

**МОДЕЛИРОВАНИЕ ОБЩЕЙ МИГРАЦИИ НАСЕЛЕНИЯ НА
ОСНОВЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ ПАНЕЛЬНЫХ ДАННЫХ***

Огородников П.И., Катаева И.Н., Самсонова Т.А.

Важнейшими условиями формирования целенаправленной миграционной политики является разработка математических моделей, позволяющих адекватно оценить реальное и потенциальное развитие миграции как в стране, так и в рамках ее отдельных территорий.

В статье предлагается рассмотреть две различные спецификации модели панельных данных как по объектам наблюдения, так и по периодам времени. Подтвердить или опровергнуть априорные варианты относительно спецификации модели предлагается с помощью теста Хаусмана проверки гипотезы от ортогональности случайных эффектов и регрессоров. Затем по оценкам индивидуальных эффектов построенной регрессионной модели панельных данных можно провести ранжирование городов и районов области по степени напряженности ситуации с миграцией. При сравнении двух городов или районов области ситуация с миграцией будет сложнее там, где выше значение индивидуального эффекта, аккумулирующего в себе влияние на уровень миграционной подвижности всех неучтенных в модели факторов.

Предложенная модель основана на расчете индивидуальных эффектов, позволяет более эффективно определять место каждой в миграционных процессах.

Миграция – сложный социально-экономический процесс, связанный, с одной стороны, с пересечением государственных границ (внешняя миграция), с другой стороны, с территориальными передвижениями в пределах одной страны (внутренняя миграция).

Миграция – это серьезная проблема, воздействующая на все стороны жизни населения. Масштабы, формы и направления миграционных потоков создают целый ряд сложных экономических, социальных, политических и морально – психологических проблем. Прежде всего, миграционное движение непосредственно приводит к изменению общей численности и структуры населения, к переменам на рынках труда и жилья, в окружающей природной среде, в производственной и потребительской сферах, в городском и коммунальном хозяйствах, в сфере социальной защиты населения и т.д. Большой приток мигрантов, с одной стороны, в какой-то мере компенсирует естественную убыль населения, с другой стороны, обостряет ситуацию на рынке труда, что влечет за собой рост безработицы среди местного населения. Приток беженцев и вынужденных переселенцев способствует распространению опасных заболеваний, росту преступности, незаконному обороту наркотиков, что создает угрозу демографической и национальной безопасности. Неоспоримой проблемой являются также нерациональные миграционные потоки между сельской и городской местностью. Необходимо отметить, что 90% в объеме чистой миграции из села в город составляет молодежь. В результате усиливается деформация возрастной структуры сельского населения. Это создает трудности в обеспечении сельского хозяйства рабочей силой.

Таким образом, важной составляющей в деятельности государства должно быть

* Работа выполнена при финансовой поддержке гранта РГНФ № 08-02-81204 а/у.

проведение целенаправленной миграционной политики, важнейшими условиями формирования которой должны быть разработка математических моделей, позволяющих адекватно оценивать реальное и потенциальное развитие миграции, в том числе в рамках ее отдельных территорий.

Для современного уровня исследования проблем миграции весьма значимым представляется подход, при котором миграционное движение анализируется при помощи математических методов и моделей, позволяющих провести более глубокий анализ и получить наиболее структурированные и математически обоснованные результаты.

Это обуславливает актуальность изучения проблемы общей миграции населения Оренбургской области, анализа ее приоритетных направлений, форм и методов ее регулирования. Исследование особенностей развития миграционных процессов в области, как одной из важнейших приграничных территорий, ее экономической и демографической обстановки позволит выделить специфические функции и задачи региональных органов управления в миграционной сфере. В большинстве случаев при статистическом анализе миграционной ситуации ограничиваются построением изолированных уравнений регрессии для отдельных характеристик миграционных процессов, не рассматривая влияние этих характеристик на миграцию в течение определенного периода времени и не учитывая изменение характера и направления этого влияния из года в год. Статистическая отчетность по социально-экономическим показателям регулярно фиксируется в заданные периоды или моменты времени. Простое объединение данных за разные годы и применение к ним стандартных математико-статистических методов не всегда оправданно. Его использование обоснованно в стабильных и неизменных условиях. Но в условиях резких изменений экономической конъюнктуры, оказывающей непосредственное влияние на миграционную подвижность, необходимо использовать модели, учитывающие эти особенности. Учитывая сложный и многогранный характер миграционных процессов, множество факторов, на них влияющих и изменяющихся во времени, следует ожидать, что более адекватным может быть их описание с помощью моделей регрессии на

На основе панельных данных, содержащих информацию по городам и районам Оренбургской области за период с 2000 по 2007 гг., проведено моделирование уровня миграционной подвижности в году t (y_t) в зависимости от следующих показателей:

- $x_{1,t}$ – удельный вес населения в трудоспособном возрасте в году t , %;
- $x_{2,t}$ – обеспеченность населения врачами в году t , на 1000 населения;
- $x_{3,t}$ – обеспеченность населения средним медицинским персоналом в году t , на 1000 населения;
- $x_{4,t}$ – обеспеченность населения больничными койками в году t , на 1000 населения;
- $x_{5,t}$ – среднемесячная номинальная начисленная заработная плата работников в году t , рублей;
- $x_{6,t}$ – площадь жилищ, приходящаяся в среднем на одного жителя в году t , кв. м;
- $x_{7,t}$ – среднесписочная численность работников в году t , человек;
- $x_{8,t}$ – оборот розничной торговли на душу населения в году t , руб.;
- $x_{9,t}$ – доля населения, ищущего работу, в общей численности среднесписочных работников в году t , %;
- $x_{10,t}$ – объем платных услуг на душу населения в году t , руб.;
- $x_{11,t}$ – число зарегистрированных преступлений в году t , на 1000 населения.

Для описания зависимости общей миграции населения y_t от объясняющих переменных $x_{1,t}$, $x_{2,t}$, $x_{3,t}$, $x_{4,t}$, $x_{5,t}$, $x_{6,t}$, $x_{7,t}$, $x_{8,t}$, $x_{9,t}$, $x_{10,t}$, $x_{11,t}$ были рассмотрены линейные регрессионные модели панельных данных следующего типа:

- модели с фиксированными эффектами;
- модели со случайными эффектами.

Общий вид модели с фиксированными эффектами имеет вид:

$$y_{it} = \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \beta_4 x_{4,it} + \beta_5 x_{5,it} + \beta_6 x_{6,it} + \varepsilon_{it}, \quad (1)$$

где индекс i указывает на номер объекта наблюдения, то есть идентифицирует город или район Оренбургской области;

α_i – индивидуальные эффекты;

ε_{it} – регрессионные остатки.

Относительно регрессионных остатков предполагается:

1) $M(\varepsilon_{it}) = 0$;

2) $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma^2_{\varepsilon}, i = j \\ 0, i \neq j \end{cases}$;

3) $\text{cov}(\varepsilon_{it}, x_{j,ts}) = 0, \quad i, j = \overline{1, n}, \quad t, s = \overline{1, T}$.

Индивидуальные эффекты в модели (1) учитывают влияние на результативный показатель всех (наблюдаемых и ненаблюдаемых) переменных, которые принимают разные значения для разных объектов, но не меняются во времени. К таким переменным относятся специфические характеристики каждого города и района. Эти характеристики могут не только оказывать влияние на результативный показатель – коэффициент миграции, но и быть коррелированными с объясняющими переменными. Если в такой ситуации не учитывать панельную структуру данных и рассматривать обычную регрессию по объединенным данным, то это приведет к смещенным оценкам параметров.

Если предполагается, что ненаблюдаемые переменные не коррелированы с остальными регрессорами, то их влияние учитывается иначе – как компоненты ошибок наблюдения. В этом случае для панельных данных используется модель со случайными эффектами, имеющая вид:

$$y_{it} = \alpha + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \beta_4 x_{4,it} + \beta_5 x_{5,it} + \beta_6 x_{6,it} + u_i + \varepsilon_{it}, \quad (2)$$

где u_i – случайное отклонение (возмущение).

К модели (2) предъявляются следующие требования:

1) $M(u_i) = 0, M(u_i^2) = \sigma_u^2$;

2) $M(\varepsilon_i) = 0, M(\varepsilon_i^2) = \sigma_\varepsilon^2$;

3) $\text{cov}(\varepsilon_{it}, u_{ij}) = 0$;

4) $\text{cov}(u_i, u_j) = \begin{cases} \sigma_u^2, i = j \\ 0, i \neq j \end{cases}$;

5) $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2, i = j \\ 0, i \neq j \end{cases}$;

6) $\text{cov}(\varepsilon_{it}, x_{j,ts}) = 0, \quad \text{cov}(u_i, x_{j,ts}) = 0; \quad i, j = \overline{1, n}, \quad t, s = \overline{1, T}$.

Для оценки параметров регрессионных моделей с фиксированными и случайными эффектами воспользуемся пакетом Stata 7.0. Результаты оценивания параметров модели с фиксированными эффектами представлены в табл. 1.

Таблица 1

Результаты модели с фиксированными эффектами

Коэффициент	Оценка	Среднекв. ошибка	t-статистика	Значимость	Доверительный интервал	
					Нижняя граница	Верхняя граница
β_1	2,162327	0,827248	0,276	0,002	-17,57227	13,24762
β_5	0,0145211	0,0051271	0,960	0,000	-0,0152605	0,0443028
β_6	8,626024	4,2985	0,529	0,000	-40,7138	23,46175
β_8	0,0042585	0,0005884	0,928	0,004	-0,013292	0,0047749
β_9	-63,70345	23,84107	-1,253	0,000	-163,7971	36,39023

Для проверки значимости уравнения регрессии с фиксированными эффектами применяется формула:

$$F(k, NT - N - k) = \frac{(TSS^{FE} - RSS^{FE}) / k}{RSS^{FE} / (NT - N - k)}, \quad (3)$$

где $TSS^{FE} = \sum_{i=1}^{31} \sum_{t=1}^{131} (y_{it} - \bar{y}_i)^2$;

RSS – сумма квадратов остатков модели с фиксированными эффектами;

N – количество объектов;

T – количество периодов времени;

k – количество переменных.

В условиях справедливости нулевой гипотезы эта статистика распределена по закону Фишера-Снедекора с k и (NT - N - k) степенями свободы. Результаты проверки гипотезы представлены в виде таблицы 2.

Таблица 2

Результаты проверки значимости уравнения регрессии с фиксированными эффектами

Показатель	Значение показателя
F-критерий	18,79
Вероятность	0,00

Анализируя полученный результат, можно сделать вывод о том, что нулевая гипотеза о незначимости всех коэффициентов уравнения регрессии отвергается, т.е. полученное уравнение регрессии с фиксированными эффектами является значимым. По данным таблицы можно сделать вывод, что значимыми являются следующие коэффициенты: удельный вес населения в трудоспособном возрасте, среднемесячная номинальная начисленная заработная плата работников, площадь жилищ, приходящаяся в среднем на одного жителя, оборот розничной торговли на душу населения, доля населения, ищущего работу, в общей численности среднесписочных работников.

Для того, чтобы доказать оправданность введения в модель фиксированных эффектов, необходимо проверить гипотезу об их различии для разных объектов наблюдения. Если индивидуальные эффекты равны между собой, то модели следует предпочесть обычную регрессию с единственной константой. Выдвигаются следующие гипотезы:

$H_0: \alpha_i = \alpha_j = \alpha$ для любых i, j;

$H_1: \alpha_i \neq \alpha_j$, хотя бы для одной пары i, j.

Для проверки нулевой гипотезы используется статистика:

$$F = \frac{(RSS - RSS^{FE}) / (N - 1)}{RSS^{FE} / (NT - N - k)}, \quad (4)$$

где RSS – сумма квадратов остатков обычной регрессии;

RSS^{FE} – сумма квадратов остатков модели с фиксированными эффектами.

В условиях справедливости нулевой гипотезы статистика распределена по Фишеру-Снедекору с числом степеней свободы $N-1$ и $N1 - N - k$.

Если нулевая гипотеза отвергается, то обычной регрессии следует предпочесть регрессионную модель с фиксированными эффектами. Результаты проверки представлены в таблице 3.

Полученные результаты подтверждают неоднородность городов и районов Оренбургской области и свидетельствуют о необходимости использования панельной структуры данных.

Таблица 3

Результаты проверки гипотезы о различии индивидуальных эффектов для разных городов и районов области

Статистика	Значение статистики
$F_{набл}$	1,46
$F_{кр}(0,05;11;271)$	0,1479

Оценки индивидуальных эффектов рассчитаны по следующей формуле:

$$\hat{\alpha}_i = \bar{y}_i - b_1 \bar{x}_{1,i} - b_5 \bar{x}_{5,i} - b_6 \bar{x}_{6,i} - b_8 \bar{x}_{8,i} - b_9 \bar{x}_{9,i}, \quad (5)$$

где $\hat{\alpha}_i$ – оценка индивидуального эффекта для i -го объекта;

b_1, b_5, b_6, b_8, b_9 – оценки соответствующих коэффициентов модели;

$\bar{y}_i, \bar{x}_{1,i}, \bar{x}_{5,i}, \bar{x}_{6,i}, \bar{x}_{8,i}, \bar{x}_{9,i}$ – средние значения соответствующих признаков в модели для i объекта за все периоды времени.

Результаты оценивания параметров регрессионной модели со случайными эффектами представлены в табл. 4.

Модель со случайными эффектами примет вид:

$$y_{it} = \alpha + 0,992x_{1,it} + 0,001x_{5,it} + 611,737x_{6,it} + 0,02x_{8,it} - 74,622x_{9,it} + u_i + \varepsilon_{it}, \quad (6)$$

Результаты проверки значимости модели со случайными эффектами с помощью критерия Вальда представлены в табл. 5.

Таблица 4

Результаты оценивания модели со случайными эффектами

Коэффициент	Оценка	Среднекв. ошибка	t-статистика	Значимость	Доверительный интервал	
					Нижняя граница	Верхняя граница
β_1	0,9922066	0,3957707	0,128	0,002	-14,21262	16,19703
β_5	0,0014655	0,0006199	0,088	0,000	-0,031109	0,03404
β_6	11,73678	4,2888	0,768	0,003	-18,22872	41,70229
β_8	0,0199903	0,0007107	4,244	0,000	-0,0292232	-0,0107574
β_9	-74,62197	29,12692	-1,405	0,001	-178,7488	29,50487

Таблица 5

Результаты проверки значимости уравнений регрессии со случайными эффектами

Показатель	Значение показателя
------------	---------------------

Критерий Вальда	124,99
Вероятность	0,00

Анализируя полученный результат, можно сделать вывод о том, что нулевая гипотеза о равенстве всех коэффициентов уравнений регрессии нулю отвергается, т.е. полученное уравнение регрессии со случайными эффектами является значимым. Как видно из табл. 4, значимыми являются коэффициенты при следующих переменных: обеспеченность населения врачами, среднесписочная численность работников, оборот розничной торговли на душу населения, объем платных услуг на душу населения.

Для сравнения регрессии по объединенным данным и регрессионной модели со случайными эффектами реализован тест Бреуша-Пагана множителей Лагранжа на гетероскедастичность. Выдвинуты следующие гипотезы:

$$H_0 : \sigma_u^2 = 0$$

$$H_1 : \sigma_u^2 > 0.$$

Для проверки нулевой гипотезы применяется следующая статистика:

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left(\sum_{t=1}^T e_{it} \right)}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right]^2, \quad (7)$$

где e_{it} – оценки регрессионных остатков в объединенной регрессии.

В условиях справедливости нулевой гипотезы статистика (7) распределена по закону Хи-квадрат с одной степенью свободы. Высокие значения статистики свидетельствуют о том, что нулевая гипотеза о возможности игнорировать индивидуальные эффекты и объединить данные должна быть отвергнута, и необходимо строить модель со случайными эффектами. Результаты проверки гипотезы с помощью пакета Stata 7.0 представлены в табл. 6.

Таблица 6

Результаты теста Бреуша-Пагана

Показатель	Значение показателя
Статистика LM	237,04
Вероятность	0,00

По данным табл. 6 можно сделать вывод, что нулевая гипотеза отвергается, следовательно, для моделирования общей миграции населения необходимо отдать предпочтение регрессионным моделям панельных данных. Таким образом, предлагается рассмотреть две различные спецификации модели панельных данных, позволяющие учесть неоднородность данных как по объектам наблюдения, так и по периодам времени.

Подтвердить или опровергнуть априорные предположения относительно спецификации модели можно с помощью теста Хаусмана проверки гипотезы об ортогональности случайных эффектов и регрессоров. Его подход основан на том, что если гипотеза об отсутствии корреляции верна, то оценки моделей с фиксированными и случайными эффектами являются состоятельными, а оценки модели со случайными эффектами – неэффективными. Поэтому при выполнении нулевой гипотезы между оценками нет систематического смещения. При альтернативной гипотезе состоятельны лишь оценки модели с фиксированными эффектами. Для проверки гипотезы используется статистика:

$$W = [b_{FE} - b_{RE}]' [Cov(b_{FE}) - Cov(b_{RE})]^{-1} [b_{FE} - b_{RE}] \quad (8)$$

где b_{FE} – оценки параметров модели с фиксированными эффектами;

b_{RE} – оценки параметров модели со случайными эффектами;

$Cov(b_{FE})$ и $Cov(b_{RE})$ – оценки ковариационных матриц для параметров моделей с фиксированными и случайными эффектами.

При выполнении нулевой гипотезы статистика (8) асимптотически подчиняется закону распределения Хи-квадрат с k степенями свободы. Если нулевая гипотеза не отвергается, то необходимо выбрать модель со случайными эффектами, в противном случае – модель с фиксированными эффектами. Результаты реализации теста Хаусмана представлены табл. 7 и 8.

Таблица 7

Результаты реализации теста Хаусмана

Показатель	Значение показателя
Статистики W	57,79
Вероятность	0,00

Таблица 8

Различия в оценках коэффициентов регрессионных моделей панельных данных

Коэффициент	Модель с фиксированными эффектами	Модель со случайными эффектами	Разность
β_1	2,162327	0,9922066	1,17012
β_5	0,0145211	0,0014655	0,013056
β_6	8,626024	11,73678	-3,11076
β_8	0,0042585	0,0199903	-0,01573
β_9	-63,70345	-74,62197	10,91852

Анализируя полученные результаты, можно сделать вывод, что нулевая гипотеза отвергается, т.е. предпочтение следует отдать модели с фиксированными эффектами, оценка которой имеет вид:

$$\hat{y}_{it} = \hat{\alpha}_i + \underset{(0.827)}{2,162} x_{1,it} + \underset{(0.005)}{0,015} x_{5,it} + \underset{(4.299)}{8,626} x_{6,it} + \underset{(0.0005)}{0,004} x_{8,it} - \underset{(23.841)}{63,703} x_{9,it} \quad (9)$$

Оценкам коэффициентов при значимых переменных можно дать следующую интерпретацию:

1) при увеличении удельного веса населения в трудоспособном возрасте в момент времени t на 1 % общая миграция населения возрастет в среднем на 2,162%;

2) при увеличении среднемесячной номинальной начисленной заработной платы работников в момент времени t на 100 рублей общая миграция населения возрастет в среднем на 0,015%;

3) при увеличении площади жилищ, приходящейся в среднем на одного жителя, в момент времени t на 1 кв. м общая миграция населения возрастет в среднем на 8,626%;

4) при увеличении оборота розничной торговли на душу населения в момент времени t на 100 рублей общая миграция населения возрастет в среднем на 0,004%;

5) при увеличении доли населения, ищущего работу, не занятого трудовой деятельностью, в общей численности среднесписочных работников в момент времени t на 1% общая миграция населения снизится в среднем на 63,703%.

По оценкам индивидуальных эффектов построенной регрессионной модели панельных данных можно провести ранжирование городов и районов области по степени напряженности ситуации с миграцией. Чем больше значение индивидуального эффекта, тем при прочих равных условиях будет наблюдаться большее значение уровня миграционной подвижности. При сравнении двух городов или районов области ситуация с миграцией будет сложнее там, где выше значение индивидуального эффекта, аккумулирующего в себе влияние на уровень миграционной подвижности всех неучтенных в модели факторов. Результаты ранжирования городов и районов Оренбургской области по степени напряженности ситуации с миграцией населения представлены в табл. 9.

Таблица 9

Результаты ранжирования административно-территориальных образований Оренбургской области по значениям индивидуальных эффектов

№ п/п	Города и районы	Значение индивидуального	№ п/п	Города и районы	Значение индивидуального эффекта
1	г. Оренбург	-553,458	25	Ясенский	-335,196
2	г. Орск	-422,05	26	Новосергеевский	-333,807
3	г. Бузулук	-411,497	27	Курманаевский	-330,835
4	г. Абдулино	-403,802	28	Пономаревский	-327,732
5	г. Ясный	-400,722	29	Шарлыкский	-327,055
6	г. Гай	-400,549	30	Матвеевский	-326,632
7	г. Новотроицк	-397,458	31	Кваркенский	-320,835
8	Оренбургский	-378,003	32	Домбаровский	-319,542
9	г. Сорочинск	-374,041	33	Саракташский	-318,244
10	г. Медногорск	-366,295	34	Кувандыкский	-316,739
11	Грачевский	-361,38	35	Соль-Илецкий	-315,156
12	г. Бугуруслан	-360,847	36	Илекский	-311,57
13	Новоорский	-359,227	37	Александровский	-310,277
14	Светлинский	-357,154	38	Перволоцкий	-307,745
15	Гайский	-353,66	39	Первомайский	-307,462
16	Октябрьский	-351,341	40	Беляевский	-306,667
17	Сорочинский	-350,144	41	Тоцкий	-301,596
18	Абдуллинский	-349,421	42	Сакмарский	-300,362
19	Бугурусланский	-348,969	43	Асекеевский	-296,091
20	Северный	-348,744	44	г. Кувандык	-294,564
21	Бузулукский	-344,391	45	Акбулакский	-286,082
22	Красногвардейский	-343,219	46	Тюльганский	-274,966
23	Ташлинский	-339,753	47	г. Соль-Илецк	-268,752
24	Адамовский	-336,352			

По результатам ранжирования можно сделать вывод, что наиболее высокий уровень миграционной подвижности наблюдается в городах области. Это объясняется тем, что в городах сосредоточены многочисленные предприятия различных направлений деятельности, отчасти обеспечивающие мигрантов рабочими местами. Более низкий уровень миграционной подвижности наблюдается в сельских районах с низким уровнем развития экономики. Результаты ранжирования наглядно демонстрируют наличие дифференциации миграции в городской и сельской местности Оренбургской области.