3.3. ПРИМЕНЕНИЕ МНОГОФАКТОРНЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

При моделировании социально-экономической динамики объективно приходится иметь дело с многофакторными зависимостями, когда значение показателя или группы показателей определяется поведением не одного, а сразу многих факторов. Действительно, если взять любой показатель социально-экономического развития и посмотреть, какие факторы оказывают на него влияние, то, пожалуй, и невозможно встретить хотя бы один из них, который не формировался бы под воздействием множества различных причин и условий. Например, цена товара – на её величину оказывает влияние объём предложения товара, объём спроса на товар, наличие конкурентов, наличие товаров-заменителей и цены на них и множество других факторов. Или численность занятых на производстве основных рабочих. Она, на первый взгляд, кажется, определяется только величиной производственного задания. Но на самом деле она зависит и от того, на каком оборудовании работает рабочий, какова организация труда на предприятии, насколько развит менеджмент, в конце концов - и от условий оплаты труда на нём.

Поэтому применительно к задаче прогнозирования социально-экономических явлений, пожалуй, труднее всего найти однофакторную зависимость, чем многофакторную. Очевидно, что однофакторная модель в смысловом отношении более точно отражает социально-экономические явления, нежели модель тренда. Но при построении однофакторной модели приходится всё же упрощать моделируемое представление действительности – вместо моделирования воздействия многих факторов на прогнозируемый показатель приходиться из этого множества отбирать только один фактор, который прогнозист посчитает наиболее существенным. Любая однофакторная модель в этой ситуации настолько условна и груба, что ее применение в прогнозировании может давать лишь очень приближенные ориентиры. Если экономист собирается выполнить точный прогноз с помощью экономико-математической модели, то он должен подобрать такую модель, которая бы лучше всего отражала суть происходящих процессов и описывала бы их. А так как практически все социально-экономические показатели формируются под воздействием множества факторов, то и модель, прогнозирующая их, также должна учитывать это – то есть, быть многофакторной.

Следовательно, от многофакторной прогнозной модели можно ожидать большей точности, чем от однофакторной модели, поскольку она вскрывает особенности и моделирует экономическую реальность более подробно.

Если обозначить прогнозируемый показатель через Y_{t}, а факторы, оказывающие влияние на него через x_{it}, то многофакторная модель будет иметь вид:

$$\hat{Y}_t = f(x_{it}). \qquad (3.3.1)$$

Многофакторные модели могут быть как линейными, так и нелинейными. Исходя из общенаучного принципа – от простого к сложному, рассмотрим вначале задачу построения линейной многофакторной модели:

$$\hat{Y}_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \ldots + a_{it} x_{it} + \ldots + a_{nt} x_{nt} . \qquad (3.3.2)$$

Для того чтобы построить такую модель, необходимо вначале отобрать факторы, которые оказывают наибольшее влияние на результирующий показатель, ведь общее количество факторов, влияющих на социально-экономические показатели очень велико. Продемонстрируем это. Чаще всего в экономической практике прогнозируются цены. Просто перечислим первую десятку факторов, которые оказывают влияние на цену, например, килограмма картофеля, который реализует один из продавцов, например, Сытного рынка Санкт-Петербурга:

1) цена, по которой был куплен картофель у поставщика;
2) стоимость аренды одного торгового места на рынке и различных сборов;
3) сорт картофеля;
4) качество клубней и их товарный вид;
5) цена, по которой продают картофель конкуренты;
6) общее количество конкурентов и объёмы их товара;
7) сезон года (осенью много других овощей, поэтому есть заменители картофеля);
8) объём спроса на картофель со стороны посетителей рынка;
9) маркетинговое сопровождение продаж (оформление павильона, вид ценника, вид и характеристики упаковочного материала, рекламные лозунги и т.п.);
10) день недели (в выходные дни посетителей больше).

Эти все факторы оказывают влияние на цену товара, но приведённый перечень далеко не исчерпывающий. Множество других факторов не вошли в число первых десяти, но это не значит, что их нет или что их влияние на цену менее значимо, чем влияние перечисленных факторов.

Поэтому, при формировании прогнозной многофакторной модели необходимо составить максимально широкий перечень факторов, оказывающих влияние на прогнозируемый показатель. После этого тщательное изучение сути происходящих процессов может помочь прогнозисту некоторую часть факторов убрать из рассмотрения постольку, поскольку они с позиций сущности социально-экономического процесса оказываются малозначимыми. Но даже и после такого профессионального анализа факторов остаётся довольно большая их часть, и модель с их использованием может быть очень громоздкой. Возникает необходимость отбора действительно наиболее значимых из них. На этом этапе рекомендуется формализовать процедуру, прибегнув к помощи корреляционного анализа.

В первом параграфе данной главы говорилось о том, что корреляционный анализ выступает лишь дополнительным инструментом исследования взаимосвязи между факторами, поскольку основной коэффициент корреляционного анализа – коэффициент парной корреляции, - характеризует степень приближения зависимости между двумя случайными факторами к линейной форме. Нелинейные зависимости он не оценивает. Но поскольку мы строим линейную многофакторную модель, то перед нами как раз и стоит задача оценки того, насколько взаимосвязь между факторами приближается к линейной. Для ответа на этот вопрос заполняют матрицу коэффициентов парной корреляции так, как это приводится в таблице 3.3.1.

Таблица 3.3.1.
Матрица коэффициентов парной корреляции

Признак	Y_t	x_{1t}	x_{2t}	…	x_{it}	…	x_{nt}
Y_t	1						
x_{1t}	r_{Yx1}	1					
x_{2t}	r_{Yx2}	r_{x2x1}	1				
…							
x_{it}	r_{Yxi}	r_{xix1}	r_{xix2}		1		
…							
x_{nt}	r_{Yxn}	r_{xnx1}	r_{xnx2}		r_{xixn}		1

Поскольку коэффициент парной корреляции симметричен, то есть: $r_{yx} = r_{xy}$, то и матрица табл. 3.3.1 симметрична относительно диагонали, значения которой, очевидно, равны единице. Поэтому заполняют либо верхнюю правую часть матрицы, либо нижнюю левую часть матрицы, как это сделано в табл. 3.3.1. Анализ этой матрицы даёт много дополнительной информации прогнозисту – как степень влияния каждого фактора на результирующий признак, так и взаимосвязь между факторами. Последнее особенно важно, поскольку наличие линейной взаимосвязи между факторами означает построение модели в условиях мультиколлинеарности, а это имеет очень неприятные последствия для вычислимости модели.

Так, если какой-либо из факторов имеет с другими факторами высокое значение коэффициентов парной корреляции, то это как раз и приведёт к вычислительным сложностям.

По значениям таблицы 3.3.1 сложно судить о том, насколько все приведённые факторы взаимосвязаны, или иначе говоря, какова множественная корреляция между факторами? Парная корреляция в таблице приведена, а вот что можно сказать о многофакторной зависимости в целом?

Для того, чтобы получить ответ на этот вопрос, вычисляют коэффициент множественной корреляции. Есть несколько способов его вычисления, отличающиеся друг от друга трудоёмкостью процесса. Укажем на самый простой из них[1].

Вначале, воспользовавшись данными табл. 3.3.1, находят определитель матрицы:

[1] Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2004. - С. 372.

$$\Delta^* = \begin{vmatrix} r_{Yx_1} & r_{Yx_2} & \dots & r_{Yx_k} & 1 \\ 1 & r_{x_2x_1} & \dots & r_{x_kx_1} & r_{Yx_1} \\ r_{x_1x_2} & 1 & \dots & r_{x_kx_2} & r_{Yx_2} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_1x_k} & r_{x_2x_k} & \dots & 1 & r_{Yx_k} \end{vmatrix}. \qquad (3.3.3)$$

Затем, убирая из матрицы все коэффициенты парной корреляции с переменной Y_t, вычисляют определитель другой матрицы:

$$\Delta = \begin{vmatrix} 1 & r_{x_2x_1} & \dots & r_{x_kx_1} \\ r_{x_1x_2} & 1 & \dots & r_{x_kx_2} \\ \dots & \dots & \dots & \dots \\ r_{x_1x_k} & r_{x_2x_k} & \dots & 1 \end{vmatrix}. \qquad (3.3.4)$$

Коэффициент множественной корреляции будет равен:

$$R = \pm\sqrt{\frac{\Delta^*}{\Delta}}. \qquad (3.3.5)$$

Для случая двухфакторной модели его рассчитать можно довольно просто через коэффициент детерминации:

$$R^2 = \frac{r_{Yx_1}^2 + r_{Yx_2}^2 - 2r_{Yx_1}r_{Yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}. \qquad (3.3.6)$$

Из вышеприведённого следует и логика подбора факторов модели, поскольку, включая или исключая тот или иной фактор из матрицы (3.3.3) и матрицы (3.3.4), и вычисляя коэффициент множественной корреляции (3.3.5), останавливаются на том наборе факторов, при котором коэффициент множественной корреляции принимает наибольшие значения.

При построении многофакторных моделей социально-экономической динамики уже на первом этапе – построении корреляционной матрицы (табл. 3.3.1), практически всегда встречаются с тем, что многие, а весьма часто и все коэффициенты парной корреляции этой матрицы близки по модулю к единице. Это явление получило название «мультиколлинеарности».

Мультиколлинеарность, как следует из самого названия проблемы, возникает тогда, когда факторы модели имеют одинаковые, монотонные относительно друг друга тенденции в динамике.

Попробуем разобраться в последствиях мультиколлинеарности для оценивания коэффициентов многофакторных моделей. Для этого воспользуемся статистическими данными таблицы 3.3.2, в которой приведены некоторые относительные показатели экономической динамики Ульяновской области[1]. Построим с помощью МНК многофакторную линейную модель. Она имеет вид:

$$Y_t = -1.8798 + 0.0756X_{1t} - 0.0896X_{2t} + 2.3768X_{3t} + 0.5135X_{4t}. \qquad (3.3.3)$$

[1] Светуньков С.Г., Благов А.А., Инешин К.А., Козлов М.А. Развитие Ульяновской области: проблемы моделирования : Ученые записки экономического факультета // Внутривузовский сборник фМГУ, 1991 г. – С.3.

Здесь Y_t - производство продукции,

X_{1t} - потребление электроэнергии,

X_{2t} - основные производственные фонды,

X_{3t} - численность занятых,

X_{4t} - фонд оплаты труда.

Если попытаться дать экономическое толкование полученным результатам, то мы вынуждены будем утверждать, что увеличение выпуска промышленной продукции было обусловлено положительным влиянием всех факторов за исключением стоимости основных производственных фондов, которая влияла отрицательно на рост производства. Отсюда следует, например, рекомендация для руководства области – для роста объёма регионального продукта необходимо уменьшать величину основных фондов предприятий. Очевидно, что полученный вывод, как и результат не имеют смысла, хотя данная модель хорошо описывает имеющийся ряд наблюдений.

Таблица 3.3.2.

Относительные значения показателей развития Ульяновской области

Производство продукции, Y_t	Потребление электроэнергии, X_{1t}	Основные производственные фонды, X_{2t}	Численность занятых, X_{3t}	Фонд оплаты труда, X_{4t}
1,0000	1,0000	1,0000	1,0000	1,0000
1,0410	0,9480	1,0994	1,0145	1,0374
1,0910	0,9800	1,2108	1,0323	1,0878
1,1390	1,1338	1,3346	1,0403	1,1267
1,1902	1,1380	1,4840	1,0565	1,1778
1,2438	1,1335	1,6759	1,0780	1,2276
1,3408	1,4320	1,8039	1,0968	1,3105
1,4293	1,5980	1,9623	1,1127	1,3978
1,4936	1,7681	2,0949	1,1109	1,5018
1,5220	2,1061	2,2011	1,1029	1,6133

Если теперь значения показателей табл.3.3.2 округлить не до четвертого знака после запятой, а до второго и вновь оценить параметры многофакторной модели по округленным значениям, она примет совсем иной вид:

$$Y_t = -1.2152 + 0.3878X_{1t} + 0.1098X_{2t} + 2.2309X_{3t} - 0.4784X_{4t}. \qquad (3.3.4)$$

Если сравнить полученные коэффициенты с коэффициентами модели (3.3.3), то можно увидеть, что свободные члены моделей отличаются друг от друга в 1,55 раза; коэффициенты при X_{1t} – в 5,1 раза; коэффициенты при X_{3t} – в 1,06 раза; а коэффициенты при переменных X_{2t} и X_{4t} вообще поменяли свои знаки на противоположные.

Если же на основе полученных теперь значений коэффициентов модели (3.3.4) попытаться провести анализ и дать соответствующие рекомендации, то из него будет следовать, что на рост производства в Ульяновской области все факторы оказывали положительное влияние за исключением оплаты труда. Поэтому, если вновь поверить результатам математических вычислений и давать рекомендации руководству Ульяновской области, то они будут такими –

перестать выплачивать заработную плату сотрудникам предприятий (X_{4t} =0), а по возможности даже брать с них плату за возможность выхода на работу (X_{4t} <0).

Выводы, которые следуют из анализа коэффициентов модели (3.3.4), как и выводы, следующие из предыдущей модели (3.3.3) оказываются абсурдными и противоречивыми. Но в данном случае мы понимаем абсурдность выводов, поскольку обратили внимание на неустойчивость модели к ошибкам округления и этим выводам доверять не будем, как не будем доверять ни одной из моделей. Ситуация, с которой мы столкнулись в этом примере является типичной ситуации мультиколлинеарности, когда полученные результаты оказались неустойчивыми к незначительным ошибкам округления и привели к бессмыслице. Для подтверждения того, что модель строилась в условиях мультиколлинеарности, необходимо вычислить значения корреляционной матрицы. Результаты расчета коэффициентов парной корреляции для факторов табл. 3.3.2 приведёны в табл. 3.3.3.

Таблица 3.3.3.
Матрица коэффициентов парной корреляции для многофакторной модели Ульяновской области

Признак	Y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}
Y_t	1				
x_{1t}	*0,947046*	1			
x_{2t}	*0,995281*	*0,9283*	1		
x_{3t}	*0,96455*	*0,8355*	*0,9736*		
x_{4t}	*0,987411*	*0,9774*	*0,9811*	*0,9194*	1

Все коэффициенты парной корреляции близки к единице, что подтверждает множественную коллинеарность и факторов, и результирующего показателя. Поэтому все те неприятности с оценкой коэффициентов модели, которые были продемонстрированы выше, как раз и вызваны наличием мультиколлинеарности.

Рассмотрим математическую постановку задачи оценки коэффициентов многофакторной модели с помощью МНК. Уравнение регрессии в матричной форме имеет вид:

$$\hat{Y} = XA\,. \qquad (3.3.5)$$

Здесь исходные факторные переменные представлены в форме матрицы:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{n1} \\ 1 & x_{12} & x_{22} & \dots & x_{n2} \\ 1 & x_{13} & x_{23} & \dots & x_{n3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1t} & x_{2t} & \dots & x_{nt} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1T} & x_{2T} & \dots & x_{nT} \end{bmatrix},$$

а коэффициенты представлены в виде вектора-столбца:

$$A = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix}.$$

Эта модель описывает реальные значения с некоторой ошибкой аппроксимации, что может быть в матричной форме выражено так:

$$Y = \hat{Y} + \varepsilon = XA + \varepsilon. \quad (3.3.6)$$

Сумма квадратов отклонений фактических отклонений от расчётных будет представлена так:

$$Q = \sum_t \varepsilon_t^2 = \varepsilon^T \varepsilon = \left(Y - XA\right)^T (Y - XA) = Y^T Y - 2A^T X^T Y + A^T X^T XA. \quad (3.3.7)$$

Дифференцируя по переменным задачи, то есть, по коэффициентам модели, и приравнивая нулю полученные значения, можно в итоге определить выражение для определения матрицы оценки коэффициентов многофакторной модели:

$$X^T Y = X^T XA. \quad (3.3.8)$$

Откуда легко получить:

$$A = X^T Y (X^T X)^{-1}. \quad (3.3.9)$$

Исходя из формальной постановки задачи, никаких проблем возникнуть не может – всё логично и легко вычисляется. Но в условиях мультиколлинеарности квадратная матрица $(X^T X)$ является слабо обусловленной, поэтому при вычислении обратной матрицы как раз и возникают вычислительные проблемы – определитель матрицы практически равен нулю, поэтому матрица является необратимой. В результате этого оценить коэффициенты модели не представляется возможным. Это - во-первых.

Если всё же вычислить определитель, используя современную технику и вычисляя его значения до двадцатого, например, знака после запятой, можно исчислить обратную матрицу и осуществить действия (3.3.9), но оценки параметров многофакторных моделей будут очень неточными из-за указанной выше причины.

В-третьих, так как оценки параметров оказываются неточными, то интерпретация влияния факторов на прогнозируемый показатель будет совершенно не той, которая есть на самом деле.

В-четвертых, оценки параметров модели оказываются крайне неустойчивыми - малейшее изменение исходных данных или даже ошибки округления, приводят к очень значительным изменениям параметров.

В-пятых, ценность таких моделей крайне низка, так как неустойчивая модель дает очень сильную вариацию своих коэффициентов, а значит и расчётных значений прогнозируемого показателя.

В-шестых, модель, с помощью которой сделана попытка описать сложное многофакторное явление, не описывает это явление. Причинно-следственные связи, выявленные на предыдущем этапе, не описываются этой моделью – чего только стоят отрицательные коэффициенты при факторах, оказывающих прямое положительное влияние на результат!

Именно поэтому построение многофакторных моделей в условиях мультиколлинеарности является предметом многолетнего внимания учёных. Учёные единодушны в том, что применение МНК к задаче оценивания коэффициентов многофакторных моделей приводит к необходимости решения системы нормальных уравнений, которая в условиях мультиколлинеарности является почти «вырожденной», а потому и неустойчивой к малейшим ошибкам округления. Поэтому, если исходить из такой постановки проблемы, понятны направления борьбы с проклятием «мультиколлинеарности» - либо устранять саму мультиколлинеарность, либо разрабатывать математические методы, устойчивые в этих условиях.

В первом направлении выделяются два подхода:

- исключают из совокупности факторов одну или несколько линейно связанных факторных переменных, чтобы вновь полученные элементы корреляционной матрицы были меньше порогового значения 0,8.

- преобразуют факторы в новые переменные, уменьшая тем самым количество переменных (факторный анализ).

Во втором направлении, совершенствовании математического аппарата оценивания коэффициентов модели используют методы регуляризации – раздел, называющийся «робастной статистикой», например, гребневой регрессии, стабилизированных оценок или «сжатых» оценок и т.п.,

Каждый из этих подходов обладает очень серьезными недостатками, которые снижают эффективность использования на практике многофакторных моделей.

В первом случае рекомендуется исключать из модели тот фактор, у которого значения парного коэффициента корреляции с другими факторами превышает по модулю «пороговое» значение, равное 0,8. Логика такого подхода базируется на том, что если этот фактор линейно связан с другими, то их влияние на результат отражает и влияние этого фактора, поэтому без него можно и обойтись. Но сама процедура исключения из совокупности факторов одной или нескольких факторных переменных весьма сомнительна - так можно последовательно прийти и к простой однофакторной линейной модели. И с экономических позиций эта процедура не может не вызвать недоумения - если специалист отобрал совокупность факторов, объясняющих динамику показателя, то уменьшение их числа только ухудшит свойства модели. А если вернуться к рассмотренному выше примеру с построением многофакторной модели динамики Ульяновской области, то анализ матрицы коэффициентов парной корреляции показывает, что все её значения превышают пороговое значение в 0,8, поэтому, следуя этой логике, надо отбросить все факторы, оставив в рассмотрении только один из них. Но ведь желание построить многофакторную модель вызвано вполне очевидным стремлением вскрыть сложную причинно-следственную связь социально-экономического явления и тем самым повысить точность прогноза, а вместо этого мы опять приходим к элементарной однофакторной модели.

Другой способ побороть мультиколлинеарность решается с помощью методов факторного анализа, когда пытаются заменить реальные переменные x_{it}

на другие факторы z_{jt} $(j>i)$, которые как бы являются скрытой причиной их динамики. Такие переменные неизвестны исследователю и называются латентными. С помощью математических преобразований (наиболее известен метод главных компонент) находят эти новые переменные и строят новую многофакторную модель зависимости переменной Y_t от новых факторов z_{jt}. При этом, однако, модель теряет какой-либо экономический смысл, так как латентные переменные в данной ситуации не существуют и являются математической абстракцией. И хотя модель становится более устойчивой к ошибкам округления, например, но для выполнения прогноза показателя Y_t необходимо вначале вычислить прогнозные значения каждой из этих новых переменных z_{jt} с помощью формул преобразования с учётом исходных переменных x_{it}. А поскольку исходные переменные наблюдаются с ошибками, то эти ошибки способствуют тому, что и новые переменные будут содержать в себе эти же ошибки. Но так как находится регрессионная зависимость между Y_t и z_{jt}, то и эта зависимость генерирует ошибку регрессии. Все эти пересчёты вносят дополнительную погрешность в прогнозирование результирующего признака и модель становится менее точна, чем хотелось бы.

Второе направление, связанное с совершенствованием математического аппарата оценивания коэффициентов, представляет собой некоторую попытку «утяжеления» задачи посредством намеренного введения «засоряющей» величины[1]. Здесь вполне допустима аналогия - если какая-либо конструкция неустойчива, то повысить ее устойчивость можно, загрузив конструкцию балластом. В случае многофакторного моделирования построенная таким образом модель, в результате такого утяжеления, будет не оптимальной, но вполне устойчивой, что уже следует признать положительным моментом. Этот путь решения проблемы мультиколлинеарности более приемлем, чем первые, однако, он требует от исследователя хорошей математической подготовки, так как аппарат методов регуляризации достаточно сложен. Тем не менее, квалифицированные специалисты предпочитают использовать именно эти методы робастной статистики.

Как это не кажется парадоксальным, но для случая нелинейной многофакторной модели эта проблема достаточно просто решается! Если прогнозист использует в качестве прогнозной модели нелинейную многофакторную модель, например, такую:

$$Y'_t = a_0 + a_1X_{1t}X_{2t} + a_2X_{1t}X_{2t}X_{3t} + a_3\,X^2_{3t},$$

то для нахождения параметров a_0, a_1, a_2 и a_3 осуществляют замену переменных и представляют модель в простой линейной форме:

$$Y'_t = a_0 + a_1X'_{1t} + a_2X'_{2t} + a_3X'_{3t},$$

где $X'_{1t} = X_{1t}X_{2t}$,

$X'_{2t} = X_{1t}X_{2t}X_{3t}$,

$X'_{3t} = X^2_{3t}$.

Для вновь полученной модели и новых переменных строят корреляционную матрицу (табл. 3.3.1) и если преобразованные новые

[1] Дадаян В., Бессараб С. Модель долгосрочного экономического роста // Российский экономический журнал. – 1992, № 4. – С 78.

переменные коррелируют друг с другом и коэффициенты парной корреляции выше по абсолютному значению, чем 0.8, то предлагается для уменьшения взаимной коррелированности переменных центрировать исходные переменные X_{it}. Действительно, корреляционная матрица новых переменных сильно меняет свои значения в сторону уменьшения коэффициентов парных корреляций. Это доказал еще в 1969 году Г.Ф.Филаретов[1]. Однако, для линейной многофакторной модели центрирование исходных данных (как и любые другие аффинные преобразования) не изменит коэффициентов корреляционной матрицы. Поэтому сложилась парадоксальная ситуация - сложные многофакторные нелинейные модели в условиях мультиколлинеарности можно построить достаточно просто, а элементарные линейные многофакторные модели в этих же условиях построить не удаётся!

Покажем, как можно получить удовлетворительные оценки МНК многофакторной модели в условиях мультиколлинеарности.

Прежде всего, запишем систему нормальных уравнений МНК для многофакторной линейной модели (3.3.2), для простоты записи опуская пределы суммирования, понимая, что оно ведётся для всех t, начиная с $t=1$, и заканчивая $t=T$:

$$\begin{cases} \sum Y_t = a_0 T + a_1 \sum x_{1t} + ... + a_n \sum x_{nt} \\ \sum Y_t x_{1t} = a_0 \sum x_{1t} + a_1 \sum x_{1t}^2 + ... + a_n \sum x_{nt} x_{1t} \\ ... \\ \sum Y_t x_{nt} = a_0 \sum x_{nt} + a_1 \sum x_{1t} x_{nt} + ... + a_n \sum x_{nt}^2 \end{cases} . \qquad (3.3.5)$$

Теперь представим эту систему уравнений в виде уравнений в отрезках наподобие того, как мы это делали в предыдущем параграфе. Получим:

$$\begin{cases} 1 = \dfrac{a_0}{\dfrac{\sum Y_t}{T}} + \dfrac{a_1}{\dfrac{\sum Y_t}{\sum x_{1t}}} + ... + \dfrac{a_n}{\dfrac{\sum Y_t}{\sum x_{nt}}} \\ 1 = \dfrac{a_0}{\dfrac{\sum Y_t x_{1t}}{\sum x_{1t}}} + \dfrac{a_1}{\dfrac{\sum Y_t x_{1t}}{\sum x_{1t}^2}} + ... + \dfrac{a_n}{\dfrac{\sum Y_t x_{1t}}{\sum x_{nt} x_{1t}}} \\ ... \\ 1 = \dfrac{a_0}{\dfrac{\sum Y_t x_{nt}}{\sum x_{nt}}} + \dfrac{a_1}{\dfrac{\sum Y_t x_{nt}}{\sum x_{1t} x_{nt}}} + ... + \dfrac{a_n}{\dfrac{\sum Y_t x_{nt}}{\sum x_{nt}^2}} \end{cases} . \qquad (3.3.6)$$

Уравнения системы МНК здесь представлены в виде уравнений гиперплоскостей в отрезках в гиперпространстве коэффициентов модели. Если в однофакторном случае оценки МНК представляют собой точку пересечения на плоскости параметров двух прямых условий МНК, поскольку неизвестных параметров всего два – a_0 и a_1, и задачу можно изобразить на плоскости, то уже при числе факторов, равном двум, число коэффициентов модели будет равно

[1] Филаретов Г.Ф. К вопросу о построении нелинейной регрессионной модели по данным пассивного эксперимента // Проблемы планирования эксперимента / Под ред. Г.К.Круга. – М.: Наука, 1969.

трем – a_0, a_1 и a_2. Графически такую задачу оценивания параметров многофакторной модели следует рассмотреть не на плоскости, а в трехмерном пространстве параметров. Действительно, число неизвестных параметров становится равным трем и их можно изобразить в качестве осей трехмерного пространства $0a_0$, $0a_1$ и $0a_2$. В этом случае условия МНК представляют собой систему из трех уравнений с тремя неизвестными, причем каждое из уравнений представляет собой не что иное, как уравнение плоскости в пространстве. Решение системы МНК в данном случае будет представлять собой точку пересечения трех плоскостей в пространстве. Координаты этой точки дают искомые с помощью МНК значения коэффициентов модели.

Воспользуемся, уравнениями в отрезках (3.3.6) для того, чтобы более тщательно изучить причины неустойчивости оценок параметров в условиях мультиколлинеарности на примере Ульяновской области, рассмотренном выше. Используя систему МНК для абсолютных значений показателей развития области, приведем эту систему к форме уравнений в отрезках. Такое преобразование системы покажет нам отрезки, которые отсекают гиперплоскости условий многофакторного МНК на осях гиперпространства параметров:

$$\begin{cases} 1 = \dfrac{a_0}{0,2278} + \dfrac{a_1}{0,433} + \dfrac{a_2}{0,853} + \dfrac{a_3}{0,1379} + \dfrac{a_4}{0,301} \\ 1 = \dfrac{a_0}{0,2195} + \dfrac{a_1}{0,449} + \dfrac{a_2}{0,876} + \dfrac{a_3}{0,1339} + \dfrac{a_4}{0,303} \\ 1 = \dfrac{a_0}{0,2198} + \dfrac{a_1}{0,445} + \dfrac{a_2}{0,876} + \dfrac{a_3}{0,1343} + \dfrac{a_4}{0,302} \\ 1 = \dfrac{a_0}{0,2266} + \dfrac{a_1}{0,434} + \dfrac{a_2}{0,856} + \dfrac{a_3}{0,1373} + \dfrac{a_4}{0,302} \\ 1 = \dfrac{a_0}{0,2229} + \dfrac{a_1}{0,441} + \dfrac{a_2}{0,867} + \dfrac{a_3}{0,1356} + \dfrac{a_4}{0,302} \end{cases}. \qquad (3.3.7)$$

Из анализа величин полученных отрезков легко убедиться в том, что отрезки, отсекаемые гиперплоскостями условий МНК на каждой из осей гиперпространства коэффициентов, практически совпадают друг с другом, а иногда и вообще совпадают, т.е. сами гиперплоскости почти параллельны друг другу. Очевидно, поэтому, что решение системы МНК, которое представляет собой точку пересечения этих практически параллельных друг другу гиперплоскостей в гиперпространстве, является чрезвычайно неустойчивым - малейшая ошибка в округлении может привести к тому, что гиперплоскости могут, переместясь незначительно, иметь новую точку их пересечения, значительно удаленную от первоначальной. Например, отрезок на оси a_4 первой гиперплоскости величиной 0.301 может стать равным 0.303 (в результате округления, например). При этом первая гиперплоскость практически незаметно изменит свое положение в пространстве, но решение системы МНК, как точка пересечения гиперплоскостей, меняется так, что искажается не только абсолютная величина коэффициентов многофакторной модели, но и сам знак этих коэффициентов, что и было продемонстрировано выше.

Следовательно, последствия мультиколлинеарности действительно вызваны неприемлемостью существующего алгоритма оценивания параметров многофакторных моделей. Исследования в этом случае должны быть направлены не на борьбу с объективно существующей реальностью - сильной коллинеарностью практически всех показателей и факторов социально-экономической динамики (что подтверждается явлением ложной корреляции), а на улучшение используемого аппарата оценивания таких моделей. Этот вывод имеет место и при построении простых линейных однофакторных моделей, правда неустойчивость полученных оценок, в отличие от многофакторного случая, не особенно заметна - в последнем случае ситуация усугубляется еще и вычислительными сложностями, поэтому и становится такой явной.

Для повышения устойчивости оценок параметров многофакторных моделей необходимо «развести» гиперплоскости системы нормальных уравнений МНК и устранить тем самым их практическую параллельность. Следовательно, надо сделать тек, чтобы отрезки на осях гиперпространств оценок параметров моделей были в максимально возможной степени отличны друг от друга, а не совпадать до такой степени, как это происходит всегда в условиях мультиколлинеарности. Лучший вариант – предложить такой вариант получения оценок МНК, при котором гиперплоскости будут перпендикулярны друг другу – это даст возможность получить устойчивые оценки коэффициентов модели[1].

Для того чтобы решить эту задачу, вспомним формулу определения угла между двумя плоскостями. Если одна плоскость записывается уравнением

$$Ax+By+Cz+D=0, \tag{3.3.8}$$

а вторая плоскость уравнением

$$A'x+B'y+C'z+D'=0, \tag{3.3.9}$$

то косинус угла между ними можно определить по формуле[2]:

$$\cos\gamma=\frac{AA'+BB'+CC''}{\sqrt{A^2+B^2+C^2}\sqrt{A'^2+B'^2+C'^2}}\,. \tag{3.3.10}$$

Если плоскости почти параллельны, то угол между ними близок к нулю и косинус угла, соответственно, близок к единице. Если же плоскости перпендикулярны друг другу, то угол между ними равен 90 градусов и косинус равен нулю.

Применим эту формулу для гиперплоскостей системы МНК (3.3.5). Косинус угла между первым и вторым уравнением будет определяться так:

$$\cos\gamma_{12}=\frac{T\sum x_{1t}+\sum x_{1t}\sum x_{1t}^2+\ldots+\sum x_{nt}\sum x_{1t}x_{nt}}{\sqrt{T^2+(\sum x_{1t})^2+\ldots+(\sum x_{nt})^2}\sqrt{(\sum x_{1t})^2+(\sum x_{1t}^2)^2+\ldots+(\sum x_{1t}x_{nt})^2}}\,. \tag{3.3.11}$$

Близок к нулю он может быть при двух условиях – когда числитель близок к нулю и когда знаменатель стремится к бесконечности.

---

[1] Светуньков С.Г.. Моделирование в условиях мультиколлинеарности // Промышленная энергетика, 1994, № 6. – С. 28 - 32

[2] Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука, 1984. - С. 85.

Если второе условие трудно выполнимо, то первое условие выполняется легко. Для этого следует произвести центрирование исходных переменных относительно их средней арифметической. Следует напомнить, что операция центрирования представляет собой такое преобразование исходного ряда *$\{x_t\}$*, при котором из каждого значения ряда вычитается средняя арифметическая этого ряда:

$$x_t^{'} = x_t - \bar{x} \,. \qquad (3.3.12)$$

Как следует из основ теории математической статистики сумма такого центрированного ряда будет равна нулю. Поэтому в формуле (3.3.11) будут равны нулю такие суммы:

$$\sum x_{1t} = 0, \ \sum x_{2t} = 0, \ \dots, \sum x_{nt} = 0. \qquad (3.3.13)$$

А так как все они являются сомножителями каждого слагаемого числителя (3.3.11), то числитель будет равен нулю, что означает перпендикулярность этих двух гиперплоскостей. В свою очередь это говорит о том, что линия пересечения этих гиперплоскостей будет определяться очень устойчиво.

Легко убедиться в том, что при условии (3.3.12) для всех факторных переменных все углы между первой гиперплоскостью и другими гиперплоскостями гиперпространства искомых коэффициентов, определяемых системой нормальных уравнений МНК, будут прямыми и первая гиперплоскость будет находиться по отношению к остальным гиперплоскостям плоскости перпендикулярно. Косинусы углов между другими гиперплоскостями системы (3.3.5), например, второй и третьей, не будут равны нулю, но они будут меньше, чем в случае использования не центрированных переменных, а это свидетельствует о том, что гиперплоскости пересекают друг друга под более тупым углом, чем прежде и точка пересечения всех гиперплоскостей будет устойчива к ошибкам округления и поступлению новой информации.

Необходимо отметить, что центрирование исходных переменных подразумевает построение не модели (3.3.2), а её модификации:

$$\hat{Y}_t - \bar{y} = a_1(x_{1t} - \bar{x}_1) + a_2(x_{2t} - \bar{x}_2) + \dots + a_{it}(x_{it} - \bar{x}_i) + \dots + a_{nt}(x_{nt} - \bar{x}_n)\,, \qquad (3.3.14)$$

в которой свободный коэффициент равен нулю.

Для того чтобы использовать исходную модель (3.3.2), необходимо найти величину этого коэффициента. Она определяется довольно просто по известным значениям коэффициентов регрессии. Для этого в первое уравнение системы МНК (3.3.5) подставляют исходные, а не центрированные переменные, в результате чего легко определяется искомое значение коэффициента:

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 - \dots - a_{it}\bar{x}_i - \dots - a_{nt} - \bar{x}_n\,, \qquad (3.3.14)$$

С учетом того, что и для нелинейных многофакторных моделей именно эта процедура центрирования является наиболее эффективной, следует отметить, что *предварительное центрирование данных при построении многофакторных моделей следует обязательно выполнять. Построенные при этом модели будут более устойчивыми.* При этом вовсе не обязательно проверять наличие или отсутствие мультиколлинеарности.

Продемонстрируем эффективность предлагаемого подхода на данных примера по Ульяновской области, приведенного ранее. Для этого следует лишь центрировать исходные данные, что и было сделано. Если теперь строить многофакторную линейную модель с помощью центрированных данных Y'_t и X'_{it}, получим следующую многофакторную модель:

$$Y'_t = 0.15633\ X'_{1t}+ 0.28614\ X'_{2t}+ 1.01790\ X'_{3t} - 0.14803\ X'_{4t}.$$

Если теперь для построения многофакторной модели воспользоваться не исходными данными, а округленными в ней значениями до второго знака после запятой (как мы делали ранее без центрирования), получим такую модель:

$$Y'_t = 0.15634\ X'_{1t} + 0.28684\ X'_{2t} + 1.01289\ X'_{3t} - 0.14855\ X'_{4t}.$$

Как видно из сравнения коэффициентов двух последних многофакторных моделей их оценки оказались достаточно устойчивыми в условиях мультиколлинеарности и изменяются лишь в четвертом или пятом знаках после запятых, то есть – значительно менее одного процента, в то время как до центрирования коэффициенты модели менялись не на проценты, а на несколько сотен процентов, в том числе, и меняя свои знаки.

Таким образом, для построения любых многофакторных моделей в условиях мультиколлинеарности следует осуществлять предварительное центрирование исходных данных, а затем использовать любой из множества методов оценивания параметров эконометрических моделей, в том числе и МНК. Тогда модели будут устойчивыми.

Но устойчивость модели, которую удаётся так просто достичь, ещё не означает того, что модель действительно описала вскрытые причинно-следственные связи. Действительно, как видно из полученных результатов для Ульяновской области, модель является устойчивой, но отрицательный коэффициент при четвёртой переменной (фонд оплаты труда) не может не смущать – вряд ли снижение оплаты труда работающих в этой области приведёт к росту регионального продукта!

К сожалению, при прогнозировании социально-экономической динамики прогнозисту приходится иметь дело с очень короткими по статистическим меркам рядами. Например, для прогнозирования динамики экономики России в 2008 году пригодными являются данные об этой динамке лишь с 1999 года, когда после знаменитого августовского дефолта 1998 года экономика России начала переживать новый этап развития. Поэтому в распоряжении прогнозиста могут быть ряды длиной в десять наблюдений и строить на таких коротких рядах устойчивые многофакторные модели, с числом факторов, например, равным пяти, статистически не имеет смысла – уж очень велико становится влияние случайных ошибок, которыми засорены исходные статистические данные. Но как же быть в этом случае, когда потребность в построении многофакторных моделей есть? Здесь может быть пригоден способ синтеза однофакторных моделей в многофакторную.

Суть его заключается в следующем.

Строится система однофакторных моделей, которые с разной степенью точности аппроксимируют результирующий показатель:

$$\begin{cases} Y_t = f_1(x_{1t}) + \varepsilon_{1t}, \\ Y_t = f_2(x_{2t}) + \varepsilon_{2t}, \\ Y_t = f_3(x_{3t}) + \varepsilon_{3t}, \\ \ldots \\ Y_t = f_n(x_{nt}) + \varepsilon_{nt}. \end{cases} \tag{3.3.14}$$

Левые части равенств равны друг другу, как равны друг другу правые части равенств этой системы. Сложим левые и правые части равенств. Получим:

$$nY_t = f_1(x_{1t}) + \varepsilon_{1t} + f_2(x_{2t}) + \varepsilon_{2t} + f_3(x_{3t}) + \varepsilon_{3t} + \ldots + f_n(x_{nt}) + \varepsilon_{nt}. \tag{3.3.15}$$

Группируя и разделив левую и правую части равенства на число n, получим:

$$Y_t = \frac{1}{n}(f_1(x_{1t}) + f_2(x_{2t}) + f_3(x_{3t}) + \ldots + f_n(x_{nt})) + \frac{1}{n}(\varepsilon_{1t} + \varepsilon_{2t} + \varepsilon_{3t} + \ldots + \varepsilon_{nt}). \tag{3.3.16}$$

То есть, результативный показатель описывается с помощью многофакторной модели такого вида:

$$\begin{gathered} \hat{Y}_t = \frac{1}{n}(f_1(x_{1t}) + f_2(x_{2t}) + f_3(x_{3t}) + \ldots + f_n(x_{nt})) = \\ \frac{1}{n}f_1(x_{1t}) + \frac{1}{n}f_2(x_{2t}) + \frac{1}{n}f_3(x_{3t}) + \ldots + \frac{1}{n}f_n(x_{nt}). \end{gathered} \tag{3.3.17}$$

ошибка аппроксимации которой равна:

$$\varepsilon = \frac{1}{n}(\varepsilon_{1t} + \varepsilon_{2t} + \varepsilon_{3t} + \ldots + \varepsilon_{nt}) = \frac{1}{n}\varepsilon_{1t} + \frac{1}{n}\varepsilon_{2t} + \frac{1}{n}\varepsilon_{3t} + \ldots + \frac{1}{n}\varepsilon_{nt}. \tag{3.3.18}$$

Многофакторная модель (3.3.17) формируется, исходя из предположения, что вклад каждого фактора, а, значит, и каждой однофакторной модели в общую дисперсию одинаков и равен *1/n*. Но если перед прогнозистом стоит задача построить многофакторную модель на основе синтеза однофакторных так, чтобы она лучше всего описывала прошлые значения, то естественно положить такое правило: модели с меньшей дисперсией учитываются в большей степени, а модели с большей дисперсией, соответственно – в меньшей степени.

Тогда возникает необходимость вычисления весовых коэффициентов v_i, которые следует задавать каждой однофакторной модели, но делать это необходимо так, чтобы сумма этих весовых коэффициентов была равна единице:

$$\sum_{i=1}^{n} v_i = 1. \tag{3.3.19}$$

Учитывая весовые коэффициенты, модель (3.3.17) может быть представлена так:

$$\hat{Y}_t = v_1 f_1(x_{1t}) + v_2 f_2(x_{2t}) + v_3 f_3(x_{3t}) + \ldots + v_n f(x_{nt}). \tag{3.3.20}$$

Остаётся только найти эти весовые коэффициенты. Одна из наиболее простых формул для расчета этих коэффициентов имеет такой вид:

$$v_i = \frac{1}{n-1}\frac{\sigma^2 - \sigma_i^2}{\sigma^2}. \tag{3.3.21}$$

Здесь σ^2_i - дисперсия i-й однофакторной модели, с помощью которых вычисляется:

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 \tag{3.3.22}$$

К числу вполне очевидных преимуществ этого метода построения многофакторных моделей следует отнести и возможность учёта степени влияния каждого из факторов на результат. Весьма часто точность аппроксимации результирующего показателя с помощью какой-либо однофакторной модели фактора i вовсе не означает, что этот фактор является важнейшим для объяснения динамики показателя. Это говорит только о том, что прогнозисту удалось подобрать хорошую аппроксимационную модель и только. Экономический анализ сути происходящих процессов позволяет экономисту утверждать, что влияние одних факторов весомее, чем других и поэтому формировать формальную многофакторную модель с помощью весов (3.3.21) будет не верно. В этом случае веса каждого фактора задаются экспертным путём с учётом условия (3.3.19) и однофакторные модели синтезируются в многофакторную (3.3.20) с помощью этих весов.

Прогнозист самостоятельно должен выбрать метод построения многофакторной модели – либо используя многофакторный МНК, осуществив предварительное центрирование исходных переменных, либо построить однофакторные модели и синтезировать их в однофакторную.

К сожалению, ни тот, ни другой метод не позволяют утверждать об успешном решении поставленной задачи – построении многофакторной зависимости, которая лучше, чем однофакторная описывает прогнозируемый процесс, а, значит, и будет давать более точные прогнозы, чем однофакторная модель. Покажем это на простом примере.

Пусть мы наблюдаем некоторый социально-экономический процесс, который описывается строго функциональной двухфакторной моделью:

$$Y_t = 2 - 0{,}5x_{1t} + 3x_{2t}. \tag{3.3.22}$$

Факторы x_{1t} и x_{2t} меняются линейно так, как это показано в табл. 3.3.4. В этой же таблице рассчитаны, в соответствии с приведённой формулой, величины результативного показателя Y_t.

Таблица 3.3.4

Данные условного примера

t	x_{1t}	x_{2t}	*Расчётное значение Y_t, полученное по (3.3.22)*
1	1	7	22,5
2	2	8	25
3	3	9	27,5
4	4	10	30
5	5	11	32,5
6	6	12	35

Теперь поставим обратную задачу, а именно - по данным табл.3.3.4, не зная то, что между переменными имеется зависимость (3.3.22), построить многофакторную модель. Центрирование исходных переменных и применение МНК позволяют получить такую систему нормальных уравнений:

$$\begin{cases} 43,75 = a_1 17,5 + a_2 17,5 \\ 43,75 = a_1 17,5 + a_2 17,5 \end{cases}. \qquad (3.3.23)$$

Как видно, система уравнений вырождена и не имеет решений. То есть, многофакторный МНК не позволяет решить поставленную задачу.

Попробуем теперь воспользоваться синтезом однофакторных моделей в многофакторную. Получим с помощью МНК две модели:

$$\hat{Y}_t = 2,5x_{1t} + 20,$$
$$\hat{Y}_t = 2,5x_{2t} + 5. \qquad (3.3.24)$$

Поскольку каждая модель описывает результирующий признак с нулевой дисперсией, то их синтез в однофакторную осуществляется с весовым коэффициентом 0,5 и позволяет получить такую модель:

$$\hat{Y}_t = 1,25x_{1t} + 1,25x_{2t} + 12,5\ . \qquad (3.3.25)$$

Если сравнить полученную модель с исходной (3.3.22), можно убедиться в том, что получена другая модель, хотя данная описывает результирующий признак с нулевой дисперсией.

Очевидно, что приведённый пример строго функциональной зависимости в реальной практике социально-экономического прогнозирования не встречается, но он наглядно демонстрирует, что имеющийся математический аппарат построения многофакторных моделей ещё далёк от совершенства.

Другой пример, который показывает сложность самой задачи многофакторного моделирования. Чаще всего в практике социально-экономического прогнозирования просто невозможно выявить и включить в модель все факторы, оказывающие влияние на результирующий признак, а это может привести к построению не верной модели. Вновь вернёмся к модели (3.3.22) и добавим в неё ещё один фактор:

$$Y_t = 2 - 0,5x_{1t} + 3x_{2t} + x_{3tt}. \qquad (3.3.26)$$

Добавим этот фактор и в таблицу исходных данных так, как это показано в табл.3.3.5.

Таблица 3.3.5

Данные условного примера

t	x_{1t}	x_{2t}	x_{3t}	*Расчётное значение Y_t, полученное по (3.3.22)*
1	1	7	2	24,5
2	2	8	4	29
3	3	9	3	30,5
4	4	10	5	35
5	5	11	7	39,5
6	6	12	4	39

Пусть теперь, для приближения задачи к реальной ситуации, будем считать, что прогнозисту не известно о влиянии второго фактора или же, что он заметил, что между первым и вторым фактором имеется линейная функциональная зависимость (коэффициент парной корреляции равен единице) и исключил один из факторов из рассмотрения. Тогда он захочет построить двухфакторную линейную модель вида:

$$\hat{Y}_t = a_0 + a_1 x_{1t} + a_3 x_{3t}\,. \qquad (3.3.27)$$

Предварительно осуществив центрирование исходных переменных, и решая полученную систему нормальных уравнений, можно определить оценки МНК и получить модель такого вида:

$$\hat{Y}_t = 20 + 2{,}5 x_{1t} + x_{3t}\,. \qquad (3.3.28)$$

Здесь, как легко убедиться, коэффициент перед первым фактором положителен, хотя влияние этого фактора, как следует из исходной модели (3.3.26), отрицательно – увеличение фактора x_{1t} в анализируемом процессе приводит к уменьшению результирующего признака, а в полученной с помощью МНК модели – моделируется противоположный результат. Очевидно, что эта модель ошибочна.

Не особенно меняет положения и попытка построить многофакторную модель с помощью синтеза однофакторных моделей.

Однофакторная модель зависимости результативного показателя от первого фактора, построенная с помощью МНК, имеет такой вид:

$$\hat{Y}_t = 3{,}1 x_{1t} + 22{,}1 \qquad (3.3.29)$$

Однофакторная модель зависимости результативного показателя от третьего фактора, также построенная с помощью МНК, имеет вид:

$$\hat{Y}_t = 2{,}77 x_{3t} + 21{,}4 \qquad (3.3.30)$$

Простая синтезированная многофакторная модель (без учёта влияния различия в дисперсиях), будет иметь вид:

$$\hat{Y}_t = 21{,}8 + 1{,}55 x_{1t} + 1{,}385 x_{3t} \qquad (3.3.31)$$

При этом опять первый фактор в модель включается с положительным коэффициентом.

Теперь, в завершение данного параграфа необходимо сказать несколько слов о попытке экономической интерпретации коэффициентов многофакторных моделей. Только что мы видели, что с учётом того, что все факторы выявить и включить в модель не удаётся, а если даже это и произошло, то вычислительные процедуры не позволяют построить абсолютно адекватную многофакторную модель. Любая многофакторная модель будет являться лишь некоторым в той или иной степени удачным приближением к описанию истинных процессов. Поэтому, получив отрицательное значение коэффициента при каком-то факторе ни в коем случае нельзя утверждать, что этот фактор отрицательно влияет на результирующий показатель. Точно также, нельзя утверждать и о положительном влиянии фактора на результат, если коэффициент при нём получился положительным. Судить о направлении и степени влиянии факторов на результирующий показатель по коэффициентам многофакторной модели также нелепо, как судить о творчестве Энрико Карузо по тому, как его изображает ваш сосед по лестничной клетке.

Многофакторные модели представляют собой лишь некоторое приближение к описанию сложных социально-экономических процессов, поэтому к результатам прогнозов с их помощью необходимо относится очень осторожно.