Марно Вербик

Модели, основанные на панельных данных

От редакции

Последние два десятилетия ознаменовались настоящим прорывом в области прикладного эконометрического инструментария. Создан и развит новый общий подход к оцениванию эконометрических моделей — ***обобщенный метод моментов*** *(GMM — «Generalized Method of Moments»); разработан коинтеграционный анализ моделей регрессии (включая* ***модель коррекции регрессионных остатков*** *— «Error Correction Model»), необходимый при построении моделей регрессии* ***по временным рядам****; предложены методы устранения тех смещений в статистических выводах, которые обусловлены ограничениями на процесс формирования представительной (случайной) выборки («Sample Selection Problem»); наконец, разработаны специальные методы построения регрессионных моделей по так называемым панельным данным («Models Based on Panel Data»). К сожалению, до настоящего времени все эти важные в прикладном плане методы в русскоязычной специальной литературе практически не представлены (или представлены крайне слабо).*

Поэтому редакционная коллегия журнала ПЭ решила создать в рамках журнала специальный раздел ***«Консультации»****, в котором будет оперативно помещаться информация об относительно новых достижениях в области прикладного эконометрического инструментария, слабо отраженных в русскоязычной учебно-методической или монографической литературе.*

В данном номере журнала мы помещаем главу из подготовленного к изданию русского перевода книги Марно Вербика «Путеводитель по современной эконометрике» (издательство «Научная книга», научный редактор перевода С.А. Айвазян, перевод В.А. Банникова), посвященную ***эконометрическому анализу панельных данных*** *и, в определенной мере,* ***проблеме селективности выборки*** *(Sample Selection Problem). Издательство «Научная книга» планирует выпустить эту книгу во втором полугодии 2006 г. Глава публикуется в том виде, в котором она будет представлена в книге.*

Совокупность панельных данных содержит повторные наблюдения для одних и тех же выборочных единиц (людей, домашних хозяйств, фирм), собранные за ряд тактов времени. Хотя панельные данные, как правило, собираются на микроэкономическом уровне, на практике все чаще и чаще начинают объединять индивидуальные временны́е ряды множества стран или множества отраслей промышленности и анализировать их одновременно. Применение повторных (для разных тактов времени) наблюдений относительно одних и тех же выборочных единиц позволяет экономистам специфицировать и оценивать более сложные и более реалистические модели, чем применение одной пространственной («cross-section») выборки или одного временно́го ряда. Неудобства имеют скорее практическую природу: поскольку мы повторно наблюдаем одни и те же выборочные единицы, то обычно больше нереалистично предполагать, что различные наблюдения независимы. Это может усложнить анализ, особенно для нелинейных и динамических моделей. Кроме того, совокупности панельных данных очень часто страдают от пропущенных наблюдений. Даже если эти наблюдения отсутствуют случайным образом (см. ниже), стандартный анализ должен быть скорректирован. Эта глава является введением в анализ панельных данных. В параграфе 10.1 представлена простая линейная модель панельных данных и в контексте этой модели обсуждены определенные преимущества по сравнению с пространственными данными или данными одномерного временно́го ряда. В параграфе 10.2 уделяется внимание так называемым моделям с фиксированными эффектами и моделям со случайными эффектами и обсуждаются проблемы, относящиеся к выбору между этими двумя основными моделями. В параграфе 10.3 приводится эмпирический пример. Введение лагированной зависимой переменной в линейную модель усложняет состоятельное оценивание и, как обсуждается в параграфе 10.4, методы инструментальных переменных или ОММ предоставляют интересные альтернативы. В параграфе 10.5 приводится эмпирический пример оценивания краткосрочных и долгосрочных динамических эластичностей спроса на рабочую силу относительно заработной платы. Другие сложности возникают, когда интересующая нас модель включает ограниченные зависимые переменные. Расширение логит, пробит и тобит моделей на случай панельных данных обсуждается в параграфе 10.6. И, наконец, в параграфе 10.7 мы обсуждаем проблемы, связанные с неполными панельными данными и смещениями, обусловленными ограничениями в способе отбора выборочных единиц[1]. Обширные обсуждения эконометрического анализа панельных данных можно найти в работах (Hsiao, 1986), (Baltagi, 1995) и (Matyas, Sevestre, 1996).

10.1 Преимущества панельных данных

Важное преимущество панельных данных по сравнению с данными одномерного временного ряда или пространственной совокупностью данных состоит в том, что панельные данные позволяют идентифицировать определенные параметры или вопросы без необходимости делать ограничительные допущения. Например, панельные данные позволяют анализировать *изменения* на *индивидуальном* уровне. Рассмотрим ситуацию, в которой средний уровень потребления повышается на 2% ежегодно. Панельные данные могут идентифицировать, является ли это повышение результатом, например, увеличения на 2% уровня потребления для всех индивидуумов или увеличения на 4% уровня потребления приблизительно для одной половины индивидуумов и никакого изменения уровня потребления для другой половины (или результатом любой другой комбинации). Таким образом, панельные данные подходят не только для моделирования или объяснения, почему выборочные единицы ведут себя по-разному, но также и для моделирования, почему конкретная выборочная единица ведет себя по-разному в различные периоды времени (например, из-за различного прошлого).

В последующем мы будем индексировать все переменные индексом i для индивидуумов[2] ($i = 1,\ldots, N$) и индексом t для периодов времени ($t = 1,\ldots, T$) В общем виде мы могли бы специфицировать линейную модель как

$$y_{it} = x'_{it}\beta_{it} + \varepsilon_{it},$$

где вектор коэффициентов β_{it} измеряет частные эффекты вектора объясняющих переменных x_{it} в период t для выборочной единицы i. Конечно, такая модель является слишком общей, чтобы быть полезной, и мы должны наложить более ограничительную структуру на вектор коэффициентов β_{it}. Стандартное предположение, используемое во многих эмпирических случаях, состоит в том, что вектор β_{it} является вектором констант для всех i и t, за исключением, возможно, свободного члена. Такую модель можно написать как

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}, \tag{10.1}$$

где x_{it} — K-мерный вектор объясняющих переменных, не включающий константу[3]. Это означает, что влияние от изменений в компонентах вектора x на y одинаковы для всех выборочных единиц и всех периодов, но средний уровень для выборочной единицы i может отличаться от среднего уровня для выборочной единицы j. Таким образом, коэффициент α_i улавливает эффекты тех переменных, кото-

[1] В общем плане эта проблема («проблема выборочной селективности») обсуждалась в параграфе 7.5 (*прим. научн. ред. пер.*).

[2] Несмотря на то, что мы ссылаемся на пространственные выборочные единицы как на индивидуумов, они могут также относиться к другим выборочным единицам, например, фирмам, странам, отраслям промышленности, домашним хозяйствам или активам.

[3] Элементы в векторе β индексируются от элемента β_1 до β_K, где первый элемент в отличие от предыдущих глав не относится к свободному члену.

рые являются специфическими для i-го индивидуума, и которые являются постоянными во времени. В стандартном случае предполагается, что остатки ε_{it} являются независимыми и одинаково распределенными по индивидуумам и времени с нулевым средним и дисперсией σ_ε^2. Если мы рассматриваем коэффициенты α_i как N фиксированных неизвестных параметров, то модель (10.1) называется стандартной **моделью с фиксированными эффектами.**

Альтернативный подход предполагает, что свободные члены индивидуумов различны, но их можно рассматривать как извлечения из распределения со средним μ и дисперсией σ_α^2. Существенное предположение здесь состоит в том, что эти извлечения являются независимыми от объясняющих переменных в векторе x_{it} (см. ниже). Это приводит к **модели со случайными эффектами**, где индивидуальные эффекты рассматриваются как случайные. Член ошибки в этой модели состоит из двух компонент: не зависящей от времени компоненты[4] α_i и остаточной компоненты ε_{it}, которая некоррелирована во времени[5]. Такую модель можно записать как

$$y_{it} = \mu + x'_{it}\beta + \alpha_i + \varepsilon_{it}, \qquad (10.2)$$

где μ обозначает свободный член.

Возможность рассматривать эффекты α_i как фиксированные параметры имеет несколько больше преимуществ, но также и некоторые неудобства. Большинство моделей панельных данных оценивается либо в предположении фиксированных эффектов, либо в предположении случайных эффектов, и мы будем обсуждать это подробно в параграфе 10.2. Но сначала в следующих двух пунктах обсудим более подробно некоторые потенциальные преимущества панельных данных.

10.1.1 Эффективность оценивания параметров

Поскольку совокупности панельных данных, как правило, обширнее, чем совокупности пространственных данных или совокупности данных одномерного временного ряда, и объясняющие переменные изменяются в двух измерениях (индивидуумы и время), а не в одном измерении, то оценки, построенные на основе панельных данных, весьма часто точнее, чем те, которые построены на основе других источников данных. Даже при одинаковых объемах выборок применение совокупности панельных данных часто будет приводить к более эффективным оценкам, чем ряд независимых совокупностей пространственных данных (где различные выборочные единицы извлекаются в каждом такте времени). Чтобы проиллюстрировать это, рассмотрим следующий специальный случай модели со случайными эффектами (10.2), в которую мы включим временные фиктивные переменные (манекены), т. е.

$$y_{it} = \mu_t + \alpha_i + \varepsilon_{it}, \qquad (10.3)$$

где каждый μ_t является неизвестным параметром, соответствующим среднему значению генеральной совокупности для такта времени t. Предположим, что мы не интересуемся средним μ_t для определенного такта времени t, а интересуемся изменением μ_t от одного такта времени к другому. Вообще дисперсия эффективной оценки для разности $\mu_t - \mu_s$ $(s \neq t)$, задается в виде

$$V\{\mu_t - \mu_s\} = V\{\mu_t\} + V\{\mu_s\} - 2\,\mathrm{cov}\{\mu_t, \mu_s\} \qquad (10.4)$$

с $\mu_t = 1/N \sum_{i=1}^{N} y_{it}$ $(t = 1,\ldots, T)$. Как правило, если используется совокупность панельных данных, то ковариация между средними μ_t и μ_s будет положительна, в частности если справедливы допущения, принятые для модели со случайными эффектами (10.2), то эта ковариация равна σ_α^2/N. Однако если используются две независимые совокупности пространственных данных, то разные периоды времени будут содержать различных индивидуумов, поэтому средние μ_t и μ_s будут иметь нулевую

[4] В модели случайных эффектов случайные величины α_i переопределены таким образом, что они имеют нулевое среднее значение.

[5] Модель иногда называется (однофакторной) моделью остаточных ошибок.

ковариацию. Другими словами, если интересуются изменениями параметра модели от одного такта времени к другому, то методы анализа панельных данных приводят к более эффективным оценкам, чем методы анализа пространственных данных, примененные к той же совокупности исходных данных.

Однако заметим, что обратное также справедливо в том смысле, что повторные пространственные данные будут более информативны, чем панельные данные, когда, например, речь идет об оценке суммы или среднего значения μ_t по совокупности нескольких тактов времени. На интуитивном уровне панельные данные могут предоставить лучшую информацию, поскольку *одни и те же* индивидуумы наблюдаются повторно. С другой стороны, наличие одних и тех же индивидуумов, а не различных, может подразумевать меньшую вариацию в объясняющих переменных и таким образом снижать эффективность построенных по ним оценок. Всесторонний анализ выбора между чисто панельными данными, чисто пространственными данными и совместной комбинацией этих двух источников данных представлен в работе (Nijman, Verbeek, 1990). Результаты работы показали, что когда речь идет об оценке параметров, определяющих эффект влияния включенных в модель экзогенных переменных, то анализ совокупности панельных данных, как правило, будет приводить к более эффективным оценкам, чем анализ, основанный на выборке пространственных данных с тем же самым числом наблюдений.

10.1.2 Идентификация параметров

Другое преимущество наличия панельных данных состоит в том, что ослабляются проблемы идентификации, и хотя такое преимущество может проявляться в разных ситуациях, во многих случаях оно включает идентификацию при наличии эндогенных регрессоров или ошибки измерения, устойчивость по отношению к не включенным в модель переменным и идентификацию индивидуальной динамики.

Начнем с последнего примера. Существует два альтернативных объяснения часто наблюдаемого явления, что индивидуумы, которые испытали некоторое событие в прошлом, более вероятно, испытают то же событие в будущем. Первое объяснение состоит в том, что факт испытания события индивидуумом изменяет его предпочтения, ограничения, и т. п. таким образом, что он более вероятно испытает такое событие в будущем. Второе объяснение говорит, что индивидуумы могут отличаться ненаблюдаемыми особенностями, которые влияют на вероятность испытания события (но испытание события не влияет на ненаблюдаемые особенности индивидуума). Хекмэн (Heckman, 1978) назвал первое объяснение истинной зависимостью состояния, а последнее — мнимой зависимостью состояния. Известный пример относится к «событию» — быть безработным. Наличие панельных данных ослабит проблему различения между истинной и мнимой зависимостью состояния, поскольку наблюдаются индивидуальные предыстории, которые можно включить в модель.

Смещение от невключения переменной возникает, если переменная, которая коррелирована с включенными переменными, не включена в модель. Классическим примером является оценивание производственных функций (Mundlak, 1961). Во многих случаях, особенно в случае малых фирм, в качестве производственных затрат в производственную функцию желательно включить качество менеджмента. Однако вообще качество управления не наблюдаемо. Предположим, что производственная функция типа Кобба-Дугласа задана в виде

$$y_{it} = \mu + x'_{it}\beta + m_i\beta_{K+1} + \varepsilon_{it}, \qquad (10.5)$$

где y_{it} обозначает логарифмический объем производства, x_{it} — K-мерный вектор логарифмических производственных затрат для фирмы i в момент времени t, а m_i обозначает качество управления для фирмы i (которое, как предполагается, является постоянным во времени). Ожидается, что ненаблюдаемая переменная m_i будет отрицательно коррелирована с другими производственными затратами в векторе x_{it}, так как высококачественное управление вероятно приведет к более эффективному использованию производственных затрат. Поэтому кроме случая $\beta_{K+1} = 0$, исключение m_i из модели (10.5) приведет к смещенным оценкам других параметров модели. Если доступны панельные данные, то такую проблему можно решить, введя специфический эффект фирмы $\alpha_i = \mu + m_i\beta_{K+1}$ и рассматривая его в качестве фиксированного неизвестного параметра. Заметим, что без дополнительной информации идентифицировать неизвестный параметр β_{K+1} невозможно; ограничение, которое идентифицирует параметр β_{K+1}, состоит во введении условия постоянной отдачи от масштаба[6].

Подобным образом в модель можно включить фиксированный временной эффект, чтобы уловить эффект всех (наблюдаемых и ненаблюдаемых) переменных, который не изменяется на индивидуальных единицах. Этим поясняется утверждение, что для панельных данных можно снизить эффекты смещения из-за невключенных переменных, или, другими словами, оценки, построенные по совокупности панельных данных, могут быть более устойчивыми к неполной спецификации модели.

И, наконец, во многих случаях панельные данные предоставляют «внутренние» инструментальные переменные для регрессоров, которые являются эндогенными переменными или переменными, подверженными ошибке измерения. Т.е., часто можно аргументировать такие преобразования исходных переменных, при которых они станут некоррелированными с остатками модели и коррелированными с самими объясняющими переменными, и никакие внешние инструментальные переменные не требуются. Например, если вектор x_{it} коррелирован с эффектом α_i, то можно утверждать, что разность $x_{it}-\bar{x}_i$, где $\bar{x}_i$ — среднее по времени для индивидуума *i*, некоррелирована с эффектом α_i и предоставляет действительную инструментальную переменную для вектора x_{it}. Более обще, оценивание модели при предположении фиксированных эффектов устраняет эффект α_i из остаточного члена и, следовательно, устраняет все связанные с этим проблемы эндогенности. Это будет проиллюстрировано в следующем параграфе. Обширное обсуждение преимуществ и ограничений панельных данных представлено в работе (Hsiao, 1985).

10.2. Статическая линейная модель

В этом параграфе мы обсудим статическую линейную модель для панельных данных. Мы начнем с двух основных моделей, модели с фиксированными эффектами и модели со случайными эффектами, и последовательно обсудим выбор между этими двумя моделями, а также обсудим альтернативные процедуры, которые можно рассматривать как промежуточные между обработкой фиксированных эффектов и обработкой случайных эффектов.

10.2.1. Модель с фиксированными эффектами

Модель с фиксированными эффектами является просто линейной моделью регрессии, в которой свободные члены изменяются по индивидуальным единицам *i*, т. е.

$$y_{it}=\alpha_i+x'_{it}\beta+\varepsilon_{it},\ \varepsilon_{it}\sim \text{НОР}(0,\sigma_\varepsilon^2),\tag{10.6}$$

где обычно предполагается, что все x_{it} независимы от всех ε_{it}. Мы можем написать это в обычной структуре регрессии включением фиктивной переменной для каждой единицы *i* в модели. Таким образом,

$$y_{it}=\sum_{j=1}^{N}\alpha_j d_{ij}+x'_{it}\beta+\varepsilon_{it},\tag{10.7}$$

где $d_{ij}=1$, если $i=j$, и $d_{ij}=0$ в противном случае. Таким образом, мы имеем множество из *N* фиктивных переменных в модели. Параметры $\alpha_1,\dots,\alpha_N$ и β можно оценить с помощью МНК в регрессии (10.7). Соответствующая оценка для вектора неизвестных параметров β называется **оценкой метода наименьших квадратов с фиктивными переменными (МНК ФП-оценкой)**. Однако, возможно, непривлекательно с вычислительной точки зрения иметь модель регрессии с таким большим количеством регрессоров. К счастью, можно вычислить оценку для вектора неизвестных параметров β более простым способом. Можно показать, что точно та же самая оценка для вектора β получается, если регрессия строится в отклонениях от индивидуальных средних. По существу,

[6] Постоянная отдача от масштаба производства подразумевает, что $\beta_{K+1}=1-(\beta_1+\dots+\beta_K)$.

это подразумевает, что сначала с помощью преобразования данных мы исключаем индивидуальные эффекты α_i. Чтобы увидеть это, сначала заметим, что

$$\bar{y}_i = \alpha_i + \bar{x}_i' \beta + \bar{\varepsilon}_i, \tag{10.8}$$

где $\bar{y}_i = T^{-1} \sum_t y_{it}$ и аналогично для других переменных. Следовательно, мы можем написать

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i). \tag{10.9}$$

Это — модель регрессии в отклонениях от индивидуальных средних и она не включает индивидуальные эффекты α_i. Преобразование, которое переводит наблюдения в отклонения от индивидуальных средних как в регрессии (10.9), называется **внутригрупповым преобразованием**. МНК-оценку для вектора неизвестных параметров β, полученную из этой преобразованной модели, часто называют **внутригрупповой МНК-оценкой**, или **оценкой с фиксированными эффектами**, и она в точности идентична МНК ФП-оценке, описанной выше. Эта оценка задается в виде

$$\beta_{ФЭ} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i). \tag{10.10}$$

Если предполагается, что *все* x_{it} независимы от *всех* ε_{it} (сравните с предположением (A2) из гл. 2), то можно показать, что оценка с фиксированными эффектами будет несмещенной для вектора неизвестных параметров β. Кроме того, если накладывается условие нормальной распределенности остатков ε_{it}, то $\beta_{ФЭ}$ также имеет нормальное распределение. Для состоятельности[7] требуется, чтобы

$$E\{(x_{it} - \bar{x}_i)\varepsilon_{it}\} = 0 \tag{10.11}$$

(сравните с предположением (A7) из гл. 2 и 5). Для этого достаточно, чтобы x_{it} не был коррелирован с ε_{it} и чтобы $\bar{x}_i$ не имел никакой корреляции с остатками модели. Эти условия в свою очередь обеспечиваются условиями

$$E\{x_{it}\varepsilon_{is}\} = 0 \text{ для всех } s, t, \tag{10.12}$$

при выполнении которых мы называем переменные в векторе x_{it} **строго экзогенными**. Строго экзогенная переменная не должна зависеть от текущих, будущих и прошлых значений остатков. Возможно, что в некоторых приложениях такое условие является ограничительным. Ясно, что оно исключает включение лагированных зависимых переменных в вектор x_{it}, но любая переменная вектора x_{it}, которая зависит от предыстории y_{it}, также нарушила бы это условие. Например, если мы объясняем предложение труда индивидуума, то мы можем захотеть включить в модель годы трудового опыта, несмотря на то, что совершенно ясно, что опыт работы зависит от трудовой предыстории человека.

Если объясняющие переменные независимы от всех остатков, то N свободных членов оцениваются несмещенно как

$$\alpha_i = \bar{y}_i - \bar{x}_i' \beta_{ФЭ}, \; i = 1, \ldots, N.$$

По предположению (10.11) эти оценки состоятельны для фиксированных эффектов α_i (по T стремящемуся к бесконечности). Причина, почему оценки α_i несостоятельны по $N \to \infty$ при фиксированном T, ясна: если T фиксировано, то индивидуальные средние $\bar{y}_i$ и $\bar{x}_i$ при возрастании числа индивидуумов никуда не сходятся.

Предполагая, что остатки ε_{it} являются независимо и одинаково распределенными (по индивидуумам и по времени) с дисперсией σ_ε^2, ковариационная матрица для оценки с фиксированными эффектами $\beta_{ФЭ}$ задается в виде

[7] Если не утверждается иное, то в этой главе мы рассматриваем состоятельность по числу индивидуумов N, стремящемуся к бесконечности. Это соответствует общей ситуации, когда мы имеем панельные данные с большим N и относительно малым T.

$$V\{\beta_{ФЭ}\} = \sigma_\varepsilon^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \qquad (10.13)$$

Если T не является большим, то применение стандартной МНК-оценки для ковариационной матрицы, основанной на внутригрупповой регрессии (10.9), будет недооценивать истинную дисперсию. Причина заключается в том, что в этой преобразованной регрессии ковариационная матрица ошибок является вырожденной (поскольку T преобразованных ошибок каждого индивидуума дают в сумме нуль), и дисперсия разности $\varepsilon_{it} - \bar{\varepsilon}_i$ равна $(T-1)/T\sigma_\varepsilon^2$, а не σ_ε^2. Состоятельная оценка для дисперсии σ_ε^2 получается как внутригрупповая остаточная сумма квадратов, деленная на множитель $N(T-1)$. Таким образом,

$$\sigma_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \alpha_i - x_{it}' \beta_{ФЭ})^2 =$$
$$= \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - y_i - (x_{it} - \bar{x}_i)' \beta_{ФЭ})^2. \qquad (10.14)$$

Можно скорректировать обычные степени свободы вычитанием K в знаменателе. Заметим, что применение стандартной ковариационной матрицы МНК в модели (10.7) с N индивидуальными фиктивными переменными (манекенами) оправдано, поскольку коррекция степеней свободы включает N дополнительных неизвестных параметров, соответствующих индивидуальным свободным членам. При слабых условиях регулярности оценка с фиксированными эффектами асимптотически нормальна, так что можно использовать обычные статистические процедуры (например, t-критерий и критерий Вальда).

По существу, модель с фиксированными эффектами сфокусирована на различиях «внутри» индивидуумов. Т.е. на объяснении, до какой степени y_{it} отличается от $\bar{y}_i$, а не на объяснении, почему $\bar{y}_i$ отличается от $\bar{y}_j$. С другой стороны, параметрические предположения о векторе β накладывают условие, что изменения в x влияют на y одинаково (при прочих равных условиях), является ли это изменением от одного такта времени к другому или изменением от одного индивидуума к другому. Однако интерпретируя результаты для регрессии с фиксированными эффектами, возможно, важно понять, что параметры идентифицируются только через внутрииндивидуальную (или, что то же, внутригрупповую) размерность данных.

10.2.2. Модели со случайными эффектами

В регрессионном анализе обычно предполагается, что все факторы, которые влияют на зависимую переменную, но которые не были включены в качестве регрессоров, соответственно могут в итоге суммироваться в случайном остаточном члене уравнения. В нашем случае это приводит к предположению, что эффекты α_i являются случайными факторами, независимо и одинаково распределенными по индивидуумам. Таким образом, мы записываем модель случайных эффектов в виде

$$y_{it} = \mu + x_{it}'\beta + \alpha_i + \varepsilon_{it},$$
$$\varepsilon_{it} \sim НОР(0, \sigma_\varepsilon^2); \qquad \alpha_i \sim НОР(0, \sigma_\alpha^2), \qquad (10.15)$$

где $\alpha_i + \varepsilon_{it}$ рассматривается как остаточный член, состоящий из двух компонент: индивидуальной специфической компоненты, которая не изменяется во времени, и компоненты остатка, которая, как предполагается, является некоррелированной во времени. Таким образом, вся корреляция остаточных членов во времени приписывается индивидуальным эффектам α_i. Предполагается, что α_i и ε_{it} взаимно независимы и независимы от x_{js} (для всех j и s). Это означает, что МНК-оценки для μ и β в модели со случайными эффектами (10.15) являются несмещенными и состоятельными. Структура компонент остатков подразумевает, что составной остаток $\alpha_i + \varepsilon_{it}$ будет иметь определенный вид автокорреляции (если только $\sigma_\alpha^2 \neq 0$) Следовательно, обычно вычисляемые стандартные ошибки для МНК-оценок некорректны, и можно получить более эффективную оценку (ОМНК-оценку), используя структуру ковариационной матрицы остатков.

Чтобы получить ОМНК-оценку[8], сначала заметим, что для индивидуального i все члены ошибок можно скомпоновать в виде $\alpha_i \iota_T + \varepsilon_i$, где $\iota_T = (1, 1, \ldots, 1)'$ размерности T и $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. Ковариационная матрица этого вектора равна (см. (Hsiao, 1986, p. 34)).

$$V\{\alpha_i \iota_T + \varepsilon_i\} = \Omega = \sigma_\alpha^2 \iota_T \iota_T' + \sigma_\varepsilon^2 I_T, \qquad (10.16)$$

где I_T — T-мерная единичная матрица. Эту ковариационную матрицу можно использовать, чтобы получить ОМНК-оценку для параметров модели со случайными эффектами (10.15). Для каждого индивидуума

мы можем преобразовать данные, умножая слева векторы $y_i = (y_{i1}, \ldots, y_{iT})'$ и т.д. на матрицу Ω^{-1} которая задается как

$$\Omega^{-1} = \sigma_\varepsilon^{-2}\left[I_T - \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}\iota_T \iota_T'\right],$$

и которую также можно записать в виде

$$\Omega^{-1} = \sigma_\varepsilon^{-2}\left[\left(I_T - \frac{1}{T}\iota_T \iota_T'\right) + \psi \frac{1}{T}\iota_T \iota_T'\right],$$

где

$$\psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}.$$

Заметив, что $I_T - (1/T)\,\iota_T \iota_T'$ преобразует данные в отклонения от индивидуальных средних, а $(1/T)\,\iota_T \iota_T'$, принимает индивидуальные средние значения, ОМНК-оценку для вектора неизвестных параметров β можно написать как

$$\beta_{\text{ОМНК}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \psi T \sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1} \times$$
$$\times \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \psi T \sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})\right), \qquad (10.17)$$

где $\bar{x} = (1/(NT))\sum_{i,t} x_{it}$ обозначает общее среднее вектора x_{it}. Легко видеть, что при $\psi = 0$ приходим к оценке с фиксированными эффектами. Поскольку $\psi \to 0$ при $T \to \infty$, то из этого следует, что для большого T оценка с фиксированными эффектами и оценка со случайными эффектами эквивалентны. Если $\psi = 1$, то ОМНК-оценка просто является МНК-оценкой (и Ω является диагональной матрицей). Из общей формулы для ОМНК-оценки можно получить, что

$$\beta_{\text{ОМНК}} = \Delta\beta \quad + (I_k - \Delta)\beta_{\text{ФЭ}},$$

где

$$\beta \quad = \left(\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'\right)^{-1}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})$$

[8] Возможно, полезно снова прочитать общее введение в ОМНК-оценивание в разд. 4.2.

является так называемой **межгрупповой оценкой** для вектора неизвестных параметров β. Она является обычной МНК-оценкой вектора параметров β в модели для индивидуальных средних

$$\bar{y}_i = \mu + \bar{x}_i'\beta + \alpha_i + \bar{\varepsilon}_{it}, \quad i = 1, \ldots, N. \tag{10.18}$$

Матрица Δ является матрицей весов, она пропорциональна обращению ковариационной матрицы оценки β_M (подробности см. в работе (Hsiao, 1986, p. 36)). Таким образом, ОМНК-оценка является матрично-взвешенным средним межгрупповой и внутригрупповой оценок, где веса зависят от соотношения дисперсий этих двух оценок (более точная оценка получает больший вес).

Межгрупповая оценка игнорирует любую внутригрупповую информацию. ОМНК-оценка при сделанных предположениях является оптимальной комбинацией внутригрупповой и межгрупповой оценок и поэтому более эффективна, чем любая из этих двух оценок в отдельности. МНК-оценка (с $\psi = 1$) также является линейной комбинацией этих двух оценок, но не является эффективной оценкой. Таким образом, как обычно, ОМНК-оценки более эффективны, чем обычные МНК-оценки. Если объясняющие переменные независимы от всех ε_{it} и всех α_i, то ОМНК-оценка является несмещенной. Она является состоятельной оценкой по N или T, или N и T, одновременно стремящимся к бесконечности, если в дополнение к условию (10.11) также справедливо, что $E\{\bar{x}_i \varepsilon_{it}\} = 0$ и наиболее важно, что

$$E\{\bar{x}_i \alpha_i\} = 0. \tag{10.19}$$

Заметим, что эти условия также требуются для состоятельности межгрупповой оценки.

Легкий способ вычисления ОМНК-оценки получается, если заметить, что ее можно определить как обычную МНК-оценку для преобразованной модели (см. гл. 4), имеющей вид

$$(y_{it} - \vartheta \bar{y}_i) = \mu(1-\vartheta) + (x_{it} - \vartheta \bar{x}_i) + u_{it}, \tag{10.20}$$

где $\vartheta = 1 - \psi^{1/2}$. Остатки в этой преобразованной регрессии являются независимо и одинаково распределенными по индивидуумам и времени. Опять заметим, что $\psi = 0$ соответствует внутригрупповой оценке ($\vartheta = 1$). В общем, фиксированная доля ϑ индивидуальных средних вычитается из данных, чтобы получить эту преобразованную модель ($0 \le \vartheta \le 1$).

Конечно, компоненты дисперсии σ_α^2 и σ_ε^2 на практике неизвестны. В таком случае мы должны использовать реализуемую ОМНК-оценку (РОМНК), где на первом шаге состоятельно оцениваются неизвестные дисперсии. Оценка дисперсии σ_ε^2 легко получается из внутригрупповых остатков, как это дано в выражении (10.14). В межгрупповой регрессии дисперсия остатка равна $\sigma_\alpha^2 + (1/T)\,\sigma_\varepsilon^2$, которую можно оценить состоятельно в виде

$$\sigma_M^2 = \frac{1}{N} \sum_{i=1}^{N} (\bar{y}_i - \mu_M - \bar{x}_i'\beta_M)^2, \tag{10.21}$$

где μ_M — межгрупповая оценка μ. Отсюда следует состоятельная оценка для дисперсии σ_α^2

$$\sigma_\alpha^2 = \sigma^2 - \frac{1}{T}\sigma_\varepsilon^2. \tag{10.22}$$

Снова возможно скорректировать эту оценку применением коррекции степеней свободы, подразумевая, что число регрессоров $K+1$ вычитается в знаменателе выражения (10.21) (см. (Hsiao, 1986, p. 38) или (Baltagi, 1995, p. 15)). Полученная РОМНК-оценка называется **оценкой со случайными эффектами** для вектора неизвестных параметров β (и μ) и ниже обозначается как $\beta_{СЭ}$.

При слабых условиях регулярности функция оценка со случайными эффектами асимптотически нормальна. Ее ковариационная матрица задается как

$$V\{\beta_{СЭ}\}=\sigma_\varepsilon^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'+\psi T\sum_{i=1}^{N}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'\right)^{-1}, \qquad (10.23)$$

которая показывает, что оценка со случайными эффектами более эффективна, чем оценка с фиксированными эффектами до тех пор, пока $\psi > 0$. Выигрыш в эффективности обусловлен применением межгрупповой вариации в данных ($\bar{x}_i - \bar{x}$). Ковариационная матрица (10.23) обычно оценивается по МНК для преобразованной модели (10.20).

В итоге мы увидели ряд оценок для вектора неизвестных параметров β. Основные две оценки следующие:

1. **Межгрупповая оценка**, использующая межгрупповую размерность данных (различия между индивидуумами), определенная как МНК-оценка для регрессии индивидуальных средних y по индивидуальным средним x (и константе). Состоятельность при $N \to \infty$ требует, чтобы выполнялись условия $E\{\bar{x}_i\ \alpha_i\} = 0$ и $E\{\bar{x}_i\ \bar{\varepsilon}_i\} = 0$. Обычно это означает, что объясняющие переменные являются строго экзогенными и некоррелированными с индивидуальным специфическим эффектом α_i.

2. **Внутригрупповая оценка с фиксированными эффектами**, использующая внутригрупповую размерность данных (различия внутри индивидуумов), определенная как МНК-оценка для регрессии в отклонениях от индивидуальных средних. Она состоятельна для вектора неизвестных параметров β при $T \to \infty$ или $N \to \infty$ при условии, что справедливо $E\{(x_{it}-\bar{x}_i)\,\varepsilon_{it}\} = 0$. И опять состоятельность требует, чтобы x-переменные были строго экзогенными, но это не налагает никаких ограничений на соотношение между α_i и x_{it}.

Другие две оценки следующие:

3. **МНК-оценка**, использующая обе размерности (внутригрупповую и межгрупповую), но не эффективно. Определяется (конечно) как МНК-оценка для исходной модели. Состоятельность при $T \to \infty$ или $N \to \infty$ требует выполнения условия $E\{x_{it}(\varepsilon_{it} + \alpha_i)\} = 0$. Состоятельность требует, чтобы объясняющие переменные были некоррелированными с α_i, но не требует наложения условия их строгой экзогенности. Требуется также, чтобы x_{it} и ε_{it} были «одновременно» некоррелированными (contemporaneously uncorrelated).

4. **РОМНК-оценка со случайными эффектами**, комбинирующая информацию из межгрупповой и внутригрупповой размерности эффективным образом. Она состоятельна при $T \to \infty$ или при $N \to \infty$ при допущениях, сформулированных для оценок в пп. 1 и 2. Ее можно определить как взвешенное среднее межгрупповой и внутригрупповой оценок или как МНК-оценку в регрессии, где переменные преобразованы к виду $y_{it} - \vartheta\bar{y}_i$, где ϑ является оценкой для $\vartheta = 1 - \psi^{1/2}$ с $\psi = \sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + T\sigma_\alpha^2)$.

10.2.3. Фиксированные эффекты или случайные?

Как рассматривать индивидуальные эффекты α_i: как фиксированные или как случайные? — вопрос нелегкий для ответа. Можно привести удивительные различия в оценках неизвестных параметров β в случаях, если T мало, а N является большим. Когда для каждого индивидуума имеется только несколько наблюдений во времени, очень важно наиболее эффективное использование данных. Самая общая точка зрения состоит в том, что обсуждение не должно касаться «истинной природы» эффектов α_i. Соответствующая интерпретация заключается в том, что подход фиксированных эффектов является условным по значениям эффектов α_i. Т.е., по существу рассматривается распределение y_{it} при заданных эффектах α_i, где эффекты α_i можно оценить. Интуитивно такая интерпретация имеет смысл, если индивидуумы в выборке «одного типа», и не могут рассматриваться как случайные извлечения из некоторой лежащей в основе генеральной совокупности. Вероятно, что такая интерпретация наиболее уместна, когда i обозначают страны, большие компании или отрасли промышленности, и мы хотим получить прогнозы для конкретной страны, компании или отрасли промышленности. Таким образом, выводы относятся только к тем эффектам, которые находятся в выборке.

Напротив, подход случайных эффектов не является условным по индивидуальным эффектам α_i, а «исключает их объединением в одно целое». В этом случае обычно мы не заинтересованы в конкретном значении эффекта α_i для некоторого индивидуума; мы просто сфокусированы на случайно выбранных индивидуумах, которые имеют определенные характеристики. Подход случайных эффектов позволяет сделать вывод относительно характеристик генеральной совокупности. Один из способов формализовать различие в подходах состоит в том, чтобы отметить, что в модели со случайными эффектами утверждается

$$E\{y_{it} \mid x_{it}\} = x'_{it}\beta, \tag{10.24}$$

тогда как в модели с фиксированными эффектами оценивается

$$E\{y_{it} \mid x_{it}, \alpha_i\} = x'_{it}\beta + \alpha_i. \tag{10.25}$$

Заметим, что коэффициенты β в этих двух условных математических ожиданиях будут одинаковыми, если только справедливо условие $E\{\alpha_i \mid x_{it}\} = 0$. Суммируя эти соображения, можно сказать, что первая причина, почему можно предпочесть оценку с фиксированными эффектами, заключается в том, что эффекты α_i представляют некоторый интерес, который имеет смысл, если число индивидуальных единиц относительно мало и имеет определенную природу. Т.е. важна идентификация индивидуальных единиц.

Однако даже если мы заинтересованы в большей генеральной совокупности индивидуальных единиц и кажется подходящей структура случайных эффектов, оценка с фиксированными эффектами может быть предпочтительнее. Причина состоит в том, что возможен случай коррелированности α_i и x_{it}, в котором подход случайных эффектов, игнорирующий эту корреляцию, приводит к несостоятельным оценкам. Мы видели это в вышеприведенном примере, в котором эффекты α_i включали качество управления и аргументировалась их коррелированность с другими производственными затратами, включенными в производственную функцию. Проблему корреляции между индивидуальными эффектами α_i и объясняющими переменными в векторе x_{it} можно решить, применив подход фиксированных эффектов, который по существу исключает эффекты α_i из модели и тем самым устраняет любые проблемы, которые могут быть связаны с этими эффектами.

Хаусман (Hausman, 1978) предложил тестирование нулевой гипотезы некоррелированности x_{it} и α_i. Общая идея **теста Хаусмана** состоит в том, что сравниваются две оценки: оценка, которая состоятельна как при нулевой гипотезе, так и при альтернативной гипотезе, и оценка, которая состоятельна (и, как правило, эффективна) только при нулевой гипотезе. Значимое различие между этими двумя оценками указывает, что нулевая гипотеза вряд ли будет справедлива. В настоящем случае предположим, что для всех *s* и *t* выполняется условие $E\{\varepsilon_{it} x_{is}\} = 0$, так что оценка с фиксированными эффектами $\beta_{ФЭ}$ является состоятельной для вектора неизвестных параметров β независимо от того, коррелированы ли x_{it} и α_i, тогда как оценка со случайными эффектами $\beta_{ФЭ}$ состоятельна и эффективна, только если x_{it} и α_i некоррелированы. Рассмотрим вектор разностей $\beta_{ФЭ} - \beta_{СЭ}$. Чтобы оценить значимость этих разностей, нам потребуется ковариационная матрица вектора разностей. В общем, требовалось бы оценить ковариационную матрицу между векторами $\beta_{ФЭ}$ и $\beta_{СЭ}$, но поскольку последняя функция оценивания эффективна при нулевой гипотезе, то можно показать, что (при нулевой гипотезе)

$$V\{\beta_{ФЭ} - \beta_{СЭ}\} = V\{\beta_{ФЭ}\} - V\{\beta_{СЭ}\}. \tag{10.26}$$

Следовательно, мы можем вычислить критическую статистику Хаусмана как

$$\xi_H = (\beta_{ФЭ} - \beta_{СЭ})'[V\{\beta_{ФЭ}\} - V\{\beta_{СЭ}\}]^{1}(\beta_{ФЭ} - \beta_{СЭ}), \tag{10.27}$$

где V — оценки истинных ковариационных матриц. При нулевой гипотезе, которая неявно говорит, что $\text{plim}(\beta_{ФЭ} - \beta_{СЭ}) = 0$, статистика ξ_H имеет асимптотическое хи-квадрат распределение с *K* степенями свободы, где *K* — число элементов в векторе β.

Таким образом, критерий Хаусмана тестирует, значимо ли различие оценок с фиксированными и случайными эффектами. В вычислительном отношении провести такое тестирование относительно легко, поскольку ковариационная матрица удовлетворяет соотношению (10.26). Важная причина, почему эти две оценки могут быть различны, заключается в существовании корреляции между x_{it} и α_i хотя другие виды неправильной спецификации также могут объяснить отклонение нулевой гипотезы (мы увидим такой пример ниже). Практическая проблема при вычислении критической статистики (10.27) состоит в том, что ковариационная матрица в квадратных скобках может быть неположительно определенной в конечных выборках, так что ее обращение нельзя вычислить. В качестве альтернативы можно проводить такое тестирование лишь для подмножества элементов в векторе β.

10.2.4. Качество подгонки данных моделью

Вычисление мер качества подгонки данных моделью в приложениях панельных данных несколько необычно. Одна из причин состоит в том, что можно по-разному оценивать важность объяснения внутригрупповой и межгрупповой вариации в данных. Другая причина заключается в том, что обычный или скорректированный («adjusted») критерии R^2 уместны только тогда, когда модель оценивается с помощью МНК.[9] Наша отправная точка состоит в определении R^2 в терминах квадрата коэффициента корреляции между фактическими и прогнозными значениями, как это представлено в параграфе 2.4 (см. соотношение (2.44)). Такое определение имеет определенное преимущество, поскольку приводит к значениям, находящимся внутри интервала [0, 1] независимо от вида функции оценивания, которая применяется для получения прогнозных значений. Напомним, что это определение соответствует стандартному определению R^2 (в терминах сумм квадратов), если модель оценивается с помощью МНК (при условии включения свободного члена). В текущем контексте полную вариацию переменной y_{it} можно записать в виде суммы внутригрупповой и межгрупповой вариаций, т.е.

$$\frac{1}{NT}\sum_{i,t}(y_{it}-\bar{y})^2=\frac{1}{NT}\sum_{i,t}(y_{it}-\bar{y}_i)^2+\frac{1}{N}\sum_{i}(\bar{y}_i-\bar{y})^2, \qquad (10.28)$$

где $\bar{y}$ обозначает общее выборочное среднее. Теперь мы можем определить альтернативные версии меры R^2 в зависимости от размерности анализируемых данных.

Например, оценка с фиксированными эффектами выбирается, чтобы наиболее полно объяснить внутригрупповую вариацию, и поэтому максимизируется «внутригрупповой R^2», заданный в виде

$$R^2_{\text{вн}}(\beta_{\text{ФЭ}})=corr^2\{\hat{y}^{\text{ФЭ}}_{it}-\hat{y}^{\text{ФЭ}}_{i},\ y_{it}-\bar{y}_i\}, \qquad (10.29)$$

где $\hat{y}^{\text{ФЭ}}_{it}-\hat{y}^{\text{ФЭ}}_{i}=(x_{it}-\bar{x}_i)'\beta_{\text{ФЭ}}$, а $corr^2$ обозначает квадрат коэффициента корреляции. Межгрупповая оценка, являясь МНК-оценкой для модели в терминах индивидуальных средних, максимизирует «межгрупповой R^2», который мы определяем как

$$R^2_{\text{меж}}(\beta_M)=corr^2\{\hat{y}^M_i,\ \bar{y}_i\}, \qquad (10.30)$$

где $\hat{y}^M_i=\bar{x}'_i\beta_M$. МНК-оценка максимизирует общую меру качества подгонки данных моделью и таким образом максимизирует общий R^2, который определяется в виде

$$R^2_{\text{общий}}(\beta)=corr^2\{\hat{y}_{it},\ y_{it}\}, \qquad (10.31)$$

где $\hat{y}_{it}=x'_{it}b$. Возможно определить внутригрупповой, межгрупповой и общий R^2 для произвольной оценки β вектора неизвестных параметров β, применяя в качестве прогнозных значений значения $\hat{y}_{it}=x'_{it}\beta$, $\hat{y}_i=(1/T)\Sigma_t\hat{y}_{it}$ и $\hat{y}=(1/(NT))\sum_{i,t}\hat{y}_{it}$, где свободные члены исключены (и неуместны).[10] При этом оценками с фиксированными эффектами игнорируется вариация, улавливае-

[9] См. соотношения, соответственно, (2.42) и (2.45) в гл. 2 (*прим. научн. ред. пер.*).

[10] Эти определения соответствуют мерам R^2, которые вычисляются в статистическом пакете программ Stata 5.0.

мая эффектами α_i. Если мы учитываем вариацию, объясненную N оцененными свободными членами α_i, то модель с фиксированными эффектами полностью «подгоняет» межгрупповую вариацию. Хотя это несколько неудовлетворительно, поскольку трудно утверждать, что фиксированные эффекты α_i объясняют вариацию между индивидуумами, они только улавливают ее. Выражаясь по-другому, если мы спрашиваем себя: почему индивидуум i в среднем потребляет больше, чем другой индивидуум, то ответ, предоставляемый эффектами α_i, есть просто: «потому, что это индивидуум i». Учитывая этот аргумент и что эффекты α_i часто не вычисляются, кажется уместным игнорировать эту часть модели.

Приняв данное выше определение в терминах квадратов коэффициентов корреляции, три определенные выше меры можно вычислить для любой из оценок, которые мы рассматривали. Если мы берем оценку со случайными эффектами, которая является (асимптотически) наиболее эффективной, при условии справедливости нашего предположения о действии случайных эффектов, то внутригрупповая, межгрупповая и общая меры R^2 обязательно меньше, чем соответствующие меры для фиксированных эффектов, межгрупповой и МНК-оценок, соответственно. Это опять подчеркивает, что меры качества подгонки данных моделью нецелесообразно использовать при выборе между альтернативными методами оценивания. Однако эти меры предоставляют возможные критерии выбора между альтернативными (потенциально не вложенными) спецификациями модели.[11]

10.2.5. Альтернативные оценки метода инструментальных переменных

Метод оценивания с фиксированными эффектами исключает из модели все, что не зависит от времени. Возможно, это высокая цена, которую следует заплатить, чтобы позволить включить в модель помимо переменных x индивидуальную специфицированную гетерогенность α_i. Скажем, мы можем интересоваться влиянием не зависящих от времени переменных (например, пола) на заработную плату индивидуума. В действительности не существует никакой потребности ограничить внимание предположениями существования только фиксированных и случайных эффектов, поскольку возможно получить оценки методом инструментальных переменных, который можно рассматривать в качестве промежуточного подхода между подходами фиксированных и случайных эффектов.

Чтобы это увидеть, прежде всего заметим, что оценку с фиксированными эффектами можно записать в виде

$$\beta_{ФЭ}=\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)=$$
$$=\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)x'_{it}\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)y_{it} \qquad (10.32)$$

Такая запись оценки показывает, что она может быть интерпретирована как оценка метода инструментальных переменных[12] для вектора неизвестных параметров β в модели

$$y_{it}=\mu+x'_{it}\beta+\alpha_i+\varepsilon_{it},$$

где каждая объясняющая переменная инструментована своим значением отклонения от индивидуального специфицированного среднего значения. То есть вектор x_{it} инструментован векторной разностью $x_{it}-\bar{x}_i$. Заметим, что по построению справедливо условие $E\{(x_{it}-\bar{x}_i)\alpha_i\}=0$ (если мы берем математические ожидания по индексам i и t), так что ИП-оценка является состоятельной при условии $E\{(x_{it}-\bar{x}_i)\varepsilon_{it}\}=0$, которое подразумевает строгую экзогенность переменных в векторе x_{it}. Ясно, если известно, что специфическая переменная в векторе x_{it} некоррелирована с эффектом α_i, то ее

[11] В рамках одного и того же метода оценивания. Речь может идти, например, о формировании набора объясняющих переменных (*прим. научн. ред. пер.*).

[12] Возможно, полезно освежить в памяти параграф 5.3, где дается общее обсуждение оценивания методом инструментальных переменных.

обеспечение инструментальной переменной не требуется; т.е. эту переменную можно использовать в качестве ее собственной инструментальной переменной. Такой способ может позволить нам оценивать также влияние переменных, не зависящих от времени.

Чтобы описать общий подход, рассмотрим линейную модель с четырьмя группами объясняющих переменных (Hausman, Taylor, 1981)

$$y_{it} = \mu + x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + w'_{1i}\gamma_1 + w'_{2i}\gamma_2 + \alpha_i + \varepsilon_{it}, \qquad (10.33)$$

где x-переменные изменяются во времени, а w-переменные не зависят от времени. Предполагается, что переменные с индексом 1 некоррелированы с эффектом α_i и со всеми членами ошибок ε_{is}. Переменные $x_{2,it}$ и w_{2i} коррелированы с эффектом α_i, но не с любым членом ошибки ε_{is}. При этих предположениях функция оценивания фиксированных эффектов была бы состоятельной для векторов неизвестных параметров β_1 и β_2, но не идентифицировала бы коэффициенты при переменных, не зависящих от времени. Кроме того, она неэффективна, поскольку вектор $x_{1,it}$ в этом случае инструментован без необходимости. Хаусман и Тэйлор (Hausman, Taylor, 1981) предложили оценивать модель (10.33) методом инструментальных переменных, используя в качестве инструментальных следующие переменные: $x_{1,it}$, w_{1i} и $x_{2,it} - \bar{x}_{2i}$, $\bar{x}_{1i}$. То есть экзогенные переменные служат в качестве их собственных инструментов, вектор $x_{2,it}$ инструментован своим отклонением от вектора индивидуальных средних (как в подходе фиксированных эффектов), а вектор w_{2i} инструментован вектором индивидуальных средних для вектора $x_{1,it}$. Очевидно, идентификация требует, чтобы число переменных в векторе $x_{1,it}$ было, по крайней мере, не меньше числа переменных в векторе w_{2i}. Полученная оценка, **оценка Хаусмана-Тэйлора**, позволяет нам оценивать эффекты переменных, не зависящих от времени, даже, несмотря на то, что изменяющиеся во времени регрессоры коррелированы с эффектом α_i. Если переменные, не зависящие от времени, предполагаются также коррелированными с эффектом α_i, то их также следует обеспечить инструментальными переменными, и мы потребуем, чтобы включалось достаточное количество переменных, зависящих от времени, которые не коррелированы с эффектом α_i. Конечно, существует прямое расширение для включения дополнительных инструментальных переменных в процедуру, которые не основаны на переменных, включенных в модель. К такому приему прямого расширения прибегают в случае пространственных данных, где не существует никаких доступных преобразований, которые могли бы быть аргументированы для предоставления обоснованных инструментальных переменных. Главное преимущество подхода Хаусмана-Тэйлора состоит в том, что не требуется применение внешних инструментальных переменных. При достаточных предположениях инструментальные переменные можно получить внутри модели. Несмотря на это важное преимущество, оценка Хаусмана-Тэйлора играет удивительно незначительную роль в текущей эмпирической работе.

Хаусман и Тэйлор также показали, что множество инструментальных переменных эквивалентно применению $x_{1,it} - \bar{x}_{1i}$, $x_{2,it} - \bar{x}_{2i}$ и $x_{1,it}$, w_{1i}. Это следует непосредственно из того факта, что взятие разных линейных комбинаций исходных инструментальных переменных не влияет на оценку. Хаусман и Тэйлор также показали, как в модели (10.33) можно использовать недиагональную ковариационную матрицу остатков, чтобы улучшить эффективность оценки. В настоящее время оценивание, как правило, проводится в рамках обобщенного метода моментов (ОММ), что мы увидим в параграфе 10.3 (см. (Arellano, Bover, 1995)).

В двух статьях, последовавших за (Hausman, Taylor, 1981), делались попытки улучшить эффективность оценки методом инструментальных переменных Хаусмана-Тэйлора, с помощью введения большего множества инструментальных переменных. В статье (Amemiya, MaCurdy, 1986) предлагается также применение *не зависящих от времени* инструментальных переменных от $x_{1,i1} - \bar{x}_{1i}$ вплоть до $x_{1,iT} - \bar{x}_{1i}$. Это требует, чтобы для каждого t выполнялось условие $E\{(x_{1,it} - \bar{x}_{1i})\alpha_i\} = 0$. Такое предположение имеет смысл, если корреляция между α_i и $x_{1,it}$ обусловлена наличием не зависящей от времени компоненты в векторе $x_{1,it}$ такой, что $E\{x_{1,it}\alpha_i\}$ для данного t не зависит от t. В статье (Breusch, Mizon, Schmidt, 1989) представлен подробный обзор литературы по этой тематике и в качестве дополнительных инструментальных переменных предлагается применение не зависящих от времени переменных от $x_{2,i1} - \bar{x}_{2i}$ до $x_{2,iT} - \bar{x}_{2i}$.

10.2.6 Альтернативные структуры остатков

В моделях со случайными эффектами и моделях с фиксированными эффектами предполагается, что присутствие α_i улавливает всю корреляцию между ненаблюдаемыми переменными в различные периоды времени. Таким образом, предполагается, что остатки ε_{it}, являются некоррелированными по индивидуумам и времени. При условии, что переменные в векторе x_{it} строго экзогенны, присутствие автокорреляции в остатках ε_{it} не приводит к несостоятельности стандартных оценок. Однако стандартные ошибки и получающиеся критерии становятся недействительными (точно так же, как в гл. 4). Кроме того, это будет означать, что оценки больше не эффективны. Например, если истинная ковариационная матрица Ω не удовлетворяет выражению (10.16), то оценка со случайными эффектами больше не соответствует РОМНК-оценке вектора неизвестных параметров β. Как мы знаем, присутствие гетероскедастичности в остатках ε_{it} или в эффектах α_i для модели со случайными эффектами имеет аналогичные последствия.

Один из способов избежать вводящих в заблуждение выводов без необходимости налагать альтернативные предположения на структуру ковариационной матрицы остатков Ω, состоит в использовании МНК-оценки для вектора неизвестных параметров β и одновременно коррекции ее стандартных ошибок в соответствии с общими формами гетероскедастичности и автокорреляции. Рассмотрим следующую модель[13]

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \tag{10.34}$$

без предположения, что u_{it} имеет некоторую структуру из компонент остатков. Состоятельность МНК-оценки

$$b = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x'_{it} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} y_{it} \tag{10.35}$$

вектора параметров β требует, чтобы выполнялось условие

$$E\{x_{it} u_{it}\} = 0. \tag{10.36}$$

Предполагая, что остатки для различных индивидуумов являются некоррелированными ($E\{u_{it} u_{js}\} = 0$ для всех $i \neq j$) ковариационную матрицу МНК-оценки можно оценить по Невье-Весту из гл. 4, т.е.

$$V\{b\} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x'_{it} \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} u_{it} u_{is} x_{it} x'_{is} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x'_{it} \right)^{-1}, \tag{10.37}$$

где u_{it} обозначает МНК-оцененный остаток. Эта оценка учитывает общие формы гетероскедастичности, так же как и автокорреляции (внутригрупповой). Если гетероскедастичность исключается априори, среднюю матрицу в выражении (10.37) можно заменить матрицей

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} u_{it} u_{is} \right) x_{it} x'_{is}, \tag{10.38}$$

где $(1/N)\sum_{i=1}^{N} u_{it} u_{is}$ состоятельная оценка для матрицы $\Omega_{ts} = E\{u_{it} u_{is}\}$.

Если бы остаток u_{it} имел не зависящую от времени компоненту α_i, которая могла бы быть коррелирована с объясняющими переменными, то оценка с фиксированными эффектами

[13] Для удобства обозначений предполагается, что в вектор x_{it} включена константа.

была бы более уместна, чем МНК-оценка, и могла бы быть использована аналогичная коррекция для гетероскедастичности и автокорреляции (в остатках ε_{it}) (Arellano, 1987). Получающееся выражение было бы подобно выражению (10.37), но каждый вектор x_{it} заменялся бы внутригрупповым преобразованием $x_{it} - \bar{x}_i$, а МНК-оцененный остаток — внутригрупповым МНК-оцененным остатком (см. (Baltagi, 1995, p. 13)).

Если нелишне специфицировать определенные предположения о форме гетероскедастичности или автокорреляции, то можно получить более эффективные оценки, чем МНК-оценка или оценка с фиксированными эффектами, используя известную структуру ковариационной матрицы остатков и применяя РОМНК или метод максимального правдоподобия. Краткий обзор ряда таких оценок, которые в вычислительном отношении являются, как правило, малопривлекательными, представлен в (Baltagi, 1995, Chapter 5). В монографии (Kmenta, 1986) предлагается относительно простая РОМНК-оценка, которая учитывает автокорреляцию первого порядка в остатках u_{it} вместе с индивидуальной специфической гетероскедастичностью, но не учитывает компоненту, зависящую от времени в остатках u_{it} (см. (Baltagi, 1996)).

10.2.7. Тестирование на наличие гетероскедастичности и автокорреляции

Большинство тестов, которые можно применить для тестирования на наличие гетероскедастичности или автокорреляции в модели со случайными эффектами, вычислительно обременительны. Для модели с фиксированными эффектами, которая по существу оценивается с помощью МНК, проведение такого тестирования относительно менее сложно. К счастью, можно использовать оценку с фиксированными эффектами, даже если мы делаем предположение о случайных эффектах, т. е. о том, что эффекты α_i являются независимо и одинаково распределенными случайными величинами, независимыми от объясняющих переменных. Поэтому и в случае модели со случайными эффектами можно использовать процедуры тестирования, как это делается в модели с фиксированными эффектами.

Довольно простое тестирование на наличие автокорреляции в модели с фиксированными эффектами основано на тесте Дарбина-Уотсона, обсужденном в гл. 4. Альтернативная гипотеза состоит в том, что

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + v_{it}, \qquad (10.39)$$

где v_{it} являются независимо и одинаково распределенными по индивидуумам и времени. Этим учитывается автокорреляция во времени с ограничением, что каждый индивидуум имеет один и тот же коэффициент автокорреляции ρ. Нулевой гипотезой при тестировании является гипотеза $H_0: \rho = 0$ против односторонней альтернативной гипотезы $\rho < 0$ или $\rho > 0$. Пусть ε_{it} обозначают *остатки внутригрупповой регрессии* (10.9) или, что эквивалентно, остатки регрессии с фиктивными переменными (10.7). Для такого случая в статье (Bhargava, Franzini, Narendranathan, 1983) предлагается следующее обобщение статистики Дарбина-Уотсона:

$$dw_p = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(\varepsilon_{it} - \varepsilon_{i,t-1})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2}. \qquad (10.40)$$

Используя такую же логику вывода, как Дарбин и Уотсон, авторы статьи смогли получить нижнюю и верхнюю границы для истинных критических значений, которые зависят только от N, T, и K. В отличие от случая «чисто» временного ряда, область неопределенности теста Дарбина-Уотсона в панельных данных является малой, особенно когда число индивидуумов в панельных данных большое. В табл. 10.1 мы представили некоторые выбранные нижние и верхние границы для истинных 5%-х критических значений, которые можно использовать для тестирования против альтернативной гипотезы наличия положительной автокорреляции. Числа в таблице подтверждают, что области неопределенности являются малыми, а также показывают, что варьирование критических значений, обусловленное изменением *K*, *N* или *T*, весьма ограничено. В модели с тремя объясняющими переменными, оцененными для 6 периодов времени, нулевая гипотеза H_0 : $\rho = 0$ отклоняется на 5 % уровне значимости, если dw_p меньше 1,859 для N = 100, или меньше 1,957 для N = 1000, против односторонней альтернативной гипотезы $\rho > 0$. Для панельных данных при больших *N* авторы статьи предложили простое правило тестирования нулевой гипотезы против альтернативной гипотезы наличия положительной автокорреляции: нулевая гипотеза отклоняется, если вычисленная статистика dw_p меньше двух. Поскольку оценка с фиксированными эффектами состоятельна и для модели со случайными эффектами, то этот тест Дарбина-Уотсона для панельных данных можно использовать также и в модели со случайными эффектами.

Таблица 10.1

5 %-ные нижняя и верхняя границы теста Дарбина-Уотсона для панельных данных

		$N = 100$		$N = 500$		$N = 1000$	
		d_l	d_U	d_L	d_U	d_L	d_U
$T = 6$	$K = 3$	1,859	1,880	1,939	1,943	1,957	1,950
	$K = 9$	1,839	1,902	1,935	1,947	1,954	1,961
$T = 10$	$K = 3$	1,891	1,904	1,952	1,954	1,967	1,968
	$K = 9$	1,878	1,916	1,949	1,957	1,965	1,970

Чтобы протестировать наличие гетероскедастичности в остатках ε_{it}, мы можем опять воспользоваться остатками модели с фиксированными эффектами ε_{it}. Вспомогательная регрессия для проведения тестирования строится в виде регрессии квадратов внутригрупповых МНК-оцененных остатков ε_{it}^2 по константе и *J* переменным z_{it}, которые, как предполагается, могут повлиять на гетероскедастичность. Такой тест является вариантом теста Бреуша-Пагана[14] на наличие гетероскедастичности, обсужденного в гл. 4. Альтернативная гипотеза для теста заключается в предположении, что

$$V\{\varepsilon_{it}\} = \sigma^2 h(z'_{it}\alpha), \tag{10.41}$$

где *h* — неизвестная, непрерывно дифференцируемая функция с условием $h(0) = 1$, а тестируемая нулевая гипотеза задается в виде H_0 : $\alpha = 0$. При нулевой гипотезе критическая статистика, вычисленная как $N(T-1)$, умноженное на R^2 вспомогательной регрессии, будет иметь асимптотическое хи-квадрат распределение с *J* степенями свободы. Альтернативный тест можно построить с помощью вычисления остатков межгрупповой регрессии, и критическая статистика равна *N*, умноженному на R^2 вспомогательной регрессии межгрупповых остатков по $\bar{z}_i$ или, более обще, по $z_{i1},\ldots, z_{iT}$. При нулевой гипотезе о гомоскедастичности остатков критическая статистика имеет асимптотическое хи-квадрат распределение со степенями свободы, равными числу переменных, включенных во вспомогательную регрессию (за исключением свободного члена). Альтернативная гипотеза такого теста является менее определенной.

10.3. Пример: объяснение индивидуальной заработной платы

В этом параграфе, чтобы оценить уравнение индивидуальной заработной платы, мы применим ряд описанных выше методов оценивания. Данные[15] взяты из Молодежной выборки национального

[14] В контексте панельных данных термин «тест Бреуша-Пагана» обычно связывается с тестом множителей Лагранжа для модели со случайными эффектами при нулевой гипотезе, что никаких индивидуальных специфических эффектов не существует $\sigma_\alpha^2 = 0$ (см. (Baltagi, 1995, Sect. 4.2.1)). В приложениях этот тест почти всегда отклоняет нулевую гипотезу.

[15] Данные, используемые в этом параграфе, доступны в MALES.

протяженного во времени обследования[16], проведенного в США, и представляют собой выборку из 545 работников-мужчин, занятых полный рабочий день, которые закончили свое обучение в 1980 г., а затем работали в течение 1980–1987 гг. Мужчины в выборке молодые, в возрасте от 17 до 23 лет (по состоянию на 1980 г.), и вышли на трудовой рынок довольно недавно, в среднем с тремя годами опыта работы на начало выборочного периода. Данные и спецификации, которые мы выбираем, аналогичны тем, что в статье (Vella, Verbeek, 1998). Логарифм заработной платы объясняется с помощью следующих переменных: времени обучения (в годах), опыта работы (в годах) и его квадрата, фиктивных переменных (манекенов) — членства в профсоюзе (состоит, не состоит), работы в общественном секторе (общественный сектор, частный сектор), семейного положения (женат, холост) и двух расовых фиктивных переменных.

Оценивание[17] проводилось с помощью межгрупповой оценки, основанной на индивидуальных средних, и с помощью внутригрупповой оценки, основанной на отклонениях от индивидуальных средних. Результаты оценивания представлены в табл. 10.2. Прежде всего следует заметить, что оценка с фиксированными эффектами (или внутригрупповая оценка) исключает из модели любые переменные, не зависящие от времени. Это означает, что в этом случае влияние времени обучения и расовых фиктивных переменных не учитываются. Различия между двумя рядами оценок кажутся существенными, и мы возвратимся к этому ниже. В следующей колонке представлены результаты МНК-оценивания, примененного к модели со случайными эффектами, в котором стандартные ошибки не скорректированы с учетом структуры компонент остатков. Последний столбец представляет результаты применения РОМНК-оценивания случайных эффектов. Как обсуждалось в пункте 2 параграфа 10.2, дисперсии компонент ошибок α_i и ε_{it} можно оценить по внутри- и межгрупповым остаткам. В частности мы, имеем $\sigma_м^2 = 0{,}1209$ и $\sigma_\varepsilon^2 = 0{,}1234$. Отсюда можно состоятельно оценить σ_α^2 как $\sigma_\alpha^2 = 0{,}1209 - 0{,}1234/8 = 0{,}1055$. Следовательно, множитель ψ оценивается как

$$\psi = \frac{0{,}1234}{0{,}1234 + 8 \times 0{,}1055} = 0{,}1276,$$

что приводит к $\vartheta = 1 - \psi^{1/2} = 0{,}6428$. Это значит, что РОМНК-оценку можно получить из преобразованной регрессии, где 0,64, умноженное на индивидуальное среднее значение, вычитается из исходных данных. Вспомним, что в МНК-оценке полагают, что $\vartheta = 0$, в то время как в оценке с фиксированными эффектами используется условие $\vartheta = 1$. Заметим, что значения МНК-оценок и оценок со случайными эффектами находятся внутри интервала с границами: межгрупповые оценки и оценки с фиксированными эффектами.

Если удовлетворяются предположения модели со случайными эффектами, то все четыре оценки в табл. 10.2 состоятельны и оценка со случайными эффектами является самой эффективной. Однако если индивидуальные эффекты α_i коррелированы с одной или более объясняющими переменными, то только оценка с фиксированными эффектами является состоятельной. Такую гипотезу можно протестировать, сравнивая межгрупповую и внутригрупповую оценки, или внутригрупповую оценку с оценкой со случайными эффектами. Оба сравнения приводят к эквивалентным тестам. Самое простое тестирование состоит в проведении теста Хаусмана, обсужденного в п. 3 параграфа 10.2, основанного на сравнении внутригрупповой оценки и оценки со случайными эффектами. Критическая статистика принимает значение, равное 31,75, и отражает различия в коэффициентах при переменных опыта работы, квадрата опыта работы и при манекенах членства в профсоюзе, семейного положения и работы в общественном секторе. При нулевой гипотезе критическая статистика подчиняется хи-квадрат распределению с 5 степенями свободы, так что нам следует отклонить нулевую гипотезу на любом разумном уровне значимости.

[16] Речь идет о «Youth Sample of the National Longitudinal Survey» (*прим. научн. ред. пер.*).

[17] Результаты оценивания в этом параграфе получены с помощью статистического пакета программ Stata 5.0.

Таблица 10.2

Результаты оценивания уравнения заработной платы мужчин, по данным за 1980–1987 гг. (в круглых скобках стандартные ошибки)

Зависимая переменная: log (wage)				
Переменная модели	Межгрупповая оценка	Оценка с фиксированными эффектами	МНК-оценка	Оценка со случайными эффектами
константа	0,490 (0,221)	—	−0,034 (0,065)	−0,104 (0,111)
время обучения	0,095 (0,011)	—	0,099 (0,005)	0,101 (0,009)
опыт работы	−0,050 (0,050)	0,116 (0,008)	0,089 (0,010)	0,112 0,008)
опыт работы в квадрате	0,0051 (0,0032)	−0,0043 (0,0006)	−0,0028 (0,0007)	−0,0041 (0,0006)
членство в профсоюзе	0,274 (0,047)	0,081 (0,019)	0,180 (0,017)	0,106 (0,018)
семейное положение	0,145 (0,041)	0,045 (0,018)	0,108 (0,016)	0,063 (0,017)
афроамериканец	−0,139 (0,049)	—	−0,144 (0,024)	−0,144 (0,048)
латино-американец	0,005 (0,043)	—	0,016 (0,021)	0,020 (0,043)
работа в общественном секторе	−0,056 (0,109)	0,035 (0,039)	0,004 (0,037)	0,030 (0,036)
внутригрупповой R^2	0,0470	0,1782	0,1679	0,1776
межгрупповой R^2	0,2196	0,0006	0,2027	0,1835
общий R^2	0,1371	0,0642	0,1866	0,1808

Семейное положение является фиктивной переменной, которая, вероятно, будет коррелирована с ненаблюдаемой гетерогенностью в эффектах α_i. Как правило, можно было бы не ожидать значимого причинного влияния семейного положения на заработную плату, поскольку манекен семейного положения обычно улавливает другие (ненаблюдаемые) различия между женатыми и холостыми рабочими. Это подтверждается результатами в таблице. Если мы исключаем индивидуальные эффекты из модели и рассматриваем оценку с фиксированными эффектами, то влияние манекена семейного положения снижается до 4,5%, тогда как, например, в случае межгрупповой оценки оно составляет почти 15%. Заметим, что влияние манекена семейного положения в подходе фиксированных эффектов идентифицируется только через людей, которые изменяют свое семейное положение в течение периода выборочного обследования. Подобные замечания можно сделать для влияния манекена членства в профсоюзе на заработную плату работника. Однако вспомним, что все оценки предполагают некоррелированность объясняющих переменных с остатками ε_{it}. Если бы такие корреляции существовали, то даже оценка с фиксированными эффектами была бы несостоятельной. В статье (Vella, Verbeek, 1998) уделяется особое внимание влиянию эндогенного статуса принадлежности к членам профсоюза на заработную плату работников этой группы и рассматриваются альтернативные, более сложные методы оценивания.

Меры качества подгонки данных моделью подтверждают, что оценка с фиксированными эффектами приводит к наибольшему внутригрупповому R^2 и таким образом насколько возможно объясняет внутригрупповую вариацию. МНК-оценка максимизирует обычный (общий) критерий R^2, в то время как оценка со случайными эффектами приводит к приемлемым значениям критериев R^2 для всех случаев. Вспомним, что стандартные ошибки МНК-оценки вводят в заблуждение, поскольку они не принимают в расчет корреляцию различных остатков. Корректные стандартные ошибки для МНК-оценки должны быть больше, чем стандартные ошибки для эффективной РОМНК-оценки, которая учитывает эти корреляции.

10.4. Динамические линейные модели

Способность моделировать индивидуальную динамику относится к главным преимуществам панельных данных. Во многих экономических моделях предполагается, что текущее поведение зависит от прошлого поведения (постоянство, формирование навыков, частичная корректировка, и т.д.)[18], поэтому во многих случаях хотелось бы оценить динамическую модель на индивидуальном уровне. Способность моделировать индивидуальную динамику с помощью панельных данных уникальна.

10.4.1. Модель авторегрессии панельных данных

Рассмотрим линейную динамическую модель с экзогенными переменными и лагированной зависимой переменной в роли регрессоров, т.е. модель

$$y_{it} = x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it},$$

где предполагается, что остатки ε_{it} являются НОП(0, σ_ε^2). Для статической модели мы проводили обсуждение состоятельности (устойчивости) и эффективности при выборе между моделями с фиксированными и случайными эффектами α_i. В динамической модели ситуация существенно отличается, поскольку лагированная зависимая переменная $y_{i,t-1}$ будет зависеть от эффекта α_i независимо от способа, с помощью которого мы анализируем эти эффекты. Чтобы проиллюстрировать проблемы, которые возникают при этом, сначала рассмотрим случай модели, где не включаются никакие экзогенные переменные, и модель представляется в виде:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, \quad |\gamma| < 1. \tag{10.42}$$

Предположим, что мы имеем наблюдения относительно переменной y_{it} для тактов времени $t = 0, 1, \ldots, T$.

Оценка с фиксированными эффектами для неизвестного параметра γ имеет вид

$$\gamma_{ФЭ} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \bar{y}_{i,-1})^2}, \tag{10.43}$$

где $\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{it}$, а $\bar{y}_{i,-1} = (1/T)\sum_{t=1}^{T} y_{i,t-1}$. Чтобы проанализировать свойства оценки $\gamma_{ФЭ}$, мы можем подставить выражение (10.42) в выражение (10.43) и получить оценку в виде

$$\gamma_{ФЭ} = \gamma + \frac{(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \bar{y}_{i,-1})^2}. \tag{10.44}$$

Однако эта оценка при $N \to \infty$ и фиксированном T смещенная и несостоятельная, поскольку последний член в правой части выражения (10.44) не имеет нулевого математического ожидания и не

[18] Об этих свойствах динамических моделей речь шла в предыдущих двух главах (*прим. науч. ред. пер.*).

сходится к нулю при N, стремящемся к бесконечности. В частности, можно показать, что (Nickell, 1981; Hsiao, 1986, p. 74)

$$\underset{N\to\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it}-\bar{\varepsilon}_i)(y_{i,t-1}-\bar{y}_{i,-1})=-\frac{\sigma_\varepsilon^2}{T^2}\frac{(T-1)-T\gamma+\gamma^T}{(1-\gamma)^2}\neq 0. \quad (10.45)$$

Таким образом, при фиксированном T мы имеем несостоятельную оценку. Заметим, что эта несостоятельность не вызывается ничем из того, что мы предполагали о эффектах α_i, поскольку они исключаются при оценивании. Проблема состоит в том, что внутригрупповая преобразованная лагированная зависимая переменная коррелирована с внутригрупповым преобразованным остатком. Если $T\to\infty$, то вероятностный предел (10.45) сходится к нулю, так что оценка с фиксированными эффектами является состоятельной для γ, если $T\to\infty$ и $N\to\infty$.

Можно было бы думать, что асимптотическое смещение для фиксированного T является весьма малым и поэтому реальной проблемы, вроде бы, нет. Конечно, это не так, поскольку для конечного T смещение едва ли можно игнорировать. Например, если истинное значение параметра γ равняется 0,5, то можно легко вычислить, что (при $N\to\infty$)

$$\begin{aligned} plim\,\gamma_{FE}&=-0{,}25, \text{ если } T=2,\\ plim\,\gamma_{FE}&=-0{,}04, \text{ если } T=3,\\ plim\,\gamma_{FE}&=0{,}33, \text{ если } T=10,\end{aligned}$$

поэтому даже для средних значений T смещение существенно. К счастью, существуют относительно легкие способы избежать таких смещений.

Для решения проблемы несостоятельности, прежде всего, начнем с другого преобразования, чтобы устранить индивидуальные эффекты α_i, в частности, мы возьмем первые разности. Это приводит к модели

$$y_{it}-y_{i,t-1}=\gamma(y_{i,t-1}-y_{i,t-2})+(\varepsilon_{it}-\varepsilon_{i,t-1}),\ t=2,\ldots,T. \quad (10.46)$$

Если мы оцениваем ее с помощью МНК, то мы не получаем состоятельную оценку для неизвестного параметра γ даже при $T\to\infty$, поскольку лагированная зависимая переменная $y_{i,t-1}$ и остатки $\varepsilon_{i,t-1}$ по определению коррелированы. Однако такая преобразованная спецификация наводит на мысль о применении метода инструментальных переменных. Например, лагированная зависимая переменная $y_{i,t-2}$ коррелирована с разностью $y_{i,t-1}-y_{i,t-2}$, но не с лагированным остатком $\varepsilon_{i,t-1}$, если только остаток ε_{it} не обнаруживает автокорреляцию (наличие который мы исключаем по предположению). Тем самым для оценивания неизвестного параметра γ можно воспользоваться методом инструментальных переменных[19]

$$\gamma_{ИП}=\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}y_{i,t-2}(y_{it}-y_{i,t-1})}{\sum_{i=1}^{N}\sum_{t=2}^{T}y_{i,t-2}(y_{i,t-1}-y_{i,t-2})} \quad (10.47)$$

Необходимое условие для состоятельности этой функции оценивания заключается в том, что

$$plim\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(\varepsilon_{it}-\varepsilon_{i,t-1})y_{i,t-2}=0 \quad (10.48)$$

для T или для N, или одновременно для T и N, стремящихся к бесконечности. Оценка (10.47) является одной из оценок Андерсона-Хсяо, предложенных в статье (Anderson, Hsiao, 1981). Авторы статьи также предложили альтернативу, где в качестве инструментальной переменной используется раз-

[19] См. параграф 5.3 для общего введения в оценивание методом инструментальных переменных.

ность $y_{i,t-2} - y_{i,t-3}$. Тогда альтернативная оценка методом инструментальных переменных будет иметь вид

$$\gamma_{\text{ИП}}^{(2)} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} (y_{i,t-2} - y_{i,t-3})(y_{it} - y_{i,t-1})}{\sum_{i=1}^{N} \sum_{t=3}^{T} (y_{i,t-2} - y_{i,t-3})(y_{it-1} - y_{i,t-2})}, \tag{10.49}$$

которая является состоятельной (при условиях регулярности), если

$$\text{plim} \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=3}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1})(y_{i,t-2} - y_{i,t-3}) = 0. \tag{10.50}$$

Состоятельность этих двух оценок гарантируется предположением, что остаток ε_{it} не имеет никакой автокорреляции.

Заметим, что для второй МИП-оценки при построении инструментальной переменной требуется дополнительный сдвиг, так что эффективное число наблюдений, используемых для оценивания, уменьшается (один такт времени «потерян»). Вопрос, какую из этих оценок следует выбрать не является, по существу, спорным. Подход, основанный на методе моментов, позволяет унифицировать эти оценки и устранить недостатки, связанные со снижением объемов выборок. На первом шаге этого подхода следует отметить, что

$$\text{plim} \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=3}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1}) y_{i,t-2} = E\{(\varepsilon_{it} - \varepsilon_{i,t-1}) y_{i,t-2}\} = 0 \tag{10.51}$$

является условием моментов (см. гл. 5). Точно так же условием моментов является

$$\text{plim} \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=3}^{T} (\varepsilon_{it} - \varepsilon_{i,t-1})(y_{i,t-2} - y_{i,t-3}) = E\{(\varepsilon_{it} - \varepsilon_{i,t-1})(y_{i,t-2} - y_{i,t-3})\} = 0. \tag{10.52}$$

Таким образом, при оценивании для обеих МИП-оценок налагается одно условие моментов. Известно, что наложение большего количества условий моментов повышает эффективность оценок (конечно, если действительны дополнительные условия). В статье (Arellano, Bond, 1991) предлагается расширить перечень инструментальных переменных с помощью введения дополнительных условий моментов, позволяя количеству этих условий изменяться с t. Для этого авторы статьи положили T фиксированным. Например, при $T = 4$ мы имеем

$$E\{(\varepsilon_{i2} - \varepsilon_{i1}) y_{i0}\} = 0,$$

как условие моментов для $t = 2$. Для $t = 3$ мы имеем

$$E\{(\varepsilon_{i3} - \varepsilon_{i2}) y_{i1}\} = 0,$$

но также справедливо, что

$$E\{(\varepsilon_{i3} - \varepsilon_{i23}) y_{i0}\} = 0.$$

Для такта времени $t = 4$ мы имеем условия трех моментов и можем ввести соответственно три инструментальных переменных

$$E\{(\varepsilon_{i4} - \varepsilon_{i3}) y_{i0}\} = 0,$$
$$E\{(\varepsilon_{i4} - \varepsilon_{i3}) y_{i1}\} = 0,$$
$$E\{(\varepsilon_{i4} - \varepsilon_{i3}) y_{i2}\} = 0.$$

Все эти условия моментов можно использовать в схеме реализации обобщенного метода моментов (ОММ). С целью построения ОММ-оценки определим

$$\Delta\varepsilon_i = \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \cdots \\ \varepsilon_{i,T} - \varepsilon_{i,T-1} \end{pmatrix} \tag{10.53}$$

как вектор преобразованных остатков, и

$$Z_i = \begin{pmatrix} [y_{i0}] & 0 & \cdots & 0 \\ 0 & [y_{i0}, y_{i1}] & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & [y_{i0}, \ldots, y_{i,T-2}] \end{pmatrix} \tag{10.54}$$

как матрицу значений инструментальных переменных[20]. Каждая строка в матрице Z_i содержит инструментальные переменные, которые правомочны для данного такта времени. Следовательно, совокупность всех условий моментов можно записать кратко в виде

$$E\{Z_i' \Delta\varepsilon_i\} = 0. \tag{10.55}$$

Заметим, что число этих условий равно $1+2+3+\cdots+T-1$. Чтобы получить ОММ-оценку, напишем это в виде

$$E\{Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})\} = 0. \tag{10.56}$$

Поскольку число «моментных» условий, как правило, будет превышать число неизвестных коэффициентов, мы оцениваем γ минимизацией квадратичного выражения в терминах соответствующих выборочных моментов (см. гл. 5), т.е.

$$\min\left[\frac{1}{N}\sum_{i=1}^{N} Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})\right]' W_N \left[\frac{1}{N}\sum_{i=1}^{N} Z_i'(\Delta y_i - \gamma\Delta y_{i,-1})\right], \tag{10.57}$$

где W_N — симметрическая положительно определенная матрица весов[21]. Дифференцируя это выражение по γ и решая полученное уравнение относительно γ, приходим к выражению

$$\begin{aligned}\gamma_{OMM} = \left(\left(\sum_{i=1}^{N} \Delta y_{i,-1}' Z_i'\right) W_N \left(\sum_{i=1}^{N} Z_i' \Delta y_{i,-1}'\right)\right)^{-1} \times \\ \times \left(\sum_{i=1}^{N} \Delta y_{i,-1}' Z_i\right) W_N \left(\sum_{i=1}^{N} Z_i' \Delta y_i'\right)\end{aligned} \tag{10.58}$$

Свойства этой оценки зависят от выбора матрицы весов W_N, несмотря на то, что хотя она будет состоятельной до тех пор, пока матрица W_N положительно определена, например, для матрицы $W_N = I$, где I — единичная матрица.

Оптимальной матрицей весов является такая матрица, которая приводит к эффективной оценке, т.е. дает наименьшую асимптотическую ковариационную матрицу для оценки γ_{OMM}. Из общей теории ОММ (см. гл. 5) мы знаем, что оптимальная матрица весов (асимптотически) пропорциональна матрице, обратной к ковариационной матрице выборочных моментов. В данном случае это означает, что оптимальная матрица весов должна удовлетворять

[20] Матрица Z_i в (10.54) имеет размерность $(T-1)\times(1+2+\cdots+(T-1))$, поскольку нули-строки, стоящие над и (или) под l-м выражением в квадратных скобках, имеют размерность l, $l = 1, 2, \ldots, T-1$ (*примеч. научн. ред. пер.*).

[21] Подстрочный индекс N отражает возможную зависимость матрицы W_N от объема выборки N, а не отражает размерность матрицы.

$$\underset{N\to\infty}{\text{plim}}\, W_N = V\{Z_i' \Delta\varepsilon_i\}^{-1} = E\{Z_i' \Delta\varepsilon_i \Delta\varepsilon_i' Z_i\}^{-1}. \tag{10.59}$$

В стандартном случае, когда на ковариационную матрицу ε_i никакие ограничения не налагаются, оптимальную матрицу весов можно оценить, используя на первом шаге состоятельную функцию оценивания γ, и заменяя оператор математического ожидания выборочным средним. Тогда оптимальная матрица весов имеет вид

$$W_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta\varepsilon_i \Delta\varepsilon_i' Z_i \right)^{-1} \tag{10.60}$$

где $\Delta\varepsilon_i$ — вектор оцененных на первом шаге остатков, например, при оценивании γ с использованием матрицы $W_N = I$.

В общем подходе ОММ не предполагается, что остатки ε_{it} являются независимо и одинаково распределенными по индивидуумам и времени, и, таким образом, оптимальная матрица весов тогда оценивается без наложения этих ограничений. Однако заметим, что отсутствие автокорреляции было необходимо, чтобы гарантировать выполнение «моментных» условий. Вместо оценивания оптимальной матрицы весов без ограничений также возможно (и потенциально желательно для малых выборок) наложить ограничение отсутствия автокорреляции в остатках ε_{it} одновременно с предположением их гомоскедастичности. Отметив, что при таких ограничениях

$$E\{\Delta\varepsilon_i \Delta\varepsilon_i'\} = \sigma_\varepsilon^2 G = \sigma_\varepsilon^2 \begin{pmatrix} 2 & -1 & 0 \cdots \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ \vdots & 0 & -1 & 2 \end{pmatrix}, \tag{10.61}$$

оптимальную матрицу весов можно определить как

$$W_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' G Z_i \right)^{-1}. \tag{10.62}$$

Заметим, что эта матрица не включает неизвестные параметры, так что оптимальную ОММ-оценку можно вычислить в рамках одного шага, если исходные остатки ε_{it}, как предполагается, являются гомоскедастичными и не обнаруживают никакой автокорреляции.

В общем, ОММ-оценка для неизвестного параметра γ асимптотически нормальна с ковариационной матрицей, заданной в виде

$$\text{plim}\left(\left(\frac{1}{N} \sum_{i=1}^{N} \Delta y_{i,-1}' Z_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta\varepsilon_i \Delta\varepsilon_i' Z_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta y_{i,-1} \right) \right)^{-1}. \tag{10.63}$$

Это следует из более общих выражений из параграфа 5.6. С независимо и одинаково распределенными остатками средний член в правой части выражения (10.63) сводится к

$$\sigma_\varepsilon^2 W_N^{opt} = \sigma_\varepsilon^2 \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' G Z_i \right)^{-1}.$$

10.4.2. Динамические модели с экзогенными переменными

Если модель к тому же содержит экзогенные переменные, то мы напишем модель в виде

$$y_{it} = x_{it}' \beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}. \tag{10.64}$$

Такую модель можно оценить также с помощью метода обобщенных инструментальных переменных или с помощью подхода ОММ. В зависимости от предположений, сделанных о переменных в векторе

x_{it}, можно построить разные совокупности дополнительных инструментальных переменных. Если переменные в векторе x_{it} *строго экзогенны* в том смысле, что они не коррелированны с любым из остатков ε_{is}, то мы также имеем, что

$$E\{x_{is}\Delta\varepsilon_{it}\}=0 \text{ для каждого } s,\ t, \tag{10.65}$$

так что к списку инструментальных переменных для уравнения первых разностей в каждый такт времени можно добавить $x_{i1},\ldots, x_{iT}$. Таким образом, число строк в матрице Z_i стало бы весьма большим. Вместо этого можно сохранить почти тот же самый уровень информации, если использовать первые разности переменных вектора x_{it} в качестве их собственных инструментальных переменных.[22] В этом случае мы налагаем «моментные» условия следующего типа

$$E\{\Delta x_{it}\Delta\varepsilon_{it}\}=0 \text{ для каждого } t. \tag{10.66}$$

Тогда матрица инструментальных переменных может быть записана в виде[23]

$$Z_i=\begin{pmatrix} [y_{i0},\Delta x'_{i2}] & 0 & \cdots & 0 \\ 0 & [y_{i0}, y_{i1},\Delta x'_{i3}] & & 0 \\ \vdots & & \ddots & 0 \\ 0 & & \cdots & [y_{i0},\ldots, y_{iT-2},\Delta x'_{iT}] \end{pmatrix} \tag{10.67}$$

Если переменные вектора x_{it} не строго экзогенны, а **предопределены**, что соответствует случаю, когда текущие и лагированные переменные в векторах x_{it} не коррелированы с текущими остатками, то мы имеем только, что $E\{x_{it}\varepsilon_{is}\}=0$ для $s\geq t$. Тогда действительными инструментальными переменными для уравнения первых разностей в период t являются только переменные $x_{i\,t-1},\ldots, x_{i1}$. Таким образом, соответствующие «моментные» условия будут иметь вид:

$$E\{x_{i,t-j}\Delta\varepsilon_{it}\}=0 \text{ для } j=1,\ldots, t-1 \text{ (для каждого } t). \tag{10.68}$$

На практике может возникнуть комбинация строго экзогенных и предопределенных x-переменных, а не один из этих двух крайних случаев. Тогда матрицу Z_i следует подкорректировать соответствующим образом. В монографии (Baltagi, 1995, Chapter 8) представлено дополнительное обсуждение и примеры.

В статье (Arellano, Bover, 1995) описывается структура объединения вышеупомянутого подхода с оцениванием методом инструментальных переменных Хаусмана, Тэйлора и др. (Hausman, Taylor, 1981, обсужденная в п. 5 параграфа 10.2). Наиболее важно, что авторы обсуждают, каким образом при оценивании можно также использовать информацию в уровнях[24]. Таким образом, в дополнение к представленным выше условиям моментов возможно также использование наличия обоснованных инструментальных переменных для уравнения уровней (10.64) или их среднего по времени (межгрупповая регрессия). Это имеет особое значение, когда коэффициент γ близок к единице (см. также статью (Blundell, Bond, 1998)).

10.4.3. Единичные корни и коинтеграция

Последняя литература показывает возрастающую интеграцию методов и идей анализа временных рядов с моделированием панельных данных, таких, например, как единичные корни и коинтеграционный анализ. Основная причина таких разработок заключается в том, что исследователи все более и бо-

[22] Мы отказываемся от потенциальной выгоды эффективности, если некоторые переменные в векторе x_{it} помогают «объяснению» лагированных эндогенных переменных.

[23] Матрица Z_i в (10.67) имеет размерность $(T-1)\times((K+1)+(K+2)+\cdots+(K+T-1))$, поскольку нули-строки, стоящие над и (или) под l-м выражением в квадратных скобках, имеют размерность $K+l$, $l=1, 2,\ldots, T-1$ (*примеч. научн. ред. пер.*).

[24] По-видимому, речь идет об информационном прошлом разного уровня глубины лагирования (*примеч. научн. ред. пер.*).

лее понимают, что пространственные данные являются полезным дополнительным источником информации, который следует использовать. Чтобы проанализировать эффект определенного политического решения, например принятия дорожного налога или налога на загрязнение окружающей среды, возможно, более полезно провести сравнение с другими странами, чем пробовать извлечь информацию об этих эффектах только из предыстории собственной страны. Объединение данных различных стран может также помочь преодолеть проблему довольно малых объемов выборок временны́х рядов, когда критерии анализа долгосрочных динамических свойств не являются достаточно мощными.

В ряде недавних статей обсуждаются проблемы единичных корней, ложных регрессий и коинтеграции в панельных данных. Следует подчеркнуть, что эти понятия являются *долгосрочными динамическими* понятиями и, как правило, приводят к проблемам вывода при T стремящемся к бесконечности. Во многих случаях, предполагая T фиксированным, а N стремящимся к бесконечности, такие проблемы обходят, по крайней мере, теоретически.

Критической проблемой при анализе временны́х рядов, зарегистрированных на некотором количестве выборочных единиц одновременно, является проблема гетерогенности этих единиц. До тех пор, пока мы рассматриваем каждый временно́й ряд (одномерный или многомерный) индивидуально и ряд имеет достаточную длину, нет никаких нарушений в применении методов временны́х рядов из гл. 8 и 9. Однако если мы объединяем ряды для разных выборочных единиц, мы должны отдавать себе отчет в том, что не все временны́е процессы имеют одни и те же свойства или описываются одними и теми же параметрами. Например, возможно, что временно́й ряд y_{it} является стационарным для страны 1, но интегрируемый порядка один для страны 2. Допуская, что все включенные переменные являются $I(1)$, предположим, что в каждой стране i переменные y_{it} и x_{it} являются коинтегрированными с параметром коинтеграции β_i. В таком случае линейная комбинация $y_{it}-\beta_i x_{it}$ является $I(0)$ для каждого i, но не существует общего параметра коинтеграции β, который приводит $y_{it}-\beta x_{it}$ к стационарности для всех i (если только параметры коинтеграции β_i не одинаковые для всех стран). Точно так же нет никакой гарантии, что пространственные средние, $\bar{y}_t=(1/N)\sum_i y_{it}$ и $\bar{x}_t$ являются коинтегрированными, даже если все лежащие в основе индивидуальные ряды коинтегрированные.

Чтобы проиллюстрировать некоторые из введенных проблем, рассмотрим модель авторегрессии

$$y_{it}=\alpha_i+\gamma_i y_{i,\,t-1}+\varepsilon_{it},$$

которую можно написать как

$$\Delta y_{it}=\alpha_i+\pi_i y_{i,\,t-1}+\varepsilon_{it},$$

где $\pi_i=\gamma_i-1$. Тогда нулевой гипотезой, что все временны́е ряды имеют единичный корень, является H_0: $\pi_i=0$ для всех i. Альтернативной гипотезой может быть гипотеза, что все ряды являются стационарными с одним и тем же параметром среднего возвращения, т.е. H_1: $\pi_i=\pi<0$ для всех i. В работах (Levin, Lin, 1992), (Quah, 1994) и (Harris, Tzavalis, 1999) альтернативная гипотеза подразумевается неявно. Менее ограниченная альтернативная гипотеза специфицируется в виде: H_1: $\pi_i<0$ для всех i, которая позволяет параметрам π_i различаться по группам и которая использовалась в работе (Im, Pesaran, Shin, 1997). Альтернативные критические статистики выводятся вместе с их асимптотическими распределениями, если $N\to\infty$ или $T\to\infty$, или одновременно $N\to\infty$ и $T\to\infty$, но обсуждение таких статистик выносится вне рамок этого текста. В любом случае центральная гипотеза состоит в том, что временны́е ряды *всех* индивидуальных выборочных единиц имеют единичный корень против альтернативной гипотезы, что *все* временны́е ряды являются стационарными. Поэтому можно было бы критиковать вышеупомянутые подходы, говоря, что возможно существование отличной от нуля вероятности, что один или более индивидуальных временны́х рядов являются стационарными, тогда как все другие имеют единичный корень или наоборот. В этом случае не удовлетворяется ни нулевая, ни альтернативная гипотеза, и неясно, желали бы мы отклонения нулевой гипотезы в результате нашего тестирования или нет. Другая техническая проблема заключается в возможности пространственной зависимости между остатками ε_{it} для разных стран, которая делает неправомерным использование совокупности упомянутых критериев.

В работах (Robertson, Symons, 1992) и (Pesaran, Smith, 1995) подчеркивалась важность параметрической гетерогенности в динамических моделях панельных данных, и анализировались потенциально серьезные смещения, которые могут возникать в результате обработки параметрически гетерогенных данных несоответствующим образом. Такие смещения особенно вводят в заблуждение в нестационарном мире, поскольку соотношения между индивидуальными временны́ми рядами могут полностью лишаться силы. Результаты по методам тестирования панельных данных на ложные регрессии и коинтеграцию относительно ограничены (см. (Kao, 1999) и (Phillips, Moon, 1999)).

10.5. Пример: эластичности спроса на труд по заработной плате

В этом разделе мы рассмотрим модель, которая объясняет спрос фирм на труд в зависимости от заработной платы, объема производства, лагированного спроса на труд и некоторых других переменных. Наша цель состоит в том, чтобы получить оценки для краткосрочных и долгосрочных динамических эластичностей спроса на труд по заработной плате в Бельгии. Данные и модели взяты из статьи (Konings, Roodhooft, 1997), в которой используются панельные данные более 3000 больших бельгийских фирм за период 1986–1994 гг. Статический спрос на труд задается моделью

$$\log L_{it} = \beta_1 + \beta_2 \log w_{it} + \beta_3 \log r_{it} + \beta_4 \log Y_{it} + \beta_5 \log w_{jt} + u_{it},$$

где L_{it} обозначает желаемую занятость на фирме i в период t (спрос на рабочую силу), w_{it} и r_{it} удельные издержки на труд и основные фонды соответственно, а Y_{it} обозначает уровень объема производства. Последняя переменная w_{jt} обозначает среднее реальной заработной платы в промышленности. Это соотношение интерпретируется как долгосрочный динамический результат, поскольку оно игнорирует издержки «настройки» (регулирования) модели.

Для краткосрочной динамики авторы статьи (Konings, Roodhooft, 1997) экспериментировали с альтернативными динамическими спецификациями. Самая простая спецификация предполагает, что

$$\log L_{it} = \beta_1 + \beta_2 \log w_{it} + \beta_3 \log r_{it} + \beta_4 \log Y_{it} + \beta_5 \log w_{jt} + \gamma \log L_{i,\,t-1} + u_{it}.$$

При оценивании величина r_{it} аппроксимировалась акционерным капиталом K_{it}, а Y_{it} добавленной стоимостью. Тогда динамическая модель, которую мы оцениваем, имеет вид

$$\log L_{it} = \beta_1 + \beta_2 \log w_{it} + \beta_3 \log K_{it} + \beta_4 \log Y_{it} + \beta_5 \log w_{jt} + \gamma \log L_{i,\,t-1} + \alpha_i + \varepsilon_{it},$$

где предполагается, что остатки состоят из двух компонент. Компонента α_i обозначает ненаблюдаемую гетерогенность фирм, специфицированную не зависящей от времени. Первое взятие разности в этом уравнении, как и в предыдущем параграфе, исключает компоненту α_i, но не приводит к уравнению, которое можно оценить состоятельно с помощью МНК. Во-первых, разность $\Delta \log L_{i,t-1}$ и разность $\Delta \varepsilon_{it}$ коррелированы (как и выше). Во вторых, ни в коем случае не очевидно, что факторные издержки заданы экзогенно. В частности, для удельных издержек на труд w_{it} можно представить несколько альтернативных ситуаций, в которых заработная плата определяется одновременно с занятостью. Например, профсоюзы могут заключить сделку с предпринимателями по заработной плате и занятости. Таким образом, мы можем ожидать, что

$$E\{\Delta \log w_{it}\, \Delta \varepsilon_{it}\} \neq 0.$$

Поэтому логарифмическая разность $\Delta \log w_{it}$ также инструментована при оценивании. Правомочные инструментальные переменные задаются переменными $\log w_{i,t-2}$, $\log w_{i,t-3}, \ldots$, подобными инструментальным переменным для логарифмической разности $\Delta \log L_{i,t-1}$. Таким образом, число доступных инструментальных переменных возрастает вместе с ростом t.

В табл. 10.3 мы представили результаты оценивания для статической и динамической моделей, обсужденных выше. Эти результаты являются подмножеством результатов, представленных в работе (Konings, Roodhooft, 1997), в которой также рассматривались модели с другими дополнительными лагированными переменными. В первом столбце приведены оценки для статической (т. е. долгосрочной динамической) функции спроса на труд. Заработная плата рассматривается как эндогенная и инструментована, как указано выше. Во второй столбец включен лагированный спрос на труд, который тоже инструментован, как описано выше. Обе спецификации также включают региональные и временны́е фиктивные переменные (манекены). Чтобы протестировать модель против неспецифицированной альтернативной гипотезы, мы можем использовать тесты на сверхидентифицирующие ограничения, как обсуждалось в гл. 5. Критические статистики, равные 29,7 и 51,66, должны сравниваться с критическими значениями из хи-квадрат распределения с 15 и 29 степенями свободы соответственно. С p-значениями, равными 0,013 и 0,006, сверхидентифицирующие ограничения на 1%-м уровне значимости отвергаются — на границе отклонения для обеих спецификаций. Значимость лагированной зависимой переменной (стандартные ошибки даны в круглых скобках) предполагает, что следует предпочесть динамическую спецификацию.

Таблица 10.3

Результаты оценивания уравнения спроса на труд
(Konings, Roodhooft, 1997)

Зависимая переменная $\log L$		
Переменные	Статическая модель	Динамическая модель
$\log L_{i,t-1}$	—	0,60 (0,045)
$\log Y_{it}$	0,021 (0,009)	0,008 (0,005)
$\log w_{it}$	–1,78 (0,60)	–0,66 (0,19)
$\log w_{jt}$	0,16 (0,07)	0,054 (0,33)
$\log K_{it}$	0,08 (0,011)	0,078 (0,006)
тест на сверх-идентифицирующие ограничения	29,7 ($p = 0{,}013$) ($df = 15$)	51,66 ($p = 0{,}006$) ($df = 29$)
число наблюдений	10599	10599

Оцененная краткосрочная динамическая эластичность по заработной плате из последнего столбца равна –0,66, в то время как долгосрочная динамическая эластичность равна –0,66/(1–0,60) = –1,6, которая близка к оценке, равной –1,78, из статической долгосрочной динамической модели. Обе эти оценки весьма высокие. Например, они предполагают, что в долгосрочной динамике увеличение заработной платы на 1% приводит к снижению на 1,6% в спросе на труд. Эти оценки намного выше, чем представлялось вначале на основе макроэкономических данных временны́х рядов. Очевидно, что возможность корректировать гетерогенность для наблюдаемых и ненаблюдаемых фирм имеет существенное влияние на оценки. Потенциальная проблема результатов в табл. 10.3 лежит в направлении структурного логического построения данных.

Во-первых, панельные данные несбалансированы (см. разд. 10.7 ниже), в то время как модель игнорирует изменения в спросе на труд, обусловленные включением или невключением фирм в выборку (например, из-за финансовых затруднений). Кроме того, занятость измеряется средним числом занятых в данном году, в то время как заработная плата (удельные издержки на труд) вычисляется в виде общих трудовых издержек, деленных на число занятых. Ясно, что тем самым игнорируется проблема сокращения среднего трудового времени рабочего, которая, возможно, в это десятилетие имела место. Например, если фирма заменяет одного рабочего, занятого полный рабочий день, двумя рабочими, занятыми неполный трудовой день, то занятость возрастает, а трудовые издержки снижаются, в то время как в действительности никаких реальных изменений не происходило. Более подробное обсуждение проблемы см. в (Konings, Roodhooft, 1997).

10.6. Модели с ограниченными зависимыми переменными

Панельные данные относительно часто используются в микроэкономических проблемах, где интересующие нас модели включают нелинейность. Дискретные или ограниченные зависимые переменные являются важным феноменом в этой области, а их комбинация с панельными данными обычно усложняет оценивание. Причина заключается в том, что для панельных данных обычно нельзя аргументировать, что различные наблюдения относительно одной и той же выборочной единицы независимы. Корреляции между различными членами ошибок, как правило, усложняют функции правдоподобия таких моделей и, следовательно, усложняют их оценивание. В этом разделе мы обсудим оценивание логит-модели, пробит-модели и тобит-модели панельных данных. Больше деталей относительно моделей панельных данных с ограниченными зависимыми переменными можно найти в статье (Maddala, 1987).

10.6.1. Модели бинарного выбора

Как и в случае пространственных данных, модель бинарного выбора обычно формулируется в терминах лежащей в основе латентной модели. Как правило, мы пишем[25]

$$C^*_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{10.69}$$

где мы наблюдаем, что $y_{it} = 1$ если $C^*_{it} > 0$, и $y_{it} = 0$ в противном случае. Например, y_{it} может служить показателем, работает или нет индивидуум *i* в период *t*. Предположим, что специфические остатки ε_{it} имеют симметричное распределение с функцией распределения $F(.)$, независимо и одинаково распределенные по индивидуумам и времени и независимые от всех x_{is}. Даже при таких допущениях присутствие эффектов α_i усложняет оценивание в обоих случаях: и когда мы рассматриваем их в качестве неизвестных фиксированных параметров, и когда мы рассматриваем их в качестве случайных остатков.

Если мы рассматриваем α_i как фиксированные неизвестные параметры, то по существу мы включаем в модель *N* фиктивных переменных. Таким образом, функция логарифма правдоподобия задается (сравните с функцией логарифма правдоподобия (7.12)) как

$$\begin{aligned}\log L(\beta, \alpha_1, \ldots, \alpha_N) = \sum_{i,j} y_{it} \log F(\alpha_i + x'_{it}\beta) + \\ + \sum_{i,t} (1 - y_{it}) \log[1 - F(\alpha_i + x'_{it}\beta)].\end{aligned} \tag{10.70}$$

Максимизация этой функции относительно β и α_i $(i = 1, \ldots, N)$ приводит к состоятельным оценкам *при условии, что число тактов времени T стремится к бесконечности*. Для фиксированного *T* и

[25] Для упрощения обозначений мы предположим, что x_{it} включает константу, всякий раз, когда это уместно.

$N \to \infty$, оценки несостоятельны. Причина заключается в том, что для фиксированного T число параметров возрастает с ростом объема выборки N и мы имеем проблему, которая известна как проблема «побочных (incidental) параметров». Т.е. любой эффект α_i можно оценить состоятельно, если только мы имеем возрастающее число наблюдений для каждого фиксированного i, следовательно, если T стремится к бесконечности. В общем, несостоятельность α_i для фиксированного T переносится на функцию оценивания вектора неизвестных параметров β.

Проблема побочных параметров, когда число параметров растет с числом наблюдений, возникает в любой модели с фиксированными эффектами, включая линейную модель. Однако для линейного случая можно исключить параметры α_i, так что вектор параметров β можно оценить состоятельно, даже при том, что все параметры α_i оценить нельзя. Тем не менее, для большинства нелинейных моделей, несостоятельность α_i также приводит к несостоятельности оценок для других параметров. К тому же заметим, что с практической точки зрения оценивание более, чем N параметров, по-видимому, не очень привлекательно, если N является довольно большим.

Несмотря на то, что *латентную* модель можно преобразовать таким образом, что индивидуальные эффекты α_i исключаются, в данном контексте это не помогает, поскольку нет никакого отображения, например, $y^*_{it} - y^*_{i,t-1}$ в наблюдаемые переменные, подобные переменным $y_{it} - y_{i,t-1}$. Альтернативная стратегия состоит в применении **условного максимального правдоподобия** (см. (Andersen, 1970), (Chamberlain, 1980)). В этом случае мы рассматриваем функцию правдоподобия условную по множеству статистик t_i, которые являются достаточными для параметров α_i. Это означает, что условный по t_i вклад в правдоподобие индивидуума больше не зависит от параметра α_i, но все еще зависит от других параметров β. В модели бинарного выбора для панельных данных существование достаточной статистики зависит от функционального вида функции распределения F, т.е., зависит от распределения специфических остатков ε_{it}.

На общем уровне напишем совместную плотность распределения случайных величин $y_{i1},..., y_{iT}$ как $f(y_{i1},..., y_{iT} \mid \alpha_i, \beta)$, которая зависит от вектора параметров β и параметра α_i. Если существует достаточная статистика t_i, то это значит, что существует статистика t_i такая, что $f(y_{i1},..., y_{iT} \mid t_i, \alpha_i, \beta) = f(y_{i1},..., y_{iT} \mid t_i, \beta)$[26], и поэтому она не зависит от параметра α_i. Следовательно, мы можем максимизировать **функцию условного правдоподобия**, основанную на $f(y_{i1},..., y_{iT} \mid t_i, \beta)$, и получить состоятельную оценку для вектора параметров β. Кроме того, мы можем использовать все относящиеся к законам распределения результаты гл. 6, только заменить в них логарифмическую функцию правдоподобия условной логарифмической функцией правдоподобия. Для линейной модели с нормальными ошибками достаточная статистика для параметра α_i есть $\bar{y}_i$. Таким образом, условное распределение y_{it} при заданном $\bar{y}_i$ не зависит от параметра α_i, и можно показать, что максимизация функции условного правдоподобия воспроизводит оценку с фиксированными эффектами для вектора неизвестных параметров β. К сожалению, этот результат автоматически не распространяется на нелинейные модели. Например, для пробит-модели было показано, что никакая достаточная статистика для параметра α_i не существует. Это означает, что мы не можем оценить пробит-модель фиксированных эффектов состоятельно для фиксированного T.

10.6.2. Логит-модель с фиксированными эффектами

Для логит-модели с фиксированными эффектами, ситуация отличается. В этой модели $t_i = \bar{y}_i$ является достаточной статистикой для параметра α_i, и состоятельное оценивание возможно в соответствии с условным максимальным правдоподобием. Следует отметить, что условное распределение $y_{i1},..., y_{iT}$ является вырожденным, если $t_i = 0$ или $t_i = 1$. Следовательно, такие индивидуумы не

[26] Математически несколько небрежная формулировка. Точнее: если f $(y_{i1},..., y_{iT}; \alpha_i, \beta)$ — совместная плотность распределения случайных величин $y_{i1},..., y_{iT}$, зависящая от параметров α_i и β, и если t_i — достаточная статистика в оценивании параметра α_i, то условная совместная плотность f $(y_{i1},... y_{iT}; \beta| t_i)$ (при условии заданного фиксированного значения t_i) не будет зависеть от α_i (*примеч. научн. ред. перевода*).

вносят свой вклад в условное правдоподобие, и надо отказаться от них при оценивании. Выражаясь иначе, их поведение полностью улавливалось бы их индивидуальным эффектом α_i. Это значит, что только индивидуумы, которые изменяют свой статус, по крайней мере, хотя бы один раз, уместны для оценивания вектора параметров β. Чтобы проиллюстрировать логит-модель с фиксированными эффектами, мы рассмотрим случай $T = 2$.

Существует два возможных исхода, условных по $t_i = 1/2$: (0, 1) и (1, 0). Условная вероятность первого исхода равна

$$P\{(0,1) \mid t_i = 1/2,\ \alpha_i, \beta\} = \frac{P\{(0,1) \mid \alpha_i, \beta\}}{P\{(0,1) \mid \alpha_i, \beta\} + P\{(1,0) \mid \alpha_i, \beta\}}. \tag{10.71}$$

Используем, что

$$P\{(0,1) \mid \alpha_i, \beta\} = P\{y_{i1} = 0 \mid \alpha_i, \beta\} P\{y_{i2} = 1 \mid \alpha_i, \beta\}$$

при[27]

$$P\{y_{i2} = 1 \mid \alpha_i, \beta\} = \frac{\exp\{\alpha_i + x'_{i2}\beta\}}{1 + \exp\{\alpha_i + x'_{i2}\beta\}}.$$

Откуда следует, что условная вероятность задается в виде

$$P\{(0,1) \mid t_i = 1/2,\ \alpha_i, \beta\} = \frac{\exp\{(x_{i2} - x_{i1})'\beta\}}{1 + \exp\{(x_{i2} - x_{i1})'\beta\}}, \tag{10.72}$$

которая действительно не зависит от α_i. Аналогично,

$$P\{(1,0) \mid t_i = 1/2,\ \alpha_i, \beta\} = \frac{1}{1 + \exp\{(x_{i2} - x_{i1})'\beta\}}. \tag{10.73}$$

Это означает, что мы можем оценить логит-модель с фиксированными эффектами для $T = 2$, используя стандартную логит-модель с $x_{i2} - x_{i1}$ в качестве объясняющих переменных и изменение в переменной y_{it} в качестве эндогенного события (с 1 для положительного изменения и с 0 для отрицательного изменения). Отметим, что в этой модели бинарного выбора с фиксированными эффектами еще более ясно, чем в линейном случае, что модель идентифицируется только через «внутригрупповую размерность» данных; от индивидуумов, которые не изменяют статус, при оценивании просто отказываются, поскольку они вообще не обеспечивают никакой информации о векторе параметров β. Для случая с большими значениями T все необходимые условные вероятности получить несколько сложнее, но в принципе они являются прямым обобщением вышеприведенного случая (см. (Chamberlain, 1980) или (Maddala, 1987)). В статье (Chamberlain, 1980) также обсуждается, как подход условного максимального правдоподобия можно обобщить на мультиномиальную логит-модель.

Если можно предположить, что эффекты α_i независимы от объясняющих переменных в векторе x_{it}, то, по-видимому, схема случайных эффектов более уместна. Она оказывается более легко реализуемой в контексте пробит-модели.

10.6.3. Пробит-модель со случайными эффектами

Начнем со спецификации с латентной переменной

$$y^*_{it} = x'_{it}\beta + u_{it}, \tag{10.74}$$

$$y_{it} = 1, \text{ если } y^*_{it} > 0$$
$$y_{it} = 0, \text{ если } y^*_{it} \le 0, \tag{10.75}$$

[27] См. выражение (7.6) в гл. 7 для логистической функции распределения.

где u_{it} — остатки с нулевым средним и дисперсией, равной единице, независимые от $(x_{i1},..., x_{iT})$. Чтобы оценить вектор параметров β методом максимального правдоподобия, мы должны сделать дополнительное предположение о совместном распределении остатков ошибок $u_{i1},..., u_{iT}$. Вклад в правдоподобие индивидуума i есть (совместная) вероятность наблюдения T исходов $y_{i1},..., y_{iT}$.Эта совместная вероятность определяется из совместного распределения латентных переменных $y^*_{i1},..., y^*_{iT}$ интегрированием по соответствующим интервалам. Таким образом, в общем, будет T интегралов, которые для оценивания следует, как правило, вычислять численно. Когда $T = 4$ или более, то оценивание методом максимального правдоподобия осуществимо. Можно обойти эту «напасть размерности» применением функций оценивания на основе моделирования, например, как описано в справочнике (Keane,1993) и статье (Weeks, 1995). Такое обсуждение выходит за рамки этого текста.

Ясно, если можно предполагать, что все u_{it} независимы, то мы имеем, что совместная условная плотность вероятностей $f(y_{i1},..., y_{iT} \mid x_{i1},..., x_{iT}, \beta) = \prod_t f(y_{it} \mid x_{it}, \beta)$, которая включает только T одномерных интегралов (как и в случае пространственных данных). Если мы делаем предположение о компонентах ошибок и предполагаем, что $u_{it} = \alpha_i + \varepsilon_{it}$,где ε_{it} не зависит от времени (и индивидуумов), то совместную условную вероятность можно написать как

$$\begin{aligned} & f(y_{i1},..., y_{iT} \mid x_{i1},..., x_{iT}, \beta) = \\ & = \int_{-\infty}^{\infty} f(y_{i1},..., y_{iT} \mid x_{i1},..., x_{iT}, \alpha_i, \beta) f(\alpha_i)\, d\alpha_i = \\ & = \int_{-\infty}^{\infty} \left[\prod_t f(y_{it} \mid x_{it}, \alpha_i, \beta)\right] f(\alpha_i)\, d\alpha_i, \end{aligned} \tag{10.76}$$

которая требует одномерного численного интегрирования. Это практически реализуемая спецификация, которая допускает коррелированность остатков для разных тактов времени, хотя и ограниченным образом. Критический момент в реализации выражения (10.76) состоит в том, что условные по α_i остатки были бы независимыми при разных t.

В принципе о распределениях α_i и ε_{it} можно делать произвольные предположения. Например, можно было бы предположить, что остатки ε_{it} независимо и одинаково распределены по нормальному закону, в то время как α_i имеет логистическое распределение. Однако это может привести к распределениям для остатков $\alpha_i + \varepsilon_{it}$, которые являются нестандартными. Например, сумма двух логистически распределенных переменных, в общем, не имеет логистического распределения. Это подразумевает, что индивидуальные вероятности, подобные $f(y_{it} \mid x_{it}, \beta)$, трудны для вычисления, и не соответствуют пространственной пробит-модели или пространственной логит-модели. Поэтому обычно следует начинать с совместного распределения остатков $u_{i1},..., u_{iT}$. Многомерное логистическое распределение имеет такое неудобство, что все корреляции должны ограничиваться 1/2 (Maddala, 1987), так что на практике это не очень привлекательно. Следовательно, самый общий подход состоит в том, чтобы начать с многомерного нормального распределения, которое приводит к **пробит-модели случайных эффектов**.

Предположим, что совместное распределение $u_{i1},..., u_{iT}$ нормально с нулевыми средними значениями, дисперсиями, равными 1, и $\operatorname{cov}\{u_{it}, u_{is}\} = \sigma^2_\alpha$, $s \neq t$. Это соответствует предположению, что α_i является НОНП $(0, \sigma^2_\alpha)$, а ε_{it} есть НОНП $(0, 1-\sigma^2_\alpha)$. Вспомним, что, как и в случае пространственных данных, нам требуется нормировка дисперсий остатков. Выбранная здесь нормировка подразумевает, что дисперсия остатка в заданный такт времени равна единице, так что оцененные коэффициенты β непосредственно сравнимы с оценками, полученными из оценивания модели по данным одного такта времени (выбранного из общих панельных данных), используя пространственное пробит максимальное правдоподобие. Для пробит-модели со случайными эффектами выражения в функции правдоподобия задаются в виде

$$f(y_{it} \mid x_{it}, \alpha_i, \beta) = \begin{cases} \Phi\left(\dfrac{x'_{it}\beta + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если } y_{it} = 1 \\ 1 - \Phi\left(\dfrac{x'_{it}\beta + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если } y_{it} = 0, \end{cases} \tag{10.77}$$

где Φ обозначает функцию распределения стандартного нормального закона. Плотность α_i задается в виде

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left\{-\frac{1}{2}\frac{\alpha_i^2}{\sigma_\alpha^2}\right\}. \tag{10.78}$$

Интеграл в выражении (10.76) следует вычислять численно, что можно сделать, используя алгоритм, описанный в статье (Butler, Moffitt 1982). Несколько пакетов программ (например, LIMDEP и Stata) имеют стандартные подпрограммы для того, чтобы оценивать пробит-модель случайных эффектов.

Можно показать (Robinson, 1982), что игнорирование взаимных корреляций остатков по времени и оценивание коэффициентов β, использующее стандартное максимальное правдоподобие пробит-модели для объединенных данных, является состоятельным, хотя и неэффективным. Кроме того, обычно вычисляемые стандартные ошибки некорректны. Однако их значения можно использовать в качестве начальных оценок в итерационной процедуре максимального правдоподобия, основанной на совместной вероятности (10.76).

10.6.4. Тобит-модели

Тобит-модель случайных эффектов очень похожа на пробит-модель случайных эффектов, единственное различие состоит в правиле наблюдения. Следовательно, мы можем быть довольно краткими. Начнем с модели

$$y_{it}^* = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{10.79}$$

вместе с тем, что

$$\begin{aligned} y_{it} &= y_{it}^*, \quad \text{если } y_{it}^* > 0 \\ y_{it} &= 0, \quad \text{если } y_{it}^* \le 0. \end{aligned} \tag{10.80}$$

Мы сделаем обычное предположение случайных эффектов, что эффекты α_i и остатки ε_{it} являются независимо и одинаково распределенными по нормальному закону с нулевыми средними значениями и дисперсиями, равными σ_α^2 и σ_ε^2 соответственно, и независимыми от $x_{i1},\dots, x_{iT}$. Используя f как общее обозначение для функции плотности или функции вероятностной меры, функцию правдоподобия можно написать, как выражение (10.76), в виде

$$f(y_{i1},\dots, y_{iT} \mid x_{i1},\dots, x_{iT}, \beta) =$$
$$= \int_{-\infty}^{\infty} \prod_t f(y_{it} \mid x_{it}, \alpha_i, \beta) f(\alpha_i)\, d\alpha_i,$$

где $f(\alpha_i)$ задается выражением (10.78), а $f(y_{it} \mid x_{it}, \alpha_i, \beta)$ задается выражением

$$f(y_{it} \mid x_{it}, \alpha_i, \beta) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\dfrac{1}{2}\dfrac{(y_{it} - x'_{it}\beta - \alpha_i)^2}{\sigma_\varepsilon^2}\right\}, & \text{если } y_{it} > 0 \\ 1 - \Phi\left(\dfrac{x'_{it}\beta + \alpha_i}{\sigma_\varepsilon}\right), & \text{если } y_{it} = 0. \end{cases} \tag{10.81}$$

Заметим, что последние два выражения аналогичны вкладам в функцию правдоподобия в случае пространственных данных, как обсуждалось в гл. 7. Единственное различие состоит во включении эффекта α_i в условном смысле.

Полностью подобным же образом можно рассмотреть другие формы цензурирования, чтобы получить, например, пробит-модель с упорядоченными случайными эффектами. Во всех случаях интегрирование по α_i должно проводиться численно.

Тобит-модель можно оценить состоятельно, так же как и усеченную модель регрессии с фиксированными эффектами, применяя обобщенный метод моментов с использованием моментных условий, представленный в статье (Honore, 1992) или статье (Honore, 1993) для динамической модели. Эти функции оценивания являются полупараметрическими в том смысле, что на вид функции распределения остатков ε_{it} никакие предположения не налагаются.

10.6.5. Динамика и проблема начальных условий

Возможность включения лагированной зависимой переменной в вышеупомянутые модели представляет экономический интерес. Например, предположим, что мы объясняем, действительно ли индивидуум является безработным (или нет) за ряд последующих месяцев. Как правило, справедливо, что индивидуумы, которые имеют более длинную предысторию находиться в состоянии безработного, менее вероятно оставят состояние безработицы. Как обсуждалось в вводном разделе этой гл. существует два объяснения этого: индивидуум с более длинной предысторией безработного может быть обескуражен в своих поисках работы, или для работодателя, возможно (по любой причине), менее привлекательно нанять его на работу. Это называется **зависимостью от статуса**: чем дольше вы находитесь в определенном состоянии, тем менее вероятно, что вы его оставите. Альтернативно, возможно, что присутствует **ненаблюдаемая гетерогенность,** такая, что индивидуумы с определенными ненаблюдаемыми характеристиками менее вероятно оставят статус безработного. То есть факт, что мы наблюдаем мнимую зависимость от статуса в данных, происходит просто из-за механизма выбора: безработные с продолжительными периодами имеют определенные ненаблюдаемые (не зависящие от времени) особенности, которые делают менее вероятным найти ему работу каким-либо образом. В обсужденных выше моделях бинарного выбора индивидуальные эффекты α_i улавливают ненаблюдаемую гетерогенность. Если мы включаем лагированную зависимую переменную, то мы можем различить вышеупомянутые два объяснения.

Рассмотрим пробит-модель со случайными эффектами, хотя подобные результаты справедливы и для случая тобит-модели со случайными эффектами. Предположим, что спецификация латентной переменной изменена на

$$y_{it}^* = x_{it}'\beta + \gamma y_{i\,t-1} + \alpha_i + \varepsilon_{it}, \qquad (10.82)$$

где $y_{it} = 1$, если $y_{it}^* > 0$,и $y_{it} = 0$ в противном случае. В этой модели $\gamma > 0$ указывает на положительную зависимость от статуса: при прочих равных условиях вероятность, что $y_{it} = 1$, больше, если $y_{i\,t-1}$ также равна единице. Рассмотрим оценивание методом максимального правдоподобия такой динамической пробит-модели со случайными эффектами, сделав те же самые предположения о распределениях, как прежде. В общем виде вклад правдоподобия индивидуума *i* имеет вид[28]

$$\begin{aligned} &f(y_{i1},\ldots, y_{iT} \mid x_{i1},\ldots,x_{iT},\beta) = \\ &= \int_{-\infty}^{\infty} f(y_{i1},\ldots, y_{iT} \mid x_{i1},\ldots,x_{iT}, \alpha_i, \beta) f(\alpha_i)\, d\alpha_i = \qquad (10.83)\\ &= \int_{-\infty}^{\infty} \left[\prod_{t=2}^{T} f(y_{it} \mid y_{i,\,t-1}, \alpha_i, \beta)\right] f(y_{i1} \mid x_{i1}, \alpha_i, \beta)\, f(\alpha_i)\, d\alpha_i, \end{aligned}$$

[28] Для удобства обозначений индекс времени определяется так, что первое наблюдение есть (y_{i1}, x_{i1}').

где

$$f(y_{it} | y_{i,t-1}, \alpha_i, \beta) = \begin{cases} \Phi\left(\dfrac{x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если } y_{it} = 1 \\ 1 - \Phi\left(\dfrac{x'_{it}\beta + \gamma y_{i,t-1} + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если } y_{it} = 0. \end{cases}$$

Это полностью аналогично стационарному случаю, а переменная $y_{i,t-1}$ просто включена как дополнительная объясняющая переменная. Однако член $f(y_{i1} | x_{i1}, \alpha_i, \beta)$ в функции правдоподобия может вызвать проблемы. Он дает вероятность наблюдения $y_{i1} = 1$ или $y_{i1} = 0$, без знания предыдущего состояния, но условную по гетерогенности ненаблюдаемого члена α_i.

Если начальное значение экзогенно в том смысле, что его распределение не зависит от α_i, то вне интеграла мы можем положить член $f(y_{i1} | x_{i1}, \alpha_i, \beta) = f(y_{i1} | x_{i1}, \beta)$. В этом случае мы можем просто рассмотреть функцию правдоподобия условную по y_{i1} и игнорировать член $f(y_{i1} | x_{i1}, \beta)$ при оценивании. Единственным последствием может быть потеря эффективности, если $f(y_{i1} | x_{i1}, \beta)$ обеспечивает информацию о векторе параметров β. Этот подход был бы уместен, если необходимым условием для всех индивидуумов было бы одинаковое начальное состояние, или если бы для индивидуумов оно назначалось случайно. Пример первой ситуации приведен в статье (Nijman, Verbeek, 1992), где моделируется «неотклик» относительно потребления, и начальный такт времени соответствует месяцу перед панельными данными, и «неотклик» необязательно наблюдался.

Однако во многих приложениях, возможно, трудно аргументировать, что начальное значение y_{i1} экзогенно и не зависит от ненаблюдаемой гетерогенности индивидуума. В таком случае мы нуждались бы в выражении для $f(y_{i1} | x_{i1}, \alpha_i, \beta)$, а это проблематично. Если процесс, который мы оцениваем, продолжался в течение многих тактов времени перед текущим выборочным тактом времени, то $f(y_{i1} | x_{i1}, \alpha_i, \beta)$ является сложной функцией, которая зависит от ненаблюдаемой предыстории индивидуальных *i*. Это означает, что, как правило, невозможно получить выражение для маргинальной вероятности $f(y_{i1} | x_{i1}, \alpha_i, \beta)$, которая является непротиворечивой с остатком модели. В работе (Heckman, 1981) предлагается приближенное решение этой **проблемы начальных условий**, которое, по-видимому, на практике работает достаточно хорошо. Это решение основано на аппроксимации маргинальной вероятности начального состояния пробит-функцией, использующей насколько возможно больше доступной предвыборочной информации, без наложения ограничений на ее коэффициенты и структурные параметры β и γ. В статье (Vella, Verbeek, 1999) приведен пример такого подхода к динамической тобит-модели со случайными эффектами. Влияние начальных условий снижается при возрастании числа выборочных тактов времени *T*, поэтому если *T* является довольно большим, то проблему можно игнорировать.

10.7. Неполные панельные данные и смещение, обусловленное выборочной селективностью

По разнообразным причинам эмпирические совокупности панельных данных часто неполные. Например, после того как прошло несколько тактов времени индивидуумы, включенные в обследование, предполагающее сбор панельных данных, могут отказаться от сотрудничества; для одних домашних хозяйств невозможно определить их новое местонахождения, другие распались; фирмы могут завершить свой бизнес или слиться с другими фирмами; а инвестиционные фонды могут закрыться. С другой стороны, фирмы могут войти в бизнес на более поздней стадии; можно извлечь обновленную выборочную информацию, компенсирующую потерянную; или панельные данные могут собираться в виде панели ротации. В панели ротации на каждом такте времени определенная доля выборочных единиц заменяется другими выборочными единицами. Последствие всех таких событий состоит в том, что получающаяся совокупность панельных данных больше не является «прямоугольной».

Несмотря на то, что общее количество индивидуумов равно N, а число тактов времени равно T, общее количество наблюдений оказывается существенно меньшим, чем NT.

Первая особенность, присущая работе с неполными панельными данными — это возникновение вычислительных проблем. Если наблюдения отсутствуют, то большинство выражений для представленных выше оценок больше просто неприемлемо. Простое «решение» состоит в том, чтобы любого индивидуума, по которому мы имеем неполную информацию, исключить из панельных данных, и работать только с полностью наблюдаемыми выборочными единицами. При таком подходе для оценивания используются только «**сбалансированная субпанель**». В вычислительном отношении этот подход заманчив, но потенциально очень неэффективен: можно «потерять» существенное количество информации. Потери в эффективности можно устранить, учитывая все имеющиеся наблюдения, включая тех индивидуумов, которые наблюдались в T периодах времени только частично. В этом случае используется «**несбалансированная субпанель**». В принципе применение несбалансированных панельных данных является прямой процедурой, но в вычислительном отношении требуются определенные корректировки формул, представленных в предыдущих разделах. Мы обсудим некоторые из корректировок в пункте 1 параграфа 10.7. К счастью, большая часть программного обеспечения, которое может обрабатывать панельные данные, также учитывает несбалансированные данные.

Другое потенциальное и еще более серьезное последствие применения неполных панельных данных заключается в опасности **смещения**, обусловленного ограничениями в отборе выборочных единиц, т.е. так называемой выборочной селективностью («selection bias»). Если индивидуумы наблюдаются неполностью из-за эндогенной причины, использование сбалансированной субпанели или несбалансированных панельных данных может привести к смещенным оценкам и вводящим в заблуждение критериям. Для пояснения этого предположим, что интересующая нас модель задается в виде

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}. \tag{10.84}$$

Кроме того, определим индикаторную переменную r_{it} («отклик») в виде: $r_{it} = 1$, если (x_{it}, y_{it}) наблюдалось, и $r_{it} = 0$ в противном случае. Наблюдения (x_{it}, y_{it}) **отсутствуют (пропущены) случайным образом**, если r_{it} не зависит от эффекта α_i и остатка ε_{it}. Это означает, что заданные условия процесса отбора выборочных единиц не влияют на условное распределение y_{it} для данного x_{it}. Если мы хотим сконцентрироваться на сбалансированной субпанели, то условия процесса отбора определяются соотношениями $r_{i1} = \cdots = r_{iT} = 1$, и мы требуем, чтобы индикатор r_{it} был независим от эффекта α_i и остатков $\varepsilon_{i1}, \ldots, \varepsilon_{iT}$. В этих случаях обычные свойства состоятельности функций оценивания не изменяются, если мы ограничиваем внимание только доступными или полными наблюдениями. Если процесс отбора зависит от МНК-оцененных остатков уравнения, то оценки случайных и фиксированных эффектов могут пострадать от «выборочного» смещения (см. гл. 7). В п. 10.7.2 описываются подробности по этой проблеме, включая некоторые простые тесты. В случаях с выборочным смещением, следует использовать альтернативные оценки, которые в вычислительном отношении являются, как правило, непривлекательными, что обсуждается в п. 10.7.3. Дополнительные детали и обсуждение методов анализа, основанного на неполных панельных данных, и вопросов «селективного» смещения можно найти в статьях (Verbeek, Nijman, 1992, 1996).

10.7.1. Оценивание со случайно пропущенными данными

Выражения для оценок с фиксированными и случайными эффектами легко обобщить на несбалансированный случай. Оценку с фиксированными эффектами, как и прежде, можно определить как МНК-оценку в линейной модели, где для каждого i (номера индивидуума) определен свой свободный член. Альтернативно, оценку вектора неизвестных параметров β можно получить непосредственным применением МНК к внутригрупповой преобразованной модели, где теперь все переменные являются отклонениями от среднего значения *по доступным наблюдениям*. Индивидуумы, которые наблюдаются только один раз, не обеспечивают никакой информации относительно β и их следует исключить из процесса оценивания. Определив «доступные средние значения» как[29]

$$\bar{y}_i = \frac{\sum_{t=1}^{T} r_{it} y_{it}}{\sum_{t=1}^{T} r_{it}}; \qquad \bar{x}_i = \frac{\sum_{t=1}^{T} r_{it} x_{it}}{\sum_{t=1}^{T} r_{it}},$$

функцию оценивания фиксированных эффектов можно кратко написать в виде

$$\beta_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} r_{it} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} r_{it} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)'. \tag{10.85}$$

Таким образом, просто все суммируется только по доступным наблюдениям.

Аналогичным образом можно обобщить оценку со случайными эффектами. Для несбалансированного случая ее можно получить в виде

$$\beta_{\text{ОМНК}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} r_{it} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \sum_{i=1}^{N} \psi_i T_i (x_{it} - \bar{x})(x_{it} - \bar{x})' \right)^{-1} \times$$
$$\times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} r_{it} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \sum_{i=1}^{N} \psi_i T_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})' \right), \tag{10.86}$$

где $T_i = \sum_{t=1}^{T} r_{it}$ обозначает число тактов времени, а когда наблюдался индивидуум i,

$$\psi_i = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2}.$$

Иначе эту функцию можно получить с помощью МНК, примененного к следующей преобразованной модели

$$(y_{it} - \vartheta_i \bar{y}_i) = \mu(1 - \vartheta_i) + (x_{it} - \vartheta_i \bar{x}_i) + u_{it}, \tag{10.87}$$

где $\vartheta_i = 1 - \psi_i^{1/2}$. Отметим, что применяемое здесь преобразование является индивидуальной спецификацией, поскольку оно зависит от числа наблюдений для индивидуального i.

По существу, более общие формулы для оценок с фиксированными и случайными эффектами характеризуются тем, что все суммы и средние значения вычисляются только по доступным наблюдениям и что T заменяется на T_i. Полностью аналогичные корректировки применяются к выражениям для ковариационных матриц этих двух оценок, которые задаются выражениями (10.13) и (10.23). Состоятельные оценки неизвестных дисперсий σ_α^2 и σ_ε^2 имеют вид

$$\sigma_\varepsilon^2 = \frac{1}{\sum_{i=1}^{N} T_i - N} \sum_{i=1}^{N} \sum_{t=1}^{T} r_{it} (y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)' \beta_{\text{ФЭ}})^2 \tag{10.88}$$

и

$$\sigma_\alpha^2 = \frac{1}{N} \sum_{i=1}^{N} \left[(\bar{y}_i - \bar{x}_i' \beta_M)^2 - \frac{1}{T_i} \sigma_\varepsilon^2 \right] \tag{10.89}$$

соответственно, где β_M — межгрупповая оценка для вектора параметров β (вычисляемая как МНК-оценка в выражении (10.18), где средние значения теперь отражают «доступные средние значения»). Поскольку эффективность оценок для σ_α^2 и σ_ε^2 асимптотически не влияет на эффективность оценки со случайными эффектами, то в вычислительном отношении можно использовать более простые состоятельные оценки для σ_α^2 и σ_ε^2. Например, можно использовать стандартные оценки, вычислен-

[29] Мы предполагаем, что $\sum_{t=1}^{T} r_{it} \geq 1$, т.е. каждый индивидуум наблюдался, по крайней мере, один раз.

ные только по остаткам, полученным из оценивания, основанного на сбалансированной субпанели, а затем использовать выражение (10.86) или (10.87), чтобы вычислить оценку со случайными эффектами.

10.7.2. Смещение, обусловленное выборочной селективностью, и некоторые простые тесты

В дополнение к обычным условиям для состоятельности оценок со случайными и фиксированными эффектами, основанных на сбалансированной субпанели или на несбалансированных панельных данных, выше предполагалось, что индикатор отклика r_{it} независим от всех ненаблюдаемых переменных в модели. Такое предположение может быть нереалистичным. Например, основанное на таком предположении объяснение работы фондов может быть некорректным из-за того, что менее вероятно продолжение существования фондов с плохой работой (Ter Horst, Nijman, Verbeek, 1998), исследование эксперимента по эффективности политики использования дохода может пострадать от смещений, если более вероятен отказ от участия в панельном обследовании людей, которые извлекают меньшую выгоду из эксперимента (Hausman, Wise, 1979), или оценивание воздействия уровня безработицы на индивидуальную заработную плату может нарушаться, если в случае увеличивающейся безработицы более вероятен уход с трудового рынка людей с относительно высокой заработной платой (Keane, Moffitt, Runkle, 1988).

Если r_{it} зависит от эффекта α_i или остатка ε_{it}, то в стандартных оценках может возникнуть смещение, которое в дальнейшем мы будем называть **селективным смещением** (см. гл. 7). Это означает, что распределение y при заданном x и условное по способу отбора выборочных единиц в выборке отличается от распределения y при заданном x (которое нас интересует). Для состоятельности оценки с фиксированными эффектами теперь требуется, чтобы

$$E\{(x_{it}-\bar{x}_i)\varepsilon_{it} \mid r_{i1},\dots,r_{iT}\}=0. \tag{10.90}$$

Это означает, что оценка с фиксированными эффектами несостоятельна, если факт, находится ли индивидуум в выборке или нет, говорит нам кое-что об ожидаемом значении остатка, который связан с x_{it}. Ясно, что если справедливо условие (10.11) и r_{it} не зависит от эффекта α_i и всех остатков ε_{is} (для данного x_{is}), то сформулированное выше условие удовлетворяется. Заметим, что отбор единиц в выборку может зависеть от эффекта α_i, не влияя на состоятельность оценки с фиксированными эффектами для вектора параметров β. Фактически, даже ε_{it} может зависеть от r_{it} до тех пор, пока их соотношение не зависит от времени (подробности см. в работах (Verbeek, Nijman, 1992, 1996)).

В дополнение к условию (10.90), условия для состоятельности оценки со случайными эффектами теперь задаются в виде

$$E\{\bar{x}_i\varepsilon_{it} \mid r_{i1},\dots,r_{iT}\}=0$$

и

$$E\{\bar{x}_i\alpha_i \mid r_{i1},\dots,r_{iT}\}=0. \tag{10.91}$$

Условия не позволяют математическому ожиданию любой компоненты остатка зависеть от индикаторов пропущенных данных (наблюдений). Если менее вероятно, что индивидуумы с определенными значениями для их ненаблюдаемой гетерогенности α_i будут наблюдаться в некотором такте времени панельных данных, то, как правило, это будет приводить к смещению оценки со случайными эффектами. Аналогично, если более вероятно, что индивидуумы с определенными возмущениями ε_{it} имеют пропущенные данные (наблюдения), то оценка со случайными эффектами, как правило, несостоятельна. Заметим, что поскольку оценка с фиксированными эффектами позволяет пропущенным данным зависеть от эффектов α_i и от возмущений ε_{it} независимо от времени, то она является более устойчивой к смещениям из-за пропущенных данных, чем оценка со случайными эффектами. Другое важное наблюдение, сделанное в статье (Verbeek, Nijman, 1992) состоит в том, что оценки для несбалансированных панельных данных не обязательно страдают меньше от смещения из-за пропущенных данных, чем оценки для сбалансированных подпанельных данных. В общем, смещения из-за пропущенных данных в оценках для несбалансированных и сбалансированных выборок не обязательно будут одинаковыми, и их относительная величина априори не известна.

В статье (Verbeek, Nijman, 1992) предлагается ряд простых тестов на смещение из-за пропусков в данных. Эти тесты основаны на вышеупомянутых наблюдениях. Во-первых, поскольку условия для состоятельности утверждают, что остатки модели должны, в том или другом смысле, не зависеть от индикаторных переменных, то их можно протестировать, просто включая некоторую функцию от $r_{i1},\ldots, r_{iT}$ в модель и проверяя ее значимость. Ясно, что нулевая гипотеза, говорящая о том, что индивидуум наблюдался в любом из тактов времени от 1 до T, не давала бы нам никакой информации о его ненаблюденных значениях в модели. Очевидно, что добавление r_{it} в модель (10.84) приводит к мультиколлинеарности, так как $r_{it}=1$ для всех наблюдений в выборке. Вместо этого следует добавить некоторые функции от $r_{i1},\ldots, r_{iT}$, такие, например, как $r_{i,t-1}$, $c_i=\prod_{t=1}^{T} r_{it}$ или $T_i=\sum_{t=1}^{T} r_{it}$, показывающие соответственно наблюдалась ли выборочная единица i в предыдущем временно́м такте, или она наблюдалась в течение всех тактов времени, или какое общее число тактов времени эта единица наблюдалась. Заметим, что в сбалансированной субпанели все значения переменных таких функций идентичны для всех индивидуумов, и таким образом включаются в свободный член. В статье (Verbeek, Nijman, 1992) предполагается, что включение c_i и T_i может обеспечить приемлемую процедуру, чтобы проверить наличие смещения из-за пропущенных данных. Отметим, что это требует, чтобы модель оценивалась в рамках схемы со случайными эффектами, поскольку внутригрупповое преобразование исключило бы c_i и T_i. Конечно, если нулевые гипотезы не отклоняются, то это еще не является основанием для их принятия, т. е. для утверждения об отсутствии смещения из-за пропущенных данных, поскольку мощность критериев может быть низкой.

Другая группа тестов основана на идее, что четыре различных оценки для моделей со случайными и фиксированными эффектами, основанные либо на сбалансированной субпанели, либо на несбалансированных панельных данных, обычно имеют различные смещения из-за пропуска данных. Поэтому сравнение этих оценок может служить показанием правдоподобия смещения из-за пропущенных данных. Однако, хотя и можно сравнить любую пару оценок (см. (Verbeek, Nijman, 1992) или (Baltagi, 1995, Section 10.5)), известно, что оценки с фиксированными и случайными эффектами могут различаться и по другим причинам, чем смещение из-за пропущенных данных (см. п. 3 параграфа 10.2). Поэтому, наиболее естественно сравнивать либо оценки с фиксированными, либо оценки со случайными эффектами, использующие сбалансированную субпанель, с их аналогами, использующими несбалансированные панельные данные. Если различные выборки, отобранные на основе индикаторов $r_{i1},\ldots, r_{iT}$, приводят к значимо различным оценкам, то процесс отбора должен говорить нам кое-что о пропущенных наблюдениях в модели. Таким образом, значимо различные оценки указывают на наличие смещения из-за пропущенных данных. Поскольку оценки, использующие несбалансированные панельные данные, эффективны внутри специфического класса оценок, то мы опять можем использовать результат Хаусмана и вывести критическую статистику, основанную на оценке со случайными эффектами, в виде (сравните с критической статистикой (10.27)),

$$\xi_{СЭ}=(\beta^{С}_{СЭ}-\beta^{Н}_{СЭ})'[V\{\beta^{С}_{СЭ}\}-V\{\beta^{Н}_{СЭ}\}]^{-1}(\beta^{С}_{СЭ}-\beta^{Н}_{СЭ}), \tag{10.92}$$

где V обозначают оценки ковариационных матриц, а надстрочные прописные буквы $С$ и $Н$ относятся к сбалансированной и несбалансированной выборке соответственно. Точно так же можно получить тест, основанный на двух оценках с фиксированными эффектами. При нулевой гипотезе критическая статистика подчиняется хи-квадрат распределению с K степенями свободы. Заметим, что неявная нулевая гипотеза для такого теста состоит в том, что $p\lim(\beta^{С}_{СЭ}-\beta^{Н}_{СЭ})=0$. Если такая гипотеза верна приближенно, или обе эти оценки страдают от смещения из-за пропущенных данных в равной степени, то тест не имеет никакой мощности[30]. Отметим, что можно проводить такое же тестирование лишь для подмножества элементов вектора β.

[30] Предложенный здесь тест реально не является тестом Хаусмана, поскольку при альтернативной гипотезе ни одна из оценок не является состоятельной. Тем не менее, тест, сам по себе, является корректным; просто, при применении в определенных обстоятельствах он может характеризоваться ограниченной мощностью.

10.7.3. Оценивание с неслучайно пропущенными данными

Как и в пространственном («cross-sectional») случае (см. параграф 7.5) смещение из-за пропущенных данных порождает проблему идентификации. В результате при наличии смещения из-за пропущенных данных, если не наложить дополнительные предположения, состоятельное оценивание параметров модели невозможно. В качестве примера предположим, что индикатор пропущенных данных r_{it} можно объяснить пробит-моделью со случайными эффектами, т. е.

$$r_{it}^* = z_{it}'\gamma + \xi_i + \eta_{it}, \tag{10.93}$$

где $r_{it} = 1$, если $r_{it}^* > 0$ и $r_{it} = 0$ в противном случае, а z_{it} — (хорошо-мотивированный) вектор экзогенных переменных, который включает x_{it}. Интересующая нас модель задается в виде

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}. \tag{10.94}$$

Предположим, что компоненты остатков модели в этих двух уравнениях имеют совместное нормальное распределение. Тем самым приходим к обобщению пространственной выборочной модели с пропущенными данными, которая рассматривалась в пункте 1 параграфа 7.4. Влияние ограничений на формирование выборки для модели (10.94) отражается в математических ожиданиях ее ненаблюдаемых компонентов, условных по экзогенным переменным и индикаторам пропуска, т. е.

$$E\{\alpha_i \mid z_{i1},\ldots, z_{iT}, r_{i1},\ldots,r_{iT}\} \tag{10.95}$$

и

$$E\{\varepsilon_{it} \mid z_{i1},\ldots, z_{iT}, r_{i1},\ldots,r_{iT}\}. \tag{10.96}$$

Можно показать (Verbeek и Nijman, 1992), что условное математическое ожидание (10.96) не зависит от времени, если $\operatorname{cov}\{\varepsilon_{it}, \eta_{it}\} = 0$, или если $z_{it}'\gamma$ не зависит от времени. Это требуется для состоятельности оценок с фиксированными эффектами. Далее условное математическое ожидание (10.95) равно нулю если $\operatorname{cov}\{\alpha_i, \xi_i\} = 0$, тогда как условное математическое ожидание (10.96) равно нулю если $\operatorname{cov}\{\varepsilon_{it}, \eta_{it}\} = 0$, так что оценка со случайными эффектами состоятельна если ненаблюдаемые переменные в основном уравнении (10.94) и в уравнении для индикатора пропуска (10.93) некоррелированы.

В общем случае оценивание относительно более сложное. В статье (Hausman, Wise, 1979) рассматривается случай, когда панельные данные включают два такта времени, и пропущенные наблюдения имеют место только на втором такте. В более общем случае применение метода максимального правдоподобия для одновременного оценивания этих двух уравнений требует численного интегрирования в пространстве размерности выше двух (чтобы с помощью интегрирования исключить эти два индивидуальных эффекта). В статьях (Nijman, Verbeek, 1992) и (Vella, Verbeek, 1999) представлены альтернативные оценки, основанные на двухшаговом методе оценивания для пространственной выборочной модели с пропущенными данными. По существу, идея состоит в том, что члены в условных математических ожиданиях (10.95) и (10.96), кроме константы, можно определить из пробит-модели (10.93), так что оценки этих членов могут включаться в основное уравнение. В статье (Wooldridge, 1995) представлены некоторые альтернативные оценки на основе несколько других предположений.

Упражнения

Упражнение 1 (линейная модель)

Рассмотрим следующую простую модель панельных данных

$$y_{it} = x_{it}\beta + \alpha_i^* + \varepsilon_{it}, \quad i = 1,\ldots, N, \quad t = 1,\ldots, T, \tag{10.97}$$

в которой β — одномерный неизвестный параметр, и предполагается, что

$$\alpha_i^* = \bar{x}_i\lambda + \alpha_i \quad A \quad \alpha_i \sim \text{НОНР}\,(0, \sigma_\alpha^2), \quad \varepsilon_{it} \sim \text{НОНР}\,(0, \sigma_\varepsilon^2),$$

взаимно независимы, и независимы от всех x_{it}, где $\bar{x}_i = (1/T)\sum_{i=1}^{T} x_{it}$.

Параметр β в модели (10.97) можно оценить с помощью оценки с фиксированными эффектами (или с помощью внутригрупповой оценки), заданной в виде

$$\beta_{ФЭ} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)^2}.$$

Как альтернатива, корреляция между остатком $\alpha_i^* + \varepsilon_{it}$ и переменной x_{it} может быть учтена с помощью применения метода инструментальных переменных.

а. Приведите выражение для МИП-оценки $\beta_{ИП}$ параметра β в модели панельных данных (10.97), используя в качестве инструментальной переменной для x_{it} переменную $x_{it} - \bar{x}_i$. Покажите, что $\beta_{ИП}$ и $\beta_{ФЭ}$ идентичны.

Другой способ исключать индивидуальные эффекты α_i^* из модели состоит во взятии первых разностей. В результате приходим к выражению

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}), \quad i = 1,\dots,N, \quad t = 1,\dots,T. \tag{10.98}$$

б. Обозначьте МНК-оценку, основанную на модели первых разностей (ПР) (10.98), через $\beta_{ПР}$. Покажите, что оценка $\beta_{ПР}$ идентична оценкам $\beta_{ИП}$ и $\beta_{ФЭ}$, если $T = 2$. Эта идентичность для $T > 2$ больше не справедлива. В таком случае какую из этих двух оценок Вы бы предпочли? Объясните. (Примечание: для дополнительного обсуждения см. (Verbeek, 1995).)

в. Рассмотрите межгрупповую оценку β_M параметра β в модели (10.97). Дайте выражение для оценки β_M и покажите, что она является несмещенной для векторного параметра $\beta + \lambda$.

г. И, наконец, предположите, что мы подставляем выражение для α_i^* в модель (10.97) и получаем

$$y_{it} = x_{it}'\beta + \bar{x}_i\lambda + \alpha_i + \varepsilon_{it}, \quad i = 1,\dots,N, \quad t = 1,\dots,T. \tag{10.99}$$

Вектор $(\beta, \lambda)'$ можно оценить с помощью ОМНК (случайные эффекты) из модели (10.99). Можно показать, что полученная таким образом оценка параметра β идентична оценке $\beta_{ФЭ}$. Означает ли это, что никакого реального различия между подходами случайных и фиксированных эффектов нет? Примечание: для дополнительного обсуждения см. (Hsiao, 1986, Sect. 3.4.2a).

Упражнение 2 (модель Хаусмана-Тейлора)

Рассмотрим следующую линейную модель панельных данных

$$y_{it} = x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + w'_{1,i}\gamma_1 + w'_{2,i}\gamma_2 + \alpha_i + \varepsilon_{it}, \tag{10.100}$$

в которой $w_{k,i}$ не зависит от времени, а $x_{k,it}$, являются объясняющими переменными, изменяющимися во времени. Переменные с индексом 1 ($x'_{1,it}$ и $w_{1,i}$) строго экзогенны в том смысле, что $E\{x_{1,it}\,\alpha_i\} = 0$, $E\{x_{1,is}\,\varepsilon_{it}\} = 0$ для всех s и t, $E\{w_{1,i}\,\alpha_i\} = 0$ и $E\{w_{1,i}\,\varepsilon_{it}\} = 0$. Также предполагается, что $E\{w_{2,i}\,\varepsilon_{it}\} = 0$ и что выполняются обычные условия регулярности (обеспечивающие состоятельность и асимптотическую нормальность).

а. При каких дополнительных предположениях, МНК, примененный к модели (10.100), обеспечивает состоятельную оценку для векторов параметров $\beta = (\beta_1, \beta_2)'$ и $\gamma = (\gamma_1, \gamma_2)'$?

б. Рассмотрите (внутригрупповую) оценку с фиксированными эффектами. При каких дополнительных предположениях она являлась бы состоятельной оценкой для вектора параметров β?

в. Рассмотрите МНК-оценку для вектора параметров β на основе регрессии в первых разностях. При каком (каких) дополнительном предположении(ях) эта оценка является состоятельной для вектора параметров β?

г. Обсудите одну или более альтернативных состоятельных оценок для векторов параметров β и γ при предположениях: $E\{x_{2,is}\varepsilon_{it}\}=0$ (для всех s и t), и $E\{w_{2,i}\varepsilon_{it}\}=0$. Каковы ограничения в этом случае на число переменных в каждой из категорий?

д. Обсудите оценивание вектора параметров β, если $x_{2,it}$ *равняется* $y_{i,t-1}$.

е. Обсудите оценивание вектора параметров β, если $x_{2,it}$ *включает* $y_{i,t-1}$.

ж. Можно ли оценить состоятельно, как вектор параметров β, так и вектор параметров γ, если $x_{2,it}$ включает $y_{i,t-1}$? Если можно, то как? В противном случае, почему нет? (В случае необходимости сделайте дополнительные предположения.)

Упражнение 3 (динамические модели и модели бинарного выбора)

Рассмотрим следующее динамическое уравнение заработной платы

$$w_{it}=x'_{it}\beta+\gamma w_{i,t-1}+\alpha_i+\varepsilon_{it}, \tag{10.101}$$

где w_{it} обозначает логарифм почасовой ставки заработной платы индивидуума, а x_{it} — вектор персональных характеристик и характеристик работы (возраст, время обучения, пол, отрасль промышленности, и т.д.).

а. Объясните на словах, почему МНК, примененный к модели (10.101), является несостоятельным.

б. Объясните также, почему оценка с фиксированными эффектами, примененная к модели (10.101), является несостоятельной при $N\to\infty$ и фиксированном T, но состоятельная при $N\to\infty$ и $T\to\infty$. (Предположите, что остатки ε_{it} являются независимо и одинаково распределенными.)

в. Объясните, почему результаты из пунктов **а** и **б** также означают, что оценка со случайными эффектами (ОМНК-оценка) для модели (10.101) будет несостоятельной и при фиксированном T.

г. Опишите простую состоятельную (при $N\to\infty$) оценку для вектора параметров β и параметра γ, предполагая, что α_i и ε_{it} являются независимо и одинаково распределенными и независимыми от всех x_{it}.

д. Опишите более эффективную оценку для вектора параметров β и параметра γ при тех же самых предположениях.

В дополнение к уравнению заработной платы предположим, что существует модель бинарного выбора, объясняющая, работает индивидуум или нет. Пусть $r_{it}=1$, если индивидуум i работал в такте времени t, и $r_{it}=0$ в противном случае. Тогда модель можно написать как

$$r^*_{it}=z'_{it}\delta+\xi_i+\eta_{it},$$

$$r_{it}=1, \text{ если } r^*_{it}>0,$$

$$r_{it}=0 \text{ в противном случае.} \tag{10.102}$$

где z_{it} — вектор персональных характеристик. Предположим что $\xi_i\sim$НОНП$(0,\sigma_\xi^2)$ и $\eta_{it}\sim$НОНП$(0,1-\sigma_\xi^2)$, взаимно независимы и независимы от всех z_{it}. Модель (10.102) можно оценить методом максимального правдоподобия.

е. Дайте выражение для вероятности того, что $r_{it}=1$, при заданных z_{it} и ξ_i.

ж. Используйте выражение из пункта **е**, чтобы получить выражение вклада индивидуума i в правдоподобие, легко поддающееся обработке в вычислительном отношении.

з. Объясните, почему невозможно рассмотреть эффекты ξ_i как фиксированные неизвестные параметры и оценить δ состоятельно (при фиксированном T) из пробит-модели с фиксированными эффектами?

С этого момента предположим, что соответствующее уравнение заработной платы является *статическим* и задается выражением (10.101) с параметром $\gamma=0$.

и. Каковы последствия для оценки со случайными эффектами модели (10.101), если η_{it} и ε_{it} коррелированы? Почему?

к. Каковы последствия для оценки с фиксированными эффектами модели (10.101), если ξ_i и α_i коррелированы (в то время как η_{it} и ε_{it} нет)? Почему?