

## МЕТОДЫ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ НА ОСНОВЕ ВРЕМЕННЫХ РЯДОВ С УЧЕТОМ ПРОСТРАНСТВЕННОЙ НЕОДНОРОДНОСТИ ДАННЫХ И НЕЛИНЕЙНОЙ ВЗАИМОСВЯЗИ МЕЖДУ ФАКТОРАМИ

При моделировании динамики экономических показателей возникают проблемы пространственной неоднородности данных и нелинейной взаимосвязи факторов. Авторами указывается на целесообразность многомерных моделей по временным рядам в форме систем одновременных уравнений, в форме моделей регрессии на основе панельных данных.

Будем рассматривать процесс построения математических моделей для прогнозирования некоторого экономического показателя. В качестве иллюстрации будем исследовать месячную динамику цен на хлеб высшего сорта в Оренбургской области за период с 1998 по 2004 гг.

Моделирование какого-либо процесса или явления на основе одномерного временного ряда характеризующего это явление показателя требует традиционных исследований на стационарность и наличие тренда. При этом если исследуемый процесс подвержен влиянию факторов, носящих периодический характер, то эта периодичность также должна быть непосредственно учтена в модели.

Для выявления периода колебаний, помимо чисто визуального исследования и анализа сущности изучаемого явления, часто используется спектральный анализ, сводящийся к выявлению максимальных и близких к максимальным пикам спектральной плотности временного ряда.

Между тем применение спектрального анализа к экономическим рядам динамики наталкивается на определенные трудности. Во-первых, анализ спектральной плотности требует в качестве своей информационной базы либо достаточно длинных стационарных рядов, либо несколько траекторий анализируемого ряда [1]. Во-вторых, статистические оценки теоретических спектральных характеристик, как правило, сильно флуктуируют и обладают весьма посредственной точностью<sup>1</sup>. В третьих, следует отметить также и эффект «эха», который иногда не позволяет сделать однозначного вывода о наличии периодической компоненты с частотой, например, именно  $\omega$ , а не  $\omega/2$  или  $\omega/3$ . И, нако-

нец, в-четвертых, от экономических показателей трудно ожидать строгой периодичности и постоянства амплитуды, которые предполагаются в спектральном анализе. В экономике каждый цикл имеет свои особенности, часто наблюдается асимметрия колебаний относительно уровня отсчета [2]. Мы можем говорить только о средней продолжительности таких циклов, но точная продолжительность будущего цикла остается неизвестной [3].

При исследовании подобных колебаний логичнее использовать нечеткое понятие цикла и метод фазового анализа [2], основная идея которого заключается в итеративном очищении ряда от маломощных колебаний и представлении временного ряда в виде сменяющихся друг друга фаз – фазовой диаграммы.

Проведенный фазовый анализ (рисунок 1), с одной стороны, подтвердил результаты, полученные спектральным анализом (пики спектральной плотности приходились на частоты, соответствующие 18 и 24-месячной цикличности), а, с другой стороны, доказал наличие именно нестрогих циклов в исследуемом ряду. Для дальнейших расчетов в качестве периода сезонной составляющей будем рассматривать два варианта  $T=18$  и  $T=24$ .

Определившись с длительностью периодической компоненты, можно использовать методологию Бокса-Дженкинса построения сезонных АРПСС моделей. В нашем случае для построения краткосрочного прогноза были идентифицированы и оценены две модели – АРПСС(1,1,1)(1,1,0) и АРПСС(1,1,1)(0,1,1) – с 24- и 18-месячной сезонностью каждая. При этом построенные прогнозные значения по моделям с сезонностью 24 месяца можно считать оптимистическим прогнозом, а по моделям с сезон-

ностью 18 месяцев – пессимистическим. Был построен обобщенный прогноз (рисунок 2).

Альтернативным методом решения задачи прогнозирования одномерного временного ряда является прогнозирование с помощью искусственных нейронных сетей. Преимущество использования нейронной сети как инструмента прогнозирования состоит в том, что здесь не нужно заблаговременно устанавливать взаимоотношения между величинами, не требуется никаких предположений относительно основного распределения совокупности, и, в отличие от многих традиционных методов, они могут работать с неполными данными. Нейронные сети особенно эффективны в тех случаях, когда исходные данные сильно коррелированы, либо неполны, либо рассматриваемой системе свойственна высокая степень нелинейности.

Прогнозирование на основе нейросетей было выполнено в пакете **Statistica Neural Networks**: использовались обобщенные регрессионные нейронные сети с 18 и 24 входами. Был построен обобщенный прогноз (рисунок 3). В целом оба метода дали сходные прогнозы на перспективу в 1 год.

Итак, прогнозирование на основе одномерных временных рядов является наиболее доступным и удобным методом прогнозирования с точки зрения минимума необходимой информации. С другой стороны, не учитываются взаимосвязи исследуемого процесса с другими процессами, их возможное влияние на него. Для получения инструментов управления интересующим нас процессом целесообразно строить модели по многомерным временным рядам, например, в форме системы одновременных уравнений (СОУ). При этом на первом этапе нам потребуется

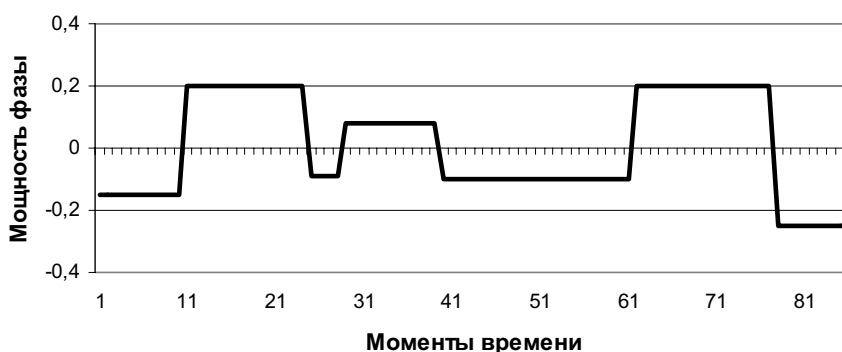


Рисунок 1. Фазовая диаграмма для ряда цены на хлеб высшего сорта (на 16 итерации)

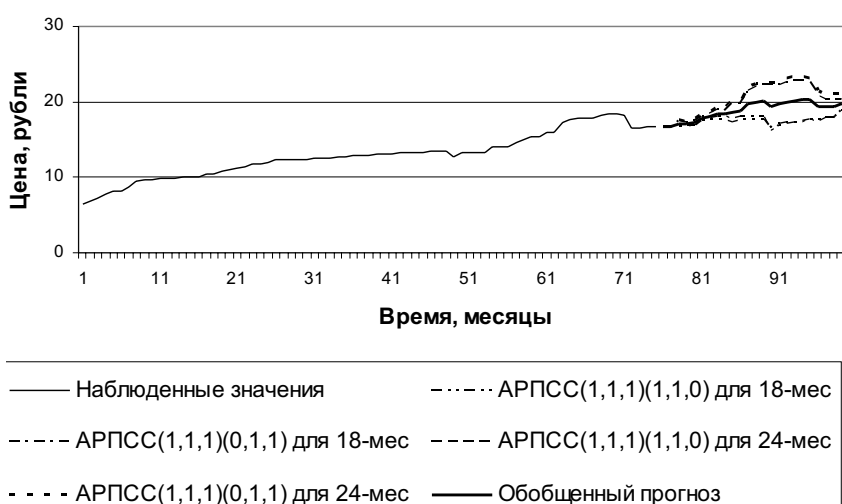


Рисунок 2. Графики наблюдаемых и прогнозных значений средней цены на хлеб высшего сорта для сезонных моделей АРПСС.

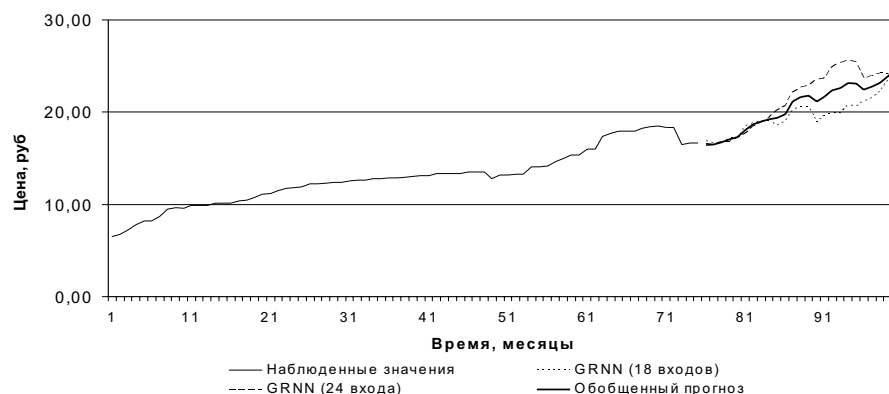


Рисунок 3. Графики наблюдаемых и полученных с использованием нейросетей прогнозных значений цены на хлеб высшего сорта

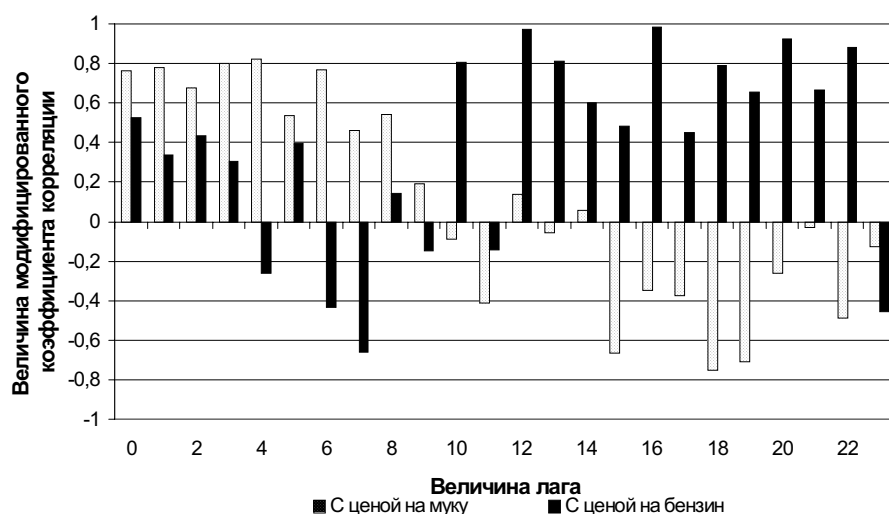


Рисунок 4. Модифицированные коэффициенты корреляции цены на хлеб со взятыми с лагом ценами на муку и бензин

установить отсутствие или существование связи между динамикой анализируемого показателя и динамикой остальных рядов, определить в случае существования связи ее тесноту. И в первую очередь нас будет интересовать связи между ценой на хлеб высшего сорта и ценами на энергоносители и муку.

Для выявления и анализа статистической связи между переменными обычно используется корреляционный анализ. Но при исследовании нестационарных процессов, представленных единственной реализацией, нарушаются условия применимости классического корреляционного анализа. Поэтому, работая с имеющими тенденцию временными рядами, следует использовать особую теорию корреляционного анализа – нетрадиционный корреляционный анализ [2]. Здесь вместо одного коэффициента

парной корреляции рассчитывается целый набор коэффициентов: модифицированный, адаптивный, условный и вероятностный коэффициенты корреляции. Безусловно, это дает намного больше информации об изучаемом явлении.

Расчет модифицированных коэффициентов корреляции показал (рисунок 4), что цена на хлеб и муку сильно коррелированы не только в текущем моменте времени (вплоть до лага 9) - наиболее сильная корреляция цены на хлеб с ценой на бензин относится к прошлому году (лаги 13-25). Важным моментом является изменение направления связи в обоих случаях.

Расчет адаптивных коэффициентов корреляции выявил неустойчивость связи между ценой на муку и ценой на бензин, а именно периодическое (19-20 месяцев) изменение на-

правления связи. Корреляционная связь между ценой на муку и ценой на хлеб значительно более устойчива во времени (рисунок 5).

Адаптивный коэффициент корреляции цены на бензин и цен на хлеб свидетельствует о наличии положительной связи между этими показателями в течение всего исследуемого промежутка времени и о существенном различии силы этой связи: если в начале исследуемого периода коэффициенты близки к 1, то в конце – только к 0,3 (рисунок 5).

Условные коэффициенты корреляции между положительными приростами цен на

муку и положительными приростами цен на бензин и приростами цен на хлеб близки к единице. Это означает, что повышение цены на муку или на бензин и повышение цен хлеб сильно связаны друг с другом. При этом согласно рассчитанным вероятностным коэффициентам, вероятность повышения цены на хлеб при условии повышения цены на муку составляет 0,96 (рисунок 6). Но вероятность снижения цены на хлеб при условии снижения цены на муку составляет всего 0,35.

Вероятность повышения цены на хлеб при условии повышения цены на бензин со-



Рисунок 5. График значений адаптивных коэффициентов корреляции ценой на хлеб и ценами на муку и бензин



Рисунок 6. Вероятностные коэффициенты корреляции между приростами рядов цен на хлеб и приростами цен на муку и бензин

ставляет 0,95. Вероятность снижения цены на хлеб при условии снижения цены на бензин значительно меньше и составляет 0,38.

Таким образом, можно говорить о наличии статистической связи между ценой на хлеб и ценами на энергоносители и муку. В то же время, очевидно, что на цену муки оказывает влияние цена зерна и также цена энергоносителей. Естественно предположить, что на цену зерна оказывает влияние величина собранного урожая и цены на энергоносители. Характер показателей и выявленные причинно-следственные связи между ними требуют построения многофакторной модели в форме СОУ.

Методом максимального правдоподобия с полной информацией в пакете EViews были получены оценки модели для Оренбургской области:

$$(1) \hat{y}_{1t} = 5,733 + 0,147 \cdot x_{1t} - 0,005 \cdot x_{2t},$$

(0,74)    (0,04)    (0,00)

$$\hat{R}^2 = 0,801, DW = 0,789;$$

$$(2) \hat{y}_{2t} = 2,442 + 0,300 \cdot x_{1t} + 1,441 \cdot y_{1t},$$

(0,82)    (0,14)    (0,39)

$$\hat{R}^2 = 0,818, DW = 0,333;$$

$$(3) \hat{y}_{3t} = 4,330 + 0,677 \cdot x_{1t} + 0,353 \cdot y_{2t-1},$$

(0,90)    (0,13)    (0,17)

$$\hat{R}^2 = 0,921, DW = 0,361;$$

где  $y_{1t}$  – цена реализации пшеницы заготовительным организациям, руб./кг;  
 $y_{2t}$  – цена муки пшеничной, руб./кг;  
 $y_{3t}$  – цена хлеба высшего сорта, руб./кг;  
 $x_{1t}$  – цена на энергоносители, руб./л;  
 $x_{2t}$  – валовый сбор пшеницы в расчете на месяц, т/чел.;  $t=1, 2, \dots, 63$ .

В данной модели предполагается постоянство рыночных цен реализации пшеницы в течение года.

Проверка остатков модели подтвердила нормальный характер их распределения, все уравнения и коэффициенты при факторных признаках значимы, коэффициенты детерминации достаточно высоки. Все это позволяет говорить о хорошем качестве модели.

С другой стороны, несмотря на то, что оцененная система одновременных уравнений правильно отражает сущность связей между

переменными (абсолютно правильные с точки зрения экономической теории и здравого смысла знаки, величины и соотношения между всеми коэффициентами), взаимодействия между ними в реальности наверняка намного более сложны, чем это предполагается линейной моделью. В таком случае естественно рассматривать прогноз интересующего нас показателя как выходной параметр некоторой нейросети. Входами теперь будут не значения этого же показателя в предшествующие моменты времени, как при использовании одного временного ряда, а значения факторов, предположительно оказывающих влияние на интересующий нас показатель, как при построении многофакторной модели.

Итак, в качестве выходов сети будем рассматривать цену хлеба высшего сорта в моменты времени  $t, t+1, t+2$ , а в качестве входов – цену муки в момент времени  $t$ , цену муки в предыдущий момент времени  $t-1$ , цену энергоносителей в момент времени  $t$ , цену энергоносителей год назад, то есть на момент времени  $t-12$ , цену реализации зерна пшеницы год назад, то есть на момент времени  $t-12$  и индикаторную переменную для месяца (принимает значение 1 для января, 2 – для февраля и т.д.).

Расчеты, как и прежде, будем проводить в пакете Statistica Neural Networks.

Наиболее подходящими оказались сети радиально-базисной функции с 9 нейронами на своем скрытом слое (RBF) и многослойный персептрон с двумя скрытыми слоями по 10 нейронов каждый (MLP). Ошибка сети на обучающем множестве составили 0,33 для RBF-сети и 0,47 для MLP-сети. Эти ошибки незначительно отличаются от соответствующих ошибок на верификационных и тестовых множествах, что свидетельствует об отсутствии переобучения.

График наблюдаемых и полученных по нейросетям значений цены на хлеб (рисунок 7), графика остатков (рисунок 8) и расчет статистики Дарбина-Уотсона дают основание предполагать наличие в обоих случаях положительной автокорреляции, что, очевидно, является следствием неучета в модели других важных факторов, оказывающих существенное влияние на цену хлеба. Между тем, максимальная относительная ошибка аппрокси-

мации не превышает 6% для RBF-сети и 7% для MLP-сети, средняя ошибка аппроксимации составила 1,71% для RBF-сети и 2,54% для MLP-сети, что говорит о хорошем качестве данных сетевых моделей и возможности их использования для прогнозирования. Кроме того, сценарные эксперименты также показали адекватную реакцию обеих сетей на изменения значений входных параметров.

Возникает вопрос, какую же сеть предпочесть? Многие различия сетей RBF и MLP связаны с различным представлением моделей: «групповым» в RBF и «плоскостным» в MLP. «Групповой» подход обуславливает гораздо меньшую способность RBF-сетей к экстраполяции выводов за область известных значений. MLP-сеть выдает более определенные решения при обработке сильно от-

клоняющихся данных, что иногда называют не критическим экстраполированием результатов. Хотя для обеих сетей необходимо проводить дообучение по мере поступления новой информации, если такое невозможно, то RBF-сеть остается как удобный инструмент только для краткосрочного прогнозирования при не слишком отклоняющихся значениях входных параметров.

Оценивая коэффициенты COY или строя нейросетевую модель, мы получали модель, учитывающую взаимосвязи факторов и временной характер данных. COY-модель описывала процесс формирования потребительской цены в среднем по области. Этот, вообще говоря, ненаблюдаемый, показатель рассчитывается как средняя взвешенная потребительских цен на изделия соответствующих

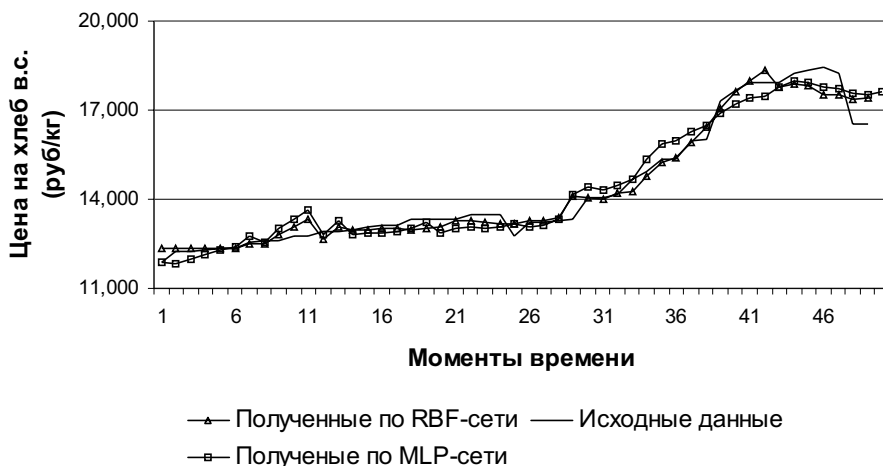


Рисунок 7. Графики наблюдаемых и полученных по нейросетям значений цены на хлеб

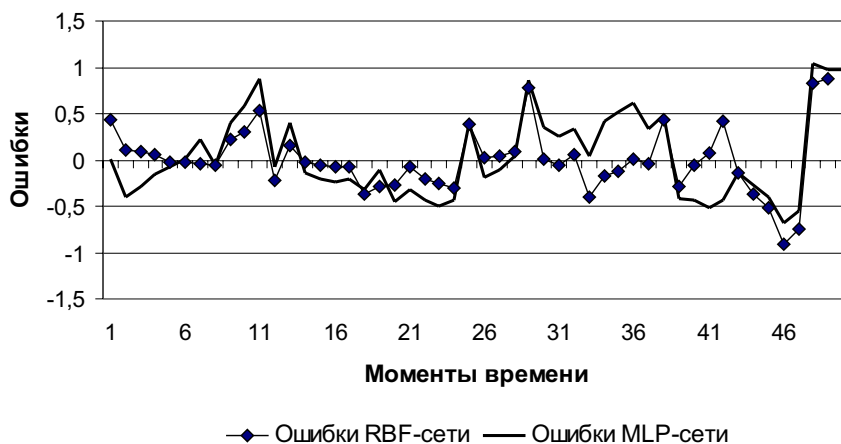


Рисунок 8. Графики отклонений наблюдаемых и полученных по нейросетям значений цены на хлеб

наименований в трех городах – Оренбурге, Орске и Бузулуке. Имея для каждого из этих объектов временные ряды характеризующих его показателей, мы переходим к совершенно иной структуре данных («панельные данные») и одновременному учету пространственного и временного аспекта исследуемого явления.

К достоинствам панельных данных относятся, во-первых, большее число наблюдений, обеспечивающее большую эффективность оценивания параметров модели. Во-вторых, возможность контроля неоднородности объектов. И, в-третьих, возможность идентифицировать эффекты, недоступные в анализе пространственных данных, например, факторы, индивидуальные для каждого объекта. Учесть эти факторы можно путем введения в модель так называемых индивидуальных эффектов [4].

Поскольку вышеназванные города вряд ли можно считать случайными выборками из большой популяции: каждый город уникален в своем роде, имеет свои собственные особенности, влияние которых учитывается с помощью параметров  $\alpha_i$ , то будем оценивать модель с фиксированными эффектами [5]:

$$y_{it} = \alpha_i + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_{it},$$

где  $y_{it}$  – цена хлеба для объекта  $i$  в момент времени  $t$ , руб./кг;

$x_{1t}$  – цена муки для объекта  $i$  в момент времени  $t$ , руб./кг;

$x_{2t}$  – цена энергоносителей для объекта  $i$  в момент времени  $t$ , руб./л;

$\alpha_i$  – индивидуальные эффекты;

$$\varepsilon_{it} \sim iid(0, \sigma^2), \text{ cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0,$$

$$i \neq j, t \neq s; i = 1, 2, 3; t = 1, 2, \dots, 63.$$

С использованием пакета Stata была получена оценка модели цены хлеба высшего сорта:

$$\hat{y}_{it} = \alpha_i + 0,76 \cdot x_{1t} + 0,50 \cdot x_{2t}, \hat{R}^2 = 0,71$$

Оценки индивидуальных эффектов значимы и приведены в таблице 1.

Таблица 1. Оценки индивидуальных эффектов

$\alpha_1$ (г. Оренбург)	$\alpha_2$ (г. Орск)	$\alpha_3$ (г. Бузулук)
3,15	1,89	0,76

Модель значима, при этом изменение цены на муку и изменение цены на энергоносители объясняет 71% вариации цены на хлеб высшего сорта. Наибольшее значение индивидуального эффекта,  $\alpha$ , следовательно, и наибольшая при прочих равных условиях цена на хлеб более высокого сорта именно в областном центре (г. Оренбурге) косвенно подтверждает, что в этом случае при установлении цены учитываются и факторы спроса на эту продукцию, по-видимому, обусловленные более высоким доходом жителей г. Оренбурга по сравнению с жителями Орска и Бузулука.

Итак, системы одновременных уравнений позволяют учитывать эндогенный характер регрессоров. Учитывая дополнительную информацию, содержащуюся в других уравнениях (ковариации остатков разных уравнений), мы получим более эффективные оценки. Однако, даже работая с СОУ, мы не застрахованы от невключения в модель факторов, которые оказывают влияние на результирующий признак, но, возможно, как в силу отсутствия соответствующей информационной базы, так и в силу ненаблюдаемости этих признаков, не могут быть включены в состав регрессоров. Пропущенные переменные могут привести к смещенным оценкам и коррелированности остатков каждого из уравнений.

С другой стороны, использование панельных данных позволяет выявлять и оценивать влияние ненаблюдаемых или явно не включенных в состав регрессоров факторов. В случае СОУ идея может заключаться в том, что неучтенные переменные должны вызывать одинаковые смещения (пропорциональные друг другу) в разных уравнениях. При использовании панельных данных предполагается, что пропущенная переменная имеет *within-* и *between-*структуру по группе. Такая структура вместе с формулировкой компонент ошибки накладывает ограничения на ковариационную матрицу остатков, что может быть использовано для идентификации модели и повышения эффективности оценок. Для получения оценок стоит использовать предложенный Балтаги [6] трехшаговый метод наименьших квадратов компонент ошибки (ЗМНККО), который асимпто-

тически эквивалентен использованному выше методу максимального правдоподобия с полной информацией.

В качестве другого преимущества панельных данных при работе с СОУ можно указать на тот факт, что они позволяют по-

лучать инструменты (при оценивании модели с помощью различных модификаций метода инструментальных переменных) внутри самой модели. Это очень важно, так как обычно поиск подходящих инструментов не является тривиальной задачей.

---

**Список использованной литературы:**

1. Айвазян, А.С. Основы эконометрики / С.А. Айвазян – М: ЮНИТИ-ДАНА, 2001. – 432 с.
2. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: учебное пособие / Ю. П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.
3. Петерс, Э. Фрактальный анализ финансовых рынков: применение теории хаоса в инвестициях и экономике. – М: Интернет-трейдинг, 2004. – 304 с.
4. Балаш, В. А. Модели линейной регрессии для панельных данных: учебное пособие для ВУЗов / В.А. Балаш, О.С. Балаш. – М., 2002. – 65 с.
5. Hsiao, C. Analysis of Panel Data / C. Hsiao. – Cambridge: Cambridge University Press, 2004. - 366 с
6. Baltagi, B.H. Econometric analysis of panel data / B.H. Baltagi. – Chichester: John Wiley & Sons, 2001 – 304 с

**Статья рекомендована к публикации 25.12.06**