На правах рукописи

МОИСЕЕВ НИКИТА АЛЕКСАНДРОВИЧ

РАЗВИТИЕ МЕТОДОЛОГИИ ПОВЫШЕНИЯ КАЧЕСТВА ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЙ

Специальность 08.00.13 –
Математические и инструментальные методы экономики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора экономических наук

Москва – 2020

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Российский экономический университет им. Г.В. Плеханова» на кафедре математических методов в экономике.

Научный консультант:	доктор экономических наук, профессор **Тихомиров Николай Петрович**
Официальные оппоненты:	**Суворов Николай Владимирович** доктор экономических наук, профессор, ФГБУН «Институт народнохозяйственного прогнозирования РАН», заведующий лабораторией Прогнозирования динамики и структуры народного хозяйства
	Щепина Ирина Наумовна доктор экономических наук, доцент, ФГБОУ ВО «Воронежский государственный университет», доцент кафедры информационных технологий и математических методов в экономике
	Ратнер Светлана Валерьевна доктор экономических наук, доцент, ФГБУН «Институт проблем управления им. В.А. Трапезникова РАН», ведущий научный сотрудник лаборатории экономической динамики и управления инновациями
Ведущая организация:	Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»

Защита диссертации состоится «18» июня 2020 г. в 13:00 на заседании диссертационного совета Д 212.196.15 на базе ФГБОУ ВО «Российский экономический университет им. Г.В. Плеханова» по адресу: 117997, г. Москва, Стремянный переулок, дом 36, ауд. 353.

С диссертацией и авторефератом можно ознакомиться в НИБЦ им. академика Л.И. Абалкина ФГБОУ ВО «РЭУ им. Г.В. Плеханова» по адресу: 117997, г. Москва, ул. Зацепа, д. 43 и на сайте организации: http://ords.rea.ru/

Автореферат разослан «___» ______________ 2020 г.

Ученый секретарь
диссертационного совета
Д 212.196.15, д.э.н., профессор — Сергей Владимирович Мхитарян

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Процесс развития любой экономической системы характеризуется постоянным изменением ее показателей во времени под воздействием разнообразных внутренних и внешних факторов. В масштабах страны такими показателями являются социальные и макроэкономические индикаторы, которые дают некоторую объективную оценку состояния экономики и общества в целом за определенный временной промежуток. Для систем управления общественным развитием важной проблемой является также предсказание будущей динамики рассматриваемых макроэкономических показателей на определенный срок. При этом точность получаемых прогнозов имеет ключевое значение, поскольку напрямую влияет на эффективность принимаемых решений органами государственного управления, иностранными и отечественными инвесторами, корпорациями и другими хозяйствующими субъектами.

В настоящее время основным средством прогнозирования макроэкономических индикаторов является эконометрический инструментарий, который достаточно хорошо зарекомендовал себя в точных науках. Однако временные ряды социально-экономических процессов зачастую характеризуются постоянной изменчивостью волатильности, уровней и направлений взаимосвязей между макроэкономическими индикаторами, что осложняет процесс обоснования и разработки эконометрических прогнозных моделей, в том числе в части оценки параметров модели и достоверности доверительных интервалов получаемых прогнозов.

Изменчивость вышеупомянутых характеристик временных рядов макроэкономических процессов ставит исследователя перед проблемой выбора длины окна наблюдений, поскольку представляется невозможным однозначно определить, какой именно временной промежуток отражает наиболее актуальные взаимосвязи между моделируемыми переменными для получения обоснованных и точных прогнозов.

Одной из важнейших задач при моделировании макроэкономических процессов является процедура спецификации регрессионного уравнения. Однако данная процедура не является унифицированной, что приводит к появлению множества методов проведения спецификации и критериев оптимальности модели. Таким образом, достаточно часто в рамках одного окна имеется возможность построить несколько регрессионных моделей, каждая из которых будет оптимальной согласно своему критерию, что ставит исследователя перед проблемой выбора наилучшей модели из набора альтернатив. В работах многих отечественных и зарубежных авторов было показано, что синтез набора моделей является более эффективным решением в данных условиях, чем осуществление выбора лишь одной модели из рассматриваемого набора. Однако существующие процедуры синтеза моделей имеют существенные недостатки, такие как невозможность расчета интервального прогноза, отсутствие адаптации к взвешиванию моделей, рассчитанных на окнах данных различной длины и т.д.

Также при моделировании макроэкономических процессов исследователь зачастую сталкивается со структурными сдвигами в анализируемых данных. Поскольку взаимосвязи между макроэкономическими индикаторами подвержены изменению по мере развития экономической системы, представляется рациональным проводить процедуру динамической спецификации для получения более точных прогнозов, при которой состав модели пересматривается по мере появления новых статистических данных. Однако при таком подходе возникает проблема расчета достоверного интервального прогноза вследствие того, что, поскольку каждый раз выбирается модель, имеющая наилучший показатель эффективности, это приводит к существенной систематической недооценке истинной дисперсии ошибок прогноза.

Наряду с вышеупомянутыми проблемами применения регрессионного инструментария к моделированию временных рядов макроэкономических процессов существует также проблема рассогласованности прогнозов при одновременном прогнозировании комплекса макроэкономических индикаторов. Поскольку практически все макроэкономические индикаторы связаны между собой функциональной или корреляционной взаимосвязью, то если прогнозировать каждый из индикаторов по отдельности – полученные прогнозы не будут удовлетворять имеющимся функциональным зависимостям и будут являться рассогласованными с этой точки зрения.

Таким образом, принимая во внимание вышеизложенное, можно сказать, что существующие методы прогнозирования и моделирования макроэкономических процессов, основанные на регрессионном анализе, не в полной мере решают рассмотренные проблемы и могут быть усовершенствованы с целью повышения качества получаемых моделей, характеризующихся более высокой точностью оценки параметров модели и надежностью доверительных интервалов. Вследствие этого, исследования, направленные на разработку новых подходов регрессионного анализа, предлагающих устойчивое решение существующих проблем при моделировании макроэкономических процессов, представляются достаточно актуальными и могут быть использованы центральными банками и руководством страны при принятии управленческих решений.

Степень научной разработанности проблемы. Разработка подходов к моделированию и прогнозированию макроэкономических процессов является популярной темой исследований ученых-эконометристов. Значительный вклад в развитие области общей теории эконометрики, на которую опирается автор в диссертационном исследовании, внесли: Айвазян С.А., Дуброва Т.А., Ершов Э.Б., Лукашин Ю.П., Мхитарян В.С., Тихомиров Н.П., Трегуб И.В., Wooldridge J.M., Hoerl A.E., Kennard R., Tibshirani R., Pearson K., Wold H., Lovell M.C., Larzelere R.E., Mulaik S.A., Wishart J., Mood A., Graybill F., Boes D., Nydick S., Андерсон Т., Бродский Б.Е., Гуреев К.А., Голубева О.С., Доугерти К., Косов В.В., Льюис К.Д., Магнус Я.Р., Катышев П.К., Михайленко К.В., Райская Н., Сергиенко Я., Френкель А., Хейс Д., Четыркин Е.Н., Orphanides A., van Norden S., Justiniano A., Primiceri G.E., Johansen S., Juselius K., Hayashi F., Hamazacebi C., Greene W.H., Glosten L.R., Jagannathan R., Fama E., Evans M., Englsted T., Engle R., Davidson R., James G.M., Cochrane J.H.

Работы, посвященные различным критериям отбора регрессионных моделей, которые используются автором для разработки предлагаемых в диссертации методов, были выполнены следующими авторами: Claeskens G., Hjort N. L., Mallows C.L., Buckland S., Burnham K., Augustin N., Akaike H., Hurvich C.M., Tsai C.L., Lee S., Karagrigoriou A.

Работы, посвященные разработке различных алгоритмов взвешивания регрессионных уравнений, которые легли в основу методов, предлагаемых автором, принадлежат: Bates J.M., Granger C.W.J., Armstrong J.S., Ramanathan R., Chatfield C., Clemen R.T., Hendry D.F., Clements M.P., Timmermann A., Stock J.H., Watson M.W., Hansen B.E., Raftery A.E., Madigan D., Hoeting J.A., Volinsky T., Fernandez C., Ley C.E., Steel M.F.J., Sala-i-Martin X., Doppelhofer G., Miller R.I., Min C.-K., Zellner A., Wright J.H., Racine J., Zhang X., Zhang J., Zou G.

Работы, посвященные проблемам спецификации регрессионных моделей и одновременному прогнозированию макропроцессов, принадлежат таким авторам, как: Shibata R., Derksen S., Keselman H.J., Schwarz G., Shehata Y.A., White P., Šidák Z.K., Ing C.-K., Chen H., Wang Y., Welsh A.H., Yee T.W., Lestari B., Budiantara I.N., Sunaryo S., Mashuri M., Chamidah N., Zain I., Ruchstuhl A., Welsh A.H., Lin X., Carroll R.J., Guo W., Antoniadis A., Spatinas T., George E.I., McCulloch R.E.

Значительный вклад в развитие теории и практики моделирования и прогнозирования временных рядов внесли: Box G.E.P., Jenkins G., Tong H., Kihoro J.M., Otieno R.O., Wafula C., Hipel K.W., McLeod A.I., Garatt A., Lee K., Pesaran M., Shin Y., Wei C.-Z., Кендалл М., Ковалева Л.Н., Faraway J., Chatfield C., Zhang G.P., Hamilton J., Baillie R.T., Ching-Fun C.

Также в процессе разработки методов, представленных в диссертации, автор опирался на законы экономической теории, широко представленные в работах следующих авторов: Макконнел К.Р., Брю С.Л., Абель Э., Бернанке Б., Балацкий Е.В., Бернштам М., Вороновицкий М.М., Кэмпбелл Р.М., Стенли Л.Б., Тарасевич Л.С., Гребенников П.И., Леусский А.И., Sims C.A., Sheshinski E., Weiss Y., Romer D., Roberds W., Runckle D., Whiteman C.H., Mishkin F., McCallum B., Cagan P.

Сформулированные в работах упомянутых выше авторов математические методы и модели макроэкономических процессов позволили проанализировать существующие подходы к процедуре проведения спецификации регрессионных моделей, взвешиванию регрессионных моделей и вычислению оценки истинного уровня значимости предикторов в уравнении. Было выявлено, что данные подходы, по мнению автора, не затрагивают следующие ключевые аспекты:

- в разработанных методах расчета истинного уровня значимости предикторов недостаточное внимание уделено неопределенности относительно истинной дисперсионно-ковариационной матрицы объясняющих переменных;
- методы взвешивания регрессионных уравнений не подразумевают алгоритма вычисления интервального прогноза, а также для большинства из них отсутствует возможность проводить взвешивание моделей, рассчитанных на окнах данных различной длины;
- недостаточно детализирована проработка методов вычисления достоверных доверительных интервалов для прогнозируемых значений

при проведении процедуры динамической спецификации или динамического взвешивания;

- не в полной мере рассмотрено одновременное прогнозирование комплекса взаимосвязанных макроэкономических индикаторов;
- недостаточное внимание уделено гибкости используемых моделей и их адаптивности к условиям изменчивости экономики и наличию структурных сдвигов.

В диссертационной работе предлагается предложить подходы к решению вышеупомянутых проблем регрессионного анализа и моделирования макроэкономических процессов с целью повышения точности получаемых прогнозов и надежности их доверительных интервалов.

Цель и задачи исследования. Целью диссертационной работы является развитие методологии повышения качества эконометрических моделей временных рядов макроэкономических индикаторов посредством достижения более высокой достоверности оценок параметров и надежности получаемых доверительных интервалов по сравнению с существующими методами.

В соответствии с целью в данной работе ставятся и решаются следующие задачи:

- Провести исследование способов спецификации уравнения регрессии, проанализировать их особенности и сформировать набор рекомендаций относительно применения того или иного метода в определенной ситуации;
- Провести анализ способов борьбы с высокой степенью мультиколлинеарности независимых переменных в регрессионных моделях и сформировать набор рекомендаций относительно применения того или иного метода в определенной ситуации;
- Разработать метод вычисления истинного уровня значимости предиктора в многофакторной линейной регрессионной модели при проведении процедуры спецификации;
- Усовершенствовать методы построения эконометрических моделей на основе взвешивания как вложенных, так и невложенных линейных регрессионных моделей;
- Разработать метод оптимизации весовых коэффициентов для взвешивания линейных регрессионных моделей, рассчитанных на окнах данных различной длины и различающихся по составу;
- Разработать метод уточнения оценки доверительного интервала

прогнозируемых значений для моделей временных рядов при проведении процедуры динамического взвешивания;

- Разработать модель взаимосвязей макроэкономических показателей, связывающую более двадцати индикаторов функциональной зависимостью;
- Разработать метод учета функциональных и корреляционных взаимосвязей между моделируемыми макроэкономическими процессами с целью повышения точности получаемых прогнозов;
- Апробировать предложенные методы и модели макроэкономических процессов на имитационных и реальных исторических данных и сравнить их эффективность с уже существующими.

Объект и предмет исследования. Объектом исследования в диссертационной работе являются временные ряды макроэкономических процессов. В качестве предмета исследования выступают методы регрессионного анализа, моделирования и прогнозирования временных рядов макроэкономических процессов.

Область исследования. Диссертационная работа выполнена в рамках п. 1.1 «Разработка и развитие математического аппарата анализа экономических систем: математической экономики, эконометрики, прикладной статистики, теории игр, оптимизации, теории принятия решений, дискретной математики и других методов, используемых в экономико-математическом моделировании» и п. 1.2. «Теория и методология экономико-математического моделирования, исследование его возможностей и диапазонов применения: теоретические и методологические вопросы отображения социально-экономических процессов и систем в виде математических, информационных и компьютерных моделей» специальности 08.00.13 «Математические и инструментальные методы экономики».

Теоретической и методологической базой исследования послужили труды российских и зарубежных ученых в области моделирования и прогнозирования временных рядов макроэкономических процессов, эконометрике и статистике.

В процессе решения поставленных в работе задач использовались статистические методы регрессионного и корреляционного анализа, моделирования временных рядов, а также табличные и графические методы представления результатов исследования.

Информационной базой исследования послужили официальные статистические данные Бюро трудовой статистики США (англ. U.S. Bureau of Labor Statistics), Совета управляющих Федеральной резервной системы США (англ. Board of Governors of the Federal Reserve System), Бюро экономического анализа США (англ. U.S. Bureau of Economic Analysis), S&P Dow Jones Indices LLC, Управления энергетической информации США (англ. U.S. Energy Information Administration), Организации экономического сотрудничества и развития (англ. Organization for Economic Co-operation and Development), Центрального банка Англии (англ. Bank of England), Международного валютного фонда (англ. International Monetary Fund), Федеральной службы государственной статистики РФ, Центрального банка РФ и других печатных и электронных СМИ по исследуемой тематике.

В диссертации были использованы следующие пакеты прикладных программ: «MS Excel», «R», «IBM SPSS Statistics», «MATLAB», «STATISTICA».

Научная новизна. В диссертационной работе предложены методы построения математических моделей временных рядов, обеспечивающие повышение точности прогнозов макроэкономических процессов и надежности их доверительных интервалов.

Предмет защиты составляют следующие результаты, полученные лично автором и содержащие элементы научной новизны:

1. Предложен численный метод расчета истинного уровня значимости предикторов (*p*-значения) в многофакторном регрессионном уравнении при проведении процедуры спецификации. Классический метод определения уровня значимости дает истинные оценки *p*-значения при условии, что все набранные потенциальные предикторы включены в модель. В случае, когда проводится процедура спецификации уравнения, традиционные *p*-значения являются смещенными, поскольку выбираются «лучшие» факторы в процессе спецификации. Разработанный метод учитывает число наблюдений в окне данных, число потенциальных объясняющих переменных и их выборочную дисперсионно-ковариационную матрицу. Проводится сравнение данного метода с существующими поправками *p*-значения (Бонферрони, Шидака, Шехата и Уайта) и показывается их смещенность при тех или иных условиях. Вычисление истинного уровня значимости позволяет сократить долю ошибок 1- ого рода при построении регрессионных моделей макроэкономических процессов, и, как следствие, существенно повысить надежность экспликативных

выводов, полученных из таких моделей.

2. Разработан метод оптимизации весовых коэффициентов для взвешенной суммы набора как вложенных, так и невложенных многофакторных моделей временного ряда, основанный на вычислении несмещенной оценки дисперсии прогноза взвешенной модели и последующей ее минимизации. Существует множество способов оценки качества регрессионных моделей. Вследствие этого при прогнозировании некоторого временного ряда социально-экономического процесса, как правило характеризующегося недостатком наблюдений, согласно различным критериям эффективности модели исследователь может получить разные по составу уравнения. В отечественной и зарубежной литературе было показано, что взвешивание моделей является более предпочтительным вариантом, чем выбор только одной модели. В связи с этим в работе предлагается новый метод оптимизации весовых коэффициентов. В результате применения данного метода вычисляется комбинированная модель, являющаяся более эффективной, чем все входящие в нее уравнения по отдельности. Также предложенный метод в среднем дает более низкую ошибку прогноза по сравнению с такими популярными методами взвешивания, как внутривыборочное взвешивание, байесовское взвешивание, взвешивание Акакике, взвешивание по критерию Маллоуса и простая средняя. Выведена формула числа степеней свободы для вычисления интервальных прогнозов, полученных в результате применения предложенного метода построения многофакторных регрессионных моделей, что недоступно для существующих методов взвешивания.

3. Показано, что при прогнозировании одних и тех же значений временного ряда макроэкономических процессов с помощью моделей, построенных на окнах данных различной длины, зависимость между точностью прогноза и длиной окна имеет не гиперболическую, а параболическую зависимость. В диссертационной работе выдвигается гипотеза, что данный эффект возникает из-за изменчивости степени и направления влияния факторов на целевую переменную во времени. Таким образом, возникает дилемма о выборе длины окна наблюдений, которую предлагается решить с помощью взвешивания регрессионных моделей, рассчитанных на окнах данных различной длины. Показывается, что предложенный метод дает более точные прогнозы, чем каждая из входящих в него моделей по отдельности, а также он является более эффективным, чем байесовское

взвешивание, взвешивание Акаике, простая средняя, выбор модели по байесовскому информационному критерию и критерию Акаике.

4. Предложен метод вычисления корректирующей поправки для среднеквадратической ошибки прогноза при проведении процедуры динамического взвешивания регрессионных моделей. Данный метод позволяет получить более полную картину относительно показателя несмещенной среднеквадратической ошибки прогноза при проведении процедуры динамической спецификации. Данная поправка существенна для повышения достоверности получаемых доверительных интервалов при прогнозировании временных рядов макроэкономических процессов.

5. Разработан метод учета функциональных и корреляционных зависимостей между макроэкономическими индикаторами для повышения точности их прогнозирования. В случае, когда необходимо комплексное прогнозирование макроэкономических индикаторов, информацию об известной функциональной зависимости между индикаторами можно использовать для увеличения точности прогнозирования каждого из них. Если прогнозировать каждый из индикаторов по отдельности, то полученные прогнозы не будут удовлетворять имеющимся функциональным зависимостям и будут являться рассогласованными с этой точки зрения. Предложенный метод учета функциональных связей и комплексного моделирования макроэкономических показателей существенно повышает точность прогноза за счет приведения получаемых прогнозов в соответствие с заданной функциональной или корреляционной связью в зависимости от формы вероятностного распределения каждого из получаемых прогнозов.

Теоретическая значимость исследования. Разработанные в диссертационном исследовании методы построения многофакторной линейной регрессии с распределенными лагами позволяют качественно улучшить прогностический потенциал классических регрессионных моделей при сохранении надежности доверительных интервалов, а также расширить экспликативные возможности регрессионных уравнений.

Практическая значимость полученных результатов заключается в возможности применения разработанных методов для построения моделей макроэкономических процессов с целью их краткосрочного, среднесрочного и долгосрочного прогнозирования. Получаемые прогнозы обладают более высокой точностью по сравнению с прогнозами, рассчитанными по уже существующим методам, а также имеют надежные доверительные интервалы. Полученные модели и прогнозы могут быть использованы для принятия оптимальных управленческих решений органами государственной власти и управления. Также результаты работы могут использоваться в учебном процессе вузов при создании и совершенствовании дисциплин «Эконометрика», «Моделирование макроэкономических процессов» и др.

Внедрение и апробация результатов работы. Результаты диссертационной работы докладывались на многих международных и всероссийских научно-практических конференциях, основными из которых являются:

- Международная научная школа-семинар имени академика С.С. Шаталина «Системное моделирование социально-экономических процессов», Воронеж, 29.09-4.10.2013 г.;

- Международная научно-практическая конференция «Интеграция отечественной науки в мировую: проблемы, тенденции и перспективы», г. Москва, 26- 29 сентября 2014 г.

- Международная научно-практическая конференция «Probability theory and mathematical statistics», г. Казань, Казанский (Приволжский) Федеральный Университет, 6-12 ноября 2017 г.;

- Международная научно-практическая конференция «12th International Conference on Computational and Financial Econometrics», г. Пиза, Пизанский Университет, 14-16 декабря 2018 г.;

- Международная научно-практическая конференция «IX Абалкинские чтения», г. Москва, РЭУ им. Г.В. Плеханова, 25-26 апреля 2019 г.;

- Международная научно-практическая конференция «13th International Conference on Computational and Financial Econometrics», г. Лондон, Лондонский Университет (Биркбек), 14-16 декабря 2019 г.

Разработанные методики были с успехом внедрены при выполнении следующих научно-исследовательских работ:

- «Анализ инфляционных процессов в российской экономике и обуславливающих их факторов», государственный контракт № 32/дБ с НИИ СП от «04» июня 2012 г.;

- «Разработка технологий выявления кризисных ситуаций и определения путей их разрешения. Создание модели опережающего стратегического управления», в рамках субсидии, полученной согласно постановлению №218 «О мерах государственной поддержки развития кооперации российских высших учебных заведений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства», совместно с ООО «ИБС Экспертиза» и ФГАОУ ВПО «Северный (Арктический) федеральный университет им. М.В. Ломоносова»;

- «Система многофакторного динамического прогнозирования биржевых котировок и индексов», внутренний грант РЭУ им. Г.В. Плеханова, приказ № 1129 от 19 ноября 2014 г.;

- «Методология и информационно-аналитические средства решения проблем пространственного развития экономики России в условиях современных реформ», Госзадание Министерства Образования и Науки, проект № 1675, 2015-2016 гг.;

- «Разработка алгоритмов машинного обучения скоринговых моделей для микрофинансовых организаций», внутренний грант РЭУ им. Г.В. Плеханова, приказ №1172 от 12.10.2016 г.;

- «Совершенствование методического инструментария краткосрочного прогнозирования социально-экономических показателей», внутренний грант РЭУ им. Г.В. Плеханова, приказ №1173 от 12.10.2016 г.;

- «Развитие регрессионного инструментария прогнозирования временных рядов макроэкономических процессов», грант Президента РФ для молодых ученых –кандидатов наук, 2017-2018 гг., договор №14.Z56.17.1169-МК.

- «Разработка системы численной оценки экономических проектов на основе межотраслевого баланса с использованием комбинированных методов компьютерной оптимизации», договор №18-310-20008\18 от «12» октября 2018 г.

Разработанные в диссертационном исследовании методы и модели используются для подготовки ежеквартальных прогнозов ключевых макроэкономических индикаторов Российской Федерации на базе Научно-исследовательского объединения РЭУ им. Г.В. Плеханова с последующей публикацией на сайте Университета.

II. ОСНОВНЫЕ ПОЛОЖЕНИЯ ДИССЕРТАЦИОННОЙ РАБОТЫ, ВЫНОСИМЫЕ НА ЗАЩИТУ

1. Метод получения несмещенных корректировок *p*-значений предикторов при проведении процедуры спецификации модели. Положим, что $\{y_t, X_t: t = 1, \dots, n\}$ является рассматриваемой выборкой действительных чисел, где y_t – целевая переменная, а $X_t = (1, x_{1t}, x_{2t}, \dots)$ – конечный вектор потенциальных объясняющих переменных. Также предположим, что можно специфицировать простую линейную регрессионную модель, выбрав подмножество объясняющих переменных $\tilde{X}_t \subset X_t$ из изначально заданного набора факторов X_t:

$$Y = \tilde{X}B + e, \tag{1}$$

где

$$\tilde{X} = \begin{pmatrix} \tilde{X}_n \\ \tilde{X}_{n-1} \\ \vdots \\ \tilde{X}_1 \end{pmatrix}, \quad Y = \begin{pmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_1 \end{pmatrix}$$

и вектор параметров может быть вычислен в явном виде:

$$B = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T Y. \tag{2}$$

Здесь, естественно, будем предполагать, что выполняются предпосылки МНК. Общепринятой практикой является вычисление двухсторонней значимости с использованием квантилей t- распределения применительно к отношению величины соответствующего коэффициента к его среднеквадратическому отклонению.

Рассмотрим набор потенциальных предикторов разрабатываемой модели $X_t = (1, x_{1t}, x_{2t}, \dots)$ и обозначим p_i как p- значение для i-ого предиктора из отобранного подмножества объясняющих переменных $\tilde{X}$. Далее обозначим массив данных $\breve{X}$, который остался после проведения процедуры отбора наилучшего подмножества для модели с добавлением к нему i-ого предиктора из $\tilde{X}$. Таким образом, $\breve{X}$ будет включать все переменные из X, которые не были включены в $\tilde{X}$ плюс i- ый предиктор, находящийся под рассмотрением. Для начала рассмотрим простейший случай, когда дисперсионно-ковариационная матрица $\breve{X}$ является диагональной матрицей, что подразумевает отсутствие корреляции между потенциальными предикторами. Тогда скорректированное p- значение для рассматриваемого предиктора может быть аналитически вычислено, как показано ниже:

$$p_{adj} = 1 - (1 - p_i)^m, \tag{3}$$

где m – число предикторов в матрице $\breve{X}$.

Формула (3) выводится от обратной вероятности и была предложена польским математиком Шидаком. Поскольку любой предиктор из матрицы $\breve{X}$ потенциально мог бы занять место рассматриваемого предиктора, для корректировки исходного p- значения необходимо вычислить вероятность того, что хотя бы одно исходное p-значение анализируемых предикторов попадет в рассматриваемый квантиль, так как при проведении процедуры спецификации регрессионного уравнения выбирается, естественно, наилучший из предикторов. Именно поэтому скорректированное p- значение весьма сильно отличается от исходного в случае, если число потенциальных объясняющих переменных достаточно велико. Таким образом, можно заключить, что в случае не проведения корректировок, вероятность совершения ошибки 1-ого рода стремительно возрастает с ростом количества потенциальных предикторов в $\breve{X}$.

Далее рассмотрим ситуацию, когда дисперсионно-ковариационная матрица $\breve{X}$ не является диагональной, т.е. данные имеют корреляционные взаимосвязи.

В данном случае предлагается прибегнуть к численной процедуре, которая базируется на машинной генерации матрицы $\breve{X}$ согласно выборочной дисперсионно-ковариационной матрице. Здесь, дополнительно к предпосылкам МНК, будем полагать следующее:

Предпосылка 1. Нормальность потенциальных предикторов, т.е. $x_{it} \sim N(m_i, \sigma_i)$.

Поскольку мы имеем возможность рассчитать только выборочную дисперсионно-ковариационную матрицу, надежность предлагаемого метода зависит от пропорции числа наблюдений и числа потенциальных объясняющих переменных в $\breve{X}$. Поэтому в данном исследовании предлагается численный метод, дающий несмещенные

p- значения путем случайной генерации не только набора данных $\breve{X}$, но также его дисперсионно-ковариационной матрицы. Для этого, при условии выполнения предпосылки 1, можно использовать обратное распределение Уишарта в качестве априорного.

$$\Sigma = W_m^{-1}\left(n-1, \breve{\Sigma}^{-1}\right) \cdot (n-1), \tag{4}$$

где n – число наблюдений, $m \leq n$ – число потенциальных объясняющих переменных в $\breve{X}$, $\breve{\Sigma}$ – выборочная дисперсионно-ковариационная матрица для набора данных $\breve{X}$.

Таким образом, имеется возможность случайным образом сгенерировать некоторую реализацию истинной дисперсионно-ковариационной матрицы для рассматриваемого набора статистических данных при условии наличия выборочной. После проведения данной процедуры также можно сгенерировать набор потенциальных объясняющих переменных $\breve{X}$ согласно некоторой полученной реализации дисперсионно-ковариационной матрицы. Для этого прибегнем к следующей процедуре. Во-первых, определим вектор-столбец независимых, идентично распределенных случайных величин Z, которые подчиняются нормальному закону распределения с нулевой средней и единичной дисперсией, что подразумевает, что

$$E(ZZ^T) = I_m. \tag{5}$$

После чего случайная реализация набора данных $\breve{X}_t$ может быть осуществлена с помощью разложения Холецкого сгенерированной дисперсионно-ковариационной матрицы Σ как показано ниже:

$$\breve{X}_t^T = SZ. \tag{6}$$

Сгенерировав новый набор случайных данных $\breve{X}$, подставим потенциальные предикторы один за другим на место рассматриваемой независимой переменной и присвоим единицу данному опыту в случае, если хотя бы один предиктор показал p-значение, меньшее, чем изначально полученное. По окончании одного опыта происходит повторная генерация дисперсионно-ковариационной матрицы Σ и последовательно нового набора данных $\breve{X}$. После чего повторяется процедура подставления предикторов в уравнение и опыту присваивается значение либо ноль, либо единица. Исправленное p- значение представляет собой отношение просуммированных присвоенных значений к общему числу проведенных имитаций. Таким образом, предлагаемая процедура дает несмещенные исправленные p- значения.

Описанная выше процедура может быть алгоритмизирована следующим образом:

1. Определить исследуемый предиктор и записать соответствующее p- значение p_i.

2. Установить счетчик *COUNT* = 0

3. DO n = 1 TO N

а). Сгенерировать Σ для $\breve{X}$ согласно формуле (4)

б). Применяя формулу (6) генерируем $\breve{X}$ согласно вновь полученной дисперсионно-ковариационной матрице Σ

в). Для созданных фиктивных данных определим значение $K = 1$, если существует хотя бы один предиктор, чье p-значение меньше p-значения исследуемого предиктора, а именно $q_i < p_i$. Иначе, определить $K = 0$

г). *COUNT* = *COUNT*+*K*

4. ENDDO

5. Скорректированное p-значение = *COUNT/N*

Проведем анализ поведения p-значений предикторов при различных формах дисперсионно-ковариационной матрицы Σ, разном числе наблюдений и числе потенциальных предикторов. Также представим сравнительный анализ предложенного метода и других широко распространенных подходов к корректировке p- значения.

Для начала исследуем случай, когда $\breve{X}$ включает в себя всего лишь две потенциальных объясняющих переменных ($m = 2$), которые имеют определенный выборочный коэффициент корреляции. На рисунке 1 представлено сравнение двух кривых исправленных p- значений, рассчитанных по наборам данных разной длины ($n = 200,\ n = 5$). В обоих случаях была произведена корректировка исходного p- значения, равного 0.05. Для получения каждой точки графика было проведено по 10 000 000 имитаций согласно предложенному алгоритму. Чем больше наблюдений имеется в рассматриваемом окне, тем более точно будет оценена истинная дисперсионно-ковариационная матрица Σ, что означает, что когда $n = 200$, мы получаем достаточно точную оценку Σ и наоборот, имеем высокий уровень неопределенности относительно Σ когда $n = 5$. Поэтому можно видеть из представленного графика, что вторая кривая показывает более низкое исправленное p- значение, чем первая, когда коэффициент корреляции невысок и наоборот, когда присутствует значительная степень корреляции между этими двумя потенциальными предикторами.

Подводя итоги сказанного выше, можем заключить следующее. Первая кривая стремится к значению 0.0975 с приближением выборочного коэффициента корреляции к нулю, что является простой поправкой, представленной в формуле (3). Когда коэффициент корреляции стремится к единице, исправленные p- значения в обоих случаях стремятся к исходному, равному 0.05. Эти кривые пересекаются в точке, где выборочный коэффициент корреляции приблизительно равен 0.6, а его функция плотности вероятности в случае $n = 5$ меняет асимметрию с положительной на отрицательную.

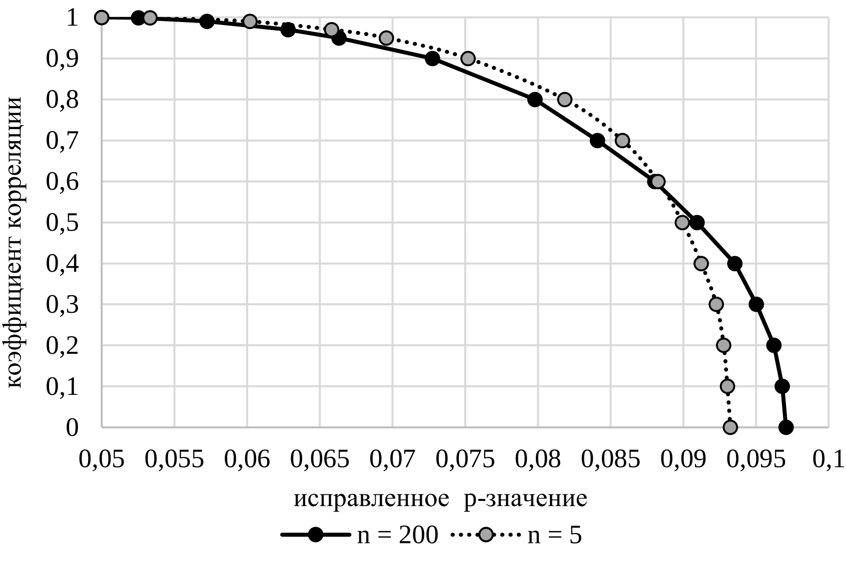

Рисунок 1 – Корректировка p-значения, равного 0.05, в случае двух потенциальных предикторов

Источник: составлено автором

Для проведения сравнительного анализа различных подходов к вычислению исправленного p-значения изобразим несмещенные

p- значения, вычисленные согласно предложенному методу (ось абсцисс) против исправленных p-значений, вычисленных согласно другим существующим методикам (ось ординат). В частности, будем сравнивать традиционный расчет p-значения, поправку Бонферрони, поправку Шехата и Уайта, простую корректировку Шидака и предложенную численную поправку p-значения. Помимо достаточно тривиальной поправки Бонферрони исследователи также прибегают к процедуре Хоммеля, методу Шидака-Холма, методу Бенджамини-Иекутиели и методу Бенджамини-Хохберга, которые хорошо себя зарекомендовали при проверке множественных гипотез. Однако упомянутые поправки в основном разрабатывались для возможности контроля доли ложных отклонений гипотез при одновременном тестировании множества гипотез, в то время как разрабатываемый в данной работе метод направлен на поправку р-значения для одной гипотезы: значимость определенного предиктора в уравнении при предварительном проведении процедуры спецификации регрессионного уравнения. Таким образом, большинство существующих методов поправок р- значения не могут быть напрямую сравнимы с предлагаемым в данной работе.

На рисунках 2 и 3 представлено сравнение рассматриваемых поправок в случае, когда $n = 30$, $m = 25$ и выборочный коэффициент корреляции принимает значения, сконцентрированные вокруг нуля. Как можно видеть, традиционный расчет p-значения показывает наихудшую точность, что не удивительно, поскольку число потенциальных предикторов достаточно высоко. Однако, альтернативные методы также демонстрируют существенное смещение, что увеличивает вероятность совершения ошибок 2-ого рода в процессе спецификации регрессионного уравнения.

Таким образом, представленный численный метод возвращает несмещенные поправки исходных p-значений для объясняющих переменных, который напрямую относится к выводам относительно степени влияния этих переменных на целевую. Проведя сравнительный анализ предложенного метода и уже существующих, можно сделать вывод, что расчет истинных p-значений может оказать помощь в ограничении числа ошибок 1-ого рода в научных исследованиях, значительно снижая количество публикаций, декларирующих ложные зависимости в качестве истинных. Предложенный метод легко алгоритмизируются и может быть интегрирован в абсолютное большинство существующих статистических пакетов.

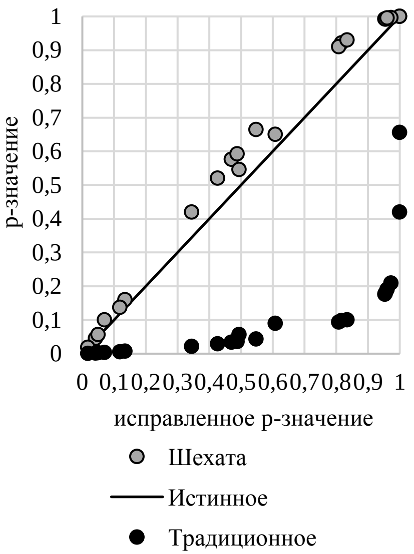

Рисунок 2 – Сравнение методов корректировок *p*- значения ($n = 30,\ m = 25$).

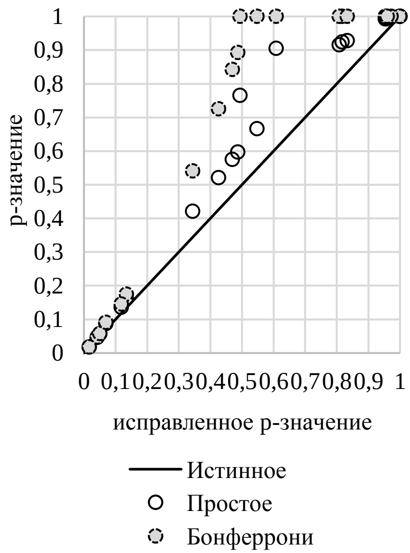

Рисунок 3 – Сравнение методов корректировок *p*- значения ($n = 30,\ m = 25$).

Источник: составлено автором

Метод взвешивания регрессионных уравнений, рассчитанных на окнах данных как одинаковой, так и различной длины. Пусть $\{y_t, X_t: t = 1, \dots, n\}$ является рассматриваемым набором действительных данных, где y_t – целевая переменная, а $X_t = (1, x_{1t}, x_{2t}, \dots)$ – конечный набор потенциальных объясняющих переменных. Также предположим, что имеется возможность рассчитать *M* регрессионных моделей, где *i*-ая модель включает в себя k_i параметров и может быть представлена следующим образом:

$$y_t = X_{it} B_i + e_{it} \text{ или} \quad (7)$$

$$\hat{y}_{it} = X_{it} B_i, \quad (8)$$

где X_{it} обозначает отобранные объясняющие переменные для *i*-ой модели, e_{it} – наблюдаемая ошибка модели за период t, B_i – вектор-столбец параметров модели, который рассчитывается как:

$$B_i = \left(X_i^T X_i\right)^{-1} X_i^T Y, \tag{9}$$

где

$$X_i = \begin{pmatrix} X_{in} \\ X_{i(n-1)} \\ \vdots \\ X_{i1} \end{pmatrix}, \quad Y = \begin{pmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_1 \end{pmatrix}.$$

Каждая модель имеет некоторую ненаблюдаемую истинную ошибку ε_{it}, обладающую следующими свойствами:

$$E\left(\varepsilon_{it} \middle| X_{it}\right) = 0, \tag{10}$$

$$E\left(\varepsilon_{it}^2 \middle| X_{it}\right) = \sigma^2(i). \tag{11}$$

Целью взвешивания набора таких моделей является построение комбинированного уравнения, рассчитываемого как показано ниже:

$$\bar{\hat{y}}_t = \sum_{i=1}^{M} \hat{y}_{it} w_i \,. \tag{12}$$

Причем легко показать, что из (12) следует точно такая же зависимость ошибок комбинированной модели от ошибок по каждой из входящих в нее подмоделей.

Здесь стоит отметить, что если выполняются предпосылки (10) и (11), то справедливо следующее:

$$E\left(\bar{\varepsilon}_t \middle| X_{1t}, X_{2t}, \dots, X_{lt}, w_1, w_2, \dots, w_l\right) = 0, \tag{13}$$

$$E\left(\bar{\varepsilon}_t^2 \middle| X_{1t}, X_{2t}, \dots, X_{lt}\right) = \bar{\sigma}^2, \tag{14}$$

где w_i является элементом вектора весовых коэффициентов, элементы которого неотрицательны и в сумме дают единицу:

$$\begin{vmatrix} \sum_{i=1}^{M} w_i = 1, \\ 0 \le w_i \le 1. \end{vmatrix} \qquad (15)$$

В работе выводится формула для несмещенной оценки среднеквадратической ожидаемой ошибки прогноза (MSFE) комбинированной модели $\bar{\hat{y}}_t$ при том, что подмодели $\hat{y}_t$ могут быть рассчитаны на окнах данных различной длины n:

$$MSFE_{n+1} = \sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\breve{e}_i^T \breve{e}_j}{tr\left(\breve{A}_i \breve{A}_j\right)} w_i w_j \left\{1 + X_{i(n+1)}\left(X_i^T X_i\right)^{-1} \breve{X}_i^T \breve{X}_j \left(X_j^T X_j\right)^{-1} X_{j(n+1)}^T\right\}, \qquad (16)$$

где $\breve{X}_i = \begin{vmatrix} 0 \\ X_i \end{vmatrix}$, а 0 обозначает нулевую матрицу, которая дополняет X_i таким образом, что число строк результирующей матрицы равняется числу строк самого длинного из рассматриваемых окон данных, $\breve{e}_i = \begin{vmatrix} 0 \\ e_i \end{vmatrix}$, 0 обозначает вектор-столбец нулей, который дополняет e_i до числа строк, равного числу наблюдений в самом длинном из рассматриваемых окон данных, $\breve{I}_{ni} = \begin{vmatrix} 0 & 0 \\ 0 & I_{ni} \end{vmatrix}$, что является дополненной нулями единичной матрицей размерности $n_i \times n_i$ до размерности самого длинного окна данных под рассмотрением, $\breve{A}_i = \breve{I}_{ni} - \breve{X}_i \left(X_i^T X_i\right)^{-1} \breve{X}_i^T$.

Тогда, процедура подбора вектора весовых коэффициентов осуществляется с помощью решения следующей оптимизационной задачи квадратичного программирования с системой ограничений.

$$MSFE_{n+1} \to \min$$
$$\begin{vmatrix} \sum_{i=1}^{M} w_i = 1, \\ 0 \le w_i \le 1. \end{vmatrix} \qquad (17)$$

В настоящее время разработано немало стандартизованных алгоритмов решения оптимизационных задач такого класса, основанных на численных методах, а при имеющихся на сегодняшний день вычислительных мощностях, доступных как корпорациям, так и частным лицам, даже при достаточно большом количестве рассматриваемых подмоделей M возможно найти численное решение за разумное время, использую любой из существующих алгоритмов.

Далее перейдем к алгоритму получения интервального с использованием комбинированной модели. Для расчета доверительного интервала для прогноза, полученного с использованием линейной регрессии, будет использоваться t- распределение.

$$\bar{\hat{y}} - T_{\alpha,r} \cdot \sqrt{MSFE_{n+1}} < y < \bar{\hat{y}} + T_{\alpha,r} \cdot \sqrt{MSFE_{n+1}}$$

где r – число степеней свободы.

В исследовании была выведена формула для числа степеней свободы для подмоделей, рассчитанных на окнах данных как различной, так и одинаковой длины.

$$r = \frac{2\left(\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\breve{e}_i^T\breve{e}_j}{tr\left(\breve{A}_i\breve{A}_j\right)}w_i w_j\breve{\theta}_{ij}\right)^2}{\sum_{a=1}^{M}\sum_{b=1}^{M}\sum_{c=1}^{M}\sum_{d=1}^{M}w_a w_b w_c w_d\breve{\theta}_{ab}\breve{\theta}_{cd}\frac{\breve{\Theta}_{abcd}tr\left(\left[\breve{A}_a\breve{A}_b\right]\circ\left[\breve{A}_c\breve{A}_d\right]\right)+\Psi_{abcd}}{tr\left(\breve{A}_a\breve{A}_b\right)tr\left(\breve{A}_c\breve{A}_d\right)}}, \tag{18}$$

где $\breve{\theta}_{ij} = 1 + X_{i(n+1)}\left(X_i^T X_i\right)^{-1}\breve{X}_i^T\breve{X}_j\left(X_j^T X_j\right)^{-1}X_{j(n+1)}^T,$

$$\breve{\Theta}_{abcd} = \left\{\left(I + P_{vec(UCOV)}\right)UCOV' \otimes UCOV'\right\}_{\langle abcd\rangle},$$

$$UCOV' = \begin{vmatrix} \frac{e_1^T e_1}{tr(A_1)} & \frac{\breve{e}_1^T\breve{e}_2}{tr\left(\breve{A}_1\breve{A}_2\right)} & \cdots & \frac{\breve{e}_1^T\breve{e}_M}{tr\left(\breve{A}_1\breve{A}_M\right)} \\ \frac{\breve{e}_2^T\breve{e}_1}{tr\left(\breve{A}_2\breve{A}_1\right)} & \frac{e_2^T e_2}{tr(A_2)} & \cdots & \frac{\breve{e}_2^T\breve{e}_M}{tr\left(\breve{A}_2\breve{A}_M\right)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\breve{e}_M^T\breve{e}_1}{tr\left(\breve{A}_M\breve{A}_1\right)} & \frac{\breve{e}_M^T\breve{e}_2}{tr\left(\breve{A}_M\breve{A}_2\right)} & \cdots & \frac{e_M^T e_M}{tr(A_M)} \end{vmatrix},$$

$$\breve{\Psi}_{abcd} = \operatorname{cov}(e_{at}e_{ct})\operatorname{cov}(e_{bt}e_{dt})\left[tr\left(\left[\breve{A}_a\breve{A}_b\right]\times\left[\breve{A}_c\breve{A}_d\right]^T\right) - tr\left(\left[\breve{A}_a\breve{A}_b\right]\circ\left[\breve{A}_c\breve{A}_d\right]\right)\right]$$
$$+ \operatorname{cov}(e_{at}e_{dt})\operatorname{cov}(e_{bt}e_{ct})\left[tr\left(\left[\breve{A}_a\breve{A}_b\right]\times\left[\breve{A}_c\breve{A}_d\right]\right) - tr\left(\left[\breve{A}_a\breve{A}_b\right]\circ\left[\breve{A}_c\breve{A}_d\right]\right)\right].$$

Для подтверждения эффективности разработанного метода в диссертационной работе было проведено имитационное и эмпирическое тестирование на реальных макроэкономических данных: Индекс Потребительских Цен (ИПЦ) в США (Q1.1913-Q2.2019, квартальные данные), ИПЦ в Великобритании (Q1.1956-Q2.2019, квартальные данные, ИПЦ в Японии (Q1.1956-Q2.2019, квартальные данные), ИПЦ в России (01.1992-07.2019, месячные данные), денежный агрегат М2 в США (Q4.1959-Q2.2019, квартальные данные), безработица в США (Q3.1948-Q2.2019, квартальные данные), денежный агрегат М2 в Японии (Q1.1955-Q2.2019, квартальные данные) и безработица в Великобритании (Q1.1956-Q2.2019, квартальные данные).

В первой части эмпирического эксперимента были рассмотрены модели авторегрессии разного порядка для построения четырех уравнений для проведения процедуры взвешивания при одинаковой длине окна наблюдений. Для того, чтобы данные переменные были стационарны показатели безработицы были преобразованы в разности первого порядка, а ИПЦ и М2 конвертированы в темпы прироста.

Таким образом, были получены четыре уравнения для взвешивания по каждому из анализируемых макроиндикаторов:

1). $I\%_t = b_{01} + b_{11}I\%_{t-1} + b_{21}I\%_{t-2} + b_{31}I\%_{t-3} + b_{41}I\%_{t-4} + e_{1t}$,

2). $I\%_t = b_{02} + b_{12}I\%_{t-1} + b_{22}I\%_{t-2} + b_{32}I\%_{t-3} + e_{2t}$,

3). $I\%_t = b_{03} + b_{13}I\%_{t-1} + b_{23}I\%_{t-2} + e_{3t}$,

4). $I\%_t = b_{04} + b_{14}I\%_{t-1} + e_{4t}$.

Сравнительный анализ способов взвешивания регрессионных моделей проводился с помощью вычисления индекса их относительной эффективности:

$$RPI_i = \frac{MSRE_i}{\min\left(MSRE_1, MSRE_2 ..., MSRE_M\right)}. \qquad (19)$$

Также к сравнению была добавлена каждая из взвешиваемых моделей по отдельности. Результаты для временного ряда ИПЦ Великобритании, М2 Японии и М2 США представлены в таблицах 1, 2 и 3 соответственно.

Таблица 1 – Сравнение методов взвешивания моделей, ИПЦ Великобритании

n	ISMA	MMA	SMA	MSFE	BMA	$m=4$	$m=3$	$m=2$	$m=1$
10	1.610	1.434	1.088	1.036	1.403	1.701	1.483	1.201	**1.000**
15	1.391	1.350	1.016	1.038	1.488	1.458	1.240	1.115	**1.000**
20	1.307	1.314	1.095	**1.000**	1.458	1.359	1.322	1.231	1.160
30	1.255	1.265	1.169	**1.000**	1.338	1.292	1.382	1.312	1.323
40	1.203	1.213	1.207	**1.000**	1.289	1.234	1.443	1.377	1.421
50	1.171	1.180	1.194	**1.000**	1.300	1.198	1.387	1.351	1.454
60	1.175	1.180	1.214	**1.000**	1.297	1.199	1.388	1.372	1.507
70	1.160	1.165	1.235	**1.000**	1.232	1.182	1.425	1.401	1.551
80	1.183	1.183	1.065	**1.000**	1.235	1.217	1.157	1.170	1.254
90	1.012	1.012	1.118	**1.000**	1.020	1.014	1.203	1.223	1.332
100	1.010	1.010	1.290	1.004	1.001	**1.000**	1.423	1.461	1.665
Mean	1.225	1.210	1.154	**1.007**	1.278	1.259	1.350	1.292	1.334

Источник: составлено автором на основе данных Организации экономического сотрудничества и развития

Таблица 2 – Сравнение методов взвешивания моделей, М2 Японии

n	ISMA	MMA	SMA	MSFE	BMA	$m=4$	$m=3$	$m=2$	$m=1$
10	1.126	**1.000**	1.062	1.070	1.098	1.079	1.167	1.193	1.688
15	1.153	1.067	1.306	**1.000**	1.026	1.022	1.182	1.438	2.169
20	1.093	1.035	1.267	**1.000**	1.009	1.023	1.117	1.532	1.911
30	1.066	1.060	1.316	**1.000**	1.011	1.019	1.083	1.523	1.951
40	1.058	1.061	1.286	**1.000**	1.040	1.079	1.071	1.534	1.834
50	1.057	1.067	1.299	**1.000**	1.103	1.107	1.073	1.636	1.800
60	1.052	1.054	1.304	**1.000**	1.097	1.078	1.069	1.790	1.727
70	1.062	1.063	1.351	**1.000**	1.091	1.082	1.080	1.914	1.770
80	1.067	1.067	1.376	**1.000**	1.085	1.084	1.083	1.981	1.775

90	1.075	1.075	1.375	**1.000**	1.093	1.093	1.093	2.009	1.738
100	1.072	1.072	1.327	**1.000**	1.092	1.092	1.092	1.847	1.689
Mean	1.080	1.056	1.297	**1.006**	1.068	1.069	1.101	1.672	1.823

Источник: составлено автором на основе данных Международного валютного фонда

Таблица 3 – Сравнение методов взвешивания моделей, М2 США

n	ISMA	MMA	SMA	MSFE	BMA	$m = 4$	$m = 3$	$m = 2$	$m = 1$
10	1.497	1.313	**1.000**	1.146	1.238	1.229	1.578	1.176	1.257
15	1.352	1.253	**1.000**	1.064	1.189	1.206	1.417	1.118	1.277
20	1.125	1.094	1.168	**1.000**	1.041	1.021	1.157	1.359	1.551
30	1.004	**1.000**	1.299	1.023	1.098	1.066	1.013	1.575	1.745
40	1.003	**1.000**	1.298	1.006	1.085	1.024	1.011	1.606	1.699
50	1.002	**1.000**	1.320	1.002	1.057	1.030	1.009	1.637	1.716
60	1.015	1.019	1.358	**1.000**	1.049	1.024	1.022	1.737	1.730
70	1.022	1.022	1.387	**1.000**	1.032	1.030	1.030	1.823	1.766
80	1.028	1.028	1.370	**1.000**	1.038	1.038	1.038	1.834	1.706
90	1.020	1.020	1.410	**1.000**	1.027	1.027	1.027	1.945	1.751
100	1.020	1.020	1.365	**1.000**	1.030	1.030	1.030	1.868	1.678
Mean	1.099	1.070	1.271	**1.022**	1.080	1.066	1.121	1.607	1.625

Источник: составлено автором на основе данных Совета управляющих Федеральной резервной системы США

Из таблиц 1-3 можно наблюдать подавляющее превосходство предлагаемого метода как над другими методами взвешивания, так и над каждой из взвешиваемых моделей по отдельности.

Во второй части эмпирического эксперимента данные макроэкономические индикаторы моделировались с помощью авторегрессии четвертого порядка (для ИПЦ России был выбран двенадцатый порядок), что означает, что для предсказания каждого индикатора были построены пять моделей по разным окнам данных, после чего проводилось их взвешивание согласно предложенным алгоритмам. Рассматривались окна данных следующих длин: $n_1 = 100, n_2 = 80,\ n_3 = 60,\ n_4 = 40,\ n_5 = 20$.

На рисунке 4 представлено сравнение усредненного индекса относительной эффективности (RPI) рассматриваемых моделей при взвешивании уравнений, построенных на окнах данных различной длины для вышеупомянутых макроэкономических показателей без проведения предварительной спецификации.

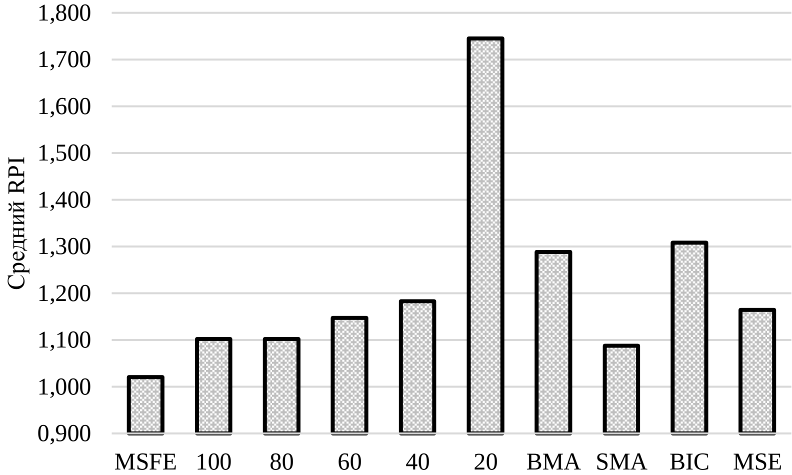

Рисунок 4 – Усредненный RPI для рассматриваемых методов, эмпирический эксперимент, без спецификации

Источник: составлено автором на основе данных Бюро трудовой статистики США, Организации экономического сотрудничества и развития, Мирового валютного фонда, Федеральной службы государственной статистики РФ и Совета управляющих Федеральной резервной системы США

Рисунок 4 наглядно демонстрирует превосходство предлагаемого в диссертационном исследовании метода взвешивания регрессионных моделей над каждой из подмоделей, а также над каждым из рассмотренных методов взвешивания или выбора наилучшей модели в условиях прогнозирования реальных исторических данных макроэкономических процессов.

Подводя итоги, следует подчеркнуть, что в результате проведенной работы был обнаружен и исправлен главный недостаток существующих методов взвешивания регрессионных моделей: невозможность получить несмещенную оценку среднеквадратической ошибки прогноза и доверительные интервалы прогнозируемых значений. При использовании MSFE взвешивания мы получаем ожидаемую ошибку прогноза для комбинированной модели и доверительный интервал с заданным уровнем значимости. Предлагаемый метод разрабатывался как для вложенных, так и для невложенных моделей, что существенно расширяет область его применения. Проведенное имитационное и эмпирическое тестирование выявило устойчивое превосходство предлагаемого метода над существующими способами взвешивания регрессионных моделей, что, как следствие, дает ему право считаться предпочтительным в использовании при анализе и прогнозировании временных рядов социально-экономических процессов. Основная причина, по которой MSFE взвешивание работает лучше остальных методов, заключается в том, что он учитывает как корреляционную структуру ошибок взвешиваемых моделей, так и корреляционную структуру их коэффициентов, что позволяет получать более эффективные весовые коэффициенты.

Метод корректировок доверительных интервалов при проведении процедуры динамического взвешивания моделей. В исследовании показывается, что процедура взвешивания регрессионных уравнений во многом схожа по сущности с процедурой спецификации, поскольку она является своего рода категоричной формой взвешивания вследствие того, что одной спецификации присваивается весовой коэффициент, равный единице, и нулевые коэффициенты всем оставшимся спецификациям.

При взвешивании моделей методом минимизации ожидаемой среднеквадратической ошибки прогноза, предложенным в исследовании, вычисляется оценка истинной среднеквадратической ошибки прогноза (MSFE) согласно формуле (16). При этом истинная ошибка интегральной модели $\bar{\sigma}^2$ является неизвестной величиной. Оценка MSFE является несмещенной только в случае статической модели, при которой однажды рассчитанные веса не подвергаются изменениям в процессе моделирования рассматриваемого временного ряда. В случае с динамическим взвешиванием MSFE является смещенной оценкой $\bar{\sigma}^2$ и дает систематически заниженные значения истинной дисперсии ошибок. Поскольку оптимизация весовых коэффициентов является нелинейной процедурой, для которой недоступно аналитическое решение при количестве взвешиваемых моделей, превышающем три, вывод несмещенной оценки истинной дисперсии в явном виде представляется невозможным. Таким образом, при неизвестном значении $\bar{\sigma}^2$ основной задачей является вычисления математического ожидания величины смещения оценки MSFE, на которое впоследствии она будет корректироваться.

Исправленное значение для MSFE в случае динамического взвешивания регрессионных моделей предлагается вычислять следующим образом:

Шаг 1. Получить несмещенную оценку дисперсионно-ковариационной матрицы истинных ошибок Σ по взвешиваемым подмоделям.

Шаг 2. Положить, что полученная на предыдущем шаге матрица Σ является истинной и выполнить N раз следующий цикл:

а). Сгенерировать по истинной дисперсионно-ковариационной матрице согласно распределению Уишарта случайную ее реализацию Ω;

б). Провести оптимизацию весовых коэффициентов для вновь сгенерированной дисперсионно-ковариационной матрицы Ω согласно оптимизационной задаче (17);

в). Рассчитать показатель MSFE согласно формуле (16) по оптимизированным весовым коэффициентам и сгенерированной дисперсионно-ковариационной матрице ошибок рассматриваемых подмоделей Ω;

г). Рассчитать показатель MSFE по оптимизированным весовым коэффициентам и истинной дисперсионно-ковариационной матрице ошибок рассматриваемых подмоделей Σ;

д). Рассчитать отношение показателя MSFE, полученного на предыдущем шаге и показателя MSFE, вычисленного в пункте в);

Шаг 3. Усреднить полученные в пункте д) отношения, что даст окончательный поправочный коэффициент.

Шаг 4. Умножить показатель MSFE для исходной модели на полученный на предыдущем шаге поправочный коэффициент.

Проведем тестирование разработанного численного метода расчета поправочного коэффициента для классической оценки дисперсии ошибки прогноза при динамическом взвешивании регрессионных уравнений. Представим имитационный эксперимент, где данные получаются посредством машинной генерации случайных величин с заданными свойствами. В данном эксперименте предположим, что целевая переменная y_t зависит только от одной объясняющей переменной $\{x_t \sim N(1{,}1),\ t = 1,..,n\}$ и генерируется согласно указанной ниже формуле:

$$y_t = x_t + \varepsilon_t ,$$

где $\varepsilon_t \sim N(0,1)$.

Далее получаемая выборка разбивалась на несколько окон данных различной длины, по которым строились и впоследствии взвешивались регрессионные модели одинакового состава. Динамическое взвешивание моделей осуществлялось согласно методу минимизации несмещенной оценки ожидаемой среднеквадратической ошибки прогноза, после чего рассчитывался поправочный коэффициент по численному методу, представленному выше, на который умножалась получаемая классическая оценка дисперсии ошибки прогноза. Для проверки эффективности работы поправочного коэффициента сравним реализованную, исправленную в соответствии с предлагаемым методом и классическую среднеквадратическую ошибку прогноза.

В таблице 4 представлены результаты имитационного эксперимента при взвешивании двух регрессионных уравнений, рассчитанных на окнах длины n_1 и n_2. Для получения реализованной среднеквадратической ошибки прогноза, исправленной и классической оценок дисперсии ошибок при каждом сочетании значений n_1 и n_2 использовалось по 10 000 имитаций. Жирным шрифтом для каждой комбинации значений n_1 и n_2 выделена та оценка средней дисперсии ошибки прогноза, которая оказалась ближе к реализованной среднеквадратической ошибке прогноза.

Таблица 4 – Реализованная, классическая и исправленная среднеквадратические ошибки прогноза при взвешивании 2-х уравнений, имитационный эксперимент

n_1	n_2	**реализованная**	**исправленная**	**классическая**
10	15	1.1829	**1.2693**	1.0405
10	20	1.1422	**1.1661**	0.9758
10	40	1.1036	**1.0489**	0.9041
10	80	1.0683	**1.0049**	0.8779
20	30	1.0885	**1.1503**	0.9981
20	60	1.0680	**1.0544**	0.9414
20	80	1.0522	**1.0308**	0.9266
20	100	1.0511	**1.0229**	0.9233

50	80	1.0461	**1.0663**	0.9765
50	100	1.0167	**1.0495**	0.9669
50	150	1.0101	**1.0252**	0.9534
50	200	1.0005	**1.0155**	0.9487
Средняя		1.0692	**1.0753**	0.9528

Источник: составлено автором

Как видно из таблицы 4, в случае динамического взвешивания двух регрессионных моделей при всех рассматриваемых сочетаниях n_1 и n_2 исправленная оценка дисперсии ошибок показывает более близкие значения к реализованной, чем классическая, которая во всех случаях ее недооценивает. Реализованная среднеквадратическая ошибка прогноза (MSRE) показывает зависимость гиперболического типа от числа наблюдений во втором окне данных n_2 при $n_1 = 10$. Классический расчет ожидаемой среднеквадратической ошибки прогноза (MSFE) также демонстрирует схожую гиперболическую связь с n_2, но при этом дает систематически заниженные оценки истинной дисперсии ошибок.

Предлагаемый метод исправленной MSFE путем умножения классической MSFE на поправочный коэффициент дает несколько завышенные оценки истинной дисперсии ошибок при n_2, относительно близком к n_1, и несколько заниженные – при n_2, существенно превышающем n_1. Схожая картина в условиях более пологой гиперболической формы наблюдается и при $n_1 = 20$. Классическая оценка истинной дисперсии ошибок при любом значении n_2 примерно на 10% недооценивает реализованную среднеквадратическую ошибку прогноза, а исправленная оценка при коротком окне данных второй модели несколько завышает оценку реализованной MSFE, при длинном – занижает. Также присутствует гиперболическая связь между рассчитанными среднеквадратическими ошибками прогноза и числом наблюдений во втором окне данных, однако в данном случае исправленная MSFE при всех рассматриваемых значениях n_2 слегка переоценивает истинную дисперсию ошибок прогноза, постепенно приближаясь к ней по мере роста n_2. Таким образом, можно сделать общий вывод о том, что при увеличении длины рассматриваемых окон данных n_1 и n_2 разница между сравниваемыми оценками и реализованной дисперсией ошибки прогноза уменьшается, что естественно, так как оценка дисперсии ошибок по каждой из взвешиваемых моделей с ростом числа наблюдений стремится к истинной.

В качестве продолжения имитационного эксперимента приведем в таблице 5 результаты применения предлагаемого поправочного коэффициента при взвешивании трех регрессионных уравнений, рассчитанных на окнах данных длины n_1, n_2 и n_3.

Таблица 5 – Реализованная, классическая и исправленная среднеквадратические ошибки прогноза при взвешивании 3-х уравнений, имитационный эксперимент

n_1	n_2	n_3	**реализованная**	**исправленная**	**классическая**
10	15	20	1.1540	**1.2177**	0.9674
10	20	40	1.0949	**1.0880**	0.8970
10	40	80	1.0708	**1.0238**	0.8702
20	30	40	1.0841	**1.1387**	0.9620
20	50	80	1.0605	**1.0559**	0.9197
20	60	100	1.0346	**1.0378**	0.9099
40	60	80	1.0207	1.0857	**0.9600**
40	80	120	1.0511	**1.0537**	0.9456
40	100	150	1.0329	**1.0377**	0.9377
Средняя			1.0671	**1.0821**	0.9299

Источник: составлено автором

По данным таблицы 5 можно видеть, что в подавляющем большинстве сочетаний n_1, n_2 и n_3 исправленная оценка дисперсии ошибок показывает более близкие значения к реализованной, чем классическая оценка. Исключение составляет только случай, когда $n_1 = 40$, $n_2 = 60$ и $n_3 = 80$, при котором разница между исправленной дисперсией и реализованной составляет 0.065, а между классической и реализованной – 0.0607. Здесь также следует отметить, что при взвешивании моделей, рассчитанных на окнах данных с незначительной разницей в длине, исправленная оценка MSFE склонна переоценивать реализованную, а со значительной разницей – недооценивать. Классическая оценка дисперсии ошибки прогноза при всех рассматриваемых сочетаниях значений n_1, n_2 и n_3 систематически недооценивает реализованную среднеквадратическую ошибку.

Таблица 6 представляет результаты имитационного эксперимента при взвешивании четырех регрессионных уравнений, рассчитанных на окнах данных различной длины n_1, n_2, n_3 и n_4.

Таблица 6 – Реализованная, классическая и исправленная среднеквадратические ошибки прогноза при взвешивании 4-х уравнений, имитационный эксперимент

n_1	n_2	n_3	n_4	**реализованная**	**исправленная**	**классическая**
10	15	20	25	1.1076	**1.2098**	0.9224
10	30	50	70	1.1042	**1.0480**	0.8532
10	40	80	120	1.0979	**1.0118**	0.8397
20	30	40	50	1.0582	**1.1328**	0.9302
20	40	60	80	1.0328	**1.0736**	0.9042
20	50	90	120	1.0416	**1.0473**	0.8972
40	50	60	70	1.0554	**1.1312**	0.9632
40	60	80	100	1.0307	**1.0847**	0.9408
40	70	100	120	1.0493	**1.0706**	0.9373
Средняя				1.0642	**1.0900**	0.9098

Источник: составлено автором

Как видно из таблицы 6, в случае взвешивания четырех моделей исправленная оценка дисперсии ошибки прогноза при любых анализируемых сочетаниях значений n_1, n_2, n_3 и n_4 является более точной оценкой реализованной среднеквадратической ошибки прогноза, чем классический расчет дисперсии ошибок. Как и в предыдущем случае взвешивания по трем уравнениям можно сделать вывод о том, что исправленная оценка MSFE несколько завышает реализованную дисперсию ошибок при плотном расположении значений длин рассматриваемых окон данных и несколько занижает при разреженном расположении n_1, n_2, n_3 и n_4. Также можно отметить тот факт, что при увеличении длин окон наблюдений классическая и исправленная оценки MSFE стремятся к реализованной среднеквадратической ошибке прогноза. При достаточно длинных окнах данные отклонения являются несущественными, однако, при моделировании временных рядов макроэкономических процессов выбор в пользу длинного окна далеко не всегда является целесообразным в силу изменчивости зависимостей между анализируемыми переменными. Вследствие этого вычисление более точной оценки дисперсии ошибки прогноза в условиях присутствия коротких окон данных представляется достаточно актуальным для получения более точных прогнозов и одновременно с этим более надежных доверительных интервалов.

В качестве заключительного этапа имитационного эксперимента рассмотрим случай, когда взвешивание происходит по пяти регрессионным уравнениям с числом наблюдений в окнах данных n_1, n_2, n_3, n_4 и n_5. Результаты представлены в таблице 7.

Таблица 7 – Реализованная, классическая и исправленная среднеквадратические ошибки прогноза при взвешивании 5-ти уравнений, имитационный эксперимент

n_1	n_2	n_3	n_4	n_5	**реализованная**	**исправленная**	**классическая**
10	20	30	40	50	1.1115	**1.1024**	0.8543
10	30	50	70	90	1.0826	**1.0373**	0.8339
10	40	70	100	130	1.0388	**1.0102**	0.8248
20	30	40	50	60	1.0634	**1.1250**	0.9092
20	40	60	80	100	1.0385	**1.0771**	0.8960
20	50	80	110	140	1.0451	**1.0426**	0.8825
Средняя					1.0633	**1.0658**	0.8668

Источник: составлено автором

Таблица 7 подтверждает эффективность предлагаемого поправочного коэффициента, поскольку исправленная оценка дисперсии ошибки прогноза во всех анализируемых случаях располагается ближе к реализованной среднеквадратической ошибке, чем классическая оценка, которая систематически занижает дисперсию ошибки прогноза. Предлагаемая исправленная оценка, как и во всех предыдущих планах имитационного эксперимента, дает завышенную оценку MSFE при плотно расположенных значениях n_1, n_2, n_3, n_4 и n_5, а при разреженных значениях – заниженную.

Для графического изображения результатов проведенного имитационного тестирования приведем гистограмму средних по показателям MSE для каждого плана эксперимента, т.е. для 2-х, 3-х, 4- х и 5-ти взвешиваемых уравнений, которую можно видеть на рисунке 5.

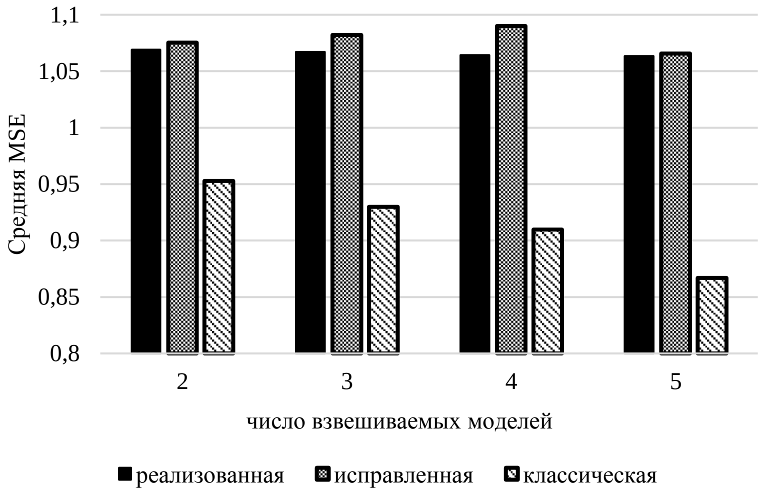

Рисунок 5 – Средние реализованная, исправленная и классическая среднеквадратические ошибки прогноза, имитационный эксперимент
Источник: составлено автором

Основным выводом, который можно сделать из рисунка 5, является увеличивающаяся разница между реализованной среднеквадратической ошибкой прогноза и классической оценкой дисперсии ошибок при добавлении подмоделей в пул взвешиваемых. При двух подмоделях классическая оценка MSFE отличалась от реализованной на 11%, при трех – на 13%, при четырех – на 14.5%, а при пяти – уже на 18.5%. Таким образом, уже при пяти альтернативных моделях классическая оценка дисперсии ошибки прогноза будет недооценивать истинную примерно на двадцать процентов, что негативным образом отразится на надежности доверительных интервалов. К примеру, 5%-ные доверительные интервалы будут пробиваться в 8% случаев. Следовательно, предлагаемый метод поправок классической оценки дисперсии ошибки прогноза позволяет существенно повысить надежность рассчитываемых доверительных интервалов, поскольку исправленная оценка при любом рассматриваемом количестве подмоделей находится в районе реализованной.

Метод учета функциональных и корреляционных зависимостей между макроэкономическими индикаторами. Пусть

$\{y_t, X_t: t = 1, \dots, n\}$ является рассматриваемым набором действительных данных, где y_t – целевая переменная, а $X_t = (1, x_{1t}, x_{2t}, \dots, x_{mt})$ – конечный набор объясняющих переменных. Тогда линейная регрессионная модель для y_t выглядит следующим образом:

$$y_t = X_t B + e_t \text{ или } \hat{y}_t = X_t B,$$

где B – вектор-столбец параметров модели, рассчитываемый по формуле (9).

Здесь отметим, что для корректной работы предлагаемого метода необходимо выполнение предпосылок МНК. Тогда функция плотности вероятности для прогнозного значения на один шаг вперед y_{n+1} является сдвинутым и масштабированным t- распределением с $\nu = n - m - 1$ степенями свободы.

$$\Psi(y_{n+1}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{MSFE_{n+1}}\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu + \frac{(y_{n+1} - \hat{y}_{n+1})^2}{MSFE_{n+1}}}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)}, \qquad (20)$$

где $\hat{y}_{n+1}$ является параметром сдвига, а $MSFE_{n+1}$ – параметром масштаба.

Далее предположим, что имеется набор целевых переменных $y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(K)}$, каждая из которых моделируется с помощью набора данных $X^{(1)}, X^{(2)}, \dots, X^{(K)}$ соответственно, а также что все рассматриваемые целевые переменные связаны между собой некоторой функциональной зависимостью:

$$y_t^{(i)} = f_i\left(y_t^{(1)}, \dots, y_t^{(i-1)}, y_t^{(i+1)}, \dots, y_t^{(K)}\right), \qquad (21)$$

где f_i обозначает функцию, которая выражает $y_t^{(i)}$ через остальные целевые переменные.

Для обеспечения надежности получаемых далее выводов предположим выполнение еще одной предпосылки:

Предпосылка 2. Отсутствие корреляции остатков рассматриваемых целевых переменных, т.е. $\operatorname{cov}\left(\varepsilon_t^{(i)}; \varepsilon_t^{(j)}\right) = 0, \forall i \neq j$. Случайные отклонения являются попарно независимыми друг от друга, что означает отсутствие систематической взаимосвязи между ошибками любых двух отдельно взятых моделей.

Тогда существует возможность провести корректировки полученных прогнозов, учитывая полученные функции плотности вероятности и известную функциональную связь между целевыми переменными. Данную процедуру предлагается проводить с помощью Метода Максимального Правдоподобия. В случае с упомянутыми выше функционально зависимыми целевыми переменными логарифм совместной плотности вероятности представляется следующим образом с использованием функции f_1:

$$\text{Log-LH} = ln\left\{\Psi_1\left(f_1\left[y_{n+1}^{(2)}, y_{n+1}^{(3)}, \dots, y_{n+1}^{(K)}\right]\right)\right\} + \\ + ln\left\{\Psi_2\left(y_{n+1}^{(2)}\right)\right\} + \dots + ln\left\{\Psi_K\left(y_{n+1}^{(K)}\right)\right\}. \qquad (22)$$

Таким образом, процедура корректировки полученных прогнозов сводится к поиску таких значений прогнозируемых случайных величин, которые максимизировали бы выражение (22). Помимо проводимых корректировок полученных прогнозов можно также получить скорректированную функцию плотности вероятности для всех целевых переменных под рассмотрением. Данную процедуру предлагается проводить с помощью вычисления маргинального распределения для анализируемой целевой переменной, которое учитывает вероятностные распределения остальных выходных переменных и функциональную связь между ними. Таким образом, согласно данному способу, скорректированная функция плотности вероятности вычисляется как показано ниже:

$$pdf\left(y_{n+1}^{(i)}\right) = \frac{\Omega\left(y_{n+1}^{(i)}\right)}{\int_{-\infty}^{\infty}\Omega\left(y_{n+1}^{(i)}\right)dy_{n+1}^{(i)}}, \qquad (23)$$

где нормировочная константа в знаменателе представляет собой интеграл функции правдоподобия по всем целевым переменным под рассмотрением, а $\Omega\left(y_{n+1}^{(i)}\right)$ вычисляется следующим образом:

$$\Omega\left(y_{n+1}^{(i)}\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Psi_1\left(y_{n+1}^{(1)}\right)\cdots\Psi_{i-1}\left(f_{i-1}\left[y_{n+1}^{(1)},y_{n+1}^{(2)},\ldots y_{n+1}^{(i-2)},y_{n+1}^{(i)}\ldots,y_t^{(K)}\right]\right)\cdot$$
$$\cdot\Psi_i\left(y_{n+1}^{(i)}\right)\cdots\Psi_K\left(y_{n+1}^{(K)}\right)dy_{n+1}^{(1)}\cdots dy_{n+1}^{(i-1)}dy_{n+1}^{(i+1)}\cdots dy_{n+1}^{(K)}.$$

Для вычисления исправленной функции плотности для корреляционно связанных процессов будем по аналогии с предыдущим случаем вычислять маргинальное распределение для анализируемой целевой переменной, которое учитывает вероятностные распределения остальных выходных переменных и корреляционную связь между ними. Таким образом, согласно данному способу, скорректированная функция плотности вероятности вычисляется как показано ниже:

$$pdf\left(y_{n+1}^{(i)}\right)=\frac{\tilde{\Omega}\left(y_{n+1}^{(i)}\right)}{\int_{-\infty}^{\infty}\tilde{\Omega}\left(y_{n+1}^{(i)}\right)dy_{n+1}^{(i)}}, \qquad (24)$$

где нормировочная константа в знаменателе представляет собой интеграл функции правдоподобия по всем целевым переменным под рассмотрением, а $\tilde{\Omega}\left(y_{n+1}^{(i)}\right)$ вычисляется следующим образом:

$$\tilde{\Omega}\left(y_{n+1}^{(i)}\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Psi_1\left(y_{n+1}^{(1)}\right)\cdot\Psi_2\left(y_{n+1}^{(2)}\right)\cdots\Psi_i\left(y_{n+1}^{(i)}\right)\cdots\Psi_K\left(y_{n+1}^{(K)}\right)\cdot$$
$$\cdot\varphi_i\left(y_{n+1}^{(i)}\middle|y_{n+1}^{(1)},\ldots,y_{n+1}^{(i-1)},y_{n+1}^{(i+1)},\ldots,y_{n+1}^{(K)}\right)dy_{n+1}^{(1)}\cdots dy_{n+1}^{(i-1)}dy_{n+1}^{(i+1)}\cdots dy_{n+1}^{(K)}.$$

В работе было проведено имитационное и эмпирическое тестирование предлагаемого метода поправок с учетом функциональной или корреляционной связи моделируемых целевых переменных. Здесь в качестве примера рассмотрим простейшее трехфакторное макроэкономическое уравнение, которое связывает показатели индекс ВВП, индекс РВВП и индекс-дефлятор ВВП таким образом, что индекс ВВП равен произведению индекса ВВП, выраженного в постоянных ценах, который отражает изменение в реальном выпуске товаров и услуг, и индекса-дефлятора ВВП, отражающего рост или падение цен в экономике. Для тестирования разработанного метода была собрана поквартальная статистика по США для данных индикаторов, начиная с Q1.1947 и заканчивая Q2.2019. Базовая система моделей представляется в виде авторегрессии четвертого порядка как показано ниже:

$$\begin{cases} \hat{I}_{pt} = b_{10} + b_{11} I_{p(t-1)} + b_{12} I_{p(t-2)} + b_{13} I_{p(t-3)} + b_{14} I_{p(t-4)}, \\ \hat{I}_{qt} = b_{20} + b_{21} I_{q(t-1)} + b_{22} I_{q(t-2)} + b_{23} I_{q(t-3)} + b_{24} I_{q(t-4)}, \\ \hat{I}_{pqt} = b_{30} + b_{31} I_{pq(t-1)} + b_{32} I_{pq(t-2)} + b_{33} I_{pq(t-3)} + b_{34} I_{pq(t-4)}. \end{cases}$$

Также добавим к сравнению модели рассматриваемых трех целевых переменных с участием всех экзогенных переменных системы выше. Таким образом, получим в некотором роде эквивалент приведенной формы системы одновременных линейных уравнений.

$$\begin{cases} \tilde{I}_{pt} = b_{10} + \sum_{i=1}^{4} b_{1i} I_{p(t-i)} + \sum_{j=1}^{4} c_{1j} I_{q(t-j)} + \sum_{k=1}^{4} d_{1k} I_{pq(t-k)}, \\ \tilde{I}_{qt} = b_{20} + \sum_{i=1}^{4} b_{2i} I_{p(t-i)} + \sum_{j=1}^{4} c_{2j} I_{q(t-j)} + \sum_{k=1}^{4} d_{2k} I_{pq(t-k)}, \\ \tilde{I}_{pqt} = b_{30} + \sum_{i=1}^{4} b_{3i} I_{p(t-i)} + \sum_{j=1}^{4} c_{3j} I_{q(t-j)} + \sum_{k=1}^{4} d_{3k} I_{pq(t-k)}. \end{cases}$$

В таблице 8 приведена среднеквадратическая реализованная ошибка прогноза по моделям, представленным выше, и моделям с поправками согласно предлагаемому методу, рассчитанным по собранной базе статистических данных.

Таблица 8 – Сравнение анализируемых методов, эмпирический эксперимент, функциональная зависимость

n	$\hat{I}_p$	$\hat{I}_q$	$\hat{I}_{pq}$	$\breve{I}_p$	$\breve{I}_q$	$\breve{I}_{pq}$	$\tilde{I}_p$	$\tilde{I}_q$	$\tilde{I}_{pq}$
20	0.9705	0.1103	0.9757	**0.9123**	**0.1064**	**0.9702**	2.4943	0.3209	2.4845
30	0.8295	0.0857	0.8598	**0.7985**	**0.0837**	**0.8362**	1.2493	0.1734	1.3475
40	0.7695	0.0775	0.8373	**0.7444**	**0.0769**	**0.7971**	1.0735	0.1281	1.1849
50	0.6878	0.0779	0.7502	**0.6592**	**0.0779**	**0.7181**	0.9584	0.1032	1.0215
60	0.6662	0.0792	0.7364	**0.6341**	**0.0791**	**0.7009**	0.8775	0.0951	0.9421
70	0.6654	0.0837	0.7683	**0.6443**	**0.0828**	**0.7217**	0.9275	0.0957	1.0054
80	0.6562	**0.0887**	0.7718	**0.6481**	**0.0887**	**0.7199**	0.8932	0.0988	0.9639

90	0.6516	**0.0866**	0.7941	**0.6461**	0.0867	**0.7388**	0.9257	0.0989	0.9981
100	0.5814	**0.0776**	0.6822	**0.5694**	0.0792	**0.6537**	0.8594	0.1081	0.8257
110	0.5294	**0.0577**	0.6628	**0.5247**	0.0581	**0.6242**	0.6032	0.0701	0.7112
120	0.5324	**0.0511**	0.6626	**0.5208**	0.0521	**0.6256**	0.5933	0.0588	0.7183

Источник: составлено автором на основе данных Бюро экономического анализа США (англ. U.S. Bureau of Economic Analysis) и Бюро трудовой статистики США

Как видно из таблицы 8, предложенный метод поправок прогнозируемых целевых переменных в подавляющем числе случаев превосходит по точности остальные модели. Исключением в данном случае являются только среднеквадратические реализованные ошибки по модели $\hat{I}_{qt}$.

Перейдем к проведению эмпирического тестирования метода поправок прогнозов макроэкономических индикаторов на основе учета корреляционных зависимостей между ними. Здесь будем рассматривать дневные данные по двум фондовым индексам США DJIA и S&P 500. С целью обеспечения стационарности моделируемых процессов преобразуем исходные данные в темпы прироста. Для удобства дальнейшего анализа обозначим целевую переменную DJIA как y_{1t}, а целевую переменную S&P 500 как y_{2t}.

В данном эксперименте моделирование рассматриваемых целевых переменных осуществляется с помощью простой авторегрессионной модели, а именно:

$$\begin{aligned} \hat{y}_{1t} &= b_{10} + b_{11} y_{1(t-1)}, \\ \hat{y}_{2t} &= b_{20} + b_{21} y_{2(t-2)}. \end{aligned} \qquad (25)$$

Здесь отметим, что во избежание высокой степени мультиколлинеарности между лаговыми переменными системы (25) и, как следствие, малой информативности поправок при такой структуре уравнений переменная y_{2t} моделируется с помощью лага $t-2$, а не $t-1$. По аналогии с предыдущими экспериментами данного раздела описанные модели тестировались на окнах данных различной длины (n) и сравнивались по показателю среднеквадратической реализованной ошибки прогноза, для расчета каждой из которых использовались дневные данные по рассматриваемым фондовым индексам США за период с 05.10.2012 по 05.10.2017.

В таблице 9 представлена среднеквадратическая реализованная ошибка прогноза для моделей (25), а также для моделей с учетом предложенных поправок и для моделей с учетом всех объясняющих переменных системы, приведенных ниже:

$$\begin{aligned} \tilde{y}_{1t} &= c_{10} + c_{11} y_{1(t-1)} + c_{12} y_{2(t-2)}, \\ \tilde{y}_{2t} &= c_{20} + c_{21} y_{1(t-1)} + c_{22} y_{2(t-2)}. \end{aligned} \tag{26}$$

По аналогии с предыдущим имитационным экспериментом для проведения поправок будем использовать две функции правдоподобия. Поправки $\breve{y}_{1.1}$ и $\breve{y}_{2.1}$ получаются посредством максимизации следующей функции правдоподобия:

$$\mathrm{LH} = \Psi_1\left(y_{n+1}^{(1)}\right) \cdot \Psi_2\left(y_{n+1}^{(2)}\right) \cdot \phi_1\left(y_{n+1}^{(1)} \middle| \, y_{n+1}^{(2)}\right), \tag{27}$$

а поправки $\breve{y}_{1.2}$ и $\breve{y}_{2.2}$ вычисляются с помощью максимизации функции правдоподобия, представленной ниже:

$$\mathrm{LH} = \Psi_1\left(y_{n+1}^{(1)}\right) \cdot \Psi_2\left(y_{n+1}^{(2)}\right) \cdot \phi_2\left(y_{n+1}^{(2)} \middle| \, y_{n+1}^{(1)}\right). \tag{28}$$

Отличием функций правдоподобия (27) и (28) является то, что в случае (27) строится регрессия для $y_{n+1}^{(1)}$ с фактором $y_{n+1}^{(2)}$, а для функции правдоподобия (28) наоборот – используется уравнение регрессии для $y_{n+1}^{(2)}$ с $y_{n+1}^{(1)}$ в качестве объясняющей переменной.

Таблица 9 – Сравнение анализируемых методов, эмпирический эксперимент, корреляционная зависимость

n	$\hat{y}_1$	$\hat{y}_2$	$\breve{y}_{1.1}$	$\breve{y}_{2.1}$	$\breve{y}_{1.2}$	$\breve{y}_{2.2}$	$\tilde{y}_1$	$\tilde{y}_2$
5	1.0316	1.1901	**0.8787**	0.9914	0.8813	**0.9731**	1.6726	1.8255
6	0.9159	1.0409	**0.7846**	0.8589	0.7929	**0.8524**	1.3907	1.4832
7	0.8381	0.9439	0.7569	0.8232	**0.7564**	**0.8107**	1.1868	1.2695
8	0.7646	0.8501	**0.7084**	0.7645	0.7134	**0.7627**	1.0383	1.1243
9	0.7322	0.8268	**0.6798**	0.7433	0.6853	**0.7418**	0.9737	1.0591
10	0.6858	0.7825	**0.6585**	0.7182	0.6603	**0.7157**	0.8821	0.9529

15	0.6429	0.7235	**0.6191**	0.6723	0.6196	**0.6701**	0.7339	0.7872
20	0.6161	0.6753	**0.5999**	**0.6424**	0.6019	0.6436	0.6789	0.7301
30	0.5799	0.6476	**0.5731**	0.6209	0.5756	**0.6208**	0.6133	0.6643
50	0.5763	0.6267	**0.5679**	**0.6111**	0.5717	0.6118	0.5968	0.6406

Источник: составлено автором на основе данных S&P Dow Jones Indices LLC

Анализируя результаты, представленные в таблице 9, можно заключить, что предлагаемый метод поправок дает меньшую по сравнению с другими методами среднеквадратическую реализованную ошибку прогноза при любом из рассматриваемых числе наблюдений. При моделировании эмпирических данных в отдельных случаях модель $\breve{y}_{1.2}$ работает несколько лучше, чем модель $\breve{y}_{1.1}$ (при числе наблюдений $n = 7$), а также $\breve{y}_{2.1}$ является более точной, чем $\breve{y}_{2.2}$ (при числе наблюдений $n = 20$ и $n = 50$).

Проведенные в диссертации имитационные и эмпирические эксперименты показывают на достаточно простых примерах практическую пользу от предлагаемых методов поправок прогнозов при одновременном моделировании временных рядов макроэкономических процессов, функционально или корреляционно зависящих друг от друга. При проведении таких поправок получаемые прогнозы являются согласованными между собой, а не «оторванными» выходами каждой из несвязанных между собой моделей, что неминуемо сказывается на их качестве. В общем и целом, разработанные методы дают значимое улучшение качества прогнозов по сравнению с регрессионными уравнениями, моделирующими каждый макроэкономический процесс по отдельности, а также работают лучше моделей, включающих все экзогенные переменные рассматриваемой системы уравнений. Данный положительный эффект достигается за счет использования незадействованной в традиционных моделях информации о виде функциональной или корреляционной зависимости между прогнозируемыми целевыми переменными. В силу того, что существует возможность функционально связать большинство макроэкономических индикаторов между собой, предлагаемый метод поправок получаемых прогнозов может считаться актуальным для комплексного прогнозирования макроэкономических систем.

III. ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

В работе предлагается численный метод расчета истинного уровня значимости предикторов при проведении процедуры спецификации регрессионного уравнения. Показывается, что при наличии некоторой системы отбора переменных в модель, традиционно вычисляемые *p*-значения предикторов являются существенно заниженными. Предлагаемый метод вычисления исправленных *p*-значений учитывает такие аспекты, как число наблюдений в окне данных, количество потенциальных объясняющих переменных, а также их дисперсионно-ковариационную матрицу. Расчет истинных *p*-значений может оказать помощь в ограничении числа ошибок как 1-ого, так и 2-ого рода в научных исследованиях, значительно снижая количество публикаций, декларирующих ложные зависимости в качестве истинных. Предложенные методы легко алгоритмизируются и могут быть интегрированы в абсолютное большинство существующих статистических пакетов.

Рассматриваются существующие методы взвешивания регрессионных уравнений и выявляются их основные недостатки. В частности, при их использовании не представляется возможным получить несмещенную ожидаемую среднеквадратическую ошибку прогноза, что значит, во-первых, что отсутствует ясное понятие об ожидаемом среднеквадратическом отклонении вычисляемого точечного прогноза и, во-вторых, нет возможности рассчитать доверительный интервал с каким-либо заданным уровнем значимости. Разрабатываемый метод оптимизации весовых коэффициентов основан на минимизации несмещенной оценки среднеквадратической ошибки прогноза (MSFE) и работает как для вложенных, так и для невложенных моделей. При использовании MSFE взвешивания мы получаем ожидаемую ошибку прогноза для комбинированной модели и доверительный интервал с заданным уровнем значимости. Проведенное имитационное и эмпирическое тестирование выявило устойчивое превосходство предлагаемого метода над существующими способами взвешивания регрессионных моделей, что, как следствие, дает ему право считаться предпочтительным в использовании при анализе и прогнозировании временных рядов социально-экономических процессов. Основная причина, по которой MSFE взвешивание работает лучше остальных методов, заключается в том, что он учитывает как корреляционную структуру ошибок взвешиваемых моделей, так и корреляционную структуру их коэффициентов, что позволяет получать более эффективные веса. Помимо этого, MSFE взвешивание точно выполняет задачу оптимизации весовых коэффициентов: выбрать такие веса, которые бы минимизировали ожидаемую ошибку прогноза.

В работе предлагается решение проблемы выбора длины окна наблюдений при моделировании временных рядов макроэкономических процессов. Эмпирически доказывается параболическая зависимость точности прогноза от числа наблюдений в окне данных по причине изменчивости зависимостей между макроэкономическими процессами во времени. В этой связи встает проблема выбора оптимального окна наблюдений, которая может быть решена либо путем выбора наилучшей согласно определенным критериям модели, либо посредством взвешивания по всему рассматриваемому набору моделей, рассчитанных на окнах данных разной длины. Разрабатывается расширение метода MSFE взвешивания регрессионных уравнений с целью получения возможности взвешивать модели, построенные на окнах данных разной длины. Данный метод основан на минимизации несмещенной оценки среднеквадратической ожидаемой ошибки прогноза. MSFE взвешивание помогает не только получить более адекватный точечный прогноз экономических показателей, но также получить доверительные интервалы для рассчитанных прогнозов с заданным уровнем значимости. Проведенные имитационные и эмпирические эксперименты подтверждают эффективность предлагаемого метода взвешивания линейных регрессионных уравнений и его применимость на практике. Главное отличие MSFE взвешивания от других заключается в том, что данный метод принимает во внимание корреляции ошибок взвешиваемых подмоделей, а также корреляции их прогнозов. Таким образом, принимая во внимание вышесказанное, можно заключить, что предлагаемый в диссертационном исследовании метод взвешивания регрессионных уравнений является самостоятельной единицей регрессионного анализа и заслуживает внимания исследователей, занимающихся прогнозированием временных рядов социально-экономических процессов.

В работе показывается, что при прогнозировании временных рядов макроэкономических процессов процедура динамической спецификации и взвешивания регрессионных моделей демонстрирует более высокую эффективность, чем одноразовая спецификация и одноразовое взвешивание, а также, что при проведении динамической спецификации или взвешивания регрессионных моделей традиционные методы расчета доверительных интервалов для полученных прогнозов дают систематически заниженные значения ширины данных интервалов. Вследствие этого возникает дилемма выбора в пользу либо более точного прогноза, либо в пользу более надежных доверительных интервалов. Предпринимается попытка решить данную проблему, а именно при сохранении повышенной точности рассчитать более надежные доверительные интервалы для прогнозируемых значений. Для этого предлагается численный метод вычисления поправочного коэффициента для классической оценки дисперсии ошибки прогноза. Предложенное решение также не дает точной оценки реализованной среднеквадратической ошибки прогноза, но тем не менее позволяет существенно повысить адекватность рассчитываемых доверительных интервалов.

В работе предлагается концептуальная модель функциональных взаимосвязей макроэкономических показателей, включающая в себя до двадцати переменных, что позволяет создавать комплексные модели, охватывающие основные макроиндикаторы государства. Показывается, что при комплексном одновременном прогнозировании временных рядов макроэкономических индикаторов можно ввести в разрабатываемую систему моделей функциональные и корреляционные взаимосвязи между моделируемыми показателями. Таким образом, получаемые прогнозы приводятся в соответствие друг с другом, что оказывает положительное влияние на точность предсказаний. Проведенные имитационные и эмпирические эксперименты показывают на достаточно простых примерах практическую пользу от предлагаемых методов поправок прогнозов при одновременном моделировании временных рядов макроэкономических процессов, функционально или корреляционно зависящих друг от друга. При проведении таких поправок получаемые прогнозы являются согласованными между собой, а не «оторванными» выходами каждой из несвязанных между собой моделей, что неминуемо сказывается на их качестве. В общем и целом, разработанные методы дают значимое улучшение качества прогнозов по сравнению с регрессионными уравнениями, моделирующими каждый макроэкономический процесс по отдельности, а также работают лучше моделей, включающих все экзогенные переменные рассматриваемой системы уравнений. Данный положительный эффект достигается за счет использования незадействованной в традиционных моделях информации о виде функциональной или корреляционной зависимости между прогнозируемыми целевыми переменными. В силу того, что существует возможность функционально связать большинство макроэкономических индикаторов между собой, предлагаемый метод поправок получаемых прогнозов может считаться актуальным для комплексного прогнозирования макроэкономических систем.

IV. ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

- *в рецензируемых научных изданиях:*

1. Moiseev, N. A. Estimation and forecast of regional competitiveness level / G. Yu. Gagarina, N. A. Moiseev, A. V. Ryzhakova, G. V. Ryzhakov // Economy of region. – 2016. – Issue 4. – Vol. 12. – P. 1040–1048. – 0,63 п.л. (авторских – 0,3 п.л.).

2. Moiseev, N. A. Improving the accuracy of macroeconomic time series forecast by incorporating functional dependencies between them / N. A. Moiseev // Ученые записки Казанского университета. Серия: Физико-математические науки. – 2018. – Т. 160. – С. 350–356. – 0,44 п.л.

3. Moiseev, N.A. Increasing the accuracy of macroeconomic time series forecast by incorporating functional and correlational dependencies between them. / N. A. Moiseev, A. Volodin // Прикладная эконометрика. – 2019. – Т. 53 (1). – С. 119–137. – 1,13 п.л. (авторских – 0,9 п.л.).

4. Моисеев, Н. А. Анализ эффективности способов спецификации уравнения регрессии / Н. А. Моисеев, А. Н. Романников // Экономический журнал. – 2017. – № 1(45). – С. 87–110. – 1,5 п.л. (авторских – 1,2 п.л.).

5. Моисеев, Н. А. Вычисление истинного уровня значимости предикторов при проведении процедуры спецификации уравнения регрессии / Н. А. Моисеев // Статистика и Экономика. – 2017. – № 3. – С. 10–20. – 0,7 п.л.

6. Моисеев, Н. А. Инновационная модель регрессионного прогноза / Н. А. Моисеев, Б. А. Ахмадеев // Инновация и инвестиции. – 2014. – № 9. – С. 123–127. – 0,31 п.л. (авторских – 0,2 п.л.).

7. Моисеев, Н. А. Инновационная экосистема как ключевой фактор экономического роста региона / Н. А. Моисеев, Б. А. Ахмадеев // Вестник Российского экономического университета им. Г.В. Плеханова. – 2016. – № 4(88). – С. 145–153. – 0,56 п.л. (авторских – 0,45 п.л.).

8. Моисеев, Н. А. Метод разложения Тейлора для увеличения прогнозной силы регрессионных уравнений / Н. А. Моисеев // Экономический журнал. – 2014. – № 3(35). – С. 51–58. – 0,56 п.л. (авторских – 0,56 п.л.).

9. Моисеев, Н. А. Метод расчета коэффициента влияния отрасли на экономику / Н. А. Моисеев, Б. А. Ахмадеев // Аудит и финансовый анализ. – 2017. – № 1. – С. 69–73. – 0,31 п.л. (авторских – 0,25 п.л.).

10. Моисеев, Н. А. Методы построения точечных прогнозов временных рядов социально–экономических показателей на региональном уровне / Н. А. Моисеев, Б. А. Ахмадеев // Экономика и предпринимательство. – 2014. – №11(52). – С. 299–302. – 0,25 п.л. (авторских – 0,15 п.л.).

11. Моисеев, Н. А. Модель пространственного динамического равновесия экономики / Н. А. Моисеев // Аудит и финансовый анализ. – 2013. – №5. – С. 134–137. – 0,25 п.л. (авторских – 0,25 п.л.).

12. Моисеев, Н. А. Повышение уровня конкурентоспособности региона посредством оптимизации бюджетной политики / Н. А. Моисеев, С. В. Манахов, О. Г. Деменко // Вестник РЭУ им. Г.В. Плеханова. – 2017. – № 2(92). – С. 169–177. – 0,56 п.л. (авторских – 0,3 п.л.).

13. Моисеев, Н. А. Сравнительный анализ эффективности методов устранения мультиколлинеарности / Н. А. Моисеев // Учет и статистика. – 2017. – № 2(46). – С. 62–73. – 0,75 п.л. (авторских – 0,75 п.л.).

14. Моисеев, Н. А. Повышение точности прогнозирования макроэкономических процессов посредством учета взаимосвязей между ними / Н. А. Моисеев // Экономический журнал. – 2017. – № 4(48). – С. 20–38. – 1,19 п.л.

15. Моисеев, Н. А. Планирование и статистическая обработка данных экспериментов в пакете R / А. С. Сорокин, Н. А. Моисеев, В. И. Митрофанов // Интеллект. Инновации. Инвестиции. – 2017. – № 1. – С. 58–64. – 0,44 п.л. (авторских – 0,3 п.л.).

- *монография:*

16. Моисеев, Н. А. Методы повышения достоверности прогнозных эконометрических исследований : монография / Н. А. Моисеев – М. : Русайнс, – 2019. – 272 с. – Текст : непосредственный. – 9,5 п.л.

- *в изданиях, индексируемых БД SCOPUS:*

17. Moiseev, N. Oil Price Predictors: Machine Learning Approach. / J. An, A. Mikhaylov, N. A. Moiseev // International Journal of Energy Economics and Policy. – 2019. – 9(5). – P. 1-6. – 0,38 п.л. (авторских – 0,15 п.л.).

18. Moiseev, N. A. Agent–based simulation of wealth, capital and asset distribution on stock markets / N. A. Moiseev, B. A. Akhmadeev //

Journal of Interdisciplinary Economics. – 2017. – Vol. 29 (2). – P. 176–196. – 1,31 п.л. (авторских – 1,0 п.л.).

19. Moiseev, N. A. Forecasting time series of economic processes by model averaging across data frames of various lengths / N. A. Moiseev // Journal of Statistical Computation and Simulation. – 2017. – № 16. – Vol. 87. – P. 3111–3131. – 1,31 п.л.

20. Moiseev, N. A. Linear model averaging by minimizing mean–squared forecast error unbiased estimator / N. A. Moiseev // Model Assisted Statistics and Applications. – 2016. – № 4. – T. 11. – P. 325–338. – 0,88 п.л.

21. Moiseev, N. A. p–Value adjustment to control type I errors in linear regression models / N. A. Moiseev // Journal of Statistical Computation and Simulation. – 2017. – № 9. – Vol. 87. – P. 1701–1711. – 0,69 п.л.

22. Moiseev, N. A. Computation of reliable interval forecast for dynamic averaging of economic time series regression models / N. A. Moiseev, O. G. Demenko, N. P. Savina // International Journal of Civil Engineering and Technology. – 2019. – № 2. – Vol. 10. – P. 1594–1602. – 0,5 п.л. (авторских – 0,3 п.л.).

23. Moiseev, N. A. Interval forecast for model averaging methods. / N. A. Moiseev, A. S. Sorokin // Model Assisted Statistics and Applications. – 2018. – 13. – P. 161-172. – 0,88 п.л. (авторских – 0,6 п.л.).

24. Moiseev, N. A. Improvement of Regression Forecasting Models / V. A. Zubakin, O. A. Kosorukov, N. A. Moiseev // Modern Applied Science. – 2015. – № 6. – Vol. 9. – P. 344–353. – 0,63 п.л. (авторских – 0,4 п.л.).