ЭКОНОМИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ, *2007, том 43, № 1, с. 76–86*

МАТЕМАТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ

ИСПОЛЬЗОВАНИЕ МЕТОДА БУТСТРЕП ДЛЯ ВОССТАНОВЛЕНИЯ МАТРИЦЫ ПЕРЕХОДОВ*

И. А. Кабанова, О. В. Староверов

(Москва)

Рассматривается метод восстановления матрицы переходов работников, находящихся в разных отраслевых группах. Прогнозирование производится при отсутствии данных о переходах работников между группами. Имеются лишь наблюдения над структурой занятого в этих группах населения для некоторого числа последовательных моментов времени. Приводится числовой пример, основанный на реальных данных, для проверки возможности прогнозирования.

ВВЕДЕНИЕ

Изучение занятости в народном хозяйстве и прогнозирование ее межотраслевых связей и структуры по сферам деятельности – одна из основных задач экономики труда. Решение задачи занятости позволит эффективнее использовать рабочую силу, перераспределять ограниченные инвестиционные средства и повысить качество жизни населения, производительность труда, конкурентоспособность отраслей на мировом рынке, достичь заданных Президентом России темпов экономического роста России.

Рынок труда в России находится в постоянном движении – с одной стороны, структурная перестройка, вызванная радикальными экономическими реформами начала 1990-х годов, до сих пор так и не завершилась, а с другой стороны, растут требования к скорости и глубине реакции рынка труда на современные условия мирового экономического устройства: быстро меняющиеся технологии и нестабильность политической ситуации в мире.

Статистические ведомства многих стран, в том числе и России, способствуют изучению занятости, регулярно предоставляют сведения о численности работников по сферам деятельности. Обычное прогнозирование структуры занятых в этом случае строится на основе либо эконометрических моделей, либо временных рядов следующим образом. Численности занятых в группах, сформированных из работников, занятых в разных отраслях народного хозяйства, образуют временные ряды, число которых равно числу рассматриваемых отраслей, и каждый ряд прогнозируется независимо от других временных рядов. Правда, в эконометрические модели добавляются еще эндогенные переменные. Одна из последних работ, встретившаяся авторам, была статья (Фиткулина, 2001). В этом же сборнике в докладе (Ликман, 2001) представлена работа, где временные ряды используются не для оценки числа занятых и безработных, а для числа прогнозирования родившихся и умерших.

Назовем основные проблемы, возникающие при таком методе прогнозирования:

– в переходный период число наблюдений бывает настолько малым, что возможно только линейное прогнозирование;

– часто сумма прогнозов численностей всех групп может быть не равна численности всего населения; для устранения этого несоответствия приходится использовать дополнительные гипотезы и допущения, что снижает достоверность получаемых результатов;

– невозможно понять связи между группами и их изменение во времени. Такой метод прогнозирования не отражает механизма взаимосвязи сфер деятельности человека.

Так появляется стремление к прогнозированию на основе проверенных моделей движения населения и трудовых ресурсов, которые учитывают связи между группами. Ведь именно взаимосвязь, т.е. межгрупповые потоки, позволяет не только лучше прогнозировать структуру, но и реально влиять на нее в нужном направлении, когда это бывает необходимо.

* Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда (проект 98-02-02006).

Следовательно, возникает задача выбора модели движения населения и трудовых ресурсов, которая отражает потоки между отраслевыми группами и обеспечена статистическими наблюдениями, позволяющими оценить все параметры модели.

Таким образом, проблема оценки взаимосвязей на основе данных о занятости по сферам народного хозяйства – одна из главных проблем статистики движения населения. К сожалению, занятость в отраслевом разрезе показывает лишь итог межотраслевых связей, а не сами эти связи, которые как раз необходимы. В результате недостатка информации, вытекающего из доступных для исследования наблюдений, сами связи не измеряются, и для их оценки приходится прибегать к подходящим, но не слишком сильным предположениям.

Цель настоящей работы – исследовать метод оценки параметров модели, учитывающей связи работников, занятых в разных сферах деятельности. Для этого нужно, в соответствии с постановкой задачи о взаимосвязи (Ли, Джадж, Зельнер, 1977), восстановить матрицу переходов по итоговым (макро) данным (численности или структуре занятых).

Следует отметить, что данные о взаимосвязях предоставляют органы статистики. Так, например, известно число переходов между регионами России (всего их 10, а вместе с Калининградской областью – 11). Но получение даже таких микроданных требует определенных затрат, и совершенно не ясно, не дешевле ли по макроданным восстановить микроданные, чем получать их статистическим путем. Поэтому вопрос об оценке переходных вероятностей в моделях движения людей по макроданным также не лишен смысла.

Наиболее исследованными являются марковские модели движения людей. Их использование в прогнозах более привлекательно как с содержательной точки зрения (они позволяют учитывать взаимосвязи между отраслями занятости (Бартоломью 1985; Староверов 1997)), так и со статистической (информация о переходах иногда дается статистическими органами). Отсюда возникает вторая задача: по восстановленным взаимосвязям составить прогноз структуры занятости и проверить его качество.

МОДЕЛЬ

Одной из самых простых, но вполне оправданных с практической и теоретической точек зрения является модель движения трудовых ресурсов, учитывающая направления переходов работников (т.е. связи между группами), и итог этих связей – численности или структуры групп населения. Марковские модели достаточно хорошо изучены теоретически и уже успешно зарекомендовали себя на практике (Бартоломью 1985; Староверов 1997). В этих моделях в качестве направлений перемещений (связи) рассматриваются вероятности перехода p_{ij} отдельного работника из группы i в группу j за время h, а итогом перемещения служат вероятности оказаться в любой из групп в момент $t = kh$.

Пусть $\nu(t)$ – наблюдаемый, поэтому случайный, вектор-столбец, каждая компонента $\nu_i(t)$ которого отражает число занятых в отрасли i и так же является случайной величиной. Обозначим математическое ожидание $\nu(t)$ через $\mathbf{x}(t)$, тогда модель подвижности трудовых ресурсов, использующая марковские цепи (процессы), будет иметь вид:

$$\mathbf{x}(t + h) = \mathbf{P}^T\mathbf{x}(t), \tag{1}$$

$$d\mathbf{x}(t)/dt = \mathbf{\Lambda}^T\mathbf{x}(t), \tag{2}$$

где (1) формула для модели с дискретным временем; h – единица времени $\mathbf{P}^T$ – переходная матрица, состоящая из $p_{ij} \geq 0$ и $\sum_j p_{ij} = 1$; (2) – формула для модели с непрерывным временем; $\mathbf{\Lambda}$ – матрица интенсивностей переходов и ее элементы $\lambda_{ij} \geq 0$ при $i \neq j$ и $\Sigma_j\lambda_{ij} = 0$. Матрицы $\mathbf{P}$ и $\mathbf{\Lambda}$ связаны соотношением $\mathbf{P} = \mathbf{I} + \Lambda h$, которое позволяет переходить от модели с дискретным временем при $h \longrightarrow 0$ к модели с непрерывным временем и наоборот. Такие модели движения трудовых ресурсов обычно называются марковскими.

Предположения первой группы, приводящие к марковским моделям (моделям с межгрупповыми связями), описаны в (Староверов 1997), но одно предположение хотелось бы здесь рассмотреть. При малых интервалах времени h матрица вероятностей переходов, задающая средние потоки между группами, близка к единичной матрице $\mathbf{I}$, поэтому матрица $\mathbf{\Lambda}$ имеет неотрицательными лишь недиагональные элементы. Пусть марковская *модель однородна в течение некоторого периода времени,* т.е. *матрица переходов* $\mathbf{P}$ *не изменяется на протяжении всего периода* h. Ниже будет показано, почему это предположение является одним из основных. Для России, особенно с учетом переходного характера ее экономики, однородность может оправдываться в течение 1 года (самое большее – в течение 2- или 3-летнего периода).

ОЦЕНКА ПАРАМЕТРОВ

В марковской модели с конечным числом групп (k) (в нашем случае – сфер деятельности или отраслей экономики) число параметров связей (межгрупповых или межотраслевых), т.е. интенсивностей переходов или переходных вероятностей, будет равным $k(k-1)$ на каждом временном шаге h. Поэтому для определения этих $k(k-1)$ параметров не хватает $2k$ наблюдений $\mathbf{x}(t+h)$ и $\mathbf{x}(t)$ из модели (1). Нужно иметь, по крайней мере, $k-1$ k-мерных наблюдений в области занятости.

Федеральная служба государственной статистики собирает сведения о численности занятых по отраслям раз в месяц. Однако далее предлагается оценивать параметры модели, учитывающей связи работников, на основе регулярно публикуемых годовых данных и сравнить их с имеющимися у Госкомстата России данными, а затем на основании полученных данных – построить прогноз ситуации на рынке труда. Возможно, такой подход окажется дешевле, и не придется ежемесячно собирать сведения. К тому же для целей прогнозирования подобный подход является более корректным, так как мы априори видим тренд, и нет необходимости проводить анализ "выбросов", которые появляются естественным образом, когда используются ежемесячные данные.

Подойдем более внимательно к вопросу о необходимых статистических данных, учитывая метод оценки параметров марковской модели. Действительно, пусть $\mathbf{P}(\mathbf{\Lambda})$ – искомая матрица переходных вероятностей (интенсивностей), которую нужно знать, так как именно в ней отражаются все взаимосвязи. Пусть $\mathbf{s}(t)$ – вектор (структура) занятых по их сферам деятельности, т.е. $\mathbf{s}(t) = \mathbf{x}(t)/N(t)$, где $N(t) = \mathbf{e}^T\mathbf{x}(t)$; $\mathbf{e}$ – единичный вектор-столбец, следовательно, $N(t)$ – суммарное число занятых. Поскольку при неизменной[1] (или мало меняющейся) во времени суммарной численности $N(t)$ векторы $\mathbf{s}(t)$, служащие оценкой своих математических ожиданий $\mathbf{p}(t)$, близки к ним (т.е. к $\mathbf{p}(t)$), то для $\mathbf{p}(t)$ из модели (1) справедливы равенства $\mathbf{p}(t+1) = \mathbf{P}^T\mathbf{p}(t)$. Подставляя в последнее равенство известные $\mathbf{s}$ вместо неизвестных $\mathbf{p}$, получим в соответствии с методом моментов для нескольких l систему уравнений:

$$\mathbf{P}^T[\mathbf{s}(t), \mathbf{s}(t+h), \ldots, \mathbf{s}(t+lh)] = \{\mathbf{s}(t+h), \mathbf{s}(t+2h), \ldots, \mathbf{s}[t+(l+1)h]\}, \tag{3}$$

из которой следует оценить элементы матрицы $\mathbf{P}$, что возможно лишь при $l \geq k$.

Система уравнений (3) учитывает выдвинутые предположения о процессе подвижности работников, которые позволяют использовать описанный метод оценки параметров модели. Отметим, что период времени между сменой экономической ситуации не бывает обычно очень большим, следовательно, l должно быть мало.

Очевидно, что основная информация о вероятностях переходов заложена в наблюдаемых структурах $\mathbf{s}(t)$, поэтому в разные моменты времени их значения могут отличаться. Такой процесс наблюдается и при серьезном изменении экономических обстоятельств, влияющих на рассматриваемые взаимосвязи (матрица $\mathbf{P}$ изменилась), т.е. движение работников изменилось, а структура занятых еще нет. Согласно (Бартоломью 1985; Староверов 1997), в условиях стабильного развития экономики, когда структуры остаются относительно постоянными, система (3) не имеет однозначного решения, поэтому и не дает искомого ответа. Однако нестабильный рынок труда в России позволяет получить большое число наблюдений до того момента, когда наблюдаемые структуры станут отличаться друг от друга. Поэтому воспользуемся тем, что процесс переходов работников из одной отрасли в другую не достиг стационарного состояния[2].

Система уравнений (3) справедлива для периода времени от t до $t + (l+1)h$ при постоянстве на нем переходных вероятностей – матрицы $\mathbf{P}$. Предыдущее требование сводилось к достаточно большому числу наблюдений, т.е. к большой длительности наблюдений, а настоящее – требует постоянства на этом интервале условий, поддерживающих неизменность переходной матрицы, которая обычно на малом интервале времени не меняется. Совместить эти кажущиеся противоречивыми требования можно, взяв достаточно малый временной шаг h.

[1] Если суммарная численность $N(t)$ значительно меняется с изменением l, то это изменение можно учесть с помощью соотношения, вытекающего из модели (2): $d\mathbf{s}(t) = [\mathbf{\Lambda}dt - \mathbf{I}(dN/N)dt]\mathbf{s}(t)$ или $\mathbf{s}(t+h) = [\mathbf{P}^T - \mathbf{I}\delta]\mathbf{s}(t)$, где δ – темп изменения общего числа занятых за период длины h. Таким образом, вместо марковской матрицы $\mathbf{P}$ нужно искать матрицу с неотрицательными элементами и с суммами, по строкам равными $1-\delta$.

[2] В противном случае мы имели бы $\mathbf{x}(t) \longrightarrow \mathbf{q}$ и, следовательно, $\mathbf{q} = \mathbf{P}^T\mathbf{q}$. Очевидно, что сейчас в России имеет место противоположная ситуация, т.е. $\mathbf{x}(t)$ далеки от устойчивой структуры $\mathbf{q}$.

Таблица 1. Структура занятых в России, %

Группы населения	Годы			
	1997	1998	1999	2000
1	31.79	30.15	30.294	30.384
2	13.66	14.08	13.664	13.383
3	54.55	55.77	56.042	56.233

Предположим, что наблюдения проводятся раз в году. Можно ли искусственно "размножить" годовые структуры на квартальные, месячные, недельные? Это можно сделать с помощью бутстреп-метода (Эфрон, 1988). Такое "размножение" не добавляет информации о переходах, но позволяет применять методы математической статистики.

Для того чтобы можно было применить бутстреп-метод для "размножения" данных в задаче определения переходной матрицы, воспользуемся более компактными, чем тривиальная, достаточными статистиками[3], сохраняющими всю информацию.

В данном случае, используя датчик случайных чисел, получаем случайные величины, характеризующие число выбывших ($m_i(t)$) и число прибывших ($n_i(t)$) работников, распределенные по биномиальному закону за период t (долю года) (см. Приложение, п. 1). Бутстреп-выборкой в этом случае являются натуральные ряды от 0 до m_i и от 0 до n_i.

Таким образом, разбивая год на две, возможно неравные, части, можно найти число занятых в промежуточные моменты года, для которого известны численности занятых в группах на начало и конец года. Продолжая данную процедуру, можно определить необходимое для оценки всех переходных вероятностей число структур, например, квартальных или ежемесячных. Отметим, что имеет смысл "размножать" данные до размера, необходимого для решения конкретной задачи. В данном случае будем исследовать три укрупненные отраслевые группы (см. ниже) и, соответственно, разбиваем год на три периода, чтобы иметь возможность решить соответствующую систему уравнений. Для марковских моделей при значительном увеличении числа отраслевых групп получаем вырожденную марковскую матрицу ранга 1[4]. Поэтому при большом числе отраслевых групп необходимо детерминировать задачу и создать укрупненные группы, затем последовательно разбивать их на более мелкие составляющие, рассматривая каждое новое разбиение укрупненной отрасли как новую систему.

РЕАЛИЗАЦИЯ МЕТОДА

На основании данных статистических ежегодников, публикуемых Федеральной службой государственной статистики, работники всех областей народного хозяйства были распределены по трем отраслям хозяйствования:

1) **"промышленность"** – работники, занятые в промышленности и строительстве;

2) **"сельское хозяйство"** – работники, занятые в сельском и лесном хозяйстве;

3) **"услуги"** – работники из остальных отраслей.

В качестве начальных данных на основании сведений, опубликованных в (РСЕ, 2002), имеем следующие структуры занятых в трех группах за четыре года (табл. 1, подробнее см. Приложение, п. 2):

[3] По определению, статистика $G = G(\mathbf{X})$ называется достаточной для модели $F = \{F(x, \theta), \theta \in \Theta\}$, если условная плотность вероятности $L(\mathbf{x}/g; \theta)$ случайного вектора $\mathbf{X} = (X_1, \ldots, X_n)$ при условии $G(\mathbf{X}) = g$ не зависит от параметра θ (Ивченко, Медведев 1984). Это свойство статистики G означает, что она содержит всю информацию о параметре θ, имеющуюся в выборке, и поэтому все заключения об этом параметре, которые можно сделать при наблюдении $\mathbf{x}$, зависят только от $g = G(\mathbf{x})$. Можно также утверждать, что все статистические выводы о модели, обладающей достаточной статистикой, в итоге формулируются в терминах этой достаточной статистики. Достаточная статистика дает оптимальный, в определенном смысле, способ представления статистических данных, что особенно важно при обработке больших массивов информации.

[4] Основываясь на марковских свойствах матрицы переходов, а именно: элементы матрицы неотрицательны, и сумма элементов любой ее строки равна единице, становится понятно, что матрица, возведенная в высокую степень, если число отраслей велико, вырождается в матрицу с одинаковыми строками – матрицу ранга 1.

Таблица 2. Структура занятых по периодам 1998 г., полученная бутстреп-методом, %

Номер реализации бутстреп-метода	Группы населения	1998 г.			1999 г.
		1 января	1 мая	1 августа	1 января
1	1	30.15	30.14	30.07	30.29
	2	14.08	14.00	13.79	13.66
	3	55.77	55.86	56.14	56.04
2	1	30.15	30.21	30.21	30.29
	2	14.08	13.86	13.79	13.66
	3	55.77	55.93	56.00	56.04
3	1	30.15	30.26	30.22	30.29
	2	14.08	13.92	13.83	13.66
	3	55.77	55.83	55.95	56.04
4	1	30.15	30.17	30.21	30.29
	2	14.08	13.93	13.83	13.66
	3	55.77	55.90	55.96	56.04

Численность занятых, вышедших из каждой группы, составляет в среднем 25% (PCE, 2002). Каждый год был разбит на три периода по четыре месяца: с 1 января по 30 апреля; с 1 мая по 31 июля; с 1 августа по 31 декабря.

В результате размножения данных с помощью бутстреп-метода были получены структуры на 1 мая и 1 августа для каждого года в четырех вариантах. Например, для 1998 г. были получены следующие дополнительные структуры (табл. 2).

Теперь искомую задачу можно свести к задаче линейного программирования (Ли, Джадж, Зельнер, 1977). Известны три способа построения системы линейных уравнений на основании полученных после использования бутстреп-метода данных:

1) переопределенная система линейных уравнений (число неизвестных меньше числа уравнений в системе); в этом случае целевая функция строится на основании принципа наименьших абсолютных отклонений и дополнительного предположения о близости матрицы переходов к единичной матрице (метод абсолютных отклонений);

2) определенная линейная система (число неизвестных совпадает с числом уравнений); целевая функция не нужна;

3) недоопределенная система (число неизвестных больше числа уравнений); целевая функция в этом случае строится на основании предположения о близости матрицы переходов к единичной матрице (метод моментов).

В результате решения задачи вторым способом получались результаты, не согласующиеся с экономическим смыслом поставленной задачи, а именно, часть вероятностей переходов была меньше нуля, а часть превосходила единицу, поэтому этот способ был нами отброшен.

Использование метода абсолютных отклонений обусловлено определенными ограничениями. Для большого числа изучаемых отраслей народного хозяйства период наблюдения необходимо будет разбить на соответствующее число отрезков времени, для того чтобы иметь данные требуемой размерности. Соответственно, для определения матрицы переходов за период матрицу за меньший отрезок времени нужно будет возвести в степень, равную числу отрезков на данном периоде. Это может привести к вырождению матрицы в матрицу ранга 1, так как в случае большого числа изучаемых отраслей степень, в которую необходимо будет возвести матрицу, полученную для меньшего отрезка времени, будет высокой[5]. Это приведет к нецелесообразности применения метода.

[5] Так как применяется марковская модель, то матрица переходов за меньший отрезок времени, которую мы найдем, будет также удовлетворять свойствам марковости: элементы матрицы неотрицательны и сумма элементов любой ее строки равна единице. При таких условиях матрица, возведенная в высокую степень, вырождается в матрицу с одинаковыми строками – матрицу ранга 1.

Таблица 3. Оценки переходных вероятностей для разных реализаций

В группу	Из группы (при реализации 1)			Из группы (при реализации 2)		
	1	2	3	1	2	3
1	0.963	0	0	0.962	0	0
2	0.016	0.19	0.206	0.016	0.19	0.207
3	0.022	0.81	0.794	0.022	0.81	0.793

Таблица 4. Вероятности переходов

Годы	1997–1998			1998–1999			1999–2000		
В группу	из группы			из группы			из группы		
	1	2	3	1	2	3	1	2	3
Метод наименьших отклонений									
1	0.778	0.349	0.025	0.727	0.270	0.041	0.586	0.218	0.142
2	0.159	0.276	0.094	0.140	0.162	0.131	0.152	0.127	0.126
3	0.063	0.375	0.881	0.133	0.568	0.828	0.261	0.654	0.732
Метод моментов									
1	0.946	0.000	0.000	0.339	0.082	0.334	0.296	0.268	0.316
2	0.005	0.301	0.178	0.017	0.816	0.030	0.136	0.192	0.118
3	0.049	0.699	0.822	0.644	0.101	0.635	0.567	0.540	0.566

В нашем случае имеются три группы отраслей, что позволяет эффективно использовать метод абсолютных отклонений.

Решения систем уравнений, полученных в результате реализации первого и третьего способов приведения задачи к задаче линейного программирования, были осуществлены с помощью симплекс-метода. Этот известный и хорошо себя зарекомендовавший способ оценки матрицы **P** имеет существенный недостаток: он, как правило, дает много нулей для переходных вероятностей, что практически никогда не встречается. Получение вместо годовых структур занятости – численности работающих через недели (месяцы, кварталы и т.д.) – приводит к переходным матрицам за меньший промежуток времени, что позволяет в силу (1) после перемножения нужного числа переходных матриц за малые промежутки времени иметь необходимые оценки переходов за нужный период времени (год), кроме того, эти матрицы будут уже без нулей.

В данном случае были получены переходные матрицы для переходов за 4 месяца, которые имели нули, не согласующиеся с содержательным смыслом. Однако для получения матрицы перехода за год эти матрицы были возведены в куб ($\mathbf{x}(T+1) = \mathbf{P}^T\mathbf{x}(T+2/3)$, тогда $\mathbf{x}(T+1) = (\mathbf{P}^T)^3\mathbf{x}(T)$). В последних матрицах уже не было нулей. Матрицы перехода за год приводятся для двух разных вариантов реализаций бутстреп-метода размножений, системы уравнений получены методом абсолютных отклонений для периода 1996–1997 гг. (табл. 3).

Однако такой подход только кажется простым, так как известно, что не всякая функция от оценок параметров формирует достаточные и достоверные данные для анализа. Например, степень оценки, как правило, не является несмещенной оценкой степени самого параметра. В рассмотренном случае также надо уменьшать смещение полученных результатов. Существуют методы, подобные упомянутому выше, уменьшения смещения (Кендалл, Стьюарт 1973), а иногда и его полного устранения (см. (Кузнецов, Орлов 1990)). Но применять процедуру возведения в степень, дающую несмещенную оценку, слишком сложно.

Таблица 5. Прогноз структуры занятости в России с 1997 г.

Матрица за 1997 г.	Метод наименьших отклонений		Метод моментов		Сравнение ошибок (метод наименьших отклонений – метод моментов), %
	прогноз	ошибка, %	прогноз	ошибка, %	
1998	30.85	2.32	30.09	–0.21	2.10
	13.94	–1.00	13.95	–0.92	0.08
	55.21	–1.00	55.96	0.35	0.65
1999	30.23	–0.21	28.47	–6.01	–5.80
	13.93	1.94	14.28	4.52	–2.58
	55.84	–0.36	57.25	2.15	–1.79
2000	29.76	–2.05	26.95	–11.31	–9.26
	13.89	3.77	14.60	9.11	–5.34
	56.35	0.21	58.45	3.95	–3.73
2001	29.39	–3.49	25.50	–16.26	–12.78
	13.85	9.29	14.91	17.62	–8.33
	56.76	–0.20	59.59	4.78	–4.58
2002	29.10	–2.53	24.13	–19.17	–16.64
	13.82	13.65	15.19	24.95	–11.30
	57.08	–1.56	60.67	4.64	–3.08

РЕЗУЛЬТАТЫ РАСЧЕТОВ

Исходные данные о структурах занятости в трех группах по России представлены в табл. 1. В результате расчетов по данным о структурах занятости за 1997, 1998 и 1999 гг. в России был получен ретропрогноз на 1998, 1999 и 2000 гг. В табл. 4 представлены переходные матрицы, полученные для разных пар структур занятости (результаты расчетов для 1990–2004 гг. представлены в Приложении, п. 3).

Из полученных данных следует, что рынок труда находится в недостаточно стабильном состоянии: переходы из отраслей промышленности в период 1997–1999 гг. падали, однако к 2000 г. они опять увеличились, а динамика связи между сельским хозяйством и услугами примерно эквивалентна. Этот факт на первый взгляд представляется естественным для современной экономической ситуации в России.

Известно, что на сегодняшний день огромная часть доходов сосредоточена в области добычи полезных ископаемых, которая по нашей классификации доходов входит в первую группу. Что касается двух оставшихся отраслей – сельского хозяйства и услуг, то ротация кадров между ними вполне обоснована. До сих пор рынок труда в сфере услуг в России не насыщен. В поисках лучших условий труда, более высокой заработной платы, льгот и т.п. работники пытаются найти оптимальное место работы именно в этих отраслях. К тому же у работников данных отраслей существует возможность перехода за счет “близости” необходимых профессиональных навыков и уровня образования.

В ходе исследования было сделано несколько реализаций бутстреп-метода восстановления матрицы переходов. В данном случае представлена матрица переходов, полученная путем усреднения результатов пяти реализаций бутстреп-метода. С помощью этих данных по информации о структурах 1997 и 1998 гг. была оценена матрица переходов за период 1997–1998 гг. для двух способов построения системы линейных уравнений и сделаны прогнозы структур занятости с 1998 на 1999–2002 гг., представленные далее (табл. 5). Матрицы связей, с помощью которых прогнозировались структуры, имеются в табл. 4.

Всего в ходе исследования были восстановлены матрицы переходов за период с 1990 по 2004 г. (см. Приложение, п. 3). Для проверки адекватности полученных результатов были построены соответствующие ретропрогнозы на период от года до пяти лет включительно, что позволило оценить прогностические возможности предложенного метода и сравнить прогностическую точность двух способов построения системы линейных уравнений.

В результате были получены следующие данные.

1. Ошибка прогнозов превосходила 10% барьер:

а) в 31.7% случаев при условии применения метода наименьших отклонений;

б) в 16.1% случаев при условии применения метода моментов.

2. Сравнение ошибок прогнозов в общем случае:

а) в 22.8% случаев ошибки метода моментов и метода наименьших отклонений равноценны;

б) в 22.8% случаев ошибки прогноза при применении метода абсолютных отклонений меньше ошибок прогноза при применении метода моментов;

в) в 54.4% случаев ошибки при применении метода моментов меньше ошибок прогноза при применении метода абсолютных отклонений.

3. Ошибка прогноза при построении ретро-прогнозов на пять лет:

а) при условии применения метода абсолютных отклонений ошибка доходила до 56%;

б) максимальная величина ошибки при условии применения метода моментов достигала 24%.

4. Ошибка прогнозов при построении ретро-прогнозов на три года:

а) максимальная величина ошибки при условии применения метода абсолютных отклонений достигала 37%;

б) при условии применения метода моментов не превосходила 13%.

Таким образом, нами было выявлено преимущество прогностических возможностей метода моментов над методом абсолютных отклонений. Дополнительно можно сделать вывод, что предположение о нестабильности рынка труда в России было верным и определение периода стационарности на уровне 2–3-летнего периода соответствует сложившейся экономической ситуации.

Анализ построенных матриц переходов показывает, что структурные изменения на рынке труда России происходили в 1993 г. и в период 1997–1998 гг., что соответствует экономической и социальной ситуации тех лет и косвенно подтверждает правомерность применения представленного метода.

ЗАКЛЮЧЕНИЕ

Марковские модели можно использовать даже в том случае, когда не хватает статистических данных или имеющиеся оценки являются грубыми. Но эта возможность не исключает необходимости совершенствовать статистику занятости, которая может дать гораздо больше данных о поведении работников.

Хотелось бы указать на ряд задач, возникших при проведении этого исследования.

Полученные прогнозы (представленные в табл. 5) и описанные результаты исследования позволяют думать, что переходные матрицы остаются неизменными в течение, по крайней мере, трех лет. Действительно, такой вывод основан на том, что точность прогноза на один, два и три года вперед адекватна допустимой ошибке прогнозирования в социальной статистике. Более качественная проверка необходима, но это предмет дальнейшего исследования.

На основании имеющихся реализаций метода размножения данных можно сделать вывод о достаточной точности полученных оценок переходных матриц. Дополнительно определить точность оценок можно или теоретически – на основе дисперсии оценок, полученных методом моментов, или используя метод Монте-Карло. При этом первый вариант оценки точности сложен, но сделанный один раз, пригоден для всех случаев. Второй вариант (метод Монте-Карло) более простой, но его придется повторить для каждого рассматриваемого случая.

Одна из задач дальнейшего исследования состоит в том, что пока нет ответа на очень простой вопрос: что может дать лучшие оценки подвижности населения – много слабо отличающихся структур или мало, но достаточно сильно разнящихся. Возможно, что поставленный вопрос имеет совершенно простой ответ: это безразлично.

ПРИЛОЖЕНИЕ

1. Из предположения относительно переходов работников из одной группы занятости в другую, следует, что число таких переходов – случайная величина, распределенная по закону Пуассона (Староверов, 1997). Число занятых, ушедших из группы, – люди, перешедшие из этой группы в другую, математически – это сумма независимых случайных величин. Известно, что сумма независимых пуассоновских случайных величин также является случайной величиной, распределенной по закону Пуассона.

Таким образом, в случае движения людей потоки из любой группы во все другие за год – это пуассоновские случайные величины, суммы которых – это численность уволившихся из групп, известные из статистики движения занятого населения. Предположение о распределении Пуассона в моделях движения кадров – это стандартный прием (см., например (Бартоломью 1985)). Для занятых в замкнутом населении[6] суммарное число вошедших во все группы равно числу вышедших, что приводит лишь к изменению структур.

Зафиксируем какую-либо группу с номером i. Тогда при известной численности m_i, покинувших группу за единицу времени (год), число покинувших ее в течение периода t (доли года, $t < 1$) будет случайной величиной, распределение которой предстоит определить.

Группа i за время t уменьшилась на $\xi_1 \sim Pu(\lambda_i t)$ человек, за время $(1 - t)$ – на $\xi_2 \sim Pu(\lambda_i(1 - t))$ человек. Всего за период из группы i вышло $\xi = \xi_1 + \xi_2 \sim Pu(\lambda_i)$ человек. Рассмотрим следующую условную вероятность:

$$P(\xi_1 = k_1, \xi_2 = k_2 | \xi = m_i) = \{[(\lambda_i t)^{k_1} e^{-\lambda_i t}/k_1!][(\lambda_i(1-t))^{k_2} e^{-\lambda_i(1-t)}/k_2!]\}/(\lambda_i^{m_i} e^{-\lambda_i}/m_i!) =$$

$$= m_i! t^{k_1}(1-t)^{m_i - k_1}/k_1!(m_i - k_1)!$$

Таким образом, мы пришли к тому, что эта условная вероятность не зависит от неизвестного параметра модели λ_i, т.е. найдена достаточная статистика. Кроме того, найдено такое распределение, при котором все численности уходов людей из группы можно наблюдать за любое время $t(t < 1)$, т.е. известно их распределение при заданном числе переходов за год. При заданной достаточной статистике это распределение будет биномиальным с известными параметрами m_i и t, т.е. $C_m^k t^k(1 - t)^{m - k}$.

Аналогично, обозначим через n_i – число работников, пришедших в группу i за год. Так как в каждом промежутке времени длиною в год известно число занятых, работающих в каждой группе до выхода из нее всех переходящих (m_i), и число приходящих (n_i) в нее, то можно рассчитать численности попаданий в группу. Действительно, имея m_i вышедших из каждой группы за годовой промежуток времени, сумма вошедших в нее из остальных групп будет равна $n_i = m_i + x_i(T + 1) - x_i(T)$ (так как предполагается, что все вышедшие из одной группы обязательно попали в другую, т.е. система групп полна). Теперь мы можем получить численность занятых во вех группах в начале и в конце промежутка времени t.

2. Данные Федеральной службы государственной статистики (структура занятых), %

Годы	Промышленность	Сельское хозяйство	Услуги
1990	42.26	13.23	44.52
1991	41.84	13.50	44.66
1992	40.53	14.34	45.13
1993	39.44	14.60	45.96
1994	37.04	15.37	47.59
1995	35.20	15.06	49.74
1996	33.72	14.42	51.86
1997	31.79	13.66	54.55
1998	30.15	14.08	55.77
1999	30.29	13.66	56.04
2000	30.38	13.38	56.23
2001	30.45	12.67	56.87
2002	29.86	12.16	57.98
2003	29.58	11.39	59.03
2004	29.24	10.70	60.06

[6] Замкнутое население – это население, численность которого не может изменяться за счет внешних по отношению к нему источников.

3. Матрицы переходов за соответствующие периоды[7]

В группу	Из группы			Из группы			Из группы		
	1	2	3	1	2	3	1	2	3
Годы	1990–1991			1991–1992			1992–1993		
Метод абсолютных отклонений									
1	0.994	0.000	0.000	0.926	0.000	0.000	0.848	0.000	0.000
2	0.003	0.315	0.204	0.040	0.313	0.200	0.065	0.215	0.215
3	0.003	0.685	0.796	0.034	0.687	0.800	0.087	0.785	0.785
Метод моментов									
1	0.582	0.116	0.355	0.959	0.000	0.000	0.914	0.015	0.023
2	0.050	0.446	0.121	0.013	0.543	0.141	0.025	0.233	0.231
3	0.369	0.438	0.524	0.027	0.457	0.859	0.061	0.752	0.746
Годы	1993–1994			1994–1995			1995–1996		
Метод абсолютных отклонений									
1	0.959	0.000	0.000	0.815	0.000	0.000	0.593	0.073	0.096
2	0.013	1.000	0.000	0.050	0.540	0.125	0.144	0.147	0.152
3	0.028	0.000	1.000	0.134	0.460	0.875	0.263	0.780	0.752
Метод моментов									
1	0.930	0.000	0.000	0.718	0.464	0.012	0.790	0.218	0.013
2	0.000	1.000	0.011	0.178	0.416	0.043	0.143	0.335	0.090
3	0.070	0.000	0.989	0.104	0.120	0.945	0.067	0.447	0.897
Годы	1996–1997			1997–1998			1998–1999		
Метод абсолютных отклонений									
1	0.963	0.000	0.000	0.778	0.349	0.025	0.727	0.270	0.041
2	0.016	0.190	0.206	0.159	0.276	0.094	0.140	0.162	0.131
3	0.022	0.810	0.794	0.063	0.375	0.881	0.133	0.568	0.828
Метод моментов									
1	0.925	0.020	0.000	0.946	0.000	0.000	0.339	0.082	0.334
2	0.038	0.851	0.000	0.005	0.301	0.178	0.017	0.816	0.030
3	0.037	0.129	1.000	0.049	0.699	0.822	0.644	0.101	0.635
Годы	1999–2000			2000–2001			2001–2002		
Метод абсолютных отклонений									
1	0.586	0.218	0.142	0.997	0.008	0.001	0.593	0.138	0.179
2	0.152	0.127	0.126	0.000	0.962	0.000	0.137	0.115	0.119
3	0.261	0.654	0.732	0.003	0.030	0.999	0.270	0.747	0.702
Метод моментов									
1	0.296	0.268	0.316	0.383	0.031	0.328	0.383	0.031	0.328
2	0.136	0.192	0.118	0.000	0.944	0.000	0.000	0.944	0.000
3	0.567	0.540	0.566	0.617	0.025	0.672	0.617	0.025	0.672
Годы	2002–2003			2003–2004					
Метод абсолютных отклонений									
1	0.467	0.183	0.224	0.391	0.216	0.252			
2	0.157	0.090	0.100	0.157	0.070	0.089			
3	0.376	0.727	0.676	0.452	0.714	0.659			
Метод моментов									
1	0.840	0.045	0.063	0.626	0.065	0.168			
2	0.014	0.497	0.086	0.145	0.482	0.013			
3	0.146	0.458	0.851	0.229	0.453	0.819			

[7] Матрицы переходов получены путем усреднения результатов пяти реализаций бутстреп-метода.

СПИСОК ЛИТЕРАТУРЫ

Бартоломью Д. (1985): Стохастические модели социальных процессов. М.: Фин. и стат.

Ивченко Г.И., Медведев Ю.И. (1984): Математическая статистика. М.: Высшая школа.

Кендалл М., Стьюарт А. (1973): Статистические выводы и связи. М.: Наука.

Кузнецов С.Е., Орлов В.И. (1990): О несмещенном статистическом оценивании степеней неизвестной матрицы. В кн.: "*Многомерный статистический анализ и вероятностное моделирование реальных процессов*". М.: Наука.

Ли Ц., Джадж Д., Зельнер А. (1977): Оценивание параметров марковских моделей по агрегированным временным рядам. М.: Статистика.

Ликман И.А. (2001): Краткосрочное прогнозирование рождаемости. Труды международной научно-практической конференции "*Социально-экономические и демографические проблемы занятости в современной России*". М.: ИМЭИ при Минэкономразвития России.

РСЕ (2002): Российский статистический ежегодник, 2002. М.: Госкомстат России.

Староверов О.В. (1997): Азы математической демографии. М.: Наука.

Фиткулина Ф.Р. (2001): Использование регрессии в прогнозировании обобщающих показателей экономики региона. Труды международной научно-практической конференции "*Социально-экономические и демографические проблемы занятости в современной России*". М.: ИМЭИ при Минэкономразвития России.

Эфрон Б. (1988): Нетрадиционные методы многомерного статистического анализа. М.: Фин. и стат.

Поступила в редакцию
26.02.2003 г.

The Use of the Boot-strep Method for Restoration of Matrix of Transit

I. A. Kabanova, $\boxed{\text{O. V. Staroverov}}$

The method of restoration of matrix of transit of employees in different industrial groups is proposed. The prognosis are given without data on the transits of employees between the groups. The authors have data only on the structure of employment in the groups for some number of consequent moments. The numerical example is given, based on the real data, for verifying the prognostic possibilities.