

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МЕТОДОЛОГИЯ ЭКОНОМИЧЕСКОЙ НАУКИ

М. В. Дубовик

Российский экономический университет имени Г. В. Плеханова,
Москва, Россия

С. Г. Дмитриев

Брянский филиал Российского экономического университета
имени Г. В. Плеханова, Брянск, Россия

Цель исследования – анализ общих методологических подходов, используемых в экономической теории и инструментах искусственного интеллекта, в частности, машинного обучения. По мнению авторов, существует аналогия между дебатами о реалистичных предположениях экономических моделей и актуальной дискуссией по поводу интерпретируемости при использовании машинного обучения в реальных условиях. Широко используемая концепция модели как «черного ящика» ставит под сомнение обоснованность применения таких моделей как в машинном обучении, так и в экономической теории. Помимо этого, проблема общественного доверия к использованию алгоритмов и выработка доказательной политики становятся труднодостижимыми без релевантной интерпретации функционирования моделей, несмотря на гипотетическую точность сделанных с их помощью прогнозов. В машинном обучении невозможно основываться исключительно на функционировании «черного ящика», упуская из виду роль человека при принятии решений. Аналогичным образом в экономической науке смысл описываемых моделями феноменов зачастую приносится в жертву лаконичности и стройности используемого математического и эконометрического аппарата. Анализ научных публикаций по методологическим проблемам экономики как науки позволил сделать вывод, что, несмотря (либо, напротив, благодаря) на эмпирический поворот в экономической теории, инструментальный подход Милтона Фридмана остается доминирующим, по меньшей мере в его латентном виде.

Ключевые слова: машинное обучение, интерпретируемость, доказательная политика, экономическая теория.

ARTIFICIAL INTELLECT AND METHODOLOGY OF ECONOMIC SCIENCE

Mayya V. Dubovik

Plekhanov Russian University of Economics, Moscow, Russia

Sergey G. Dmitriev

Bryansk branch of the Plekhanov Russian University of Economics,
Bryansk, Russia

The goal of the research is to analyze general methodological approaches used in economic theory and tools of artificial intellect, in particular machine learning. According to the authors there is a certain similarity between debates about realistic assumptions of economic models and the acute discussion concerning interpretability in using machine learning in real conditions. The widely used concept of the model as a 'black box' casts doubt on validity of using such models both in machine learning and in economic theory. Apart from that, the issue of public trust to applying algorithms and elaboration of evidence-based policy become hardly attainable without relevant interpretation of model functioning, in spite of hypothetic accuracy of forecasts made with their help. In machine learning it is impossible to rely only on 'black box' functioning and neglect the role of man in decision-making. In the same way in economic science the essence of phenomena described by models is often sacrificed to laconism and harmony of mathematic and econometric tool set. Analysis of academic publications on methodological problems in economics as a science allows us to draw a conclusion that in spite of (or due to) the empiric turn in economic theory Milton Friedman instrumental approach is still prevailing, at least in its latent form.

Keywords: machine learning, interpretability, evidence-based policy, economic theory.

Введение

Начиная разговор о методологии экономической науки, сложно избежать упреков в многословии, банальности и занудстве, итогом которых чаще всего являются бесполезные выводы [1. – С. 6]. Упомянутые слова Роя Харрода тем более справедливы, когда речь заходит о предполагаемом влиянии искусственного интеллекта на экономику как науку и практику. Если целью усилий экономистов в конечном счете должен стать экономический рост, то в условиях продолжающихся дебатов о причинах и самой сути экономического роста [5] в настоящее время практически невозможно убедительно утверждать, что искусственный интеллект окажет ожидаемое воздействие на динамику экономического развития.

Тем не менее инструментарий искусственного интеллекта, в частности машинное обучение, которым уже располагают экономисты-исследователи, придает актуальность спорам об экономической методологии. И в данном случае мы не будем говорить об алгоритмизации принятия управленческих решений, экономической политике и регулировании и т. д., возможности для которых предоставляет машинное обучение. Мы считаем целесообразным проследить аналогию между машинным обучением и методологией экономической науки, как бы неприязненно к ней ни относились сами экономисты [1. – С. 11]. Перефразируя слова Д. Родрика о рандомистах, можно сказать, что необходимо знать, как действует лекарство, прежде чем прописывать его пациенту [7]. Несмотря на то, что такая аналогия может показаться надуманной, последние семьдесят лет развития экономической методологии преподнесли много уроков для теоретиков и практиков машинного обучения, что может позволить избежать им паралича анализа и добиться прогресса в вопросе интерпретируемости [10]. Ценность предлагаемой аналогии для экономистов очевидна.

Результаты

Одной из самых цитируемых (и влиятельных) работ в области экономической методологии является эссе Милтона Фридмана о позитивной экономической теории [8]. Для него экономические теории состоят из двух частей: модели и области ее применения. По его мнению, модель можно рассматривать как набор предположений, а область применения – это диапазон случаев, на которых модель может быть оценена на предмет точности прогнозирования. В схеме Фридмана прогнозы – это дедуктивные выводы из модели. Ученый обращается к философии науки с более широким обсуждением выбора теории. По мысли Фридмана, точность прогнозирования в рамках модели/теории является основным критерием выбора между моделями/теориями. Такой подход близок инструментализму в отношении научных теорий. Вместо того, чтобы беспокоиться об истинности или ложности научных теорий, инструменталисты утверждают, что теории, которые лучше работают в соответствии с некоторыми показателями, например, точностью прогнозирования, должны быть предпочтительнее тех, которые этого не делают.

Отметим, что мало кто из экономических методологов и исследователей придерживается такого рода инструменталистского подхода. Многие экономисты-методологи уверены, что эссе, апеллирующее к предсказательной силе без конкретизации предсказаний, написанное автором, который защищает нереалистичные предположения, прежде чем подвергать сомнению предположения своих оппонентов, не должно быть основой методологии какой-либо дисциплины.

Научное понимание роли экономических моделей находится под влиянием двух ключевых концепций: инструментализма и реализма. С одной стороны, интуитивно может показаться необходимым, чтобы экономические модели отражали сложные причинно-следственные механизмы, действующие в реальном мире,

чтобы обеспечить высокую описательную точность. С другой стороны, включение упрощенных предположений в модели может быть обусловлено не только аналитической легкостью, но и другими прагматическими соображениями, что может привести к ограничениям в точности моделей.

Вопрос о том, как найти баланс между прогностической точностью и описательной реалистичностью моделей, остается нерешенным и вызывает дискуссии в сфере экономического моделирования. Подход к выбору между использованием точных, но нереалистичных моделей и менее точных, но более реалистичных моделей представляет собой сложную проблему, требующую внимательного анализа и оценки целей и контекста конкретного исследования. Таким образом, поиск оптимального баланса между точностью и реализмом в экономических моделях остается актуальным направлением исследований в экономической науке.

У научного реализма/инструментализма есть два измерения: аксиологическое и эпистемологическое. Д. Рейсс описывает аксиологическое измерение с помощью метафоры «зеркала природы». Рейсс утверждает, что дескриптивные теории и модели можно условно отнести к реализму. В свою очередь инструментальное направление призвано снабдить исследователя либо практика необходимым набором инструментов, релевантных их целям. В рамках данной парадигмы описание и понимание, которые являются целями реалистичного направления, носят вспомогательный характер (в лучшем случае). Проблема познания истины в качестве цели деятельности отсутствует, к этому понятию инструменталисты обращаются только в том случае, если это может помочь достижению целей [12].

В целом Д. Рейсс выступает в пользу инструментализма в экономике, поскольку практикующие специалисты заботятся о результатах истинных/причинных пред-

положений, а не об истине и причинно-следственной связи.

Похожих взглядов придерживается и Д. Хаусман, утверждая, что хорошая гипотеза дает достоверные и значимые предсказания относительно класса явлений, которые она призвана объяснить [13], реконструируя тем самым аргументы М. Фридмана в упомянутом эссе. Хаусман утверждает, что теории и/или модели применяются, чтобы направлять нас в новых обстоятельствах, и их следует оценивать соответствующим образом. Учитывая точку зрения Фридмана на цели науки, по мысли Хаусмана, не было бы смысла исследовать предположения теории, если бы можно было провести полную оценку ее эффективности по отношению к явлениям, которые она призвана объяснить, что (по определению) невозможно.

Взгляды Д. Хаусмана интуитивно понятны специалистам по машинному обучению, например, при рассмотрении поведения модели как странных точек в пространстве признаков. Так, например, для какого-то приложения мы выберем древовидную модель вместо полиномиальной регрессии, если распределение наших данных широкое и регрессия асимптотическая в некоторых направлениях.

Интересный подход предложен в работе Я. Куорикоски, А. Лехтинен и К. Маркионни, которые делят предположения экономических моделей на три категории: существенные предположения, предположения Галилея и предположения управляемости. Существенные предположения определяют набор причинных факторов, которые во взаимодействии составляют причинный механизм, относительно которого разработчик модели пытается сделать важные утверждения. Предположения Галилея изолируют причинный механизм, смоделированный существенными предположениями. Предположения об управляемости – это гипотезы о том, что какая-то функция линейна или проста и т. д. [9].

Методология Фридмана близка к машинному обучению, рассматриваемому

как инженерная задача. Фридман пишет об экономической теории так же, как ученые, работающие с данными. Инженеры по машинному обучению говорят о моделях «черного ящика».

Модели машинного обучения – это ассоциативные правила между переменными, которые могут отражать или не отражать реальные отношения, а область применения моделей определяется распределениями данных, аналогичными обучающим данным. Как мы отметили, Фридман считает, что экономическая теория – это построение моделей, а не объяснение взаимосвязей между экономическими переменными или выделение причинных механизмов.

В начале XXI в. методология экономической теории оказывается под влиянием эмпирического поворота. Доля эмпирических статей существенно увеличивается с течением времени практически по всем направлениям. Как отмечает Д. Ангрист, уже сейчас доля цитирований эмпирических исследований достигает 50 процентов [6]. Если в прошлом экономическая методология была сосредоточена на формулировании научных законов, то сегодняшняя методологическая работа – на причинности и оценке. Попутно заметим, что цитированная статья Ангриста написана с помощью машинного обучения, хотя, несмотря на широкое распространение, модели машинного обучения по-прежнему остаются «черными ящиками».

Использование концепции «черного ящика» широко распространено в самых разнообразных сферах деятельности: от академической философии до кибернетики. Насколько мы можем проследить генезис этой метафоры, одним из первых ее употребил Б. Латур, описывая функционирование кибернетических устройств, слишком сложных для понимания рядового пользователя, которому известна только входящая и исходящая информация, а что происходит внутри «черного ящика» остается невыясненным/непонятым [3. – С. 25].

На уровне практикующего специалиста по машинному обучению либо экономиста-исследователя не имеет значения, каково окончательное обоснование того или иного метода или методологии, т. е. то, что «под капотом» (вспомним название работы Д. Хаусмана). Все, что имеет значение для практикующего специалиста, это достигнет ли метод или методология целей прогнозирования с достаточной степенью точности и надежности.

Предиктор «черного ящика» – это неясная модель интеллектуального анализа данных и машинного обучения, внутренние данные которой либо неизвестны наблюдателю, либо известны, но не поддаются интерпретации людьми. Интерпретировать – это значит придавать значение или объяснять и представлять в понятных терминах некоторые концепции. Таким образом, в интеллектуальном анализе данных и машинном обучении интерпретируемость определяется как способность объяснять или предоставлять смысл в понятных для человека терминах. Эти определения неявно предполагают, что понятия, выраженные в понятных терминах, составляющих объяснение, являются самостоятельными и не нуждаются в дальнейших объяснениях. По сути, объяснение – это «интерфейс» между людьми и лицом, принимающим решения, который одновременно является точным представителем лица, принимающего решения, и понятен другим пользователям.

Объяснение является ключевым элементом в области искусственного интеллекта, поскольку оно позволяет людям понять работу «черного ящика» и принимать обоснованные решения на основе его выводов. Важно отметить, что интерпретируемость моделей машинного обучения становится все более актуальной в свете потребности в объяснении принципов работы алгоритмов и прогнозов, особенно в таких критически важных областях, как здравоохранение, финансы и юстиция. Обеспечение прозрачности и понятности алгоритмов «черного ящика» остается

важным направлением исследований в сфере искусственного интеллекта и машинного обучения для обеспечения справедливости, безопасности и доверия в цифровом мире. Например, мы можем представить, что пользователь получил модель (например, через Application programming interface) без какой-либо подробной информации о том, как она была построена, кроме знания ее входных/выходных данных, т. е. классическую модель «черного ящика». При этом модель может быть сколь угодно сложной, однако сложность не должна приноситься в жертву интерпретируемости и простоте модели, поскольку в действительности любые попытки упростить модель, сделав ее более понятной, на самом деле только приводят к большему беспорядку, подобно тому, как спонтанные процессы вызывают рост энтропии [4. – С. 12].

Сосредоточиваясь на интерпретируемости, мы можем неявно предполагать, что интерпретируемость является средством достижения целей точности, установления причинности и т. п. целей машинного обучения. Это должно вызывать подозрения, поскольку интерпретируемость зависит от предметной области и не имеет единого определения, что затрудняет использование названной концепции при попытке решить проблемы, которые послужили причиной призыва к интерпретируемости (т. е. проблему «черного ящика»). Более того, поскольку не существует адекватного описания самой концепции, невозможно оценить, обеспечивают ли конкретные технические характеристики формальные определения этой концепции.

Одной из целей использования исследователями алгоритмов машинного обучения является создание причинно-следственного вывода (Causal Inference), отражающего каузальную связь факторов. Эмпирический поворот в экономике усилил интерес исследователей к использованию инструментов машинного обучения, хотя эта тенденция ярко выражена во многих направлениях научного знания.

Мы полагаем, что методология экономической науки, восходящая к взглядам М. Фридмана, и машинное обучение помимо прочего обладают еще одним общим свойством: оба феномена проблематичны с точки зрения общественного доверия к их применению. Общественное признание алгоритмов машинного обучения требует, чтобы конечные пользователи имели некоторое представление об их (алгоритмов) принципах и особенностях функционирования, что повышает значимость проблемы интерпретации. Здесь мы сталкиваемся с необходимостью разделения профессионального и общественного, а зачастую и обывательского дискурсов. Однако на практике провести их четкое разграничение затруднительно в связи с тем, что представители, например, академического сообщества, занимающиеся проблемой интерпретируемости искусственного интеллекта, выступая в средствах массовой информации, вступают на поле обывательских представлений, по сравнению с которыми нечеткая концепция «черного ящика» может оказаться образцом определенности. Это, безусловно, не способствует росту доверия к использованию машинного обучения, результаты которого предположительно можно применять для выработки доказательной политики, а равно и к идее, что интерпретируемость должна способствовать признанию обществом в целом алгоритмов и доверию к их использованию. Все же настоятельно необходимо уделять повышенное внимание нечетким границам между общественной и академической сферами при обращении к общественным потребностям. Это вполне справедливо и в отношении методологии экономической науки, сосредоточенной на прогнозах, но не способной объяснить текущие обстоятельства.

Обсуждение

Проблемой, заслуживающей отдельного рассмотрения, является доверие к прогнозам, при этом необходимо отделять его от доверия к моделям. Однако и то и другое

напрямую зависит от того, насколько человек понимает поведение модели для того, чтобы предпринимать на ее основе какие-либо действия, последствия которых описываются прогнозом. Это приводит нас к мысли, что невозможно основываться исключительно на функционировании «черного ящика», упуская из виду роль человека при принятии решений. Когда машинное обучение используется, например, для медицинской диагностики или выявления угрозы терроризма, прогнозы не могут основываться на принципе «черного ящика», поскольку последствия могут быть катастрофическими.

Применению искусственного интеллекта, минимизирующего участие человека, подкрепленному когнитивными и поведенческими рутинными высших приматов, должен быть противопоставлен техноскептицизм [11]. Без этого широкое использование искусственного интеллекта и машинного обучения подорвет начавшийся переход к доказательной политике, который и без того вызывает многочисленные сомнения, особенно после фармацевтических (и нефармацевтических) мер по препятствованию распространения пандемии SARS-CoV-2.

Заключение

Проведенный нами анализ источников по истории науки не дает поводов для оптимизма в отношении необходимой смены методологической парадигмы экономической науки. Новые методологические подходы, особенно опирающиеся на страдающие от проблемы интерпретируемости инструменты искусственного интеллекта, еще находятся в стадии становления, а отказ от парадигмы, согласно Т. Куну, означает отказ от науки. Критика инструменталистского подхода, пренебрегающего точным (насколько это достижимо) описанием окружающей реальности, никогда не достигала своей цели. Добавим, что она и не может ее достичь, поскольку, во-первых, критикуемый подход, восходящий к упомянутому эссе М. Фридмана, попросту отвергает необходимость соответствия модели «природа», а во-вторых, никогда ранее смена научной парадигмы (или разрешение теоретических разногласий) не осуществлялась посредством сравнения теории/модели, разработанной в рамках определенной методологии, с тем, что называется окружающим миром [2. – С. 108, 106; 3. – С. 167].

Список литературы

1. Блауг М. Методология экономической науки, или Как экономисты объясняют : пер. с англ. / науч. ред. В. С. Автономов. – М. : НП «Журнал Вопросы экономики», 2004.
2. Кун Т. Структура научных революций : пер. с англ. / сост. В. Ю. Кузнецов. – М. : ООО «Издательство АСТ», 2003.
3. Латур Б. Наука в действии: следуя за учеными и инженерами внутри общества. – СПб. : Изд-во Европейского университета в Санкт-Петербурге, 2013.
4. Ло Д. После метода: беспорядок и социальная наука / пер. с англ. С. Гавриленко, А. Писарева, П. Хановой; науч. ред. перевода С. Гавриленко. – М. : Изд-во Института Гайдара, 2015.
5. Хелпман Э. Загадка экономического роста / пер. с англ. А. Калинина; под ред. М. Ханаевой, Е. Синельниковой. – М. : Изд-во Института Гайдара, 2011.
6. Angrist J. et al. Economic Research Evolves: Fields and Styles // American Economic Review. – 2017. – Vol. 107. – N 5. – P. 293–297.
7. Favereau J. On the Analogy between Field Experiments in Economics and Clinical Trials in Medicine // Journal of Economic Methodology. – 2016. – Vol. 23. – N 2. – P. 203–222.
8. Friedman M. Essays in Positive Economics. – University of Chicago Press, 1953.

9. Kuorikoski J., Lehtinen A., Marchionni C. Economic Modelling as Robustness Analysis // *The British Journal for the Philosophy of Science*. – 2010. – N 61. – P. 541–567.
10. Mullins B. Economic Methodology Meets Interpretable Machine Learning – Introduction. – URL: https://bcmullins.github.io/economic_methodology_interpretable_ml_intro/ (дата обращения: 27.02.2024).
11. Newman J. Mapping the Discourse on Evidence-Based Policy, Artificial Intelligence, and the Ethical Practice of Policy Analysis // *Journal of European Public Policy*. – 2023. – Vol. 30. – Issue 9. – P. 1839–1859.
12. Reiss J. Idealization and the Aims of Economics: Three Cheers for Instrumentalism // *Economics & Philosophy*. – 2012. – Vol. 28. – N 3. – P. 363–383. – URL: <https://doi.org/10.1017/S0266267112000284>
13. *The Philosophy of Economics: An Anthology*. – 2nd edition / edited by D. M. Hausman. – Cambridge : Cambridge University Press, 2002. – P. 17–21.

References

1. Blaug M. *Metodologiya ekonomicheskoy nauki, ili Kak ekonomisty obyasnyayut* [Methodology of Economic Science or How Economists Explain], translated from English, scientific editor by V. S. Avtonomov. Moscow, NP 'Journal Economic Issues', 2004. (In Russ.).
2. Koon T. *Struktura nauchnyh revolyutsiy* [Structure of Scientific Revolutions], translated from English, compiled by V. Yu. Kuznetsov. Moscow, Publishing House AST, 2003. (In Russ.).
3. Latur B. *Nauka v deystvii: sleduya za uchenymi i inzhenerami vnutri obshchestva* [Science in Action: Following Scientists and Engineers inside Society]. Saint Petersburg, Publishing House of the European University in Saint Petersburg, 2013. (In Russ.).
4. Lo D. *Posle metoda: besporyadok i sotsialnaya nauka* [After Method: Disorder and Social Science], translated from English by S. Gavrilenko, A. Pisarev, P. Hanova; scientific translation editor by S. Gavrilenko. Moscow, Publishing House of the Gaidar Institute, 2015. (In Russ.).
5. Helpman E. *Zagadka ekonomicheskogo rosta* [Mystery of Economic Growth], translated from English by A. Kalinin; edited by M. Hanaeva, E. Sinelnikova. Moscow, Publishing House of the Gaidar Institute, 2011. (In Russ.).
6. Angrist J. et al. Economic Research Evolves: Fields and Styles. *American Economic Review*, 2017, Vol. 107, No. 5, pp. 293–297.
7. Favereau J. On the Analogy between Field Experiments in Economics and Clinical Trials in Medicine. *Journal of Economic Methodology*, 2016, Vol. 23, No. 2, pp. 203–222.
8. Friedman M. *Essays in Positive Economics*. University of Chicago Press, 1953.
9. Kuorikoski J., Lehtinen A., Marchionni C. Economic Modelling as Robustness Analysis. *The British Journal for the Philosophy of Science*, 2010, No. 61, pp. 541–567.
10. Mullins B. Economic Methodology Meets Interpretable Machine Learning – Introduction. Available at: https://bcmullins.github.io/economic_methodology_interpretable_ml_intro/ (accessed 27.02.2024).
11. Newman J. Mapping the Discourse on Evidence-Based Policy, Artificial Intelligence, and the Ethical Practice of Policy Analysis. *Journal of European Public Policy*, 2023, Vol. 30, Issue 9, pp. 1839–1859.

12. Reiss J. Idealization and the Aims of Economics: Three Cheers for Instrumentalism. *Economics & Philosophy*, 2012, Vol. 28, No. 3, pp. 363–383. Available at: <https://doi.org/10.1017/S0266267112000284>

13. The Philosophy of Economics: An Anthology. 2nd edition, edited by D. M. Hausman. Cambridge, Cambridge University Press, 2002, pp. 17–21.

Поступила: 29.03.2024

Принята к печати: 17.06.2024

Сведения об авторах

Майя Валериановна Дубовик

доктор экономических наук, доцент,
профессор кафедры экономической теории
РЭУ им. Г. В. Плеханова.

Адрес: ФГБОУ ВО «Российский экономический
университет имени Г. В. Плеханова», 109992,
Москва, Стремянный пер., д. 36.
E-mail: mvdubovik@list.ru

Сергей Геннадьевич Дмитриев

кандидат экономических наук, научный
сотрудник кафедры экономики, таможенного
дела, информационных технологий
и дисциплин естественно-научного цикла
Брянского филиала РЭУ им. Г. В. Плеханова.

Адрес: Брянский филиал ФГБОУ ВО
«Российский экономический университет
имени Г. В. Плеханова», 241050, Брянская
область, Брянск, ул. Бежицкая, д. 8.
E-mail: sergey.g.dmitrieff@gmail.com

Information about the authors

Mayya V. Dubovik

Doctor of Economics, Assistant Professor,
Professor of the Department
for Economic Theory of the PRUE.

Address: Plekhanov Russian University
of Economics, 36 Stremyanny Lane,
Moscow, 109992, Russian Federation.
E-mail: mvdubovik@list.ru

Sergey G. Dmitriev

PhD, Researcher of the Department
for Economics, Customs, Information
Technology and Disciplines
of the Natural Science Cycle
of Bryansk branch of the PRUE.

Address: Bryansk branch of the Plekhanov
Russian University of Economics,
8 Bezhitskaya Str., Bryansk, Bryansk Region,
241050, Russian Federation.
E-mail: sergey.g.dmitrieff@gmail.com