

щения к платному здравоохранению для лиц этой группы (составляющей 11% населения) является отсутствие соответствующей бесплатной услуги. Но необходимость выживания вынуждает их компенсировать эти расходы, отказываясь в большинстве случаев (как видно из табл. 7) от покупки назначенных врачом лекарств.

Стремление дать своим детям хорошее высшее образование является определяющим мотивом в поведении самой немногочисленной (5,5% населения) группы, представленной третьим кластером (III). Именно оплата обучения кого-либо из членов семьи в высших учебных заведениях выступает отличительной чертой и одновременно главной статьёй семейных расходов этих домохозяйств. Как видно на рис. 2, ради этого им приходится время от времени отказываться даже от покупки личных вещей, предметов домашнего обихода и каких-то развлечений, несмотря на то, что их денежные доходы и имущественная обеспеченность находятся на уровне выше среднего. Поэтому они готовы воспользоваться и государственным кредитом для оплаты обучения в вузах, если бы была такая возможность. Показателен и тот факт, что образованность этой группы существенно отличается от всей совокупности: здесь нет людей с образованием ниже среднего, а доля лиц с высшим образованием в этом кластере предстает наибольшей по сравнению с другими. В структуре расходов этой группы домохозяйств присутствует также оплата медицинских услуг и покупка лекарств, хотя иногда им приходится отказываться от приобретения назначенных врачом лекарственных препаратов из-за недостатка денег, а льготные категории граждан, вошедших в эту группу, не всегда могли воспользоваться своим правом по большей части из-за отсутствия лекарства в аптеке.

Вовлеченность в сферу платного социального обслуживания людей, объединенных в пятый кластер (V) (15% населения), характеризуется следующими чертами:

1) главным побудительным мотивом обращения к платным формам медицинского обслуживания для этих людей является надежда на более высокое качество услуг. Для сравнения: соотношение причин "отсутствие бесплатной услуги" и "более высокое качество" в этом типе предстает как 0,7 : 1, тогда как в III группе как 1,5 : 1, а в IV — 3 : 1.

2) значительную часть социальных расходов этой группы лиц составляют затраты, связанные с обучением детей в школах.

3) материальные возможности этих людей (следует обратить внимание, что в пятый кластер вошли относительно обеспеченные люди, семейный месячный доход которых в 1,3 раза превышает соответствующий средний показатель) позволяют дать своим детям высшее образование, однако при наличии соответствующей возможности они не прочь прибегнуть к услугам и системы государственного кредитования. На рис. 2 видно, что и эта группа лиц время от времени не может себе позволить из-за недостатка денег определенные затраты, например, наиболее "дорогие" формы досуга (поездка в отпуск, к родственникам и т.д.), хотя отказ от других статей расходов происходит достаточно редко.

Выводы. Результаты проведенного исследования свидетельствуют не только о *широком распространении платных форм* медицинского обслуживания и образования, но и в значительной мере *принудительном характере* такой тенденции. В первую очередь такой вывод относится к медицинскому обслуживанию. Сложившиеся цены на рынке медицинских и образовательных услуг достаточно высоки относительно среднедушевых денежных доходов населения, поэтому большинство платных медицинских и образовательных услуг доступны только бес-

печенным слоям населения. В случае наличия потребности в платных услугах образования и здравоохранения даже среднеобеспеченные семьи испытывают серьезные материальные затруднения. Дифференциация по расходам на образование и здравоохранение практически в 2 раза выше, чем дифференциация по всему комплексу расходов. В совокупности эти два фактора заставляют иначе оценивать уровень жизни и его дифференциацию, а также степень распространения бедности. Государственная политика предоставления льгот на оплату дорогостоящих медицинских услуг и лекарств не решает данных проблем, поскольку в 50-60% случаев льготные категории граждан не могут реализовать имеющиеся у них льготы из-за мизерного финансирования социальных программ. На фоне снижающегося уровня жизни населения России такое положение существенно усугубляет и без того высокий уровень дифференциации уровня благосостояния жителей страны и является дополнительным фактором, усиливающим социальную несправедливость в обществе.

Александр **КРЫШТАНОВСКИЙ** **Методы анализа временных рядов**

Данные "Мониторинга", отмечающего 7 лет со дня проведения первого исследования (первый опрос был в марте 1993 г.), представляют редкий, во многом уникальный материал для социологического анализа динамики процессов, происходивших в нашем обществе в эти годы. Однако методы анализа временных изменений имеют свой математический аппарат, свою терминологию, которые существенно отличаются от традиционных для социологии.

Обычные модели, которые строятся при анализе социологических данных, выявляют закономерности на уровне поведения (вербального или реального) отдельных респондентов. Например, мы можем анализировать электроральные предпочтения *респондента* в зависимости от каких-то социально-демографических либо иных показателей *респондента* же. Единицей анализа при этом выступает отдельный респондент, его ответы или определенные характеристики.

При изучении временной динамики на основе данных повторных социологических исследований у нас нет возможности рассматривать изменения во времени, которые произошли с отдельным респондентом. В фокусе анализа находятся характеристики определенных социальных групп, динамика которых, собственно, и изучается. (Заметим, что этого не происходит при анализе данных панельных исследований, когда есть возможность рассматривать изменения на уровне отдельного респондента.)

Первой проблемой, возникающей при переходе от одной единицы анализа (респондента) к другой (социальной группе), является проблема построения индикаторов, характеризующих эту новую единицу анализа. При анализе данных на уровне респондента, как правило, этой проблемы не возникает, поскольку функцию индикаторов, фиксирующих характеристики респондента, выполняют либо непосредственно ответы на вопросы анкеты, либо какие-то построенные на их основе индексы. Поскольку в нашем распоряжении имеются лишь те же самые ответы на вопросы, при переходе к новым единицам анализа новые индикаторы можно конструировать только из них. В отдельных случаях это происходит вполне просто и естественно. Так, исходя из ответов на вопросы о возрасте респондентов можно построить новый показатель — средний возраст социальной группы. В других случаях построение пока-

зателей, характеризующих социальную группу, может стать нетривиальной задачей. В ряде статей "Мониторинга" можно найти примеры показателей, построенных для различных исследовательских задач*.

Другой проблемой, возникающей при попытке применения методов анализа временных рядов к данным мониторинга, является то, что не все исследования этого проекта проводились через равные промежутки времени. Большинство же математических методов поиска закономерностей в данных, изменяющихся во времени, основаны на жестком требовании равенства интервалов между замерами. Как правило, требование равенства интервалов входит даже в определение того, что такое "временной ряд"**. Нарушение этого требования в данных мониторинга достаточно легко устраняется, поскольку опросы проводились только с интервалом в один либо в два месяца. Таким образом, удалив ряд исследований, можно прийти к одинаковости интервалов — двум месяцам***.

Общая модель временного ряда. Одна из распространенных форм представления данных временного ряда традиционно выглядит следующим образом****:

$$Y = Tr + S + C + u \quad (1)$$

где Y — значение анализируемого показателя;

Tr — значение тренда — плавнo меняющийся, нециклический компонент, описывающий чистое влияние долговременных факторов, эффект которых сказывается постепенно;

S — сезонный компонент, описывающий регулярные изменения в течении заданного периода (года, месяца, недели и т.п.). Последовательность циклов предполагается почти точно повторяющейся;

C — циклический компонент, описывающий длительные периоды подъемов и спадов. Сами циклы могут меняться по протяженности и амплитуде;

u — ошибка, т.е. отклонения реальных данных от суммы первых трех, закономерных составляющих.

Далее мы рассмотрим методы выделения каждой из указанных закономерных составляющих, хотя, разумеется, в реальных данных не обязательно одновременно присутствуют все три.

Оценка тренда. Одним из естественных обобщенных показателей для каждого из опросов, проводившихся в рамках мониторинга, является показатель среднедушевого дохода. Социальной группой при этом выступает в целом объект исследования, т.е. взрослое население России. Вопрос о суммарном доходе семьи присутствовал во всех исследованиях мониторинга, за исключением трех опросов: ноябрь 1997 г., март и май 1998 г., в которых фиксировался не суммарный, а среднедушевой доход. Поскольку во всех исследованиях фиксировалось число членов семьи, проживающих вместе и ведущих общее хозяйство, то для всех исследований можно получить значение

* См., например: *Левада Ю.* Индексы социальных настроений в "норме" и в кризисе // Мониторинг общественного мнения: Экономические и социальные перемены. 1998. № 6. С. 7–13; *Урнов М.* Некоторые факторы адаптации российского общества к ситуации после августовского кризиса 1998 г. // Там же. 1999. № 2. С. 7–11.

** *Ostrom C.W.* Time Series Analysis: Regression Techniques // Sage University Series Paper on Quantitative Applications in the Social Sciences. Newbury Park (CA): Sage, 1990.

*** Не рассматривались опросы за апрель, июнь, август, октябрь, декабрь 1993 г., апрель и июнь 1994 г. Таким образом, из 48 массивов данных мониторинга (на ноябрь 1999 г.) не включались в рассмотрение семь массивов.

**** *Тюрин Ю.Н., Макаров А.А.* Статистический анализ данных на компьютере. М.: ИНФРА-М, 1998. С. 354.

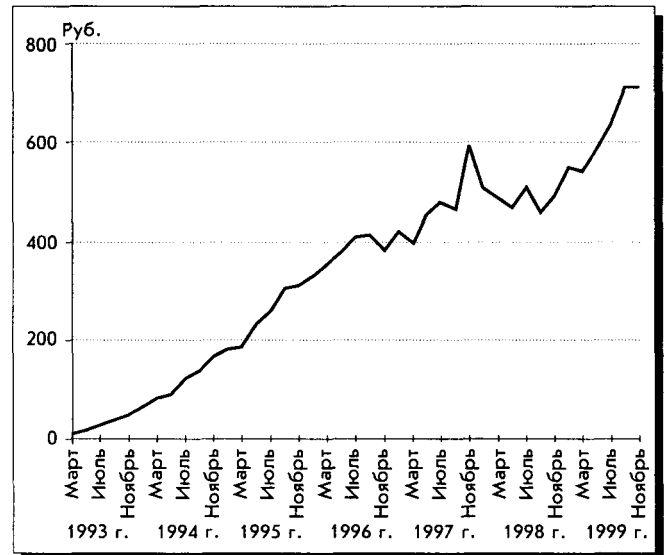


Рис. 1. Динамика среднедушевого дохода

показателя среднедушевого дохода. На рис. 1 показан график изменения этой характеристики.

Обращает на себя внимание всплеск (пик "вверх") в ноябре 1997 г., что можно объяснить уже отмечавшимся изменением формы вопроса, и пик "вниз" в ноябре 1998 г. — последствиями финансовых катаклизмов августа 1998 г.

Выявление тренда в данном временном ряду естественно искать, используя методы регрессионного анализа. При этом можно ограничиться линейной регрессией, учитывая очевидную схожесть графика с прямой. Необходимо, однако, предварительно произвести преобразование времени, которое является в нашем случае независимой переменной. Для этого можно использовать число месяцев (лет, дней), прошедших с начала первого исследования, взяв за точку отсчета первый месяц (за 0 или 1).

Проведенная таким образом регрессия дает очень высокий коэффициент детерминации — 0,956. Иными словами, почти на 96% динамика среднедушевого дохода объясняется просто тем, что со временем доход линейно возрастает. Уравнение регрессии выглядит следующим образом:

$$Y = 9,49 + 8,32 \times T, \quad (2)$$

где Y — значение среднедушевого дохода (в руб.);

T — номер месяца, при условии, что март 1993 г. принят за 0.

Столь высокий коэффициент детерминации, скорее всего, вызовет удивление у любого социолога — обычно в социологических моделях регрессия не дает значений этого показателя выше, чем 0,3–0,4. На этот счет есть два соображения. Во-первых, традиционный подход к анализу социологических данных подразумевает, что как единица анализа используется респондент. При этом на основе регрессионной либо какой-то иной модели мы описываем поведение отдельного респондента. В результате получается, что все специфическое многообразие поведения, оценок, мнений, ситуаций человека только в относительно небольшой степени детерминируется теми показателями, которыми мы хотим их объяснить. В нашем же случае мы перешли от респондента как единицы анализа к обществу в целом, уничтожив, по сути, специфику отдельного респондента, заменив показатель личного дохода на обобщенную среднюю по массиву. Потеряв ту вариативность, которая присуща отдельным респондентам, мы кардинально уменьшили дисперсию изучаемого признака, и, как следствие, даже простая регрессионная мо-

дель дала очень высокое значение коэффициента детерминации.

Второй причиной появления столь высокого значения коэффициента детерминации является то, что он может неверно вычисляться при использовании традиционного регрессионного подхода к анализу временной динамики. Дело в том, что при применении регрессионного анализа к изменяющимся во времени данным часто происходит нарушение одного из ограничений самого метода. Регрессионный анализ предполагает, что ошибки (обозначенные как u и v в формуле (1) независимы друг от друга. Иными словами, то, насколько модель ошибается в описании среднедушевого дохода в момент времени T , никак не связано с тем, насколько велика ошибка предсказания для момента времени $T+1$. Это ограничение носит название "требование отсутствия автокорреляции". Так вот, в реальных социологических показателях автокорреляция часто присутствует. В частности, как будет показано далее, это произошло и в случае с динамикой среднедушевого дохода. Ниже мы рассмотрим приемы корректного вычисления коэффициента детерминации при наличии автокорреляции. В любом случае, значение коэффициента становится меньше, и порой существенно меньше.

Мы выявили сильную трендовую составляющую в динамике показателя среднедушевого дохода. Для того чтобы искать влияние других факторов, необходимо вычесть фактор тренда как уже выявленной закономерной составляющей. Такого рода вычитание называется в регрессионном анализе получением остатков. Логика анализа именно остатков состоит в том, что наряду с уже выявленной закономерностью в поведении анализируемой переменной могут быть и другие закономерности, обнаружению которых первая будет мешать. Дальнейший анализ временного ряда строится именно на анализе остатков. На рис. 2 показан график остатков, полученных вычитанием из графика на рис. 1 трендовой составляющей.

Автокорреляция. Как уже отмечалось, одним из нарушений ограничений метода регрессионного анализа при анализе рядов временной динамики является наличие в данных автокорреляции, т.е. зависимости данных в моменты времени T и $T+1$. На самом деле наличие такого рода зависимости, например, в отношении данных среднедушевого дохода (см. рис. 1) вполне очевидно. Действительно, рост дохода в каждом следующем опросе идет не от нуля, а от достигнутого его уровня в настоящий момент. Таким образом, то значение дохода, которое будет в следующий момент времени, существенно связано с сегодняшним значением.

Однако эта зависимость определяется фактором тренда, который мы уже установили, и можно ожидать, что после того, как тренд из данных удален (см. рис. 2), т.е. при переходе к анализу остатков, автокорреляция будет отсутствовать. Как это ни удивительно, на самом деле это не так. На рис. 3 помещена диаграмма автокорреляционной функции, показывающей значения коэффициентов корреляции остатков показателя среднедушевого дохода для соседних моментов времени (сдвиг, называемый лагом) на один, два и многие другие моменты времени. Из графика видно, что коэффициент корреляции для двух соседних моментов времени достаточно велик (0,655), но значим он и для сдвига на два (0,517) и на три (0,350) момента времени. Иными словами, то, насколько отклоняется среднедушевой доход от линейной зависимости в настоящее время, существенно влияет на то, насколько он будет отклоняться от этой модели через 2, 4 и 6 месяцев. Такое длительное "последствие" может определяться двумя различными ситуациями.

Во-первых, то, что значение показателя в момент времени T влияет на значение в момент $T+1$, а значение в

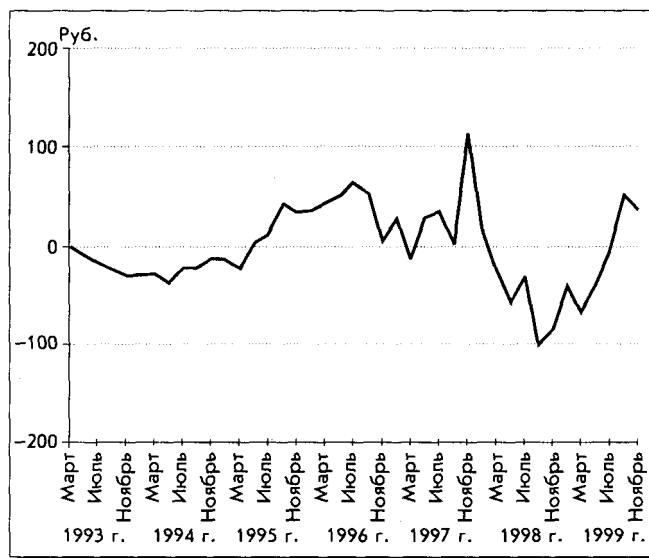


Рис. 2. График разностей (остатков) между значениями среднедушевого дохода и линейным трендом

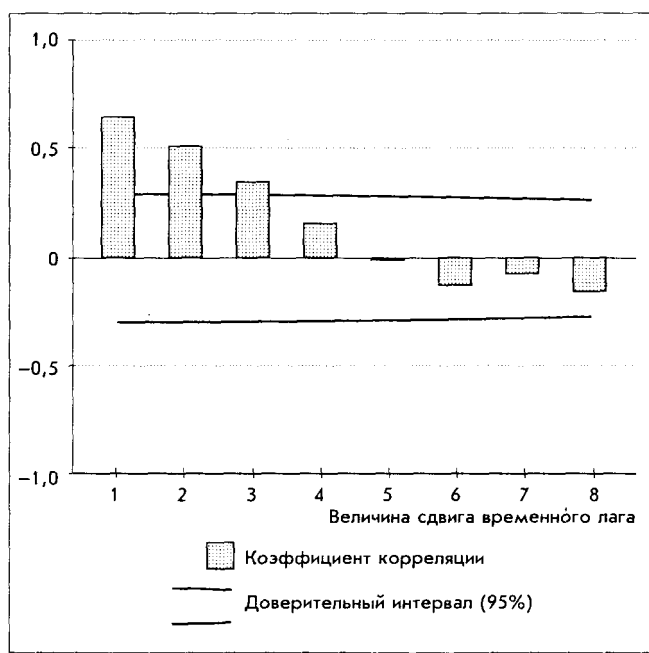


Рис. 3. Автокорреляционная функция для остатков показателя среднедушевого дохода

момент $T+1$ влияет на $T+2$, необходимым образом приводит нас к тому, что значение в момент времени T будет влиять и на $T+1$, и на $T+2$, и на $T+3$, и т.д. Хотя такого рода влияние должно ослабевать. Это, так сказать, естественное последствие, и объясняется оно тем, что значение в каждый из моментов времени просто связано со значением в предыдущий момент.

Во-вторых, возможным объяснением может быть ситуация, когда значение показателя в момент времени T влияет на значение в момент $T+2$ не только через $T+1$, но и непосредственно. Например, успеваемость на первом курсе института влияет на успеваемость на втором курсе (поскольку от того, как усвоены предметы первого курса зависит качество усвоения предметов, изучаемых на втором курсе). Однако, так как на первом курсе не только ус-

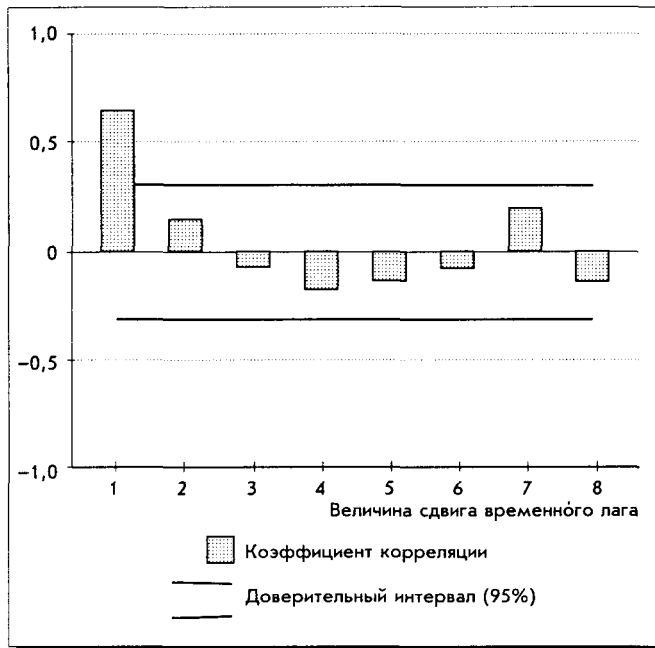


Рис. 4. Частная автокорреляционная функция для остатков показателя среднедушевого дохода

ваиваются определенные знания, но и формируется "умение учиться" в институте, то через этот "канал" успеваемость на первом курсе влияет на успеваемость на третьем курсе и далее.

В том случае, когда значения показателя в момент времени T непосредственно влияют только на значения показателя в момент времени $T+1$, а на значения в момент времени $T+2$ влияют только через $T+1$, тогда говорят о наличии автокорреляции первого порядка. Если же есть не только опосредованное влияние T на $T+2$, $T+3$, то говорят об автокорреляции второго, третьего порядков. Надо сказать, что в социологии, как правило, приходится иметь дело с автокорреляциями первого порядка.

Для определения порядка автокорреляции используется частная автокорреляционная функция (рис. 4).

На рис. 4 видно, что только первый коэффициент корреляции является значимым. Это свидетельствует о том, что мы имеем дело с процессом, в котором есть автокорреляция первого порядка.

Итак, мы установили, что в динамике показателя среднедушевого дохода есть автокорреляция. Следовательно, непосредственно применять регрессионный анализ для получения кривой тренда мы не имеем права. Для адекватного применения регрессии в такой ситуации используются так называемые модели авторегрессии. На их основе вычисляются регрессионные коэффициенты и коэффициент детерминации с учетом наличия автокорреляции. Есть несколько различных подходов к решению этой задачи. Часть из них реализована в пакете SPSS (это модели Бича-Маккиннона, Кокрена-Оркута, Прайса-Уинстена).

Метод Кокрена-Оркута вычисления зависимости среднедушевого дохода от времени с учетом автокорреляции дает следующее уравнение:

$$Y = 13,97 + 1,81 \times T + 0,79 \times (\text{доход в предыдущем опросе}), \quad (3)$$

где Y — размер среднедушевого дохода;

T — число месяцев, прошедших с марта 1993 г.

На рис. 5 показаны графики реальных значений среднедушевого дохода, значений, предсказываемых на основе модели линейной регрессии и модели регрессии с уче-

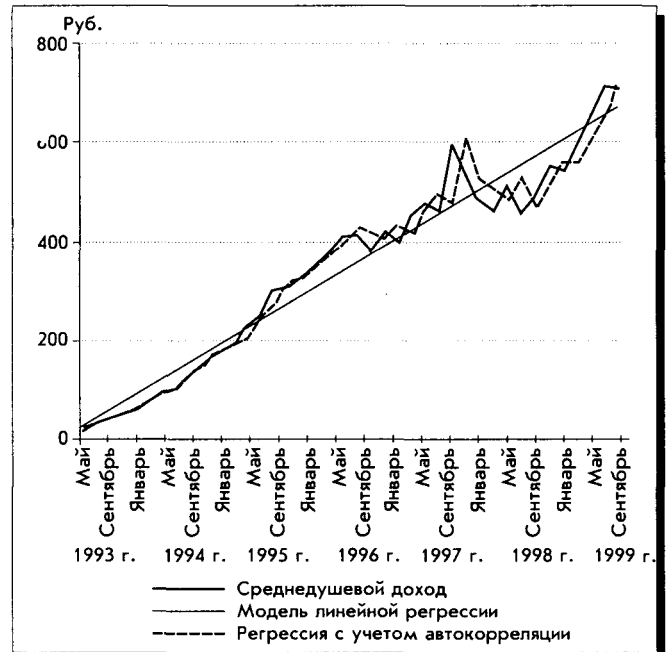


Рис. 5. График изменения среднедушевого дохода и различных моделей описания его изменений

том автокорреляции. Коэффициент качества (R^2) для последней модели равен 0,98. Казалось бы, он не сильно выше, чем для линейной модели, однако, как уже отмечалось, значению R^2 для линейной модели при наличии автокорреляции нельзя доверять. Действительно, на рис. 5 видно, насколько линейная модель хуже описывает реальные данные.

Обычно поиски аналитических зависимостей типа уравнения (3) преследуют две цели. Во-первых, это объяснение поведения изучаемой переменной (среднедушевого дохода) в зависимости от поведения независимой переменной (в данном случае — времени). В этом смысле уравнение (3) представляется весьма полезным, поскольку из него следует, что за период с 1993 г. значение среднедушевого дохода населения почти полностью определялось линейным ростом во времени и тем, насколько доход отклонялся от этого линейного роста в предыдущий момент времени.

Во-вторых, модели типа уравнения (3) используются для предсказания значений анализируемой переменной в будущем. По отношению к анализу данных временной динамики этот второй вариант использования уравнения (3) не кажется очень полезным. Все дело в том, что единицами анализа у нас являются данные по отдельным исследованиям, т.е. таких единиц в проводимом анализе всего 41 (по числу опросов). По этой причине точность получаемых предсказаний размера среднедушевого дохода оказывается невелика. 95%-ный доверительный интервал составляет приблизительно ± 70 руб. (около $\pm 10\%$ от величины среднедушевого дохода). С учетом этих замечаний можно спрогнозировать значения среднедушевого дохода на основании уравнения (2) для 2000 г.:

Месяц опроса	Значение среднедушевого дохода (руб.)
Январь	726
Март	739
Май	754
Июль	769
Сентябрь	784
Ноябрь	800

Значение индекса настроения

Месяц	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.	1998 г.	1999 г.	Средне-месячные значения индекса
Январь	0	-7,2	-25,7	-24,9	-16,2	-13,8	-22,5	-18,4
Март	-16,0	-10,1	-24,1	-23,1	-22,1	-10,7	-21,7	-18,3
Май	-1,0	-15,6	-10,6	-4,5	-12,7	-16,2	-11,2	-10,2
Июль	-8,07	-8,2	-11,0	-1,4	-3,0	-7,7	-17,0	-8,0
Сентябрь	-4,6	-4,2	-15,8	-4,6	-3,5	-44,1	-11,3	-12,6
Ноябрь	-5,0	-13,5	-10,3	-16,3	-6,1	-22,6	-4,2	-11,2
Среднегодовые значения индекса	-7,0	-9,8	-16,3	-12,4	-10,6	-19,2	-14,7	-13,0

Разумеется, предсказываемые значения вычисляются исходя из предположения, что тенденции, определяющие значение среднедушевого дохода, останутся теми же, что были за последние годы, а также что не произойдет каких-либо неожиданностей, похожих на август 1998 г.*

Сезонные колебания. Следующим фактором, который может определять колебания изучаемого показателя, является сезонный фактор. Для демонстрации сезонных колебаний возьмем показатель индекса настроения (табл. 1)**.

Данные табл. 1 показывают, что практически по всем годам минимальное значение индекса настроения приходится на январь—март. Далее значение возрастает, достигая максимума в июле, а затем постепенно убывает. При этом интересно, что среднее значение индекса для сентября ниже июльского исключительно за счет экстремально низкого его значения в сентябре 1998 г. В это время мы сталкиваемся, по всей видимости, с некими "форсмажорными" обстоятельствами, которые никак не связаны с общими закономерностями изменения индекса настроения. В связи с этим, если сосредоточиться именно на поиске закономерностей, можно попытаться заменить значение за сентябрь 1998 г. на среднее значение индекса в сентябре за другие годы.

В табл. 2 показаны средние значения индекса настроения по месяцам исследования с учетом этой корректировки.

На основании данных о значениях индекса настроения можно провести регрессионный анализ, чтобы найти некоторую закономерность, фиксирующую динамику индекса в виде функции. Наилучшей кривой регрессии, которую нам удалось найти, является следующая кубическая кривая:

$$Y = -5,9 - 0,56 T + 0,013 T^2 + 0,001 T^3,$$

где Y — значение индекса настроения;

T — месяц исследования, при условии, что март 1993 г. принят за 0.

Отметим, что коэффициент детерминации невелик и составляет 0,063. Действительно, на рис. 6 видно, что регрессионная кривая достаточно плохо описывает реальные данные.

На основании определенной регрессионной зависимости вычисляется *сезонный коэффициент*, который показывает, во сколько раз реальное значение анализируемого показателя больше (меньше) того, которое предсказыва-

* Мы не обсуждаем здесь вопрос о причинах расхождения данных о среднедушевом доходе, даваемых опросами "Мониторинга" и органами государственной статистики, и приводим предсказываемые возможные результаты опросов.

** *Левада Ю.* Индексы социальных настроений в "норме" и в кризисе // Мониторинг общественного мнения: Экономические и социальные перемены. 1998. № 6. С. 7.

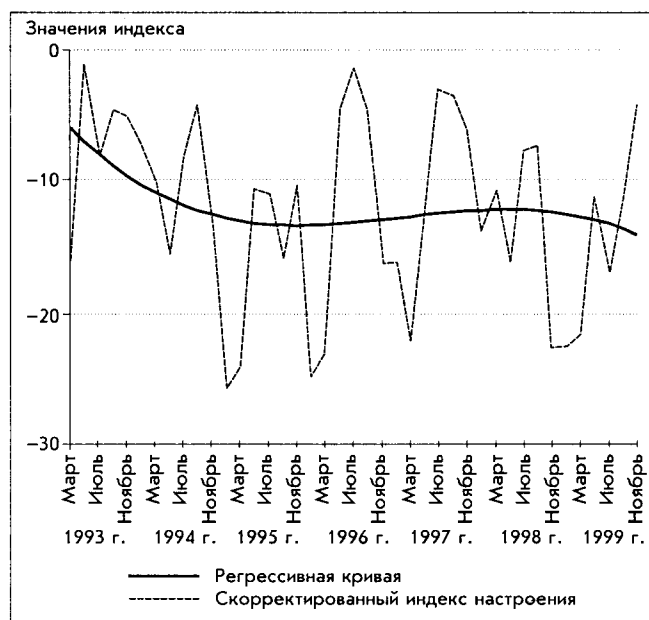


Рис. 6. Скоректированный индекс настроения и оптимальная регрессионная кривая

ется регрессионной кривой. Приведем значения сезонных коэффициентов для показателя настроения:

Месяц	Значение сезонного индекса (%)
Январь	145,60
Март	164,56
Май	83,41
Июль	68,25
Сентябрь	58,59
Ноябрь	89,25

Таблица 2

Средние значения индекса настроения за 1993–1999 гг. с учетом его корректировки в сентябре 1998 г.

Месяц	Среднее значение индекса	Стандартное отклонение
Январь	-18,4	7,28
Март	-18,3	5,94
Май	-10,2	5,61
Июль	-8,0	5,15
Сентябрь	-7,3	4,62
Ноябрь	-11,2	6,76

Представляется, что эти данные показывают наличие систематического смещения настроения в зависимости от месяца опроса. Однако стоит проверить наличие различий статистически. Наиболее адекватным для этой цели является дисперсионный анализ, который покажет наличие (или отсутствия) статистически значимых различий средних значений индекса по месяцам исследования. В нашем случае дисперсионный анализ показывает, что с вероятностью 90% статистически значимые различия есть.

В такой ситуации вполне естественно скорректировать наши данные с учетом сезонного смещения. Такая корректировка позволит искать детерминанты индекса настроения, отсеяв найденное сезонное смещение. Для проведения сезонного выравнивания каждое значение исходного ряда следует разделить на соответствующий индекс, а полученный результат умножить на 100. Результаты сезонного сглаживания представлены на рис. 7.

На рис. 7 визуально не видно, что после удаления сезонного компонента поведение индекса настроения стало более закономерным, однако, на самом деле, это не так. Значение коэффициента детерминации для регрессии увеличилось более чем вдвое и составило после проделанной корректировки 0,145.

Взаимосвязанные временные ряды. Под взаимосвязанными временными рядами понимают такие ряды, в которых уровни одного ряда в какой-либо степени определяют уровни другого ряда*. Далее мы рассмотрим то, насколько изменение значений среднедушевого дохода влияет на значение индекса "поддержки реформ".

Вопрос о потенциале поддержки реформ измеряется в "Мониторинге" следующим вопросом:

Как Вы считаете, рыночные реформы сейчас нужно продолжать или их следует прекратить?

1. Нужно продолжать.
2. Следует прекратить.
3. Затрудняюсь ответить.

Вопрос этот задавался начиная с марта 1994 г.

Важным представляется вопрос о детерминантах показателя поддержки реформ. Первым, вполне очевидным на первый взгляд показателем, определяющим поддержку/отрицание реформ, представляется доход респондента. Действительно, если взглянуть на значения среднего размера общего дохода в группах поддерживающих и не поддерживающих реформы, то разница в средних доходах составляет более 1300 руб. В табл. 3 приводятся значения среднего дохода в двух данных группах по исследованию ноября 1999 г.

Использование регрессионного анализа с зависимой переменной "доход" и независимой — "готовность поддерживать реформы" дает значение коэффициента детерминации R^2 (0,05), т.е. 5% информации о доходе объясняется поддержкой/неподдержкой реформ. Эта величина значима с вероятностью $P > 0,999$. Таким образом, представляется, что поддержка курса на продолжение реформ в определенной степени объясняется просто материальным положением респондента — чем выше доход, тем больше вероятность поддержки реформ. Однако более тщательный анализ показывает, что это не совсем так. На доход, равно как и на показатель поддержки/неподдержки реформ, влияют также и такие характеристики, как пол, возраст, образование, число членов семьи, место жительства (тип населенного пункта), регион. При элиминировании этих переменных разность в доходах между поддерживающими/не поддерживающими продолжение ре-

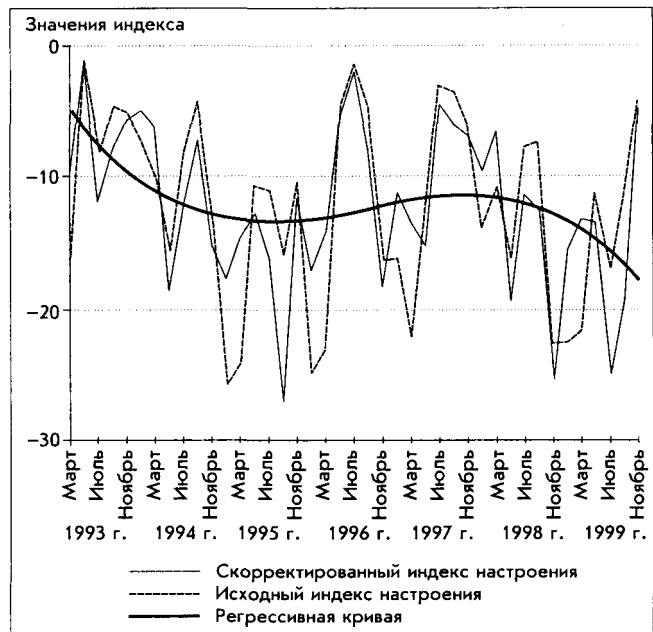


Рис. 7. Исходный и скорректированный по сезонным колебаниям индекс настроения

форм составляет около 200 руб. Для опроса за ноябрь 1999 г. разность остается значимой.

Мы проверили с этой точки зрения шесть исследований мониторинга за 1999 г. В четырех из них разность является значимой, а в двух — нет. Общий вывод, который можно сделать из этого, состоит в том, что на уровне рассмотрения отдельных респондентов есть, хотя и достаточно слабая (на пределе значимости), зависимость между уровнем дохода респондента и степенью поддержки курса на продолжение реформ. Рассмотрим теперь этот вопрос не на уровне респондента, а на уровне общества в целом. Есть ли взаимосвязь между поддержкой обществом курса реформ и размером дохода?

В упоминавшейся статье Ю.Левады для фиксации изменения значений показателя поддержки реформ предлагается использовать следующий индекс: разность между долей тех, кто дал первый вариант ответа, и долей тех, кто дал второй вариант ответа на вопрос. Материалы статьи показывают эффективность использования данного индекса, однако вполне очевидно, что в индексе никак не учитывается число затруднившихся ответить на вопрос (индекс не фиксирует долю населения — тех, кто поддерживает (не поддерживает) реформы. На рис. 8 показаны графики изменения индекса и доли респондентов, поддерживающих реформы, с 19 марта 1994 г. по ноябрь 1999 г.

Поведение двух рассматриваемых показателей на рис. 8 достаточно синхронно. Об этом свидетельствует и высокий коэффициент корреляции показателей (0,96). Таким образом, при такой высокой корреляции все равно, какой из показателей рассматривать. Однако уже на рис. 8 видно, что вариабельность индекса существенно больше, чем вариабельность просто доли респондентов, поддерживающих реформы. Об этом свидетельствуют и значения коэффициентов вариации (отношение стандартного отклонения к среднему арифметическому). Значения коэффициентов равны 1,85 и 0,15 соответственно. Таким образом, использование индекса гораздо предпочтительнее, поскольку его вариация гораздо выше, и, следовательно, он несет гораздо больше информации о динамике показателя поддержки реформ.

* Глинский В.В., Ионин В.Г. Статистический анализ. М.: Феликс, 1998. С. 183.

Средние значения показателя суммарного дохода семьи в группах поддерживающих и не поддерживающих продолжение реформ

Вариант ответа	Среднее значение суммарного дохода (руб.)	Число опрошенных	Стандартное отклонение
Рыночные реформы нужно продолжать	2677	661	3765
Рыночные реформы следует прекратить	1363	596	1205

Для изучения детерминант, определяющих изменение значений индекса, сначала необходимо выяснить, нет ли в его поведении сезонного фактора. Дисперсионный анализ показывает, что с вероятностью $P > 0,99$ значимых различий средних по месяцам исследования нет, следовательно, сезонный фактор отсутствует.

Отметим, что при попытке объяснить значения индекса поддержки реформ только как функцию от времени, получить значимые величины коэффициента детерминации не удается. Таким образом, здесь мы имеем дело с процессом, в котором и трендовая, и сезонная составляющие отсутствуют, но, быть может, изменение значения индекса поддержки реформ связано с изменением какого-то другого фактора? В качестве такого фактора исследуем показатель среднедушевого дохода.

Вычисление простого коэффициента корреляции показывает, что между значением индекса поддержки реформ и величиной среднедушевого дохода значимой корреляции нет ($r = -0,12$, $P = 0,48$). Следовательно, казалось бы, можно сделать вывод об отсутствии взаимосвязи между поддержкой реформ и величиной дохода.

Однако, как было показано ранее, изменение дохода во времени имеет сильную трендовую составляющую. Можно предположить, что изменение поддержки реформ связано не с наличием этой трендовой, т.е. в некотором смысле закономерной составляющей, а с отклонениями от нее. Иными словами, респондент меняет свое отношение к продолжению реформ не в том случае, если происходит как бы "запланированное" увеличение его (респондента) дохода, а в том случае, если этот "план" не выполняется. Таким образом, для проверки данной модели необходимо посмотреть на корреляцию индекса поддержки реформ и остатков от вычитания из значений среднедушевого дохода линейного компонента.

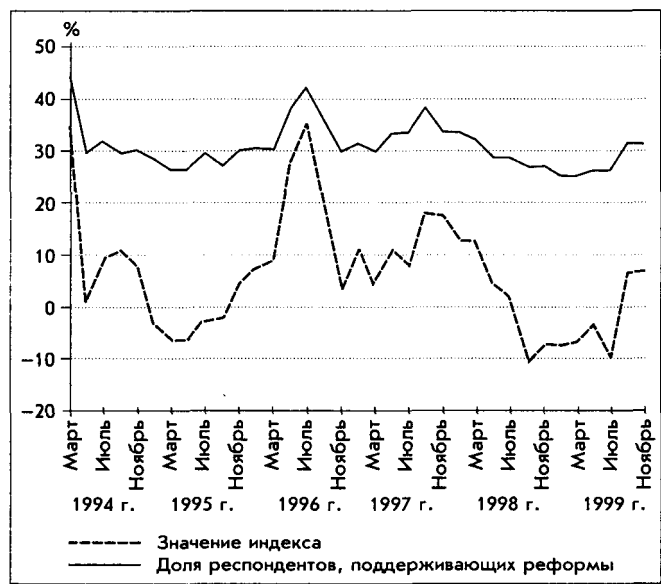


Рис. 8. Изменение значений индекса поддержки реформ и доли респондентов, считающих, что реформы следует продолжать

Действительно, коэффициент корреляции для этих двух показателей равен 0,5 и значим с вероятностью $P > 0,99$. Следовательно, увеличение среднедушевого дохода свыше ожидаемого линейного роста приводит к увеличению поддержки реформ, а его уменьшение — к уменьшению поддержки.

С выявлением факта такого рода связи встает и следующий вопрос. Насколько долговременны результаты воздействия превышения зарплаты над средним ростом на увеличение уровня поддержки реформ? Выявлению такого "последствия" служит взаимокорреляционная (кросс-корреляционная) функция. Данная функция представляет собой просто значения коэффициентов корреляции, при том, что один из временных рядов сдвигается по времени относительно другого. На рис. 9 показана взаимокорреляционная функция для индекса поддержки реформ и остатков для показателя среднедушевого дохода.

На рис. 9 видно, что значимые значения коэффициента корреляции сохраняются и при сдвиге на пять исследований! Иными словами, создается впечатление, что превышение роста дохода над уровнем ожиданий респондента сказывается на его поддержке реформ еще в течении почти года.

Этому факту может быть дано два существенно различных объяснения. Во-первых, может быть, действительно память респондента — достаточно сильный фактор и "нелинейное" изменение дохода долго помнится и оказывает влияние на поддержку (неподдержку) реформ еще длительное время. Другое объяснение кроется в ранее обнаруженном факте автокорреляции роста среднедушевого дохода. Иными словами, респондент не помнит прошлого воздействия роста дохода и реагирует только на сегодняшнее его увеличение (уменьшение). Одна-

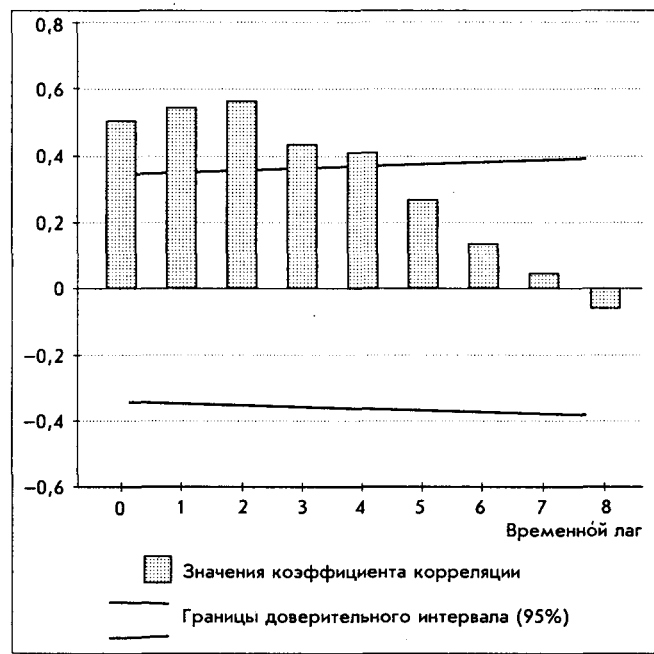


Рис. 9. Взаимокорреляционная функция индекса поддержки реформ и остатков показателя среднедушевого дохода

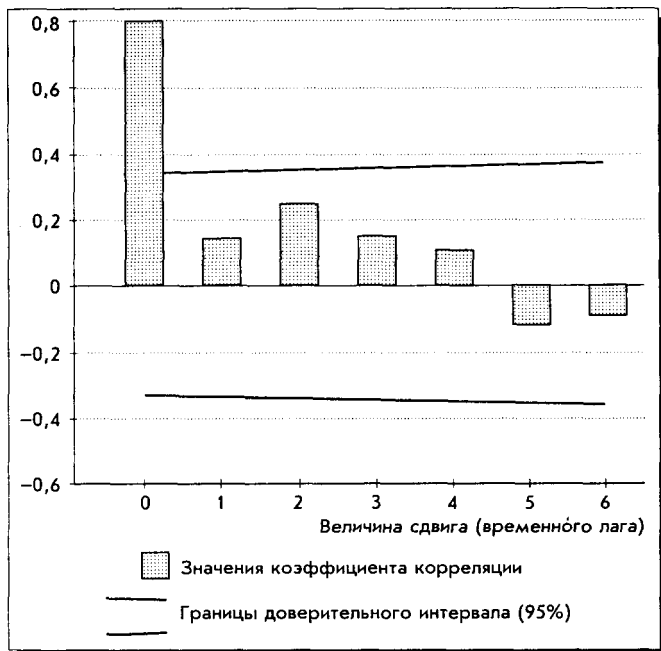


Рис. 10. Взаимокорреляционная функция для остатков от уравнения (3) и значения индекса поддержки реформ

ко поскольку у дохода есть автокорреляция, то как бы сам доход "помнит" свое увеличение (уменьшение) и в следующий период его (дохода) значение связано с предыдущим значением. Таким образом, связь уровня поддержки реформ с предыдущим значением дохода объясняется просто тем, что сам сегодняшний доход связан с предыдущим доходом, а наличие связи сегодняшнего дохода с сегодняшним уровнем поддержки реформ мы уже установили.

Важно отметить, что два предложенных объяснения не исключают друг друга. Может существовать как то, так и другое (и память респондента, и автокорреляция среднедушевого дохода). Как выяснить, какая из этих моделей (или обе одновременно) в реальности работает? Прежде всего, поскольку ранее мы установили наличие автокорреляции у изменения доходов, то фактически наличие второй модели уже зафиксировано (уравнение (3)).

Остается удалить из показателя роста дохода величину автокорреляционного влияния и посмотреть, как после этого будут взаимосвязаны величины дохода и индекса поддержки реформ. На рис. 10 показаны значения коэффициентов корреляции (взаимокорреляционная функция) для остатков от уравнения (3) и значений индекса поддержки реформ.

На рис. 10 убедительно показано, что после удаления автокорреляционной составляющей и тренда из показателя среднедушевого дохода остается только одна значимая корреляция — между сегодняшним уровнем поддержки реформ и сегодняшним отклонением дохода от среднего уровня (значение коэффициента корреляции — 0,8, $P > 0,99$). Таким образом, модель влияния прошлого повышения дохода на будущую поддержку реформ не подтверждается.

Заключение. В данной статье продемонстрированы возможности использования статистического аппарата анализа временных рядов для работы с данными "Мониторинга". Представляется, что даже рассмотренные пока простейшие модели дают интересные и подчас неожиданные результаты. В дальнейших материалах мы постараемся дать обзор более сложных методов и подходов к изучению временной динамики социальных процессов.

Владимир ЗВОНОВСКИЙ Голосуем списком.. Списком избирателей

В ходе последней предвыборной кампании в очередной раз поднимался вопрос о точности социологических прогнозов, основанных на опросах избирателей. Принимая всю ответственность за возможные ошибки такого прогнозирования на себя, отечественные полстеры старательно обходят вопрос об обоснованности самого критерия достоверности официальных результатов выборов. Будучи совершенно оправданным с этической точки зрения, уход от этой проблемы не кажется вполне адекватным реальной ситуации, которая сформировалась на сегодняшний день в политической сфере жизни нашего общества. Попытки исследований возможных нарушений, или, выразимся точнее, корректировок волеизъявления граждан, во время или после голосования делались как отечественными исследователями, так и зарубежными. Наибольшую известность приобрела работа С.Собянина и В.Суховольского "Демократия, ограниченная фальсификациями", вышедшая в 1995 г. В ней авторы применяют одну из статистических теорий к распределению голосов по официальным данным во многих регионах России и приходят к выводу о несоответствии этих данных теории.

Основной задачей при проведении опроса 33 тыс. человек на выходе из избирательных участков в день парламентских выборов, 19 декабря 1999 г., в 20 городах России было выявить характер влияния организаторов выборов на их результат. При этом обнаружилось поразительное расхождение с официальными результатами в ряде регионов страны (например, в Башкирии), что, по мнению наблюдателей, позволяет говорить о грандиозных фальсификациях при голосовании.

Автор данной работы не собирается уличать кого бы то ни было в фальсификациях. Он лишь ставит проблему и задает вопросы и, поскольку на эти вопросы ответов нет, предлагает при прогнозировании результатов голосования учитывать воздействия на эти результаты не только избирателей, но и организаторов выборов. Что касается, собственно, фальсификаций, автор полагает, что двумя основными их источниками являются: 1) готовность избирателей смириться с возможностями подтасовок и шире — их склонность перекладывать ответственность за результаты выборов с себя на внешние обстоятельства; 2) некомпетентность организаторов выборов, заключающаяся в том, что они не в состоянии выполнить одни требования законодательства и не нарушить другие. Они проводят выборы как умеют.

Так, анализ списков избирателей Самарского района г. Самары — самого малочисленного в городе — показал, что от 1,0 до 1,5% от числа всех зарегистрированных избирателей занесены в список дважды. Наиболее распространенные ошибки "двойного" учета: 1) вышедшая замуж женщина находится в списке и под своей новой, и под своей девичьей фамилией; 2) адрес избирателя, проживающего в так называемых угловых домах, записан в обеих формах (например, ул. Фрунзе, 25/27 и ул. Пионерская, 27/25). Таким образом, число избирателей искусственно завышено не менее, чем на 1%. Данный факт в большей степени свидетельствует о низкой квалификации сотрудников избирательных комиссий, чем об их злом умысле и целенаправленной фальсификации списков избирателей.

Одним из трудно решаемых вопросов, стоящих перед измерителями рейтингов, является активность избирателей. До сих пор прогнозируемая активность отклоняется от официальных результатов больше, чем любой другой параметр (доля голосующих за того или иного кандидата или партию). Чаще всего ожидаемая активность завышается. В научной литературе существует множество объ-