

Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Национальный исследовательский университет  
Высшая школа экономики»

*На правах рукописи*

Потанин Богдан Станиславович

**Обобщение регрессионных моделей с эндогенным  
переключением и случайным отбором на многомерный  
случай**

РЕЗЮМЕ

диссертации на соискание ученой степени  
кандидата экономических наук

Научный руководитель:

доцент, кандидат физико-  
математических наук Коссова  
Елена Владимировна

JEL: C34

Москва – 2020

## Неслучайный отбор как проблема эконометрического анализа

Впервые проблема **неслучайного отбора** (selection bias<sup>1</sup>) была поднята А. Роем<sup>2</sup>. Он изучал влияние выбора профессии на распределение доходов и навыков среди различных профессиональных групп (Roy, 1951). В рамках предложенной им модели индивиды выбирают работу не случайно, а исходя из своих способностей и спроса на них. В результате навыки и доходы оказываются распределены внутри профессиональных групп не так, как между представителями всего общества. Это распределение может зависеть от количества профессий и совместного распределения навыков, а также от их влияния на оплату труда.

На базе данной модели возник обширный пласт литературы, посвященной проблемам идентификации и методам статистического оценивания параметров экономических моделей в условиях неслучайного отбора<sup>3</sup>.

Неслучайный отбор, в зависимости от **механизма формирования выборки** (sample selection), может приводить к **цензурированию** (censoring) или **усечению** (truncation). В обоих случаях предполагается существование некоторого **правила, при соблюдении которого наблюдения попадают в выборку** (sample selection rule). Классическим примером является наблюдение заработной платы индивидов<sup>4</sup>: на вопрос о зарплате отвечают лишь занятые респонденты (Heckman & Killingsworth, 1987). Ключевое различие между усеченными и цензурированными выборками заключается в том, что “при использовании усеченной выборки нельзя

---

<sup>1</sup> Данный термин используется в англоязычной литературе как для указания на неслучайный характер формирования выборки, так и для обозначения смещения в статистических оценках параметров, вызванного использованием некорректного эконометрического инструментария, должным образом не учитывающего неслучайный отбор наблюдений. Смещение в оценках, возникающее в последнем случае будем именовать **смещением отбора** или **систематической ошибкой отбора**, а использование методов, позволяющих её избежать — **коррекцией смещения (систематической ошибки) отбора** или **борьбой со смещением (систематической ошибкой) отбора**.

<sup>2</sup> В контексте модели Роя неслучайный отбор зачастую именуют **самоотбором** (self-selection), тем самым подчеркивая, что индивид самостоятельно выбирает профессию исходя из соображений максимизации дохода. Самоотбор, в свою очередь, является частным случаем цензурирования.

<sup>3</sup> Обзор соответствующих исследований в контексте их соотношения с моделью Роя, а также её математическая формализация представлены, например, в (Heckman & Taber, 2010) и (Heckman & Honore, 1990).

<sup>4</sup> Если выборка была сформирована на основе опроса, проведенного лишь среди работающей части населения, то оценить вероятность попадания зарплаты случайно взятого индивида в данную выборку невозможно, в результате чего она оказывается усеченной. Если же выборка была собрана посредством опроса случайно выбранных индивидов из всего общества, то может возникнуть возможность оценить вероятность занятости, а значит и наблюдаемости заработной платы, что сделает выборку цензурированной.

воспользоваться имеющейся информацией для того, чтобы оценить вероятность доступности наблюдения” (Heckman J. , 1976, p. 478). При этом, в контексте регрессионного анализа, речь идет, как правило, о наблюдаемости зависимых переменных. Тогда под усечением подразумевается, что в выборку попадают лишь наблюдения с принадлежащими определенному множеству значениями зависимой переменной, не совпадающему с её носителем. При цензурировании попадание наблюдений в выборку не зависит от значений зависимой переменной, однако часть самих этих значений может быть ненаблюдаема. Отметим, что в диссертационном исследовании рассматриваются методы работы лишь с цензурированными выборками.

Подходы к статистическому оцениванию параметров, не учитывающие цензурированный или усеченный характер выборки, могут давать несостоятельные оценки. При этом выборки, подверженные смещению отбора, встречаются в экономических исследованиях довольно часто и, как правило, возникают либо в результате самоотбора, либо по причине того, что те, кто собирают данные, устанавливают определенные критерии попадания наблюдений в выборку (Heckman J. , 1979, стр. 1), (Хекман, 2013, стр. 130). Соответствующие обстоятельства обусловили, продолжающееся вплоть до настоящего времени, интенсивное развитие **эконометрических методов оценивания в условиях неслучайного отбора наблюдений** (sample selection models). Рассмотрим основные направления и результаты работ в области разработки данных методов, релевантные теме диссертационного исследования.

### Краткий обзор методологии статистического оценивания параметров экономических моделей в условиях многомерного неслучайного отбора наблюдений

Классические методы коррекции смещения отбора рассматривают случаи, когда существует одно условие отбора, определяемое значением бинарного уравнения, и одно целевое уравнение (Heckman J. , 1979). Однако, некоторые

задачи требуют рассмотрения более сложных селективных механизмов. Так, например, уравнение отбора может быть не бинарным, а порядковым (Kugler, 1987), (Vella, 1993), непрерывным (Garen, 1984) или категориальным (Jeffrey & McFadden, 1984). Последний случай эквивалентен ситуации, когда отбор наблюдения определяется несколькими бинарными правилами, т.е. значением многомерной бинарной переменной. Например, зарплата может наблюдаться лишь для индивидов, которые не только работают, но и согласились раскрыть информацию о своих трудовых доходах.

Из-за сложностей реализации метода максимального правдоподобия, работы, рассматривающие многомерный селективный механизм, встречаются редко и в основном ограничиваются двумерным случаем, а также предположением о независимости правил отбора (Vella, 1998).

Исследования, учитывающие возможную связь между уравнениями отбора, также редки и, как правило, ограничиваются рассмотрением не более трех уравнений с непрерывным, (Poirier, 1980), (Cinzia, 2009), (Ogundimu & Hutton, 2016) и бинарным (Rosenmana, Mandal, Tennekoon, & Hill, 2010) целевыми уравнениями. В других работах при оценивании подобных моделей используют непараметрические двухшаговые процедуры (De Luca & Peracchi, 2012), (Das, Newey, & Vella, 2003), не позволяющие реконструировать совместное распределение случайных ошибок, тем самым существенно сужая горизонты экономической интерпретации полученных результатов. При этом, в работах (Ogundimu & Hutton, 2016) и (De Luca & Peracchi, 2012) не поднимаются проблемы состоятельности оценок предлагаемых двухшаговых методов оценивания, а также вывода их асимптотического распределения и состоятельной оценки асимптотической ковариационной матрицы, что ограничивает возможность использования соответствующих подходов в силу затруднений, связанных с тестированием гипотез о параметрах.

Таким образом, возникает необходимость в разработке методов эконометрического оценивания параметров моделей в условиях многомерной механизма отбора наблюдений. При этом желательно, чтобы соответствующие

методы позволяли оценивать параметры совместного распределения случайных ошибок.

В рамках данной работы предлагается обобщение классического, а также некоторых полупараметрических подходов к оцениванию параметров эконометрических моделей в условиях неслучайного отбора: рассматривается несколько условий отбора и несколько форм целевого уравнения. Задача ставится в максимально общей постановке. Какое именно целевое уравнение имеет место для изучаемого показателя, определяется комбинацией правил отбора. В частном случае, для некоторых из них, наблюдение целевого показателя может отсутствовать. Также, могут отсутствовать наблюдения по некоторым комбинациям самих правил, что соответствует схеме последовательного принятия решений.

## Цели и задачи исследования

В качестве **объекта** исследования выступают эконометрические модели, в рамках которых наблюдаемость зависимых переменных определяется соблюдением нескольких правил отбора. **Предметом** исследования являются методы, позволяющие оценивать параметры этих моделей. **Цель** исследования заключается в разработке соответствующих методов для случая произвольного конечного числа правил отбора. Для достижения поставленной цели необходимо решить следующие **задачи**.

1. Вывести ряд свойств усеченного многомерного нормального распределения и распределения, предложенного (Gallant & Nychka, 1987). Эти свойства необходимы для получения оценок при помощи предлагаемых в диссертационном исследовании методов, а также для расчетов, с использованием полученных оценок, предельных эффектов, стандартных ошибок, условных математических ожиданий зависимых переменных и предсказанных значений.

2. Описать основные параметрические и полупараметрические методы коррекции смещения отбора наблюдений в регрессионных моделях, многомерные обобщения которых будут предложены в работе.

3. Обобщить отобранные методы на многомерный случай, проанализировав свойства получаемых с их помощью оценок.

4. Осуществить программную реализацию предложенных методов, что позволит использовать их для анализа как симулированных, так и реальных данных.

5. Проанализировать точность оценок предложенных методов на симулированных данных при различных предположениях о спецификации регрессионных уравнений, в частности, о совместном распределении случайных ошибок и о наличии уникальных переменных в оцениваемых уравнениях (так называемых «ограничениях исключения»).

6. Применить предложенные методы для оценивания параметров эконометрических моделей на реальных данных.

## Результаты исследования

Эконометрический инструментарий оценивания в условиях многомерного неслучайного отбора наблюдений был расширен за счет предложенных в диссертационном исследовании обобщений ряда параметрических и полупараметрических методов борьбы с систематической ошибкой отбора.

Во-первых, были предложены обобщения метода Хекмана и модели с переключением при наличии произвольного числа уравнений отбора. Первый из предложенных подходов подразумевает максимизацию функции правдоподобия, а второй является многомерным аналогом двухшаговой процедуры Хекмана. Для оценок двухшаговой процедуры выведена состоятельная оценка ковариационной матрицы. Также, были предложены способы оценивания предельных эффектов и условных математических ожиданий зависимых переменных с использованием оценок данных моделей.

Во-вторых, был обобщен на многомерный случай полу-непараметрический метод Галланта и Нички, подразумевающий максимизацию функции правдоподобия с заменой истинной функции плотности на аппроксимирующую функцию в форме Эрмита. При этом был выведен ряд свойств и формул, касающихся распределения (Gallant & Nychka, 1987), функция плотности которого используется для аппроксимации<sup>5</sup>. Полученные результаты полезны не только для оценивания моделей с неслучайным отбором наблюдений и могут быть использованы для параметрического и полу-непараметрического оценивания эконометрических моделей, требующих максимизации функции правдоподобия.

В-третьих, на случай произвольного числа уравнений отбора был обобщен полупараметрический метод Ньюи, предполагающий аппроксимацию неизвестного условного математического ожидания случайной ошибки полиномом от сглаживающих функций, в качестве аргументов которых выступают линейные индексы уравнений отбора.

Точность оценок предложенных методов была проанализирована в рамках численных экспериментов, осуществленных на симулированных данных. При этом в каждом из проведенных экспериментов рассматривался процесс генерации данных, предполагавший наличие двух уравнений отбора. Обозначим основные результаты.

Во-первых, обобщенные параметрические методы обеспечивают существенно более точные оценки чем их классические одномерные аналоги и метод наименьших квадратов.

Во-вторых, точность параметрической обобщенной двухшаговой процедуры заметно снижается при отсутствии ограничений исключения (exclusion restrictions). Вероятно, это связано с тем, что, как и в одномерном случае, используемые на втором шаге обобщенные обратные отношения Миллса достаточно близки к линейной функции, что может приводить к заметному снижению эффективности оценок вследствие мультиколлинеарности.

---

<sup>5</sup> Рассчитывать усеченные, условные и маргинальные функции плотности, а также моменты данного многомерного распределения можно при помощи предложенного диссертантом пакета "hpa" в статистической среде R.

В-третьих, параметрические обобщенные методы не уступают в точности полупараметрическим при совместных распределениях случайных ошибок, отличающихся от многомерного нормального скошенностью и островершинностью. Однако, если распределение является бимодальным, то полупараметрические методы демонстрируют заметное преимущество над параметрическими.

В-четвертых, обобщенный метод Ньюи, предполагающий использование для аппроксимации смещения отбора обычных одномерных отношений Миллса с перекрестными коэффициентами, не уступает в точности спецификации с обобщенными отношениями Миллса. При этом первый способ чрезвычайно упрощает процедуру оценивания: вместо того, чтобы оценивать полупараметрически систему бинарных уравнений, каждое из них можно оценить любым полупараметрическим способом отдельно. Обратим внимание, что спецификация, использующая обычные отношения Миллса без перекрестных коэффициентов, в соответствии с полученными результатами, значительно уступает двум указанным альтернативам при бимодальном распределении случайных ошибок.

В-пятых, обобщенный метод Галланта и Нички заметно уступает в точности другим подходам коррекции с многомерным смещением отбора, если случайные ошибки имеют распределение студента или бета. Однако, при бимодальном распределении относительная точность его оценок значительно возрастает. Обратим внимание, что в работе приводится результат применения данного метода лишь с использованием малых степеней полинома. В будущем планируется осуществление более эффективной программной реализации данного метода, которая позволит достаточно быстро и точно получать оценки с использованием полиномов большей степени.

Предложенные параметрические методы учета многомерного неслучайного отбора наблюдений были применены для оценивания влияния высшего образования и брака на зарплату. Согласно полученным оценкам брак и высшее образование являются эндогенными по отношению к уравнению заработной платы.



Корреляция случайных ошибок между уравнения брака и зарплаты оказалась положительной, свидетельствуя в пользу существования ненаблюдаемых характеристик, положительно влияющих как на вероятность брака, так и на зарплату. Кроме того, было обнаружено отрицательное влияние брака на зарплату, возможно вследствие того, что холостые мужчины могут позволить себе потратить на поиск работы больше времени, так как не должны заботиться о том, чтобы постоянно поддерживать финансовую стабильность своей семьи. Наконец, обратим внимание, что, вероятно, вследствие смещения отбора, оценка эффекта брака на зарплату, полученная при помощи метод наименьших квадратов и классического метода Хекмана, оказалась положительной.

Как и в ряде предшествовавших исследований, осуществленных на данных по США, оценка эффекта образования на зарплату оказалась больше, при условии учета его эндогенности. При этом, обнаруженная отрицательная корреляция между случайными ошибками уравнений зарплаты и образования может быть вызвана тем, что в исследовании не были дифференцированы различные уровни высшего образования, зависящие от качества высшего учебного заведения, которое окончил индивид.

Наконец, уравнение заработной платы было оценено при допущении о том, что источником неслучайного отбора может являться не только занятость индивидов, но и их склонность к раскрытию информации о своих трудовых доходах. Результаты исследования свидетельствуют в пользу наличия данного дополнительного источника смещения отбора, однако обусловленное им смещение в оценках, на используемой выборке оказалось несущественным.

### Научная новизна исследования

В рамках данной работы был осуществлен следующий вклад в развитие эконометрических методов оценивания моделей с многомерным смещением отбора наблюдений:

1. Предложены многомерные обобщения метода Хекмана и модели с переключением, оцениваемые с помощью метода максимального правдоподобия и двухшаговой процедуры.

Для соответствующих методов выведены выражения для предельных эффектов и условных метаматематических ожиданий целевой переменной, что позволяет предсказывать значения зависимой переменной при условии различных комбинаций правил отбора. Обоснована состоятельность и асимптотическая нормальность оценок двухшаговой процедуры, а также предложена состоятельная оценка ковариационной матрицы оценок данного метода.

2. Предложено многомерное обобщение полу-непараметрического метода Галланта и Нички (Gallant & Nychka, 1987) и, с опорой на идею (De Luca & Peracchi, 2012), полупараметрического метода Ньюи (Newey, 2009). Для метода Галланта и Нички выведены формулы для расчета условных усеченных математических ожиданий случайный ошибок, позволяющие, при помощи использования численного дифференцирования, рассчитывать как условные математические ожидания зависимой переменной, так и предельные эффекты.

3. Проанализирована на симулированных данных точность оценок предложенных методов.

В частности, продемонстрирована устойчивость оценок, полученных с помощью обобщения параметрических методов коррекции смещения отбора, к нарушению допущения о совместном нормальном распределении случайных ошибок.

Научная новизна исследования с точки зрения применения соответствующей методологии к анализу реальных данных заключается в следующем:

1. Получена оценка отдачи от образования на российском рынке труда с учетом неслучайного отбора индивидов не только в число занятых, но и в число тех, кто оканчивает высшее учебное заведение.

Результаты исследования свидетельствуют в пользу того, что образование является эндогенным по отношению к уравнению заработной платы. Без учета неслучайного отбора эффект образования может быть недооценен, что согласуется с результатами ряда предшествовавших исследований, осуществленных на данных рынка труда США с использованием метода инструментальных переменных.

2. Показано, что семейный статус является эндогенной переменной по отношению к заработной плате мужчин в России.

Полученные в рамках проведенного анализа результаты свидетельствуют в пользу того, что без учета эндогенности брака по отношению к уравнению заработной платы смещение в оценке эффекта брака на зарплату может оказаться весьма существенным, вплоть до изменения знака с сохранением значимости соответствующего коэффициента.

3. Найдены свидетельства в пользу того, что отказ от ответа на вопрос о заработной плате может являться дополнительным источником смещения при оценивании уравнения заработной платы. Однако, по данным РМЭЗ за 2016 год возникающее смещение не существенно.

Отметим, что выведенные в рамках данного исследования выражения для условных, усечённых и маргинальных функций плотности и распределения, а также моментов распределения (Gallant & Nychka, 1987), могут быть использованы для полу-непараметрического обобщения обширного класса эконометрических моделей, оценивание которых осуществляется при помощи метода максимального правдоподобия.

## Структура исследования

Исследование имеет следующую структуру. Во второй главе рассматриваются основные подходы к оцениванию параметров экономических

моделей в условиях неслучайного отбора. В третьей главе<sup>6</sup> осуществляется обобщение метода Хекмана при допущении о многомерном нормальном распределении случайных ошибок на случай нескольких уравнений отбора и дифференциации функциональной формы целевого уравнения в зависимости от различных комбинаций правил отбора. Четвертая глава посвящена обобщению<sup>7</sup> метода Галланта и Нички (Gallant & Nychka, 1987) и метода Ньюи (Newey, 2009). В пятой главе изучается точности оценок предложенных в диссертационной работе методов на симулированных данных. В шестой главе рассматриваются примеры практического использования предложенных методов.

## Публикация результатов исследования

Результаты диссертационного исследования были опубликованы в ведущих отечественных и международных экономических журналах, входящих в систему Scopus:

1. Kossova, E., & Potanin, B. (2018). Heckman method and switching regression model multivariate generalization. *Applied Econometrics*, 50, 114-143.

**Характеристики статьи:** Scopus Q4; Объем — 1.3 авторских листа.

2. Potanin, B. (2019). Estimating the Effect of Higher Education on an Employee's Wage. *Studies on Russian Economic Development*, 30(3), 319-326.

**Характеристики статьи:** Scopus Q3; Объем — 0.7 авторских листа.

3. Kossova, E., Potanin, B., & Sheluntcova, M. (2020). Estimating effect of marriage on male wages in Russia. *Journal of Economic Studies*, 47(7). Принята к печати.

**Характеристики статьи:** Scopus Q1; Объем — 1 авторский лист.

4. Kossova, E., Potanin, B., & Kupriyanova, L. (2020). Parametric and semiparametric multivariate sample selection models estimators accuracy comparative analysis on simulated data. *Applied Econometrics*. Принята к печати.

---

<sup>6</sup> Основан на статье (Kossova & Potanin, 2018).

<sup>7</sup> Приводится обобщение на случай произвольного числа уравнений отбора методов, рассмотренных в (Kossova, Kupriyanova, & Potanin, 2020).

**Характеристики статьи:** Scopus Q4; Объем — 1.1 авторского листа.

Также, диссертант принимал участие в следующих международных конференциях с докладами по теме диссертации:

1. V Международная конференция «Modern Econometric Tools and Applications – META2018» (Нижний Новгород). Доклад: Estimating the effect of marriage on male wages.
2. XIX Апрельская международная научная конференция (Москва). Доклад: Обобщение метода Хекмана на случай произвольного числа уравнений отбора.
3. XIX Апрельская международная научная конференция (Москва). Доклад: Оценка влияние типа политического режима на приток прямых иностранных инвестиций при помощи многомерной свич-пробит модели с фиксированными эффектами.
4. XIX Апрельская международная научная конференция (Москва). Доклад: Применение многомерной иерархической свич-пробит модели для анализа дискриминации замужних женщин на российском рынке труда по данным РМЭЗ за 2016 год.
5. 6th International Conference on Modern Econometric Tools and Applications (Nizhny Novgorod). Доклад: Comparative Analysis of Parametric, Semi-parametric and SemiNonparametric Sample Selection Models with Application to Engel's Curve Parameters Estimation.
6. XX Апрельская международная научная конференция НИУ ВШЭ (Москва). Доклад: Оценивание эффекта высшего образования на зарплату в условиях неслучайного отбора.

## Список литературы

---

- Arellano-Valle, R. B., Branco, M. D., & Genton, M. G. (2006). A Unified View on Skewed Distributions Arising from Selections. *The Canadian Journal of Statistics*, 34(4), 581-601.
- Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Communications in Statistics - Theory and Methods*, 19(1), 197-203.  
doi:<https://doi.org/10.1080/03610929008830197>

- Cinzia, D. N. (2009). Sample selection correction in panel data models when selectivity is due to two sources. *Institute of Public Policy and Public Choice (POLIS) Working Papers*.
- Copas, J. B., & Li, H. G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59, 55-95.
- Das, M., Newey, K., & Vella, F. (2003). Nonparametric Estimation of Sample Selection Models. *The Review of Economic Studies*, 1(1), 33-58.
- De Luca, G., & Peracchi, F. (2012). Estimating Engel curves under unit and item nonresponse. *Journal of Applied Econometrics*, 27(7), 1076-1099.
- Gallant, A., & Nychka, D. (1987). Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55(2), 363-390.
- Garen, J. (1984). The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5), 1199-1218.
- Heckman, J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic Social Measurement*, 5(4), 475-492.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.
- Heckman, J. J., & Taber, C. (2010). Roy model. *Durlauf S.N., Blume L.E. (eds) Microeconometrics. The New Palgrave Economics Collection. Palgrave Macmillan.*  
doi:10.1057/9780230280816\_27
- Heckman, J., & Honore, B. E. (1990). The Empirical Content of the Roy Model. *Econometrica*, 1121-1149.
- Heckman, J., & Killingsworth, M. (1987). Female labor supply: a survey. *O. Ashenfelter and R. Layard (eds.) Handbook of Labor Economics*, 1, 103-204.
- Jeffrey, D., & McFadden, D. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2), 345-362.
- Kossova, E. V., Kupriyanova, L. A., & Potanin, B. S. (2020). Parametric and semiparametric multivariate sample selection models estimators accuracy comparative analysis on simulated data. *Applied Econometrics (принята к печати)*.
- Kossova, E., & Potanin, B. (2018). Heckman method and switching regression model multivariate generalization. *Applied Econometrics*, 50, 114-143.
- Kossova, E., Potanin, B., & Sheluntcova, M. (2020). Estimating effect of marriage on male wages in Russia. *Journal of Economic Studies (принята к печати)*, 47(7).
- Kugler, J. E. (1987). The earnings impact of training duration in a developing country. *Journal of Human Resources*, 22(2), 228-247.
- Manjunath, B. G., & Wilhelm, S. (2012). Moments Calculation For the Doubly Truncated Multivariate Normal Density. *arXiv*.
- Newey, W. K. (2009). Two-step series estimation of sample selection models. *The Econometrics Journal*, 12, 217-229.
- Ogundimu, E. O., & Hutton, J. L. (2016). A unified approach to multilevel sample selection models. *Communications in Statistics - Theory and Methods*, 45(9), 2592-2611.

- Poirier, D. L. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, 12(2), 209-217.
- Potanin, B. (2019). Estimating the Effect of Higher Education on an Employee's Wage. *Studies on Russian Economic Development*, 30(3), 319-326.
- Rosenmana, R., Mandal, B., Tennekoon, V., & Hill, L. G. (2010). Estimating treatment effectiveness with sample selection. *Working Paper*.
- Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3, 135-146.
- Vella, F. (1993). A simple estimator for models with censored endogenous regressors. *International Economic Review*, 34(2), 441-457.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, 34(2), 127-169.
- Хекман, Д. (2013). Смещение селективной выборки как ошибка спецификации. (Д. И. Переводчики: И. П. Станкевич, Ред.) *Прикладная Эконометрика*, 31(3), 127-137.