

УДК 519.862.6

DOI: 10.34670/AR.2023.53.28.004

Особенности применения модели Хекмана в социально-экономических исследованиях

Алтухова Ирина Анатольевна

Кандидат педагогических наук, доцент,
доцент кафедры экономики и цифровых технологий,
Дальневосточный институт управления – филиал РАНХиГС,
680000, Российская Федерация, Хабаровск, ул. Муравьева-Амурского, 33;
e-mail: altukhova-ia@ranepa.ru

Волкова Валерия Вадимовна

Студент,
Дальневосточный институт управления – филиал РАНХиГС,
680000, Российская Федерация, Хабаровск, ул. Муравьева-Амурского, 33;
e-mail: lerka007.v@mail.ru

Никитина Елизавета Ростиславовна

Студент,
Дальневосточный институт управления – филиал РАНХиГС,
680000, Российская Федерация, Хабаровск, ул. Муравьева-Амурского, 33;
e-mail: lisunenok@mail.ru

Пынько Люсьена Евгеньевна

Кандидат экономических наук,
доцент кафедры экономики и цифровых технологий,
Дальневосточный институт управления – филиал РАНХиГС,
680000, Российская Федерация, Хабаровск, ул. Муравьева-Амурского, 33;
e-mail: Lusiena_03@mail.ru

Аннотация

Модель Хекмана известна как двухшаговая модель, которая позволяет учитывать скрытые переменные в анализе данных. Модель была предложена экономистом Джеймсом Хекманом и нашла применение в различных областях, включая экономику, социологию, психологию, политологию и др. В такого рода моделях оценивается вероятность наличия скрытой переменной, которая затем используется для корректировки оценок результирующего моделирования. В исследовании современных социально-экономических процессов, часто сталкиваются с влиянием скрытых факторов на результирующие показатели, что искажает результаты исследований и прогнозы по ним. При этом точность прогнозирования важна в принятии решений. В моделях панельных данных и в системах эконометрических уравнений, вносить ненаблюдаемые скрытые

непрерывные переменные нецелесообразно, это затрудняет и искажает результаты исследований. Современный нейросетевой анализ не всегда способен отражать структурное влияние отдельных факторов на результирующий показатель. Модель Хекмана позволяет получить более точные оценки влияния наблюдаемых факторов на зависимую переменную, принимая во внимание влияние скрытых переменных, что учитывается в обосновании и выводах по модели. При этом модель имеет отдельные ограничения в исследованиях.

Для цитирования в научных исследованиях

Алтухова И.А., Волкова В.В., Никитина Е.Р., Пынько Л.Е. Особенности применения модели Хекмана в социально-экономических исследованиях // Экономика: вчера, сегодня, завтра. 2023. Том 13. № 10А. С. 15-22. DOI: 10.34670/AR.2023.53.28.004

Ключевые слова

Двухшаговая модель, модель Хекмана, зависимая переменная, корректировка, ненаблюдаемые переменные, скрытые переменные, скрытые непрерывные переменные, эконометрический анализ.

Введение. Предпосылки применения модели Хекмана

При исследовании экономических и социальных процессов, а также исследованиях в других отраслях знаний, например, в психологии, распространение получила модель Хекмана, известная как двухшаговая модель. Модель Хекмана является статистической моделью, которая используется для анализа данных, при этом, в модели результаты наблюдения могут быть смещены, то есть данные содержат скрытую переменную [Rabe-Hesketh, 2006; Могильчак, 2015]. Сама модель была предложена в 1979 году экономистом Джеймсом Хекманом, и интересна тем, что многие социально-экономические наблюдения в их статистическом измерении не раскрывают причинно-следственные связи, влияющие на их результирующие значения, и, тем более, плохо поддаются классическому эконометрическому анализу и прогнозированию средствами МНК – оценок (оценок метода наименьших квадратов). Для корректного прогнозирования отдельных социально-экономических показателей необходимо учитывать скрытые факторы влияния на них [Орлова, 2021; Puhani, 2000].

В современных направлениях научного знания, в научных исследованиях, в экспертном оценивании и т.п. исследователи сталкиваются со множеством флуктуаций в отобранных наблюдениях (для анализа и моделирования). Очень часто при построении модели исходят из рациональных предпосылок влияния факторов на нее. Но при этом ученые-исследователи в своем анализе приходят к выводам, что на модель влияют не только выбранные переменные, но и включенные в них скрытые переменные, которые не были явно учтены на начальном этапе моделирования. Именно скрытые переменные вызывают флуктуации в модели, либо в оценивании модели указывают на ее полностью случайную природу, когда доля дисперсии зависимой переменной, объясняется случайными факторами, а не включенными в модель независимыми (объясняющими) переменными.

В научном анализе при исследовании социально-экономических, психологических моделей, существуют различные теории, иногда диаметрально противоположные, которые оценивают один и тот же фактор с позиции разных групп наблюдений. Наглядным примером могут служить

подходы к исследованию модели рационального экономического поведения (А. Тверски, Д. Канеман, Р. Талер, М. Фридман и др.), в котором рациональность экономических агентов в теории отличается от их фактического поведения. Подобные вариации иногда уводят исследователей в сторону от истины.

Например, теория оценки предпочтений потребителей: насколько потребительские предпочтения рациональны и осмыслены самими потребителями – этот исследуемый фактор часто приводит к неожиданным результатам, если используются данные анонимных опросов, исследование которых, в свою очередь, показывает спонтанность и иррациональность выбора части потребителей. Отсюда возможны ошибки в маркетинговых стратегиях, или наоборот, чрезмерное использование манипуляций сознанием потребителей отдельными маркетинговыми приемами.

Приведем другой пример, исследуется влияние образования на заработную плату. При этом исследователи располагают информацией, что люди, которые имеют высшее образование, могут иметь более высокую мотивацию работать и зарабатывать больше денег, чем люди без высшего образования. Это может привести к смещению результатов, искажению оценок влияния образования на реальную заработную плату. Учитывая современные тенденции, которые усиливают значимость компетенций работников, их уровня квалификации, необходимо спроектировать модель таким образом, чтобы включить в нее скрытые факторы, влияющие на уровень заработной платы и минимизировать искажение оценок.

Условия построения модели Хекмана

При исследовании на первом шаге модели Хекмана оценивается вероятность того, что наблюдение будет иметь значение скрытой переменной. Эта вероятность затем используется во втором шаге для корректировки оценок параметров, полученных в регрессии зависимой переменной на наблюдаемые факторы [Heckman, 2005].

Кроме того, модель Хекмана может использоваться для анализа данных с самостоятельным отбором, когда наблюдения сами выбираются для участия в исследовании, что может привести к смещению результатов [там же; Юдин, 2022].

Возвращаясь к вопросу исследования модели влияния высшего образования на заработную плату, отметим порядок исследования: на первом шаге этой модели оценивается вероятность того, что человек имеет высокую мотивацию работать и зарабатывать больше. Затем найденная вероятность используется на втором шаге для корректировки оценок параметров модели: полученные в регрессии значения зависимой переменной и их влияние на наблюдаемые факторы, такие как образование, опыт работы, возраст и т.д.

Результаты анализа с помощью модели Хекмана показывают, что влияние образования на заработную плату значительно уменьшается после учета скрытой переменной – мотивации человека работать и зарабатывать больше. Это может означать, что не образование само по себе влияет на заработную плату, а скрытая переменная – мотивация человека работать и зарабатывать больше, которая может быть связана с образованием и другими факторами, в т. ч. получение нескольких квалификаций. При этом скрытыми факторами могут выступать нравственно-религиозные убеждения, когда работники видят значимость не в накоплении денег, а, например, в семейном благополучии, которое предполагает большую часть времени отводить семье, а не работе. Возможны флуктуации, связанные с использованием отдельными индивидами моделей быстрого обогащения. Тогда само исследование влияния образования на

заработную плату становится несостоятельным, т.к. возникает потребность исследовать нравственно-социальные и правовые аспекты общественной выборки. Отметим, что общество, в котором продвигаются модели быстрого обогащения, подвергается угрозам развитию нравственно-здоровых норм, и, отчасти, представляет угрозу стабильности (безопасности) самого государства [Могильчак, 2015].

В современных реалиях при исследовании социальных, экономических, психологических, политических и др. процессов, учитывая те события, которые в настоящее время формируются в мире, трудности в их прогнозировании, отсутствие возможностей точного планирования, модель Хекмана весьма актуальна и уместна в исследованиях.

Анализ данных с помощью модели Хекмана позволяет более точно оценить влияние наблюдаемых факторов на зависимую переменную, учитывая влияние скрытых переменных и смещение результатов.

Основной целью исследования было сравнить модели отбора типа Хекмана с альтернативными подходами в эконометрических исследованиях. Модель Хекмана включает в себя два отдельных уравнения – одно, фокусирующееся на отборе в выборку (наблюдаемый результат – уравнение отбора выборки), и основное уравнение, связывающее представляющие интерес ковариаты с результатом [Cubedo, 2002; Heckman, 1974, 1979, 2006].

Два уравнения Хекмана для двух скрытых реакций y_i^* (результат) и s_i^* (переменная склонности к выбору) могут быть сформулированы следующим образом [Rabe-Hesketh, 2006]:

$$\begin{aligned} y_i^* &= x_i^* \beta + \mu_i \\ s_i^* &= z_i^* \gamma + v_i, \end{aligned} \tag{1}$$

где y_i^* и s_i^* – ненаблюдаемые скрытые непрерывные переменные, x_i^* и z_i^* – векторы переменных-предикторов. В общем случае предполагается, что x является подмножеством z , что означает, что все факторы, предсказывающие основной интересующий результат (y), также предсказывают выбор s , μ и v – нормально распределенные ошибки, а β – представляет вектор основного параметра [Heckman, 1974, 1979]. Переменные результата будут наблюдаться, если скрытая склонность к отбору превышает ноль, то есть:

$$s_i = \begin{cases} 1 & \text{если } s_i^* > 0 \\ 0 & \text{если } s_i^* \leq 0 \end{cases} \tag{2}$$

Основная идея модели Хекмана заключается в том, что теоретически представляется довольно вероятным, что ненаблюдаемые или не измеряемые факторы могут влиять как на результат y , так и на вероятность выбора s ; эти неизмеренные факторы будут содержаться в остатках как уравнения y_i^* , так и уравнения s_i^* . Принимая во внимание отбор в основную выборку, ожидаемое значение результата в основном уравнении определяется как:

$$E(y|z, v) = x\beta + E(\mu|v) \tag{3}$$

Учитывая, что ковариаты x и v совместно определяют отбор в выборку, $cov(x, v|s = 1)$ в целом отличен от нуля, так что бета-оценки будут как предвзятыми, так и противоречивыми, если μ и v коррелируют [там же]. Эту корреляцию легко оценить эмпирически, подобрав независимые модели для y и s и вычислив ковариацию между двумя остаточными членами. Модель Хекмана показывает, что это смещение может быть исправлено путем вычисления

ожидаемого значения v , обусловленного z и находящегося в выборке, а также путем включения этого термина в основную эмпирическую модель [Muhamad Safiih, 2007]. Согласованные оценки могут быть получены путем совместной оценки максимального правдоподобия на первом этапе с использованием пробит-модели, а также основного уравнения, представляющего интерес, включая ожидаемое значение остатков уравнения выбора [Mahmudah, 2009].

Альтернативные эконометрические и нейросетевые модели

В контексте еще одного популярного направления эконометрики – использования панельных данных в анализе, отметим, что построение порядковых пробит моделей с несколькими категориями отклика довольно затруднительно. То есть модель Хекмана оптимально применять, как самостоятельную модель, позволяющую оценить и скорректировать влияние на результирующий фактор ненаблюдаемых и не измеряемых факторов. Сами же ненаблюдаемые факторы могут быть ожидаемыми, но могут быть и предвзятыми. В панельных данных исследуются модели с фиксированными и случайными эффектами, и в них независимые факторы измеримы. Вносить в такие модели ненаблюдаемые скрытые непрерывные переменные нецелесообразно, т. к. это затрудняет и искажает результаты исследований. Но, принимая во внимание, что для исследуемых социально-экономических процессов возможно построить модель панельных данных, оценить ее на наличие фиксированных эффектов, а затем, вернувшись к исходным статистическим наблюдениям, построить для них отдельную модель Хекмана для перепроверки отдельных гипотез (ожидаемых или противоречивых). Тогда будет реализована система по исследованию интересующего социально-экономического процесса с перекрестной перепроверкой отдельных гипотез и, в этом случае, модель Хекмана может существенно дополнить результаты исследования панельных данных.

Современные направления в моделировании, которые используют нейросетевое прогнозирование и машинное обучение, позволяют строить относительно точные прогнозы в исследовании интересующего социально-экономического процесса. Но, в зависимости от того, какие методы используются в разработке нейросетей, на основе каких прототипов происходило обучение нейронных сетей, какие методы машинного обучения были заложены, часто исследователи, получая прогнозы, не имеют к ним обоснования или пояснения. Почему нейросеть определяет именно такой интервал прогноза? Что, в свою очередь, не позволяет выполнить объективный и полный анализ изучаемого процесса или явления. Авторы, исследуя модель влияния образования на мотивацию работников зарабатывать посредством построения простого перцептрона в KNIME, получили неточный прогноз, и главное, отсутствие обоснования влияния скрытых факторов на зависимую переменную, т.е. нейросеть не обосновала смещение результатов в исследовании. Процесс же эконометрического анализа этой модели оказался лучшим, но достаточно объемным, что не позволяет в полной мере его отразить в рамках одной статьи.

Заключение

В результате исследования пришли к следующим выводам:

Использование модели Хекмана в системах эконометрических уравнений также затруднительно [Puhani, 2000]. Ввиду своей трудоемкости при реализации решения системы с включением в нее анализа условий Хекмана, т. к. включение в систему приводит к усложнению

и громоздкости модели исследования, предполагаем, что данный подход в исследованиях нецелесообразен.

Оптимально использование модели Хекмана, как автономной модели.

Модель Хекмана широко используется в экономических и социологических исследованиях, в которых возникает проблема смещения выборки [там же]. Выводы, полученные с помощью модели Хекмана, могут быть более точными и надежными, чем выводы, полученные без учета смещения выборки. Однако, как и любая статистическая модель, модель Хекмана имеет свои ограничения и требует определенных предположений и условий для корректной работы.

Библиография

1. Могиляк Е.Л. Выборочный метод в эмпирическом социологическом исследовании. Екатеринбург, 2015. 120 с.
2. Орлова Е.В. Эконометрическая методология исследования систем. М.: ИНФРА-М, 2021. 221 с.
3. Юдин С.В. Информационно-статистические методы решения эконометрических, социологических и психометрических задач. М.: ИНФРА-М, 2022. 199 с.
4. Cubedo M., Oller J.M. Hypothesis testing: Model Approach to selection // Magazine Statistical Planning and Inference. 2002. 108. P. 3-21.
5. Heckman J. Sample Bias as a specification error // *Econometrica*. 1979. 47. P. 152-161.
6. Heckman J. Structural Equations, Treatment Effects, and Econometric Policy Evaluation // *Econometrica*. 2005. Vol. 73. No. 3. P. 669-738.
7. Heckman J.J. Shadow Wages, Market Wages and labor supply // *Econometrics*. 1974. 42. P. 679-693.
8. Heckman J., Lochner L., Todd P. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond // *Handbook of the Economics of Education*. 2006. Vol. 1. P. 307-458.
9. Mahmudah U. et al. Parametric sample and technology. Yogyakarta, 2009.
10. Muhamad Safih L.M. Fuzzy Semi-parametric from A Sample Choice Models. Penang, 2007.
11. Puhani P. The Heckman Correction for Sample Selection and Its Critique // *Journal of Economic Surveys*. 2000. Vol. 14. Is. 1. P. 53-68.
12. Rabe-Hesketh S. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables // *The Stata Journal*. 2005. 6 (3). P. 285-308.

Features of the application of the Heckman model in socio-economic research

Irina A. Altukhova

PhD in Pedagogy, Associate Professor,
Associate Professor of the Department of Economics and Digital Technologies,
Far Eastern Institute of Management – Branch of RANEPА,
680000, 33, Russia, Murav'eva-Amurskogo str.,
Khabarovsk, Russian Federation;
e-mail: altukhova-ia@ranepa.ru

Valeriya V. Volkova

Student,
Far Eastern Institute of Management – Branch of RANEPА,
680000, 33, Russia, Murav'eva-Amurskogo str.,
Khabarovsk, Russian Federation;
e-mail: lerka007.v@mail.ru

Elizaveta R. Nikitina

Student,
Far Eastern Institute of Management – Branch of RANEP,
680000, 33, Russia, Murav'eva-Amurskogo str.,
Khabarovsk, Russian Federation;
e-mail: lisunenok@mail.ru

Ljus'ena E. Pyn'ko

PhD in Economics,
Associate Professor of the Department of Economics and Digital Technologies,
Far Eastern Institute of Management – Branch of RANEP,
680000, 33, Russia, Murav'eva-Amurskogo str.,
Khabarovsk, Russian Federation;
e-mail: Lusiena_03@mail.ru

Abstract

The Heckman model is known as a two-step model that allows you to consider hidden variables in data analysis. The model was proposed by economist James Heckman and has found application in various fields, including economics, sociology, psychology, political science, etc. In such models, the probability of the presence of a hidden variable is estimated, which is then used to adjust the estimates of the resulting simulation. In the study of modern socio-economic processes, they often encounter the influence of hidden factors on the resulting indicators, which distorts the results of research and forecasts for them. At the same time, the accuracy of forecasting is important in decision-making. In panel data models and systems of econometric equations, it is impractical to introduce unobservable hidden continuous variables, this complicates and distorts research results. Modern neural network analysis is not always able to reflect the structural influence of individual factors on the resulting indicator. The Heckman model allows us to obtain more accurate estimates of the influence of observed factors on the dependent variable, considering the influence of hidden variables, which is taken into account in the justification and conclusions of the model. At the same time, the model has some limitations in research.

For citation

Altukhova I.A., Volkova V.V., Nikitina E.R., Pyn'ko L.E. (2023) Osobennosti primeneniya modeli Khekmana v sotsial'no-ekonomicheskikh issledovaniyakh [Features of the application of the Heckman model in socio-economic research]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 13 (10A), pp. 15-22. DOI: 10.34670/AR.2023.53.28.004

Keywords

Two-step model, Heckman model, dependent variable, adjustment, unobservable variables, hidden variables, hidden continuous variables, econometric analysis.

References

1. Cubedo M., Oller J.M. (2002) Hypothesis testing: Model Approach to selection. *Magazine Statistical Planning and Inference*, 108, pp. 3-21.
2. Heckman J. (1979) Sample B sampleias as a specification error. *Econometrica*, 47, pp. 152-161.

3. Heckman J. (2005) Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73, 3, pp. 669-738.
4. Heckman J., Lochner L., Todd P. (2006) Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. *Handbook of the Economics of Education*, 1, pp. 307-458.
5. Heckman J.J. (1974) Shadow Wages, Market Wages and labor supply. *Econometrics*, 42, pp. 679-693.
6. Mahmudah U. et al. (2009) *Parametric sample and technology*. Yogyakarta.
7. Mogil'chak E.L. (2015) *Vyborochnyi metod v empiricheskoy sotsiologicheskoy issledovanii* [Sampling method in empirical sociological research]. Yekaterinburg.
8. Muhamad Safiih L.M. (2007) *Fuzzy Semi-parametric from A Sample Choice Models*. Penang.
9. Orlova E.V. (2021) *Ekonomicheskaya metodologiya issledovaniya sistem* [Econometric methodology for systems research]. Moscow: INFRA-M Publ.
10. Puhani P. (2000) The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, 14, 1, pp. 53-68.
11. Rabe-Hesketh S. (2005) Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *The Stata Journal*, 6 (3), pp. 285-308.
12. Yudin S.V. (2022) *Informatsionno-statisticheskie metody resheniya ekonomicheskikh, sotsiologicheskikh i psikhometricheskikh zadach* [Information and statistical methods for solving econometric, sociological and psychometric problems]. Moscow: INFRA-M Publ.