

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 неслучайным 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 bias¹) была поднята А. Роем². 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 отбора³.

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 индивидов⁴: 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

¹ 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**.

² 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.

³ 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).

⁴ 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 носителем⁵. При цензурировании попадание наблюдений в выборку не зависит от значений зависимой переменной, однако часть самих этих значений может не наблюдаться. Например, наблюдения по зарплате будут отсутствовать в выборке для тех, кто не работает. В результате механизм, определяющий занятость индивида, будет выступать в качестве правила отбора наблюдений. 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.

⁵ Носителем (support) дискретной (непрерывной) случайной величины называется множество значений, для которых вероятность (плотность) не обращается в ноль.

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.” (Коссова & Потанин, 2018).

Кроме того, зачастую, исследователей может интересовать оценивание нескольких форм целевого уравнений, каждая из которых определяется комбинацией правил отбора. Например, зарплата, которую индивид будет получать при условии наличия (отсутствия) у него высшего образования, может наблюдаться только если он (не) окончил высшее учебное заведение. При этом целевые уравнения, определяющие значения зарплат, могут различаться, поскольку наличие высшего образования может, например, влиять на отдачу от характеристик индивида. Такая модель может рассматриваться как регрессия с эндогенным переключением, подразумевающая существование правила, при соблюдении которого наблюдаются значения одного целевого уравнения, а при несоблюдении — другого (Lee, 1978). При этом может возникнуть необходимость в учете дополнительных правил отбора наблюдений, например, вследствие того, что еще одним условием наблюдаемости зарплаты, независимо от наличия высшего образования, является занятость индивида.

“Из-за сложностей реализации метода максимального правдоподобия, работы, рассматривающие многомерный механизм отбора наблюдений, встречаются редко и в основном ограничиваются двумерным случаем, а также предположением о независимости правил отбора (Vella, 1998).” (Коссова & Потанин, 2018).

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 (Коссова & Потанин, 2018). 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 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 многокритериального механизма отбора наблюдений и нескольких форм целевых уравнений. При этом желательно, чтобы соответствующие методы позволяли оценивать параметры совместного распределения случайных ошибок, что существенно расширит возможности интерпретации результатов.

В рамках данной работы предлагается несколько новых эконометрических методов оценивания регрессионных моделей, предполагающих наличие произвольного числа правил отбора и нескольких форм целевого уравнения. Предложенные методы обобщают классический, а также некоторые

полупараметрические подходы к оцениванию параметров эконометрических моделей в условиях неслучайного отбора и с учетом эндогенного переключения. 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. Предложить эконометрические методы, позволяющие оценивать параметры моделей при наличии произвольного конечного числа правил отбора

наблюдений, задающих вид целевого уравнений. Также, необходимо проанализировав свойства получаемых с помощью предложенных методов 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 (присутствующих лишь в данном уравнении) в оцениваемых уравнениях (так называемых «ограничениях исключения»).

6. Применить предложенные методы для оценивания параметров эконометрических моделей на реальных данных, подверженных действию нескольких правил отбора наблюдений.

Результаты исследования

Эконометрический инструментарий оценивания в условиях многомерного неслучайного отбора наблюдений был расширен за счет предложенных в диссертационном исследовании методов, обобщающих ряд параметрических и полупараметрических подходов к устранению смещения отбора.

Во-первых, были предложены обобщения метода Хекмана и модели с эндогенным 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 аппроксимации⁶. 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.

⁶ 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 Стьюдента 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 в диссертационном исследовании параметрические методы учета многомерного неслучайного отбора наблюдений были применены для оценивания влияния высшего образования и брака на зарплату по данным российского мониторинга экономического положения и здоровья населения

(РМЭЗ). Согласно полученным оценкам, брак и высшее образование являются эндогенными по отношению к уравнению заработной платы: условное распределение случайной ошибки уравнения заработной платы зависит от вероятности того, что индивид состоит в браке и имеет высшее образование.

Корреляция между случайными ошибками уравнений 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, что брак отрицательно влияет 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 метода 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 (не принималось во внимание, что некоторые индивиды оканчивали ведущие вузы, а некоторые — нет), 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 параметрические методы, обобщающие подход, предложенный Хекманом, к оцениванию модели со смещением отбора и модели с переключением на случай многокритериального отбора наблюдений и нескольких форм целевого уравнения зависимой переменной. Для соответствующих методов выведены выражения для предельных эффектов и условных математических ожиданий целевой переменной, что позволяет предсказывать значения зависимой переменной при различных комбинациях правил отбора. Обоснована состоятельность и асимптотическая нормальность оценок двухшаговой процедуры оценивания, а также предложена состоятельная оценка ковариационной матрицы оценок параметров модели.

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 случайных 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 обобщенных 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 зарплаты 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.

Результаты диссертационного исследования, выносимые на защиту

Кратко обозначим основные эконометрические методы, предложенные в рамках диссертационного исследования. В силу ограничений, налагаемых

допустимым объемом резюме, многие из выведенных в диссертационном исследовании формул не приводятся и могут быть найдены в полной версии текста.

Предлагаемая модель с многокритериальным эндогенным переключением

“Предположим, что существует m правил отбора наблюдений целевого показателя, задаваемых значениями бинарных переменных $z_{si}, s \in \{1...m\}$. Если для наблюдения i правило s выполнено, бинарная переменная z_{si} принимает значение, равное 1, если не выполнено, то (-1) . Число всех возможных комбинаций правил составляет 2^m . Перенумеруем те комбинации правил, для которых доступны значения основной переменной. Будем считать, что всего таких комбинаций r , где $(1 \leq r \leq 2^m)$. Тогда все наблюдения, за исключением тех, для которых значения основной переменной не доступны, разбиваются на r групп в зависимости от имевшей место комбинации правил. Будем считать, что номер группы совпадает с номером соответствующей ей комбинации. Оставшиеся наблюдения образуют группу, которой присвоим номер ноль. Данная группа может быть пустой, если значения основной переменной доступны для всех наблюдений. Для каждого наблюдения i определим значение функции индекса $g_i = g(z_{1i}, \dots, z_{mi}) = s$ ($0 \leq s \leq r$), равное номеру группы, к которой относится наблюдение i .” (Коссова & Потанин, 2018).

“Рассмотрим ситуацию, когда в регрессионном уравнении, связывающем зависимую переменную y_i^* и набор объясняющих переменных x_i , вектор параметров и случайная ошибка ϵ_i зависят от группы g_i , в которую попало i -е наблюдение, т.е. уравнение имеет вид $y_i^* = x_i' \beta_{g_i} + \epsilon_{i,g_i}$.” (Коссова & Потанин, 2018).

“Уравнения отбора и определяемое ими основное уравнение могут быть зависимыми из-за коррелированности входящих в них случайных ошибок, что и порождает проблему селективности. Предположим, что совместное распределение случайных ошибок является нормальным.” (Коссова & Потанин, 2018).

“Формально, модель может быть записана следующим образом:” (Коссова & Потанин, 2018).

$$\begin{aligned}
 y_i^* &= x_i' \beta_{g_i} + \epsilon_{i,g_i}, \\
 z_{si}^* &= w_{si}' \gamma_s + u_{si}, s \in \{1 \dots m\}, \\
 z_{si} &= \begin{cases} 1, & \text{если } z_{si}^* \geq 0, \text{ т.е. } u_{si} \geq -w_{si}' \gamma_s, \\ -1, & \text{если } z_{si}^* < 0, \text{ т.е. } u_{si} < -w_{si}' \gamma_s. \end{cases} \\
 g_i &= g(z_{1i}, \dots, z_{mi}), g_i \in \{0 \dots r\}, \\
 u_i' &= (u_{1i}, \dots, u_{mi})', \\
 (\epsilon_{i,g_i}, u_i') &\sim N \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma_{g_i} \right), \text{ где } \Sigma_{g_i} = \begin{bmatrix} \sigma_{g_i}^2 & \rho_{1,g_i} \sigma & \rho_{2,g_i} \sigma & \dots & \rho_{m,g_i} \sigma \\ \rho_{1,g_i} \sigma & 1 & \rho_{12} & \dots & \rho_{1m} \\ \rho_{2,g_i} \sigma & \rho_{12} & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{m,g_i} \sigma & \rho_{1m} & \dots & \dots & 1 \end{bmatrix}; \\
 y_i &= \begin{cases} y_i^*, & \text{если } g_i > 0, \\ \text{не наблюдаем}, & \text{если } g_i = 0, \end{cases} \quad i \in \{1 \dots n\}.
 \end{aligned}$$

“Как и в одномерном случае, x_i и w_{is} — векторы значений объясняющих факторов для наблюдения i основного уравнения и уравнения отбора s , β_{g_i} и γ_s — векторы коэффициентов, ϵ_{i,g_i} и u_{si} — случайные ошибки, n — число наблюдений.” (Коссова & Потанин, 2018).

Обозначим:

$$z_i = \begin{bmatrix} z_{1i} \\ \vdots \\ z_{mi} \end{bmatrix}, \tilde{z}_i = \begin{bmatrix} \tilde{z}_{1i} \\ \vdots \\ \tilde{z}_{mi} \end{bmatrix} = \begin{bmatrix} z_{1i} w_{1i}' \gamma_1 \\ \vdots \\ z_{mi} w_{mi}' \gamma_m \end{bmatrix}, u_i = \begin{bmatrix} -z_{1i} u_{1i} \\ \vdots \\ -z_{mi} u_{mi} \end{bmatrix} \text{ И } \rho_{g_i} = \begin{bmatrix} \rho_{1,g_i} \\ \vdots \\ \rho_{m,g_i} \end{bmatrix}.$$

“Используя свойства усеченного многомерного нормального распределения можно показать, что условные математическое ожидание и дисперсия целевой переменной будут иметь вид:” (Коссова & Потанин, 2018).

$$\begin{aligned}
 E(y_i | z_{1i}, \dots, z_{mi}) &= x_i' \beta_{g_i} + \sum_{j=1}^m \sigma_{g_i} \rho_{j,g_i} \lambda_j^{\tilde{u}_i}(\tilde{z}_i) z_{ji}, \\
 D(y_i | z_{1i}, \dots, z_{mi}) &= \sigma_{g_i}^2 \left(1 - \sum_{k=1}^m \rho_{k,g_i}^2 \tilde{z}_{ki} \lambda_k^{\tilde{u}_i}(\tilde{z}_i) + \right. \\
 &\quad \left. + \sum_{k=1}^m \tilde{z}_{ki} \rho_{k,g_i} \sum_{j \neq k} z_{ji} (\rho_{j,g_i} - \rho_{kj} \rho_{k,g_i}) \Lambda_{kj}^{\tilde{u}_i}(\tilde{z}_i) - \left(\sum_{k=1}^m \tilde{z}_{ki} \rho_{k,g_i} \lambda_k^{\tilde{u}_i}(\tilde{z}_i) \right)^2 \right).
 \end{aligned}$$

где:

$$\lambda^{\tilde{u}_i}(x) = \frac{\nabla F_{\tilde{u}_i}(x)}{F_{\tilde{u}_i}(x)}, \quad \Lambda^{\tilde{u}_i}(x) = \frac{H(F_{\tilde{u}_i}(x))}{F_{\tilde{u}_i}(x)}.$$

Отметим, что при единственном уравнении отбора ($m=1$), если $g(1) > 0$ и $g(-1) = 0$ данная модель совпадает с моделью Хекмана, если же $g(1) > 0$ и $g(-1) > 0$ — с моделью с эндогенным переключением. Также, при $m=2$ и $g(1,1) > 0, g(1,-1) = g(-1,1) = g(-1,-1) = 0$ предложенная модель совпадает с той, что рассматривались (Poirier, 1980) (De Luca & Peracchi, 2012). Наконец, если $m=3$, а также $g(1,1,1) > 0$ и равняется нулю при всех иных комбинациях правил отбора, то предлагаемая модель совпадает с моделью (Ogundimu & Hutton, 2016).

Кратко обозначим предложенные в диссертационном исследовании методы, позволяющие оценивать параметры описанной модели.

Параметрическая коррекция многокритериального смещения отбора с помощью метода максимального правдоподобия

Оценки параметров модели можно получить за счет максимизации следующей функции правдоподобия:

$$L(\Sigma_1, \dots, \Sigma_r, \beta_1, \dots, \beta_r, \gamma_1, \dots, \gamma_m) = \prod_{i=1}^n r_i,$$

$$r_i = \begin{cases} F_{u_i | \epsilon_{g_i} = y_i - x_i' \beta_{g_i}}(z_{1i} w_{1i}' \gamma_1, \dots, z_{mi} w_{mi}' \gamma_m) f_{\epsilon_{g_i}}(y_i - x_i' \beta_{g_i}), & \text{для } i: g_i > 0, \\ F_{u_i}(z_{1i} w_{1i}' \gamma_1, \dots, z_{mi} w_{mi}' \gamma_m), & \text{для } i: g_i = 0. \end{cases},$$

где F_{u_i} — функция распределения случайного вектора \tilde{u}_i , а $F_{u_i | \epsilon_{g_i} = y_i - x_i' \beta_{g_i}}$ — его функция распределения при условии, что $\epsilon_{g_i} = y_i - x_i' \beta_{g_i}$. (Коссова & Потанин, 2018).

Оценки, полученные методом максимального правдоподобия, окажутся состоятельными, эффективными и асимптотически нормальными при условии соблюдения допущения о совместном нормальном распределении случайных ошибок и выполнении ряда условий относительно функции правдоподобия. При этом, результаты анализа симулированных данных свидетельствуют в пользу того,

что точность оценок максимального правдоподобия при нарушении допущения о совместной нормальности может оказаться сопоставимой с точностью оценок, получаемых непараметрическими и полу-параметрическими методами коррекции многокритериального смещения отбора.

Параметрическая коррекция многокритериального смещения отбора с помощью двухшаговой процедуры

“Согласно теореме о двухшаговой процедуре оценивания (Murphy & Torpe, 1985) можно получить состоятельные оценки параметров, если оценивать уравнение для условного математического ожидания целевой переменной с помощью МНК, используя вместо обобщенных отношений Миллса их состоятельные оценки. В свою очередь, состоятельные оценки обобщенных отношений Миллса можно получить, оценив систему бинарных уравнений, задающих правила отбора наблюдений. Таким образом, оценивание разбивается на два этапа:” (Коссова & Потанин, 2018).

“1. На первом шаге оценивается система бинарных уравнений отбора, находятся оценки γ_s и ρ_{sk} , $s, k \in \{1...m\}$. Рассчитываются оценки обобщенных отношений Миллса⁷, то есть $\lambda_k^{\tilde{u}_i}(\tilde{z}_i)$, где $i \in \{1, \dots, n\}$.” (Коссова & Потанин, 2018).

“2. На втором шаге, как и в процедуре Хекмана, МНК оцениваются уравнения для условного математического ожидания целевой переменной для каждого значения $g_i = c$ ($c \in \{1...r\}$) по n_c наблюдениям, где n_c — количество наблюдений, относящихся к группе c , т.е. $\{i : g_i = c\}$.” (Коссова & Потанин, 2018).

$$y_i = x_i' \beta_c + \sum_{j=1}^m \beta_{\lambda_j} (\lambda_j^{\tilde{u}_i}(\tilde{z}_i) z_{ji}) + v_i, \text{ где } \beta_{\lambda_j} = \sigma_c \rho_{j,c},$$

“с заменой обратных отношений Миллса $\lambda_j^{\tilde{u}_i}(\tilde{z}_i)$ их состоятельными оценками $\lambda_j^{\tilde{u}_i}(\hat{\tilde{z}}_i)$, полученными на предыдущем шаге.” (Коссова & Потанин, 2018).

⁷ В случае, если некоторые комбинации правил не наблюдаются, оценивается система бинарных уравнений, учитывающая селективное смещение

Состоятельные оценки дисперсии и корреляций случайных ошибок могут быть рассчитаны по формулам:

$$\hat{\sigma}_c^2 = \frac{1}{n_c} \left(e^T e + \sum_{i=1}^{n_c} \left(\sum_{k=1}^m \beta_{\lambda_k}^2 \hat{z}_{ki} \hat{\lambda}_k^{\hat{u}_i}(\hat{z}_i) - \sum_{k=1}^m \beta_{\lambda_k} z_{ki} \sum_{j \neq k} z_{ji} (\beta_{\lambda_j} - \hat{\rho}_{kj} \beta_{\lambda_k}) \hat{\Lambda}_{kj}^{\hat{u}_i}(\hat{z}_i) + \left(\sum_{k=1}^m z_{ki} \beta_{\lambda_k} \hat{\lambda}_k^{\hat{u}_i}(\hat{z}_i) \right)^2 \right) \right),$$

$$\hat{\rho}_{k,c} = \frac{\hat{\beta}_{\lambda_k}}{\sigma_c}, \quad 1 \leq k \leq m.$$

Также, была получена состоятельная оценка ковариационной матрицы оценок предложенной двухшаговой процедуры:

$$As.Cov(\hat{\beta}) = \hat{\sigma}_c^2 (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \left[(I - \hat{\Delta}) + \hat{\Gamma} As.Cov \left(\begin{bmatrix} \hat{\gamma} \\ \hat{\rho} \end{bmatrix} \right) \hat{\Gamma}^T \right] \tilde{X} (\tilde{X}^T \tilde{X})^{-1},$$

где, в соответствии с выведенными в приложении диссертационного исследования свойствами усеченного многомерного нормального распределения:

$$\Gamma = J \left(\sum_{k=1}^m \rho_{k,c} z_k \circ \lambda_k^{\hat{u}}(\tilde{z}) \right),$$

$$\Theta^{\hat{u}_i}(x) = \frac{D^3(F_{\hat{u}_i}(x))}{F_{\hat{u}_i}(x)},$$

$$(\Gamma_{\gamma_{ks}})_i = W_{ik} \left(\sum_{j \neq k} \rho_{j,c} z_{ji} z_{ki} (\Lambda_{kj}^{\hat{u}_i}(\tilde{z}_i) - \lambda_k^{\hat{u}_i}(\tilde{z}_i) \lambda_j^{\hat{u}_i}(\tilde{z}_i)) - \rho_{k,c} (z_{ki} \lambda_k^{\hat{u}_i}(\tilde{z}_i) + (\lambda_k^{\hat{u}_i}(\tilde{z}_i))^2 + \sum_{j \neq k} \rho_{kj} \Lambda_{kj}^{\hat{u}_i}(\tilde{z}_i) z_{ki} z_{ji}) \right),$$

$$\begin{aligned} (\Gamma_{\rho_{lk}})_i &= z_{ki} \rho_{l,c} \left(\left(\frac{z_{li} z_{ki} \rho_{lk} \tilde{z}_{ki} - \tilde{z}_{li}}{1 - \rho_{lk}^2} - \lambda_l^{\hat{u}_i}(\tilde{z}_i) \right) \Lambda_{lk}^{\hat{u}_i}(\tilde{z}_i) + \sum_{j \neq l, k} \frac{z_{li} z_{ji} (\rho_{lk} \rho_{kj} - \rho_{lj})}{1 - \rho_{lk}^2} \Theta_{lj}^{\hat{u}_i}(\tilde{z}_i) \right) + \\ &+ z_{li} \rho_{k,c} \left(\left(\frac{z_{li} z_{ki} \rho_{lk} \tilde{z}_{li} - \tilde{z}_{ki}}{1 - \rho_{lk}^2} - \lambda_k^{\hat{u}_i}(\tilde{z}_i) \right) \Lambda_{lk}^{\hat{u}_i}(\tilde{z}_i) + \sum_{j \neq l, k} \frac{z_{ki} z_{ji} (\rho_{lk} \rho_{lj} - \rho_{kj})}{1 - \rho_{lk}^2} \Theta_{lj}^{\hat{u}_i}(\tilde{z}_i) \right) + \\ &+ \sum_{j \neq l, k} z_{ji} z_{li} z_{ki} \rho_{j,c} (\Theta_{lj}^{\hat{u}_i}(\tilde{z}_i) - \lambda_j^{\hat{u}_i}(\tilde{z}_i) \Lambda_{lk}^{\hat{u}_i}(\tilde{z}_i)). \end{aligned}$$

“где $\hat{\Gamma}$ это матрица Γ , рассчитанная при $\hat{z}, \hat{\rho}$ и $\hat{\sigma}$.” (Коссова & Потанин, 2018).

Поскольку, как было показано ранее, оценки данного метода являются асимптотически нормальными, то выведенная состоятельная оценка ковариационной матрицы позволяет тестировать гипотезы об оцениваемых параметрах модели.

Предложенная двухшаговая процедура оценивания даёт состоятельные и асимптотически нормальные оценки. Однако, они будут менее эффективны чем оценки, полученные с использованием метода максимального правдоподобия. С другой стороны, по результатам анализа симулированных данных было

обнаружено, что оценки двухшаговой процедуры обладают большей устойчивостью к нарушению допущения о нормальности. При этом их точность существенно снижается если не соблюдены ограничения исключений.

Полу-непараметрическая коррекция многокритериального смещения отбора

Состоятельность оценок описанных выше методов зависит от соблюдения допущения о совместном нормальном распределении случайных ошибок. Если же данное допущение не соблюдается, то оценки могут быть получены за счет максимизации функции правдоподобия с заменой функций плотности и распределения на аппроксимирующие их функции, предложенные (Gallant & Nychka, 1987). Например, при отсутствии эндогенного переключения⁸, получаем следующую функцию правдоподобия:

$$L(\alpha^{(-1)}, \mu, \sigma, \beta, \gamma) = \prod_{i=1}^n r_i.$$

$$r_i = \begin{cases} \bar{F}_{u_i | \epsilon_{g_i} = y_i - x_i' \beta}(-w'_{1i} \gamma_1, \dots, -w'_{mi} \gamma_m; \infty, \dots, \infty) f_{\epsilon_{g_i}}(y_i - x_i' \beta), & \text{для } i: g_i > 0 \\ F_{u_i}(-w'_{1i} \gamma_1, \dots, -w'_{mi} \gamma_m), & \text{для } i: g_i = 0 \end{cases},$$

где, в соответствии с выведенными в диссертационном исследовании формулами:

$$\begin{aligned} \bar{F}_{u_i | \epsilon_{g_i} = y_i - x_i' \beta}(-w'_{1i} \gamma_1, \dots, -w'_{mi} \gamma_m; \infty, \dots, \infty) &= \frac{1}{\psi_1} \prod_{r_0=1}^m \Phi \left(\frac{w'_{r_0 i} \gamma_{r_0} + \mu_{r_0+1}}{\sigma_{r_0+1}} \right) * \\ &* \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_{m+1})} \alpha_{(j_1, \dots, j_{m+1})} (y_i - x_i' \beta)^{i_1 + j_1} \prod_{r_1=2}^{m+1} \mathcal{M}_{TR}(i_{r_1} + j_{r_1}; -w'_{r_1 i} \gamma_{r_1}, \infty, \mu_{r_1}, \sigma_{r_1}) \end{aligned},$$

$$\begin{aligned} F_{u_i}(-w'_{1i} \gamma_1, \dots, -w'_{mi} \gamma_m) &= \frac{1}{\psi_2} \prod_{r_0=1}^{m_d} \left(1 - \Phi \left(\frac{w'_{r_0 i} \gamma_{r_0} + \mu_{r_0+1}}{\sigma_{r_0+1}} \right) \right) * \\ &* \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_{m+1})} \alpha_{(j_1, \dots, j_{m+1})} \mathcal{M}(i_1 + j_1; \mu_1, \sigma_1) \prod_{r_1=2}^{m+1} \mathcal{M}_{TR}(i_{r_1} + j_{r_1}; -w'_{r_1 i} \gamma_{r_1}, \infty, \mu_{r_1}, \sigma_{r_1}) \end{aligned},$$

⁸ Значение целевой переменной наблюдается лишь при $(z_{i_1} = \dots = z_{i_m} = 1)$, то есть $g(1, \dots, 1) > 0$ и равняется нулю — в противном случае.

$$f_{\epsilon_{g_i}}(y_i - x_i' \beta) = \frac{1}{\psi_3} \phi \left(\frac{y_i - x_i' \beta}{\sigma_1} \right) \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_m)} \alpha_{(j_1, \dots, j_m)} (y_i - x_i' \beta)^{i_1 + j_1} * \\ * \prod_{r=2}^{m+1} \mathcal{M}(i_{r_2} + j_{r_2}; \mu_{r_2}, \sigma_{r_2})$$

а также:

$$\psi_1 = \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_{m+1})} \alpha_{(j_1, \dots, j_{m+1})} (y_i - x_i' \beta)^{i_1 + j_1} \prod_{r=2}^{m+1} \mathcal{M}(i_{r_2} + j_{r_2}; \mu_{r_2}, \sigma_{r_2}), \\ \psi_2 = \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_{m+1})} \alpha_{(j_1, \dots, j_{m+1})} \prod_{r=2}^{m+1} \mathcal{M}(i_r + j_r; \mu_r, \sigma_r), \\ \psi_3 = \sum_{i_1=0}^{K_1} \dots \sum_{i_{m+1}=0}^{K_{m+1}} \sum_{j_1=0}^{K_1} \dots \sum_{j_{m+1}=0}^{K_{m+1}} \alpha_{(i_1, \dots, i_{m+1})} \alpha_{(j_1, \dots, j_{m+1})} \prod_{r=1}^{m+1} \mathcal{M}(i_r + j_r; \mu_r, \sigma_r), \\ \alpha^{(-1)} = \left(\alpha_{(1, \dots, 0)}, \dots, \alpha_{(K_1, \dots, K_{m+1})} \right), K_1, \dots, K_{m+1} \in (N \cup \{0\}).$$

где $\mathcal{M}_{TR}(i_r + j_r; \underline{x}, \bar{x}, \mu_r, \sigma_r)$ это момент порядка $(i_r + j_r)$ усеченной снизу в точке \underline{x} и сверху в точке \bar{x} нормальной случайной величины с математическим ожиданием μ_r и дисперсией σ_r^2 , причем $\mathcal{M}_{TR}(i_r + j_r; \mu_r, \sigma_r) = \mathcal{M}_{TR}(i_r + j_r; -\infty, \infty, \mu_r, \sigma_r)$. Через ϕ и Φ обозначаются функции плотности и распределения стандартного нормального распределения соответственно.

При этом, с целью обеспечения идентифицируемости параметров, полагается $\alpha(0, \dots, 0) = 1$, а также, в каждом бинарном уравнении исключается константа и фиксируется на единичном значении один из коэффициентов при регрессорах. Отметим, что сохраняя состоятельность оценок β и γ , можно положить любое фиксированное значение для μ_k и любое неотрицательное число для σ_k , где $k \in \{1, \dots, m+1\}$.

При максимизации функции правдоподобия с использованием аппроксимирующих функций, предложенных, можно получить состоятельные оценки параметров модели независимо от соблюдения допущения о совместном нормальном распределении случайных ошибок. Однако, асимптотическое распределение оценок неизвестно, вследствие чего для тестирования гипотез о параметрах необходимо пользоваться бутстрапированными статистиками.

Результаты анализа симулированных данных говорят о том, что данный метод обладает наибольшим преимуществом перед аналогами в случае, если частные распределения случайных ошибок являются мультимодальными.

Полупараметрическая коррекция многокритериального смещения отбора с использованием двухшаговой процедуры

Второй подход, позволяющий оценивать параметры модели с многокритериальным неслучайным отбором наблюдений при нарушении допущения о совместном нормальном распределении случайных ошибок, развивает идеи метода, предложенного (Newey, 2009).

Условное математическое ожидание целевой переменной имеет вид:

$$\begin{aligned} E(y_i) &= E(y_i^* | z_{1i}, \dots, z_{mi}) = E(y_i^* | -z_{1i}u_{1i} \leq z_{1i}w'_{1i}\gamma_1, \dots, -z_{mi}u_{mi} \leq z_{mi}w'_{mi}\gamma_m) = \\ &= x'_i\beta_{g_i} + E(\epsilon_{i,g_i} | -z_{1i}u_{1i} \leq z_{1i}w'_{1i}\gamma_1, \dots, -z_{mi}u_{mi} \leq z_{mi}w'_{mi}\gamma_m) = \\ &= x'_i\beta_{g_i} + g^*(w'_{1i}\gamma_1, \dots, w'_{mi}\gamma_m) \end{aligned}$$

Опираясь на подход (De Luca & Peracchi, 2012) аппроксимировать неизвестную функцию можно полиномами от обобщённых обратных отношений Миллса:

$$g_k(w'_{1i}\gamma_1, \dots, w'_{mi}\gamma_m) = \sum_{t=1}^k \sum_{j=1}^m \tau_t^{(v)} \lambda_j^{(-u_i)} (w'_{1i}\gamma_1, \dots, w'_{mi}\gamma_m)^t + \tau_0.$$

При этом оценки $w'_{ji}\gamma_j$, где $j \in \{1, \dots, m\}$, и корреляций между случайными ошибками, необходимые для расчета оценок $\lambda_j^{(-u_i)}(w'_{1i}\gamma_1, \dots, w'_{mi}\gamma_m)$, можно получить на первом шаге, оценивая систему бинарных уравнений полу-непараметрическим методом.

На втором шаге, с использованием полученных на первом шаге оценок элементов аппроксимирующей функции, осуществляется МНК оценивание параметров β и τ . При этом оптимальная степень полинома подбирается с помощью кросс валидации по k , например, методом leave-one-out. Тестирование гипотез о параметрах можно осуществлять при помощи бутстрапа.

Доказательство состоятельности оценок данного метода, технически сложно и в работе не приводится. В исследовании (Newey, 2009) имеется доказательство для одномерного случая, а (De Luca & Peracchi, 2012) приводят обобщение метода в случае двух уравнений отбора без вывода свойств его оценок. Однако, по результатам анализа симулированных данных в диссертационном исследовании удалось продемонстрировать достаточно высокую точность оценок предложенного метода для случая двухкритериального отбора.

Расчет предельных эффектов с использованием предложенных методов

В экономических задачах необходимо иметь возможность количественно оценивать влияние регрессоров на зависимую переменную. Если математическое ожидание целевой переменной нелинейно зависит от объясняющих факторов, должны быть вычислены предельные эффекты.

В работе получены выражения для предельных эффектов условного математического ожидания в моделях, предполагающих совместное нормальное распределение случайных ошибок:

$$\frac{\partial E(y_i | z_{1i}, \dots, z_{mi})}{\partial \psi} = \beta_\psi + \sum_{j=1}^m Y_\psi^j(\tilde{z}_i),$$

где:

$$Y_\psi^j(\tilde{z}_i) = z_{ji} \sigma_{g_i} \rho_{j,g_i} \sum_{k=1}^m z_{ki} \gamma_\psi^k \frac{\partial \lambda_j^{\tilde{u}_i}(\tilde{z}_i)}{\partial \tilde{z}_k} = z_{ji} \sigma_{g_i} \rho_{j,g_i} \left(\sum_{k \neq j} z_{ki} \gamma_\psi^k (\Lambda_{jk}^{\tilde{u}_i}(\tilde{z}_i) - \lambda_j^{\tilde{u}_i}(\tilde{z}_i) \lambda_k^{\tilde{u}_i}(\tilde{z}_i)) - \right. \\ \left. - \gamma_\psi^j \sigma_{g_i} \rho_{j,g_i} (\tilde{z}_{ji} \lambda_j^{\tilde{u}_i}(\tilde{z}_i) + (\lambda_j^{\tilde{u}_i}(\tilde{z}_i))^2 + \sum_{k \neq j} z_{ji} z_{ki} \rho_{jk} \Lambda_{jk}^{\tilde{u}_i}(\tilde{z}_i)) \right).$$

Предельные эффекты при применении предложенного непараметрического метода, развивающего идеи (Gallant & Nychka, 1987), нетрудно рассчитать при помощи численных методов дифференцирования, например, с помощью метода Ричардсона (Richardson, 1911). При этом, для получения оценок условных математических ожиданий зависимой переменной, а также ковариационной матрицы случайных ошибок, необходимо воспользоваться выведенной в приложении диссертационного исследования формулой (где $k_i \in N$):

$$\begin{aligned}
E\left(\prod_{t=1}^m \left(\xi_t^{\bar{\alpha}, \bar{\beta}}\right)^{k_t}\right) &= \frac{1}{\psi \bar{F}_{\xi}(\bar{\alpha}, \bar{\beta})} \prod_{t=1}^m (\Phi(\bar{\beta}_t; \mu_t, \sigma_t) - \Phi(\bar{\alpha}_t; \mu_t, \sigma_t))^* \\
& * \sum_{i_1=0}^{K_1} \dots \sum_{i_m=0}^{K_m} \sum_{j_1=0}^{K_1} \dots \sum_{j_m=0}^{K_m} \alpha_{(i_1, \dots, i_m)} \alpha_{(j_1, \dots, j_m)} \prod_{r=1}^m \mathcal{M}_{TR}(i_r + j_r + k_r; \bar{\alpha}_r, \bar{\beta}_r; \mu_r, \sigma_r) = , \\
& = \frac{1}{\tilde{\psi}} \sum_{i_1=0}^{K_1} \dots \sum_{i_m=0}^{K_m} \sum_{j_1=0}^{K_1} \dots \sum_{j_m=0}^{K_m} \alpha_{(i_1, \dots, i_m)} \alpha_{(j_1, \dots, j_m)} \prod_{r=1}^m \mathcal{M}_{TR}(i_r + j_r + k_r; \bar{\alpha}_r, \bar{\beta}_r; \mu_r, \sigma_r) \\
\tilde{\psi} &= \sum_{i_1=0}^{K_1} \dots \sum_{i_m=0}^{K_m} \sum_{j_1=0}^{K_1} \dots \sum_{j_m=0}^{K_m} \alpha_{(i_1, \dots, i_m)} \alpha_{(j_1, \dots, j_m)} \prod_{r=1}^m \mathcal{M}_{TR}(i_r + j_r; \bar{\alpha}_r, \bar{\beta}_r; \mu_r, \sigma_r) .
\end{aligned}$$

Публикация результатов исследования

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

1. Коссова, Е. В., & Потанин, Б. С. (2018). Обобщение метода Хекмана и модели с переключением на случай произвольного числа уравнений отбора. Прикладная эконометрика, 50, 114-143.

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

2. Потанин Б. С. Оценка влияния высшего образования на заработную плату работника // Проблемы прогнозирования. 2019. Т. 3. С. 118-126.

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. Коссова Е. В., Куприянова Л. А., Потанин Б. С. Сравнение точности оценок параметрических и полупараметрических методов коррекции многомерного смещения отбора // Прикладная эконометрика. 2020. Т. 57. Принята к печати.

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. VI Международная конференция «Modern Econometric Tools and Applications – META2019» (Нижний Новгород). 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.
- Kugler, J. E. (1987). The earnings impact of training duration in a developing country. *Journal of Human Resources*, 22(2), 228-247.
- Lee, L. (1978). Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review*, 19, 415-433.
- Manjunath, B. G., & Wilhelm, S. (2012). Moments Calculation For the Doubly Truncated Multivariate Normal Density. *arXiv*.
- Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, 3, 370-379.
- 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.
- Potinin, B. (2019). Estimating the Effect of Higher Education on an Employee's Wage. *Studies on Russian Economic Development*, 30(3), 319-326.
- Richardson, L. F. (1911). The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society A*, 210, 307-357.
- 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.
- Коссова, Е. В., & Потанин, Б. С. (2018). Обобщение метода Хекмана и модели с переключением на случай произвольного числа уравнений отбора. *Прикладная эконометрика*, 50, 114-143.
- Коссова, Е. В., Куприянова, Л. А., & Потанин, Б. С. (2020). Сравнение точности оценок параметрических и полупараметрических методов коррекции многомерного смещения отбора. *Прикладная эконометрика*, 57, 119-139.
- Потанин, Б. С. (2019). Оценка влияния высшего образования на заработную плату работника. *Проблемы прогнозирования*, 3, 118-126.
- Хекман, Д. (2013). Смещение селективной выборки как ошибка спецификации. *Прикладная Эконометрика, Переводчики: Катыйшев П. К., Малахов Д. И., Станкевич И. П.*, 31(3), 127-137.