 33 000 уравнений). Это производится для того, чтобы произвести 500 генераций случайной величины по стандартному нормальному распределению, которые будут являться объясняющими факторами в каждой модели  $j$  для динамики зависимой переменной – субиндекса  $k$ . В качестве конечного прогноза по модели случайного блуждания для субиндекса  $k$  берется среднеарифметическое значение прогнозов по 500 оцененным моделям.

В этом случае экономическая интерпретация применения моделей случайного блуждания заключается в следующем. Модели случайного

блуждания, отдаленно имитируя метод Монте-Карло, дают возможность получить асимптотическую оценку стохастических колебаний цен на товары и услуги, которые не могут улавливаться и описываться объясняющими факторами инфляции.

### *Модели многофакторных регрессий (OLS)*

$$\Delta y_t^k = \beta_0 + A(L^l)Z_t^k + \varepsilon_t^k, \quad (15.1)$$

$$k \in \overline{1, 66}, \quad (15.2)$$

$$l \in \{1, 2, \dots, 6\}; \quad (15.3)$$

$\Delta y_t^k$  – прирост логарифма субиндекса ИПЦ  $k$ ,  
 $(L^l)Z_t^k$  – вектор объясняющих факторов субиндекса инфляции  $k$  с лагом  $l$ ,  
 $\varepsilon_t^k$  – случайные ошибки, характерные для уравнения субиндекса инфляции  $k$ .

Всего оценивается 66 моделей множественных регрессий, то есть для каждого субиндекса  $k$ . Роль многофакторных регрессионных моделей в SSCIF заключается в том, что данные модели позволяют количественно оценивать степень влияния того или иного фактора инфляции на динамику цен на потребительские товары и услуги.

### *Байесовские авторегрессионные модели (BVAR)*

$$Y_t^g = X_t^g A + Z_t^g B + \varepsilon_t, \quad (16.1)$$

$$g \in \overline{1, 13}; \quad (16.2)$$

$Y_t^g$  – эндогенные факторы, вектор прироста логарифмов субиндексов ИПЦ, входящие в группу  $g$ ;  
 $X_t^g$  – матрица лаговых значений эндогенных факторов;  
 $Z_t^g$  – вектор экзогенных объясняющих факторов, характерных для группы  $g$ ;  
 $\varepsilon_t$  – вектор случайных ошибок.

Для оценки моделей BVAR (16.1 – 16.2) 66 субиндексов ИПЦ были экспертным путем разделены на 13 групп. Предполагается, что в общую группу включены цены на товары и услуги, которые могут иметь определенное взаимовлияние друг на друга, проявляемое микроэкономическими свойствами субститутности и комплиментарности рыночных товаров и услуг. Таким образом, всего в рамках SSCIF



оценивается 13 моделей BVAR на основе априорного распределения «Litterman/Minnesota».

Преимущества и причины использования BVAR-моделей в SSCIF связаны с тем, что данные классы моделей наряду с оценкой влияния экзогенных факторов инфляции на динамику потребительских цен позволяют моделировать лагированное взаимовлияние цен на товары и услуги в рамках условной группы  $g$ , что при прогнозировании дает возможность учитывать вышеупомянутые микроэкономические свойства субститутности и комплиментарности товаров и услуг.

#### 4. Обсуждение эмпирических результатов

Согласно результатам прикладного применения SSCIF в рамках данного исследования были реализованы «классический» и селективный подход к комбинации псевдо-вне-выборочных прогнозов инфляции Казахстана на 9 месяцев вперед на четырех различных тестовых периодах времени.

Как показали расчеты, точность псевдо вне-выборочных прогнозов инфляции Казахстана на тестовых периодах ни по одному из четырех классов моделей, имплементированных в SSCIF, не демонстрируют однозначного преимущества над другими моделями в робастности. Как показано на рисунке 1 и 2 Приложения, на различных тестовых периодах преобладает точность разных модельных подходов прогнозирования. Так, модель, определенная на текущем тестовом периоде как «чемпион», на другом периоде прогнозирования может оказаться «аутсайдером». Это связано с тем, что исторические наблюдения множества финансовых и макроэкономических показателей, в частности показателя инфляции, могут отражать экономические процессы, не содержащие ярко выраженной и четкой закономерной зависимости лишь от конкретного набора предикторов, которые были включены в выбранную модель-чемпион. Другими словами, наличие структурных сдвигов и шоковых явлений в экономических процессах демонстрирует, что количественная оценка степени зависимости целевого показателя от определенного набора объясняющих переменных, которая может быть максимально точно описана на историческом периоде и соответствовать всем статическим тестам, не гарантирует сохранения степени выявленных связей на горизонте вне-выборочного прогноза.

При этом результат «классического» комбинирования данных прогнозов на всех участках тестирования демонстрирует стабильную точность выше среднего, а в некоторых случаях обладает максимальной точностью прогнозов по сравнению с одиночными моделями. Таким образом, на основе прогнозирования инфляции в Казахстане было определено, что «классический» подход в комбинировании различных модельных техник

значительно улучшает качество итоговых прогнозов, что соответствует результатам других зарубежных исследований.

Вместе с тем, результаты селективного комбинирования псевдо-вне-выборочных прогнозов инфляции Казахстана на тестовых участках, как показано на рисунке 3 и 4 Приложения, продемонстрировали большую точность по сравнению с «классическим» комбинированием. Данные рисунка 4 отражают, что в большинстве случаев точность селективного комбинирования прогнозов оказываются выше точности «классического» комбинирования на протяжении всего прогнозного периода. При этом на более близких к исходной прогнозной точке периодах точность селективного комбинирования всегда превышает точность «классического» комбинирования.

Таким образом, применение SSCIF в НБК для краткосрочного прогнозирования инфляции является вполне оправданным с точки зрения получения более точных и робастных прогнозных результатов. Помимо эмпирических экспериментальных псевдо-вне-выборочных прогнозов в пользу этого указывают рабочие прогнозы инфляции с октября 2016 года, показанные на рисунке 5 Приложения.

## 5. Заключение

В работе комплексно описана процедура «классического» комбинирования прогнозов, которая в настоящее время является одним из передовых подходов прогнозирования временных рядов и перекрестных данных. Помимо этого, в связи с выявленными недостатками данного метода построения прогнозов исходных данных, автор предложил расширить «классическое» комбинирования до селективного комбинирования прогнозов. При этом в работе также представлено детальное теоретическое описание предлагаемого метода селективного комбинирования прогнозов.

Примером практического использования разработанного автором метода селективного комбинирования прогнозов является система селективно-комбинированного прогноза инфляции, или the System of Selective Combined Inflation Forecast (SSCIF), которая применяется в Национальном Банке Казахстана для получения сценарных краткосрочных прогнозов динамики потребительских цен в Казахстане. Автором в рамках данной работы представлено детальное описание SSCIF, включая описание применяемых классов моделей, исходных данных и других технических аспектов. Помимо этого, на примере экспериментальных расчетов согласно SSCIF продемонстрированы эмпирические результаты прогнозирования инфляции в Казахстане отдельно по одиночным классам моделей, также по итогам «классического» и селективного комбинирования. Так, на основании полученных результатов проиллюстрировано преимущество селективного

комбинирования в части получения более точных и робастных вне-выборочных прогнозов.

Потенциал дальнейшего развития SSCIF связан с улучшением статистических и прогнозных качеств исходных моделей предсказания субиндексов ИПЦ, а также включением в число применяемых прогнозных техник других классов и семейств моделей, оценка которых является оправданной и возможной в текущих условиях.

Также стоит отметить, что потенциал применения в целом метода селективного комбинирования прогнозов не ограничен его использованием только для предсказания динамики ИПЦ. Так, предлагаемый метод селективного комбинирования может быть применен для прогноза широкого набора, как макроэкономических и финансовых показателей различных стран, так и для предсказания других неэкономических показателей.

**Список использованных источников**

Akdogan K., Baser S., Chadwick M.G., Ertug D., Hulagu T., Kosem S., Ogunc F., Ozmen, M.U., Tekatli N. (2012). Short term inflation forecasting models for Turkey and a forecast combination analysis. Central Bank of the Republic of Turkey, Working paper N12/09.

Assenmacher-Wesche K., Pesaran M. (2008). Forecasting the Swiss economy using VECX models: An exercise in forecast combination across models and observation windows. *National Institute Economic Review*. 203(1). 91-108.

Bates J., Granger C. (1969). The combination of forecasts. *Operations Research Quarterly*. 20. pp. 451-468.

Bjornland H.C., Jore A.S., Smith C., Thorsrud L.A. (2008) Improving and evaluating short term forecasts at the Norges Bank. *Norges Bank Staff Memo*. 2008 N 4.

Granger C. and Ramanathan R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, Vol. 3, pp. 197-204.

Sancetta A. (2006). *Online Forecast Combination for Dependent Heterogeneous Data*. University of Cambridge. mimeo

Stock J. and Watson M. (2004). Combining forecasts in a seven-country data set. *Journal of Forecasting*. 23. pp. 405-430.

Timmermann A. (2006). Forecast combinations. in: G. Elliot, C.W.J. Granger and, A. Timmermann. *Handbook of Economic Forecasting*. Vol. 1.

Timmermann A., Aiolfi M. (2006). Persistence in Forecasting Performance and Conditional Combination Strategies, *Journal of Econometrics*. 135. pp. 31-53.

Yang Y. (2004). Combining Forecasting Procedures: Some Theoretical Results. *Econometric Theory* 20, 176-222.

Андреев А. (2016). Прогнозирования инфляции методом комбинирования в Банке России. Серия докладов об экономических исследованиях. Август 2016, №14

Шитиков В., Мастицкий С. (2017). Классификация, регрессия и другие алгоритмы Data Mining с использованием R.

Таблица 1. Исходные данные SSCIF

Субиндексы цен на продовольственные товары (13 субиндексов)	Субиндексы цен на непродовольственные товары (23 субиндекса)	Субиндексы цен на услуги (30 субиндексов)	Экзогенные переменные
Хлебобулочные изделия и крупы	Материалы для изготовления одежды	Чистка, ремонт и прокат одежды	Цена на нефть марки Brent
Мясо	Верхняя одежда	Ремонт и прокат обуви	Номинальный обменный курс USD/KZT
Рыба и морепродукты	Другие предметы одежды и аксессуары одежды	Фактическая арендная плата за жилье	Номинальный обменный курс RUB/KZT
Молочные изделия, сыр и яйца	Ботинки, туфли и прочая обувь	Услуги по обслуживанию и ремонту жилых помещений	Номинальный обменный курс USD/RUB
Масла и жиры	Материалы для обслуживания и ремонта жилых помещений	Водоснабжение	Продовольственная инфляция в России
Фрукты	Твердое топливо	Сбор мусора	Непродовольственная инфляция в России
Овощи	Мебель, предметы домашнего обихода, ковры и другие покрытия для пола, их ремонт	Канализация	Сервисная инфляция в России
Сахар, джем, мед, шоколад и кондитерские изделия	Текстильные изделия, используемые в домашнем хозяйстве	Прочие услуги, связанные с содержанием жилых помещений	Цены в промышленности в Казахстане
Продукты питания, не отнесенные к другим категориям	Бытовые приборы	Электроэнергия	Цены в обрабатывающей промышленности в Казахстане
Кофе, чай и какао	Стеклянные изделия, столовые приборы и домашняя утварь	Газ	Тенговая денежная масса в Казахстане
Минеральная вода, прохладительные напитки, фруктовые и овощные соки	Инструменты и приспособления, используемые в быту и садоводстве	Тепловая энергия	Реальный денежный доход в Казахстане
Алкогольные напитки	Товары и услуги, используемые для ведения домашнего хозяйства	Амбулаторные услуги	Индекс FAO Cereals
Табачные изделия	Медикаменты, лечебное оборудование и аппаратура	Услуги больниц	
	Покупка автотранспортных средств	Техническое обслуживание и ремонт личных транспортных средств	
	Запасные части и принадлежности для личных транспортных средств	Прочие услуги, связанные с личными транспортными средствами	
	Горюче-смазочные материалы для личных транспортных средств	Услуги транспорта	
	Аудиовизуальное оборудование и фотоаппаратура, оборудование для обработки информации	Связь	
	Другие крупные товары длительного пользования для организации отдыха и культурных мероприятий	Услуги в области отдыха, развлечений и культуры	
	Другие товары и оборудование для отдыха, спорта, садоводства и домашние животные	Организация комплексного отдыха	
	Газеты, книги и канцелярские товары	Дошкольное и начальное образование	
	Электрические приборы личного пользования	Среднее образование	
	Прочие предметы, приборы и товары личного пользования	Продолженное среднее образование	
	Предметы личного ухода, не отнесенные к другим категориям	Высшее образование	
		Образование, не подразделенное по ступеням	
		Услуги общественного питания	
		Гостиничное обслуживание	
		Услуги парикмахерских и заведений личного обслуживания	
		Страхование	
		Финансовые услуги, не отнесенные к другим категориям	
		Прочие услуги, не отнесенные к другим категориям	

Источник: составлено автором по данным Комитета по статистике Министерства национальной экономики Казахстана, Национального Банка Казахстана, Федеральной службы по государственной статистике России, терминал Bloomberg

Рисунок 1. Псевдо-вне-выборочные прогнозы инфляции (ИПЦ в % м/м) в Казахстане на различных тестовых участках фактической выборки по SSCIF

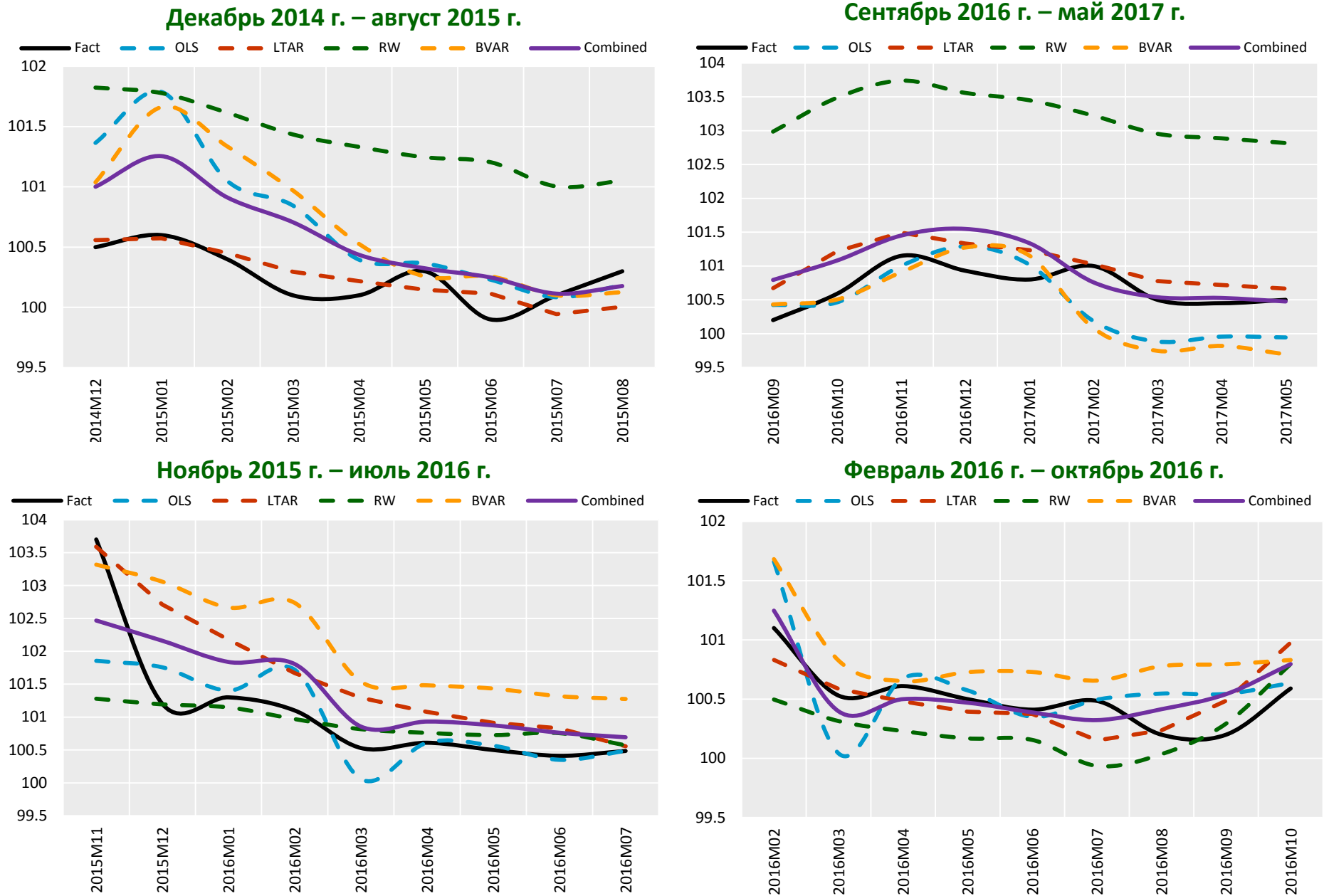


Рисунок 2. Количественная оценка точности псевдо-вне-выборочных прогнозов инфляции (ИПЦ в % м/м) в Казахстане на различных тестовых участках фактической выборки по SSCIF

**Декабрь 2014 г. – август 2015 г.**

	OLS	LTAR	RW	BVAR	Combined
MAPE	0.47	0.14	1.12	0.48	0.34
MAE	0.61	0.16	1.15	0.61	0.42
RMSE	0.47	0.14	1.13	0.49	0.35

**Сентябрь 2016 г. – май 2017 г.**

	OLS	LTAR	RW	BVAR	Combined
MAPE	0.39	0.33	2.47	0.48	0.32
MAE	0.45	0.37	2.56	0.56	0.40
RMSE	0.40	0.34	2.55	0.49	0.32

**Ноябрь 2015 г. – июль 2016 г.**

	OLS	LTAR	RW	BVAR	Combined
MAPE	0.41	0.57	0.42	1.06	0.55
MAE	0.69	0.71	0.83	1.17	0.65
RMSE	0.41	0.58	0.42	1.08	0.56

**Февраль 2016 г. – октябрь 2016 г.**

	OLS	LTAR	RW	BVAR	Combined
MAPE	0.22	0.18	0.31	0.34	0.15
MAE	0.30	0.22	0.35	0.39	0.18
RMSE	0.22	0.18	0.31	0.34	0.15

Цветовая шкала точности  менее точно более точно

Рисунок 3. Сравнение «классического» и селективного комбинирования псевдо-вне-выборочных прогнозов инфляции (ИПЦ в % м/м) в Казахстане на различных тестовых участках фактической выборки по SSCIF

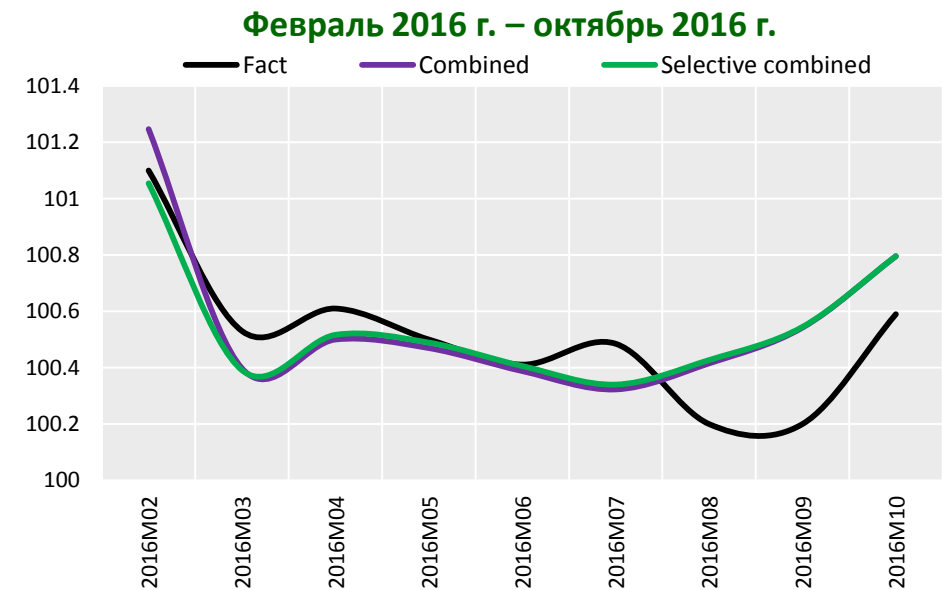
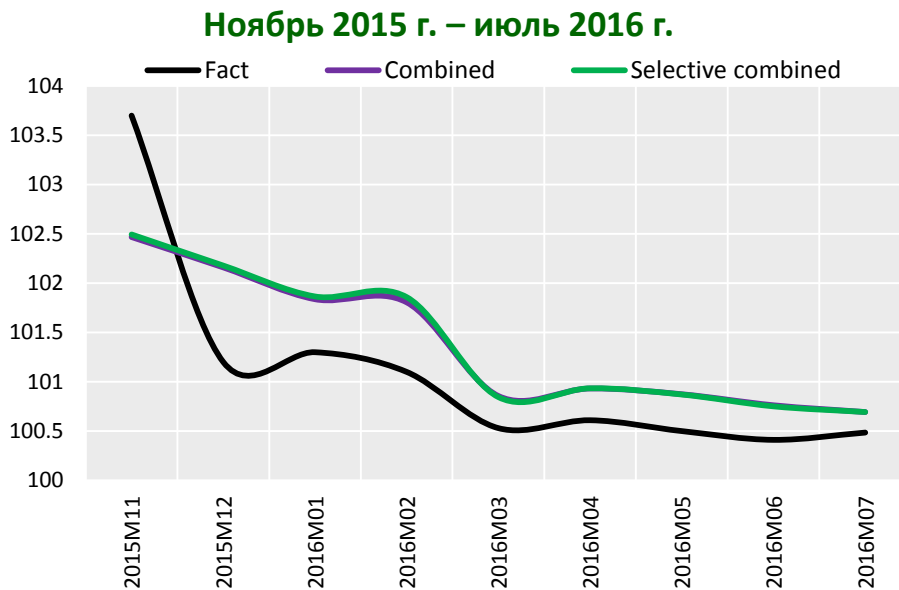
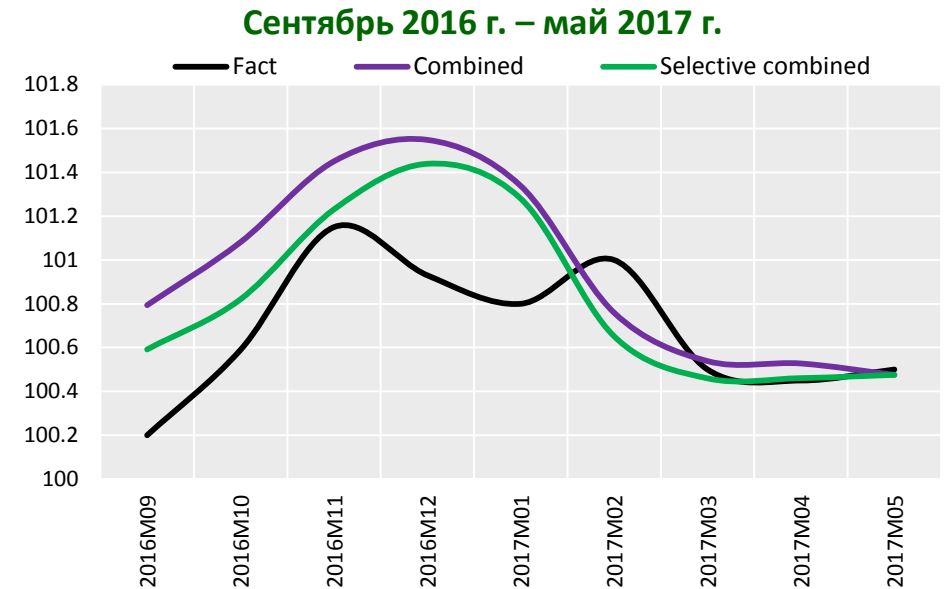
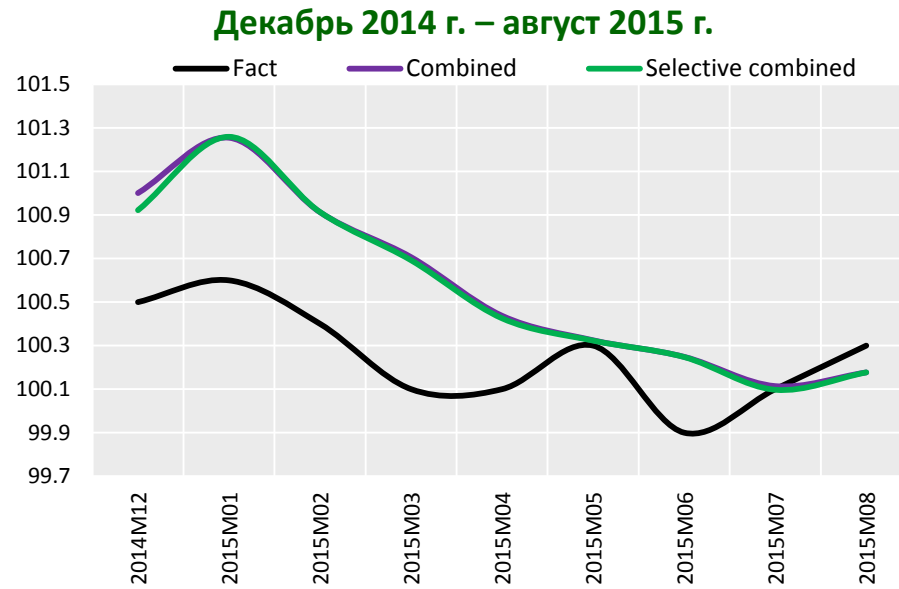




Рисунок 4. Количественная оценка точности «классического» и селективного комбинирования псевдо-вне-выборочных прогнозов инфляции (ИПЦ в % м/м) в Казахстане на различных тестовых участках фактической выборки по SSCIF

Декабрь 2014 г. – август 2015 г.

Сентябрь 2016 г. – май 2017 г.

Ноябрь 2015 г. – июль 2016 г.

Февраль 2016 г. – октябрь 2016 г.

Горизонт прогноза (месяцев)	RMSE	
	Combined	Selective combined
1	0.501	0.422
2	0.578	0.541
3	0.556	0.532
4	0.569	0.547
5	0.523	0.503
6	0.440	0.423
7	0.427	0.412
8	0.375	0.361
9	0.347	0.334

Горизонт прогноза (месяцев)	RMSE	
	Combined	Selective combined
1	0.594	0.392
2	0.543	0.311
3	0.462	0.234
4	0.501	0.303
5	0.508	0.339
6	0.464	0.340
7	0.403	0.297
8	0.362	0.262
9	0.325	0.235

Горизонт прогноза (месяцев)	RMSE	
	Combined	Selective combined
1	1.232	1.206
2	1.096	1.092
3	0.910	0.917
4	0.859	0.877
5	0.752	0.765
6	0.681	0.691
7	0.637	0.645
8	0.601	0.607
9	0.557	0.563

Горизонт прогноза (месяцев)	RMSE	
	Combined	Selective combined
1	0.147	0.045
2	0.140	0.092
3	0.130	0.093
4	0.105	0.072
5	0.089	0.059
6	0.101	0.073
7	0.117	0.095
8	0.145	0.126
9	0.152	0.135

Цветовая шкала точности  менее точно более точно

Рисунок 5. Сравнение факта и реального вне-выборочного прогноза инфляции (ИПЦ в % м/м) в Казахстане по SSCIF с октября 2016 г. по август 2017 г.

