

## Вопросы к экзамену по дисциплине «ЭКОНОМЕТРИКА»:

1. Этапы построения эконометрических моделей
2. Построение парной линейной регрессии методом наименьших квадратов
3. Парная нелинейная регрессия. Оценка параметров
4. Построение линейной регрессии в MS EXCEL. Входные и выходные параметры функции ЛИНЕЙН
5. Оценка существенности (значимости) параметров уравнения регрессии
6. Интервалы прогноза по линейному уравнению регрессии. Построение доверительных интервалов
7. Множественная регрессия. Отбор факторов при построении множественной регрессии
8. Матрица парных корреляций. Мультиколлинеарность
9. Оценка параметров уравнения множественной регрессии
10. Построение производственной функции Кобба-Дугласа в MS EXCEL
11. Уравнение множественной регрессии в стандартизованном масштабе. Оценка коэффициентов  $B_i$
12. Переход от уравнения множественной регрессии в натуральном масштабе к уравнению в стандартизованном масштабе и обратно
13. Частные уравнения регрессии
14. Множественная корреляция
15. Частные коэффициенты корреляции
16. Оценка надежности результатов множественной регрессии и корреляции. Частный  $F_{xi}$
17. Сравнение двух регрессий. Тест Чоу
18. Фиктивные переменные в уравнении множественной регрессии
19. Система одновременных уравнений. Структурная и приведенная форма модели
20. Проблемы идентификации между СФМ и ПФМ. Достаточное и необходимое условие идентификации
21. Косвенный МНК
22. Двухшаговый МНК
23. Предпосылки применения метода наименьших квадратов
24. Тест ранговой корреляции Спирмена о наличии гетероскедатичности
25. Тест Годфелда-Квандта о наличии гетероскедатичности
26. Модели с распределенными лагами. Модель Койка
27. Модели Ш.Алмон



### 3. Парная нелинейная регрессия. Оценка параметров.

Для оценки параметров нелинейных моделей используются два подхода. Первый подход основан на *линеаризации* модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде *линейного соотношения* между *преобразованными* переменными.

Для линеаризации модели в рамках первого подхода могут использоваться как модели, нелинейные по переменным, так и нелинейные по параметрам.

Если модель **нелинейна по переменным (по объясняющим переменным х)**, то введение новых переменных ее можно свести к линейной модели, для оценки параметров которой можно использовать обычный метод наименьших квадратов. Н-р: полиномиальная, обратная.

Более сложной проблемой является **нелинейность модели по параметрам (по оцениваемым коэф-там)**, т.к. непосредственное применение метода наименьших квадратов для их оценивания невозможно. К числу таких моделей можно отнести, н-р, мультипликативную модель, экспоненциальную модель. В ряде случаев путем подходящих преобразований эти модели удается привести к линейной форме, н-р логарифмированием.

Второй подход обычно применяется в случае, когда подобрать соответствующее линеаризирующее преобразование не удается. В этом случае применяются методы *нелинейной оптимизации* на основе исходных переменных.

### 7. Множественная регрессия. Отбор факторов при построении множественной регрессии.

Множественная регрессия (МР) широко исп-я в решении проблем спроса, доходности акций, издержек пр-ва и других вопросах. Основная цель МР- построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также их совокупное воздействие на моделируемый показатель.

$Y_i = Y_{теор}(x_{i1}, x_{i2}) + \epsilon_i$   $Y_{теор}(x_{i1}, x_{i2}) = a_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_p x_{ip}$   
 $S(a, b_1, b_2, \dots, b_p) = \sum_{i=1}^n (y_i - a - b_1x_{i1} - b_2x_{i2} - \dots - b_px_{ip})^2 \rightarrow \min$  а,  $b_1, b_2, \dots, b_p$  - решение задачи.

Решение задачи следует из нбх условия минимума функций многих переменных. Производная в точке минимума д.б. равна 0.

(1)  $\partial S / \partial a = (a - \sum_{i=1}^n y_i) = 0$   $\partial S / \partial b_1 = (b_1 - \sum_{i=1}^n x_{i1}) = 0$   $\partial S / \partial b_2 = (b_2 - \sum_{i=1}^n x_{i2}) = 0$   $\partial S / \partial b_p = (b_p - \sum_{i=1}^n x_{ip}) = 0$   
 (2)  $\partial S / \partial b_1 = (a - \sum_{i=1}^n y_i + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + \dots + b_p \sum_{i=1}^n x_{ip}) = 0$   
 (3)  $\partial S / \partial b_2 = (a - \sum_{i=1}^n y_i + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + \dots + b_p \sum_{i=1}^n x_{ip}) = 0$

(1); (2); (3) - система нормальных уравнений.  
 $\sum Y_i = a \cdot n + b_1 \cdot \sum x_{i1} + b_2 \cdot \sum x_{i2} + \dots + b_p \cdot \sum x_{ip}$   
 $\sum (Y_i x_{i1}) = a \cdot \sum x_{i1} + b_1 \cdot \sum x_{i1}^2 + b_2 \cdot \sum x_{i1} x_{i2} + \dots + b_p \cdot \sum x_{i1} x_{ip}$   
 $\sum (Y_i x_{i2}) = a \cdot \sum x_{i2} + b_1 \cdot \sum x_{i1} x_{i2} + b_2 \cdot \sum x_{i2}^2 + \dots + b_p \cdot \sum x_{i2} x_{ip}$   
 $\dots$   
 $\sum (Y_i x_{ip}) = a \cdot \sum x_{ip} + b_1 \cdot \sum x_{i1} x_{ip} + b_2 \cdot \sum x_{i2} x_{ip} + \dots + b_p \cdot \sum x_{ip}^2$   
 $d = \sum (Y_i x_{i1})$   
 $e = \sum (Y_i x_{i2})$   
 $f = \sum (Y_i x_{ip})$   
 $A = \sum x_{i1}^2; \sum x_{i1} x_{i2}; \sum x_{i1} x_{ip}; \sum x_{i2}^2; \sum x_{i2} x_{ip}; \sum x_{ip}^2$   
 $a^* = \sum Y_i$   
 $x = b_1^* \quad b_2^* \quad \dots \quad b_p^*$   
 $d = A^* x; A^{-1}; A^{-1} d = x$

$x$  и  $d$  - векторы, причем  $x$  - вектор неизвестных коэф-тов  
 # 1 шаг: сформировать матрицу  $A$ , сформировать столбец  $d$ , 2 шаг: сделать обратную матрицу, 3 шаг: полученную матрицу умножаем на матрицу  $d$ , получаем  $x$ . 4 шаг: проверяем с помощью сервиса  $an-z$  данных регрессии.

Замечание: также как в парной регрессии коэффициент ур-ния множественной регрессии  $m$  вычисляют 2-мя способами: 1.  $\chi^2$  линейную ф-ю. 2. Сервис -  $an$  данных -> регрессия (более предпочтительный способ) коэффициенты вычисляются и располагаются более естественно.

Правило получения хорошей модели: 1)  $f_{факт} > f_{табл}$ . 2) вероятность или значение д.б. < 0,05.  $Y_{теор}(x_{i1}, x_{i2}) = a + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \dots$  - наиболее точная.

Факторы, включ-ые во МР, д. отвечать след-щим требованиям: 1 д.б. количественно измеримы. Если нбх-мо включить в модель качественный фактор, не имеющий кол-го измерения, то ему нужно придать кол-ную определенность (# модели стоимости объектов недвижимости учитывается место нахождения недвижимости, и районы м.б. проранжированы) 2. Факторы не д.б. интеркоррелированы и находиться в точной функциональной связи.

Система нормальных ур-ий  $m$  может оказаться плохо обусловленной и повлечет неустойчивость и ненадежность оценок коэффициентов регрессии если включаются в модель факторы с высокой интеркорреляцией, когда  $R_{y_{i1}} < R_{y_{i2}}$  для зависимости  $y = a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \epsilon$ . Если  $m$  у факторов существует высокая корреляция, то нельзя определить их изолированное влияние на результат-й показатель и параметры ур-я регрессии оказываются неинтерпретируемыми. Так в ур-ии  $y = a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \epsilon$ , что факторы  $x_1, x_2$  незав-мы др. от др-га, т.е.  $r_{x_1 x_2} = 0$ . Тогда  $m$  говорит, что параметр  $b_1$  измеряет силу влияния фактора  $x_1$  на результат  $y$  при неизменном значении фактора  $x_2$ . Если же  $r_{x_1 x_2} = 1$ , то с изменением фактора  $x_1$  фактор  $x_2$  может оставаться неизменным. Отсюда  $b_1$  и  $b_2$  нельзя интерпретировать как показатели раздельного влияния  $x_1$  и  $x_2$  на  $y$ .

### 8. Матрица парных корреляций. Мультиколлинеарность.

По величине парных коэф-тов корреляции обнаруживается явная коллинеарность факторов. Наиболее трудности в использовании множественной регрессии - при наличии мультиколлинеарности факторов, когда более чем 2 фактора связаны между собой линейной зависимостью, т.е. имеет место совокупное воздействие факторов друг на друга. В результате вариации в исходных данных перестает быть полностью независимой, и нельзя оценить воздействие каждого фактора в отдельности. Включение в модель мультиколлинеарных факторов нежелательно в силу последствий:

- затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл;
- оценки параметров ненадежны, обнаруживаются большие стандартные ошибки и меняются с изменением объема наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами.

Если же факторы не коррелированы между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы были бы равны нулю.

Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю.

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежны результаты множественной регрессии, и наоборот.

Через коэффициенты множественной детерминации можно найти переменные, ответственные за мультиколлинеарность факторов. Для этого в качестве зависимой переменной рассматривается каждый из факторов. Чем ближе значение коэффициента множественной детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Сравнивая между собой коэффициенты множественной детерминации факторов можно выделить переменные, ответственные за мультиколлинеарность, следовательно, можно решать проблему отбора факторов, оставляя в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

Одним из путей учета внутренней корреляции факторов является переход к совмещенным уравнениям регрессии, т.е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Так, если  $y = f(x_1, x_2, x_3)$ , то возможно построение следующего совмещенного уравнения:

$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_1 x_2 + b_5 x_1 x_3 + b_6 x_2 x_3 + b_7 x_1 x_2 x_3$   
 Решению проблемы устранения мультиколлинеарности факторов может помочь и переход к уравнениям приведенной формы.

Отбор факторов, включаемых в регрессию, является одним из важнейших этапов практического использования методов регрессии. Подходы к отбору факторов на основе показателей корреляции могут быть разные. Каждый из этих методов по-своему решает проблему отбора факторов, давая в целом близкие результаты - отсев факторов из полного его набора (метод исключения), дополнительное введение фактора (метод включения), исключение ранее введенного фактора (шаговый регрессионный анализ).

На первый взгляд может показаться, что матрица парных коэффициентов корреляции играет главную роль в отборе факторов. Вместе с тем вследствие взаимодействия факторов парные коэффициенты корреляции не могут в полной мере решать вопрос о целесообразности включения в модель того или иного фактора. Эту роль выполняют показатели частной корреляции, оценивающие в чистом виде тесноту связи фактора с результатом. Матрица частных коэффициентов корреляции наиболее широко используется в процедуре отсева факторов.

### 10. Построение производственной функции Кобба-Дугласа в MS EXCEL.

- у нас имеются исходные данные: годы (или что-либо другое),  $Y, K, L$

- строим еше столбцы:  $\ln Y_i, \ln K_i, \ln L_i$  Утеор  
 $Y_{теор}(K_i, L_i) = A \cdot K_i^\alpha \cdot L_i^{1-\alpha}$  в степени  $\alpha$   $1^* L_i$  в степени  $\alpha$  2  
 $y_{теор}(x_{i1}, x_{i2}) = a + b_1 \cdot x_{i1} + b_2 \cdot x_{i2}$  со звезд  $x_{i1}^2 + b_2$  со звезд  $x_{i2}^2$  (1)  
 $\ln Y_i = \ln A + \alpha \ln K_i + (1-\alpha) \ln L_i$  в степени  $\alpha$  2

$\ln Y_i = \ln A + \alpha \ln K_i + (1-\alpha) \ln L_i$  (2)  
 Сравним (1) и (2): обозначим через  $a$  и  $b$  и получим уравнение множественной регрессии.

- сервис-анализ данных-регрессия: находим  $a$  со звезд  $= \ln A$ ,  $b_1$  со звезд  $= \alpha$ ,  $b_2$  со звезд  $= 1-\alpha$   
 - подставляем и находим Утеор  
 Чтобы построить диаграмму нужно построить таблицу вида...

Ki / Li	5	25	45
10	5,25184766	19,25641	30,94933
40	7,254780778	26,60036	42,75269
70	8,265424355	30,30599	48,70845
100	8,981845819	32,93282	52,93035
130	9,548177635	35,00933	56,26777

-затем мастер диаграмм-поверхность-1

### 9. Оценка параметров уравнения множественной регрессии (МР).

Оцениваются, как и в парной регрессии, методом наименьших квадратов (МНК). При его применении строится система нормальных уравнений, решение которой и позволяет получить оценки параметров регрессии.

Так, для уравнения  $y = a + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \epsilon$  система нормальных уравнений составит:

$$\sum_{i=1}^n y_i = n \cdot a + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + \dots + b_p \sum_{i=1}^n x_{ip}$$

$$\sum_{i=1}^n y_i x_{i1} = a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + b_p \sum_{i=1}^n x_{i1} x_{ip}$$

$$\sum_{i=1}^n y_i x_{ip} = a \sum_{i=1}^n x_{ip} + b_1 \sum_{i=1}^n x_{i1} x_{ip} + b_2 \sum_{i=1}^n x_{i2} x_{ip} + \dots + b_p \sum_{i=1}^n x_{ip}^2$$

Ее решение может быть получено методом определителей:  $a = \Delta_a / \Delta$ ,  $b_1 = \Delta_{b_1} / \Delta$ ,  $\dots$ ,  $b_p = \Delta_{b_p} / \Delta$ .

Где  $\Delta$  - определитель системы;  $\Delta_a, \Delta_{b_1}, \dots, \Delta_{b_p}$  - частные определители

$$\Delta = \begin{vmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i2} x_{i1} & \dots & \sum x_{ip} x_{i1} \\ \sum x_{i2} & \sum x_{i1} x_{i2} & \sum x_{i2}^2 & \dots & \sum x_{ip} x_{i2} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1} x_{ip} & \sum x_{i2} x_{ip} & \dots & \sum x_{ip}^2 \end{vmatrix}$$

$\Delta_a, \Delta_{b_1}, \dots, \Delta_{b_p}$  получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Возможен иной подход к определению параметров, когда на основе матрицы парных коэффициентов корреляции строится уравнение регрессии в стандартизованном масштабе:

$$t_y = B_1 t_{x1} + B_2 t_{x2} + \dots + B_p t_{xp} + \epsilon$$

Где  $t_y, t_{x1}, t_{x2}, \dots, t_{xp}$  - стандартизованные переменные:  $t_y = (y - \bar{y}) / \sigma_y$ ,  $t_{x1} = (x_1 - \bar{x}_1) / \sigma_{x1}$ , для которых среднее значение равно нулю:  $t_y, t_{x1}, \dots, t_{xp} = 0$ , а ср. квадратичное отклонение равно единице:  $\sigma_y^2 = \sigma_{x1}^2 = 1$ ;  $B_1, B_2, \dots, B_p$  - стандартизованные коэффициенты регрессии.

Применяя МНК к уравнению МР в стандартизованном масштабе, после соответствующих преобразований получим систему нормальных уравнений вида

$$R_{yx1} = B_1 + B_2 R_{x2x1} + B_3 R_{x3x1} + \dots + B_p R_{xpx1}$$

$$R_{yx2} = B_1 R_{x2x1} + B_2 + B_3 R_{x3x2} + \dots + B_p R_{xpx2}$$

$$\dots$$

$$R_{yxp} = B_1 R_{xpx1} + B_2 R_{xpx2} + B_3 R_{xpx3} + \dots + B_p$$

Решая ее методом определителей, найдем параметры - стандартизованные коэффициенты регрессии ( $B$ -коэффициенты). Они показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор  $x_i$  изменится на одну сигму при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии  $B_i$  сравнимы между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсевах факторов - из модели исключаются факторы с наименьшим значением  $|B_j|$

### 11. Уравнение множественной регрессии в стандартизованном масштабе. Оценка коэффициентов $b_i$ (бэтта).

$\sum_{i=1}^n X_{i1} X_{i2} X_{i3} Y_i$   $t_{x1} = (X_{i1} - \bar{X}_1) / \sigma_{X1}$   $t_{x2} = (Y_i - \bar{Y}) / \sigma_Y$

Среднее = 0  
 Ср кв отклонение (сигма) = 1  
 Сигма  $i$  = Корень из  $(\sum_{i=1}^n (X_{i1} - \bar{X}_1)^2) / (n-1)$   
 С пом преобразований  $t_{x1}$  среднее перешло в 0.

По правилу 3х сигм почти вся выборка находится в интервале  $(a-3\sigma; a+3\sigma)$

Случайная величина, у которой  $a = 0$ ,  $\sigma = 1$ , называется стандартизованной. Переменные отличаются формой графика. Т.е. для каждой стандартной переменной существует график, но они отл-ся формой. Можно построить МР: в кач  $Y$  берем  $t_y$ . Столбцы:  $t_{x1}, t_{x2}, t_{x3}$

Пар-ры  $U_r$ -я множественной регрессии оценив-ся с пом МНК. При его примен-нии строится система нормальных ур-ний, реше-е кот-го позволяет получить оценки параметров регрессии.

Иной подход к определению параметров множеств регрессии - на основе матрицы парных коэф-тов корреляции строится ур-е регрессии в стандартизованном масштабе:  $t_y = B_1 t_{x1} + B_2 t_{x2} + B_3 t_{x3} + 0$ ,

Где  $t$  - стандартизованные переменные, например  $t_x = (x_i - \bar{x}_i) / \sigma_{x_i}$  / сигма  $x_i$ , для кот-ых средн знач-е ( $t_x$  средн) равно нулю (потому свободный член = 0), а среднее квадратичное отклонение (сигма) = 1;  $B$  - стандартизованные коэф-ты регрессии.

Применяя МНК к ур-ю множеств регрессии в стандартизов-ом масштабе, после преобразований получим систему вида

$$r_{y,x1} = B_1 + B_2 r_{x2x1} + B_3 r_{x3x1}$$

$$r_{y,x2} = B_1 r_{x2x1} + B_2 + B_3 r_{x3x2}$$

$$r_{y,x3} = B_1 r_{x3x1} + B_2 r_{x3x2} + B_3$$

Из этой системы можно найти коэф-ты  $B$ . Они показ-ют на сколько сигм изменится в среднем рез-тат, если соотв-щий фактор  $x_i$  изменится на одну сигму при неизменном среднем уровне др факторов. В силу того, что все пере-ые  $B$  сравнимы между собой (в отличие от коэф-тов «чистой» регрессии), после этого сравнения можно ранжировать факторы по силе их воздействия на результат.

### 12. Переход от уравнения множественной регрессии в натуральном масштабе к уравнению в стандартизованном масштабе и обратно.

На основе матрицы парных коэффициентов корреляции строится уравнение регрессии в стандартизованном масштабе:

$$t_y = b_1 \cdot t_{x1} + b_2 \cdot t_{x2} + \dots + b_p \cdot t_{xp} + E$$

Где  $t_{y_i}, t_{x1}, t_{xp}$  – стандартизованные переменные;  $t_y = (y - \bar{y}) / \sigma_y$ ,  $t_{x1} = (x_1 - \bar{x}_1) / \sigma_{x1}$ ,

для которых среднее значение равно нулю;  $t_y, t_{x1} = 0$ , а ср. квадратическое отклонение равно единице:  $\sigma_y = \sigma_{x1} = 1$ ;

$\beta$  – стандартизованные коэффициенты регрессии.

Применяя МНК к уравнению МР в стандартизованном масштабе, после соответствующих преобразований получим систему нормальных уравнений вида

$$R_{yx1} = B_1 + B_2 \cdot R_{x2x1} + B_3 \cdot R_{x3x1} + \dots + B_p \cdot R_{xpx1},$$

$$R_{yx2} = B_1 \cdot R_{x2x1} + B_2 + B_3 \cdot R_{x3x2} + \dots + B_p \cdot R_{xpx2},$$

$$\dots$$

$$R_{xpx} = B_1 \cdot R_{xpx1} + B_2 \cdot R_{xpx2} + B_3 \cdot R_{x3px} + \dots + B_p$$

Решая ее методом определителей, найдем параметры – стандартизованные коэффициенты регрессии (В-коэффициенты). Для этого: сервис-анализ данных-корреляция, получаем матрицу парных корреляций А. Для трех ур-й с тремя неизвестными получая матрица 4\*4. Последний столбец – d-столбец свободных членов. Для нахождения коэф-тов бета умножаем d-столбец на подматрицу А (3\*3) с пом ф-и МУМНОЖ. Выделяем ячейки для получения вектора коэф-тов и заполняем их с пом F2+CTRL+SHIFT+ENTER. Полученные коэф-ты вставляем в модель и получаем ур-е регрессии в стандартизованном масштабе.

В парной зависимости стандартизованный коэффициент регрессии есть ни что иное, как линейный коэффициент корреляции  $r_{yx}$ . Подобно тому, как в парной зависимости коэффициенты регрессии и корреляции связаны между собой, так и во множественной регрессии коэффициенты «чистой» регрессии  $b_i$  связаны со стандартизованными коэффициентами регрессии  $B_i$ , а именно:

$$b_i = B_i \cdot (\sigma_{y_i} / \sigma_{x_i})$$

Это позволяет от уравнения регрессии в стандартизованном масштабе

$$t_y = B_1 \cdot t_{x1} + B_2 \cdot t_{x2} + \dots + B_p \cdot t_{xp}$$

Переходить к уравнению регрессии в натуральном масштабе переменных:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p$$

Параметр а определяется как  $a = y - b_1 \cdot x_1 - b_2 \cdot x_2 - \dots - b_p \cdot x_p$

### 13. Частные уравнения регрессии

На основе линейного уравнения множественной регрессии:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon,$$

могут быть найдены частные уравнения регрессии:

$$y_{x1} = a_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p,$$

$$y_{x2} = a_2 + b_1 \cdot x_1 + \dots + b_p \cdot x_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

**т.е. уравнения регрессии, которые связывают результирующий признак с соответствующими факторами х при закреплении других учитываемых во множественной регрессии факторов на среднем уровне. Частные уравнения регрессии имеют следующий вид:**

$$y_{x1} = a_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon_1,$$

$$y_{x2} = a_2 + b_1 \cdot x_1 + \dots + b_p \cdot x_p + \epsilon_2,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

При построении в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т.е. имеем:

$$y_{x1} = a_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon_1,$$

$$y_{x2} = a_2 + b_1 \cdot x_1 + \dots + b_p \cdot x_p + \epsilon_2,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

$$y_{xp} = a_p + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} + \epsilon_p,$$

$$\dots$$

частные коэф-ты корреляции. При нелинейной вз/св - частные индексы детерминации.

Частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель доказывается величиной корреляции частной корреляции. Частные коэффициенты (индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии, в основном их используют на стадии формирования модели. **Показатели частной корреляции** представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель. **Пример:** предположим, что зависимость объема продукции от затрат труда  $x_1$  характеризуется уравнением:

$$Y_{x1} = 27,5 + 3,5 \cdot x_1,$$

парный коэф-т корреляции  $r_{yx1} = 0,58$ .

Подставив в это уравнение факт значение  $x_1$ , найдем теоретич величины объема продукции  $Y_{x1}$  и величину остаточной дисперсии  $S^2$ :

$$S^2_{yx1} = \sum (y_i - y_{x1})^2 / n,$$

Включив в уравнение регрессии доп фактор  $x_2$  – технич оснащенность производства, получим ур-е регрессии вида:

$$Y_{x1x2} = 20,2 + 2,8 \cdot x_1 + 0,2 \cdot x_2.$$

Предположим, что  $S^2_{yx1x2} = 3,7$ , а  $S^2_{yx1} = 6$ . Чем больше число факторов включено в модель, тем меньше величина остаточной дисперсии. Сокращение остат дисперсии за счет доп включения фактора  $x_2$  составит  $6 - 3,7 = 2,3$ . Чем больше доля этого сокращения в остаточной вариации до введения доп фактора, т.е. в  $S^2_{yx1}$ , тем теснее связь между  $y$  и  $x_2$  при постоянном действии фактора  $x_1$ .

Корень квадратный из этой величины и есть индекс частной корреляции, показывающий в чистом виде тесноту связи  $y$  и  $x_2$ . Следовательно влияние фактора  $x_2$  на рез-т у определяется по формуле:

$$r_{yx2, x1} = \sqrt{\frac{S^2_{yx1} - S^2_{yx1x2}}{S^2_{yx1}}},$$

а чистое влияние  $x_1$ :

$$r_{yx1, x2} = \sqrt{\frac{S^2_{yx2} - S^2_{yx1x2}}{S^2_{yx2}}},$$

Если выразить остат дисперсию через показатель детерминации  $S^2_{ост} = \sigma^2_y \cdot (1 - r^2)$ . Соответственно формула примет вид:

$$r_{yx1, x2} = \sqrt{\frac{S^2_{yx1} - S^2_{yx1x2}}{S^2_{yx1}}} = \sqrt{\frac{\sigma^2_y (1 - r_{yx1, x2}^2)}{\sigma^2_y (1 - r_{yx1}^2)}} = \sqrt{\frac{1 - r_{yx1, x2}^2}{1 - r_{yx1}^2}},$$

для  $x_1$ :  $r_{yx2, x1} = \sqrt{\frac{S^2_{yx2} - S^2_{yx1x2}}{S^2_{yx2}}} = \sqrt{\frac{\sigma^2_y (1 - r_{yx2, x1}^2)}{\sigma^2_y (1 - r_{yx2}^2)}} = \sqrt{\frac{1 - r_{yx2, x1}^2}{1 - r_{yx2}^2}}$ ,

Рассмотренные показатели частной корреляции принято называть коэффициентами (индексами) частной корреляции 1-го порядка, ибо они фиксируют тесноту связи двух переменных при закреплении одного фактора. Если рассматривается регрессия с числом факторов  $p$ , то возможно частные коэффициенты корреляции не только 1-го, но и 2-го, 3-го и .. (p-1) порядка, т.е. влияние фактора  $x_1$  можно оценить при разных условиях независимости действия других факторов:

$r_{yx1, x2}$  - при постоянном действии фактора  $x_2$ ;  
 $r_{yx1, x2x3}$  - ... факторов  $x_2, x_3$ ;  
 $r_{yx1, x2, \dots, xp}$  - ... всех факторов.

В практических исследованиях предпочтение отдают показателям частной корреляции самого высокого порядка, т.к. они являются дополнением к уравнению множественной регрессии.

### 16. Оценка надежности результатов множественной регрессии и корреляции. Частный $F_{\alpha}$

С помощью F-критерия Фишера опред значимость уравнения множественной регрессии в целом, как и в парной регрессии.

(1)  $F_{\alpha} = \frac{D_{факт}}{D_{ост}} = \frac{R^2 / 1 - R^2}{(n - m) / m}$ ;  $D$  - дисперсия факторная и остаточная.  $D_{факт}$  - факторная сумма квадратов на одну степень свободы,  $D_{ост}$  - остаточная сумма квадратов на одну степень свободы.  $R^2$  - коэф-т множественной детерминации.  $m$  - число параметров при переменных  $x$  (в линейной регрессии совпадает с числом включенных в модель факторов),  $n$  - число наблюдений.

С помощью F-критерия Фишера опред-ся значимость уравнения множественной регрессии в целом. Формула частного критерия Фишера:

$$F_{\alpha} = \frac{R^2 \cdot y_{x1} \cdot x_m - R^2 \cdot y_{x1} \cdot x_{m-1} \cdot x_{m+1} \cdot x_m}{(1 - R^2 \cdot y_{x1} \cdot x_m) \cdot ((n - m) / 1)};$$

$R^2_{yx1, x_m}$  - коэффициент множественной детерминации для регрессии с полным набором факторов.  $R^2_{yx1, x_{m-1}}$ ;  $x_{m+1} \cdot x_m$  - для ур-я множеств-й регрессии без включения в модель фактора  $x_m$ . Частный F критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом.

Если  $F_{\alpha} > F_{табл}$  при  $\alpha = 0,05$  (заданном)  $v_1 = n - m - 1$ ;  $v_2 = 1$ , то включение i-го фактора статистически оправдано. Если  $F_{\alpha} < F_{табл}$  - то не оправдано.

С помощью частного F критерия  $m$ . проверить значимость всех коэф-тов регрессии предлагая, что каждый соотв-щий фактор  $x_i$  вводился в ур-е множ-й регр последним.

### 14. Множественная корреляция

Множественная корреляция оценивает ур-е множеств-й регрессии. Хар-т тесноту связи рассматриваемого набора факторов с исследуемым признаком (влияние факторов на результат).

Показатель множественной корреляции  $M_6$  найден как индекс множественной корреляции:  $R_{yx1, \dots, xp} = \sqrt{\frac{S^2_{ост} - S^2_{ост} \cdot \sigma^2_{y_{x1, \dots, xp}}}{S^2_{ост}}}$ ,  $\sigma^2_{y_{x1, \dots, xp}}$  - дисперсия результирующего признака,  $S^2_{ост}$  - остат дисп для ур-я  $y = f(x_1, \dots, x_p)$ .  $R_{yx1, \dots, xp}$   $M_6$  от 0 до 1, чем ближе к 1 тем теснее связь.

Можно польз-ся след формулой индекса множественной корреляции при линейн зав-ти:  $R_{yx1, \dots, xp} = \sqrt{\frac{\sum \beta_i^2 \cdot r_{yx1, \dots, xp}^2}{\sum \beta_i^2}}$ .  $\beta_i$  - стандартизованные коэф-ты регрессии,  $r_{yx1, \dots, xp}$  - парные к-ты корреляции рез-та с кажд фак-ром.

Формула индекса множественной корреляции для линейн регр-и получ назв-е линейн к-та множеств корреляции (совокуп-го коэф-та корреляции), кот можно опр-ть ч/з матрицу парных к-тов коррел-ции.  $R_{yx1, \dots, xp}$  - корень из (1 -  $\Delta K / \Delta K_{11}$ ),  $\Delta K$  - опр-л матрицы парных к-тов корреляции,  $\Delta K_{11}$  - опр-л матрицы межфакт-й корреляции. Для ур-я  $y = a + b_1 \cdot x_1 + b_2 \cdot x_2$

$$\Delta K = \begin{vmatrix} r_{yx1} & r_{yx2} & r_{yx3} & r_{yx4} \\ r_{x1x2} & r_{x1x3} & r_{x1x4} & r_{x1x5} \\ r_{x2x1} & r_{x2x3} & r_{x2y} & r_{x2x1} & r_{x2x3} \\ r_{x3x1} & r_{x3x2} & r_{x3y} & r_{x3x1} & r_{x3x2} & r_{x3x3} \\ r_{x4} & r_{x2} & r_{x3} & r_{x4} & r_{x2} & r_{x3} & r_{x4} \end{vmatrix}$$

Множественный коэф-т корреляции  $(r_{yx1, \dots, xp}) = \sqrt{\frac{\sum (x_{i1} - x_{1cp}) \cdot (x_{i2} - x_{2cp})}{\sum (x_{i1} - x_{1cp})^2 \cdot \sum (x_{i2} - x_{2cp})^2}}$

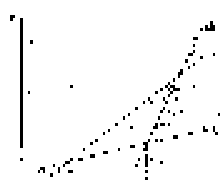
### 17. Сравнение 2-х регрессий. Тест Чоу.

Пусть момент (период) времени  $t^*$  сопровождается значительными изменениями ряда факторов, оказывающих сильное воздействие на

изучаемый показатель  $y$ . Чаще всего эти изменения вызваны изменениями в общеэкономической ситуации. Если исследуемый временной ряд включает в себя соответствующий момент (период) времени, то одной из задач его изучения становится выяснение вопроса о том, значимо ли повлияли общие структурные изменения на характер этой тенденции.

Если это влияние значимо; то для моделирования тенденции данного временного ряда следует использовать кусочно-линейные модели регрессий, т. е. разделить исходную совокупность на две подсовокупности (до момента времени  $t^*$  и после) и построить отдельно по каждой подсовокупности уравнения линейной регрессии (на рис. этим уравнениям соответствуют прямые (1) и (2)).

Если структурные изменения незначительно повлияли на характер тенденции ряда  $y$ , то ее можно описать с помощью одного для всей совокупности данных уравнения тренда (прямая (3)).



Каждый из описанных выше подходов имеет свои положительные и отрицательные стороны. При построении кусочно-линейной модели происходит снижение остаточной суммы квадратов по сравнению с одним для всей совокупности уравнением тренда. Однако разделение исходной совокупности на две части ведет к потере числа наблюдений и, следовательно, к снижению числа степеней свободы в каждом уравнении кусочно-линейной модели.

Очевидно, что выбор одной из двух моделей (кусочно-линейной или одного уравнения тренда) будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от одного уравнения регрессии к кусочно-линейной модели.

Основные обозначения для алгоритма теста Чоу

№ уравнения	Вид уравнения	Число наблюдений в совокупности	Остаточная сумма квадратов	Число параметров в уравнении	Число степеней свободы
Кусочно-линейная модель					
(1)	$Y = a_1 + b_1 \cdot t$	$n_1$	$C_{1ост}$	$k_1$	$n_1 - k_1$
(2)	$Y = a_2 + b_2 \cdot t$	$n_2$	$C_{2ост}$	$k_2$	$n_2 - k_2$
Уравнение тренда по всей совокупности					
(3)	$Y = a_3 + b_3 \cdot t$	$n$	$C_{3ост}$	$k_3$	$n - k_3 = (n_1 + n_2) - k_3$

Тест ЧОУ предполагает расчет параметров уравнений трендов, графики которых изображены на рис. прямыми (1), (2) и (3).

Выдвинем гипотезу о структурной стабильности тенденции изучаемого временного ряда.

Остаточную сумму квадратов по кусочно-линейной модели (Скл ост) можно найти как сумму  $C_{1ост}$  и  $C_{2ост}$ :

$$C_{ост} = C_{1ост} + C_{2ост} = (n_1 - k_1) + (n_2 - k_2) = (n - k) - k_3$$

Тогда сокращение остаточной дисперсии при переходе от одного уравнения тренда к кусочно-линейной модели можно определить следующим образом: из  $C_{ост} = C_{3ост} - C_{кл}$  ост:

$$C_{ост} - C_{кл} = (n - k) - k_3 - (n - k) + k_3 = 0$$

Соответствующее ей число степеней свободы составит:  $(n_1 - k_1) + (n_2 - k_2) = (n - k) - k_3$

Тогда сокращение остаточной дисперсии при переходе от одного уравнения тренда к кусочно-линейной модели можно определить следующим образом: из  $C_{ост} = C_{3ост} - C_{кл}$  ост:

$$C_{ост} - C_{кл} = (n - k) - k_3 - (n - k) + k_3 = 0$$

Если  $F_{\alpha} > F_{табл}$ , то гипотеза о структурной стабильности тенденции отклоняется, а влияние структурных изменений на динамику изучаемого показателя признают значимым. Выбираем кусочно-линейную модель.

Если  $F_{\alpha} < F_{табл}$ , то нет оснований отклонять гипотезу о структур

поддаются количественной оценке им нужно присвоить им цифровые метки, т.е. преобразовать качественные переменные в количественные = «Фиктивные переменные», или «структурные переменные».

	A	B	C	D	E
1	i	X1i возраст машины	X2i марка	Y к-во дней работы без ремонта	Yтеог
2	1	1	Москвич	10	
3	2	2	Жигули	150	
...	...	...	...	...	...
21	20	10	Тойота	1000	

Если (C2= «Москвич»;1; если (C2= «Жигули»;2;...)). В новом столбце названия заменяются цифрами.  
Сервис/анализ данных/регрессия – Находим У теор.  
У теор (x1i,x2i) = a\*+b1\*\*x1i+b2\*\*x2i...

Можно улучшить качество уравнения за счет введения дополнительных фиктивных переменных. Количество переменных = число градаций -1. Т.е. пусть марок машин 4. Тогда вводим 4-1 = 3 фиктивные переменные.  
X22i = 1, если марка=2  
= 0, в противном случае;  
X23i=1, если марка = 3,  
= 0, в противном случае  
X24i = 1, если марка =4,  
= 0, в противном случае.

Т.е. вместо столбца «С» вводим три новых столбца X22i X23i X24i. Т.е. вместо второй переменной вводим три фиктивные переменные.

У теор = a\*+b1\*\*x1i+b22\*\*x22i+b23\*\*x23i+b24\*\*x24i.  
Чтобы модель была хорошая, нужно, чтобы Fфакт был больше Fтабл, при ε=0,05 и 1-ε=0,95 при v1=m, v2=n-m-1

Если в регрессии получаются такие данные:

	Р значения
У пересечение	a* 0,02
Переменная 1	b1* 0,54
Переменная 2	b2* 0,06
Переменная 3	b3* 0,08
Переменная 4	b4*

То 1му коэффициенту можно верить с вероятностью 1-0,02= 0,98, 2й коэффициент незначим, т.к. вероятность очень низкая 1-0,54 = 0,46. Если все коэффициенты меньше 0,05, то модель хорошая.

Также для улучшения модели включаем логарифмы: вместо У теор находим ln У теор по той же формуле. Еще более точное значение можно получить.  
ln У теор = a\*+b1\*\*x1i+b22\*\*x22i+b23\*\*x23i+b24\*\*x2i^2

Модель улучшается, когда значение Rквадрат (из таблицы регрессии) улучшается (приближается к 1). При этом значения У теор при фиктивных переменных приближены к реальным значениям.

Среди моделей с ФП наибольшими прогностическими возможностями обладают модели, в которых зависимая переменная у рассматривается как функция ряда экономических факторов xi и фиктивных переменных zi (отражают различия в формировании результативного признака по отдельным группам единиц совокупности, т.е. в результате неоднородной структуры пространственного или временного характера.)

### 19. Системы одновременных (взаимозависимых, совместных) уравнений. Структурная и приведенная форма модели.

**Структурная форма модели** – система одновременных уравнений: одни и те же зависимые пер-е в одних ур-ях входят в левую часть, а в др – в правую часть системы, т.е. одни и те же пре-е (у) одновременно рассм-ся как зависимые в одних ур-ях, и как независимые в др.

**СФМ** содержит эндогенные (у-зависимые пер-е, их число = числу ур-й в системе) и экзогенные пер-е (х-предопределенные пер-е, влияющие на эндогенные, но независимые от них).

**Простейшая СФМ** имеет вид: система ур-й: y1=b1y2+a11x1+ε1; y2=b21y1+a22x2+ε2. СФМ позволяет увидеть влияние изм-й любой экзогенной пер-й на знач-я эндогенной. СФМ в правой части содержит коэф-ты: при у – bi, при х – aj, которые наз-ся структурными коэф-ми модели. Все пер-е выражены в отклонениях от среднего ур-я, т.е. под х и у подразумевается, соответственно, x=x-хкр, у=y-укр. Следовательно, нет свободных членов.

Т.к. исп-е МНК для оценивания стр-х коэф-тов невозможно (смещенные и несостоятельные оценки), СФМ преобразуется в ПФМ.

ПФМ представляет собой систему линейных функций эндогенных пер-х от экзогенных. Коэф-ты ПФМ представляют собой нелинейные функции коэф-тов СФМ. Для СФМ вида: система Ур-й: y1=b1y2+a11x1 и y2=b21y1+a22x2; ПФМ имеет вид: система Ур-й: y1=сигма11\*x1+сигма12\*x2 и y2=сигма21\*x1+сигма22\*x2, где сигмаij выражена из aj и bi. Для примера найдем первое Ур-е из ПФМ. Выразим из первого Ур-я СФМ y2, y2=(y1-a11x1)/b12. Подставим значение y2 во второе Ур-е СФМ и получим: (y1-a11x1)/b12=b21y1+a22x2. Из данного равенства выражаем y1=[a11/(1-b12\*b21)]\*x1+[a22\*b12/(1-b12\*b21)]\*x2. Пусть [a11/(1-b12\*b21)]=сигма11, а [a22\*b12/(1-b12\*b21)]=сигма12, тогда получим Ур-е ПФМ вида y1=сигма11\*x1+сигма12\*x2 (первое Ур-е системы ПФМ). Аналогично находится второе Ур-е системы ПФМ.

ПФМ хотя и позволяет получить знач-я эндогенных пер-х через знач-я экзогенных, аналитически уступает СФМ, т.к. в ней отсутствуют оценки взаимосвязей между эндогенными пер-ми.

### 20. Проблемы идентификации м/у СФМ и ПФМ. Достаточное и необходимое условие идентификации.

При переходе от ПФМ к СФМ сталкиваются с проблемой идентификации.

Идентификация - это единственность соответствия м/у приведенной и структурной формами модели.

В шир смысле - это соотв-е нек модели реальному об-ту. n(n-1)+n\*m=n(n-1+m). В общем сл кол-во переменных 1 больше чем 2. Проблема восстановления коэффициентов 1 модели по коэфф-там 2й. = проблема спецификации. Коэф-ты приведенной формы модели всегда известны, их м найти МНК.

Если кол-во перем-ых в СФМ > чем в ПФМ, то модель неидентифицируема, если их кол-во равно, то однозначно идентифицируема, если кол-во перем-ых в СФМ < чем в ПФМ, то имеет место неоднозначность, т.е. мы можем найти коэф-ты СФМ разными способами. Для опр-я коэф-тов просто идентифицируем ур-й примен-ся коэф-тов наим квадратов, для опр-я коэф-тов сверхидентифицируемых ур-й примен-ся двухшаговый метод МНК.

1. Необходимое ус-е - счетное правило идентификации.

Рассм-ся кажд ур-е. Обознач-ся через H кол-во эндогенных переменных, вход-щих в дан ур-е; D кол-во экзогенных перем-ых, вход-щих в дан ур-е.

Если D+1<H, то ур-е неидентифицируемо.

Если D+1=H, ур-е явл подозрительным на то, что оно точно идентифицируемо.

Если D+1>H, ур-е явл сверхидентифицируемым.

Пример (система из 3-х ур-ний).

y1 = b12\*y2 + b13\*y3 + a11\*x1 + a12\*x2 здесь H=3 D=4-2=2

y2 = b21\*y1 + a21\*x1 + a22\*x2 + a23\*x3 здесь H=2 D=1

y3 = b31\*y1 + b32\*y2 + a33\*x3 + a34\*x4 здесь H=3 D=2

Если в в системе ур-ний все ур-я явл просто идентифицируемы, то система наз-ся просто идентифицируемой. Если хотя бы одно сверхидентифицируемо вся система

сверхидентифицируема, если хотя бы одно неидентифицируемо вся система неидентифицируема.

2. Достаточное ус-е идентификации системы (модели).

Рассм-ся кажд ур-е и для него составл-ся матрица, состоящая из коэф-тов при

переменных, отсут-щих в дан ур-нии. Определитель этой матрицы не д.б. равен нулю

и ее ранг д.б. равен кол-ву эндогенных перем-ых, входящих в систему ур-ний за вычетом 1-цы. Тогда данное ур-е просто идентифицируемо. Например, для 1-го ур-я из

вышеприведенной системы: |A|= | a23 0 |

| A33 a34 | = a23\*a34-0\*a33 это

скорее всего не равно нулю

Ранг матрицы это максимальное кол-во линейно независимых строк/ столбцов,

значит ранг A=2. Кол-во эндогенных перем-ых = 3 (3-1=2), значит ур-е явл точно идентифицируемым.

21. Косвенный МНК

КМНК прим-ся в случае точно идентифицируемой структурной модели.

Этапы примен-я:

1. По структур-й форме модели формальным образом выписывается приведенная форма модели.

2. Для каждого ур-я приведенной формы модели обычным МНК оцен-ся приведенный коэф-ты.

3. Коэф-ты прив-ой формы модели транс-ся в параметры структурной модели.

Пример:

Y1=δ11x1+δ12x2+ε1,

Y2=δ21x1+δ22x2+ε2.

Y1

Y2

X1

X2

1 2 5 1 3

2 3 6 2 1

3 4 7 3 2

4 5 8 2 5

5 6 5 4 6

Среднее 4 6,2 2,4 3,4

Приведенная форма модели составит:

Y1=δ11x1+δ12x2+ε1,

Y2=δ21x1+δ22x2+ε2.

Где ε1, ε2 – случай-е ошибки приведенной формы модели

Для каждого ур-я приведенной формы модели прим-ем традиционный МНК и опр-ем коэф-ты δ (которые становятся числами). Т.о. приведенная форма модели имеет вид:

y1=0,852x1+0,373x2+ε1

y2=-0,072x1-0,0055x2+ε2

Далее переходим от приведенной формы модели к структур-й. Для этого из первого ур-я приведенной формы модели надо искл-ть x2, выразив его из второго ур-я приведенной формы и подставив в первое:

X2=(y2-сигма20-сигма21\*x1)/сигма22

Подставляем в уравнение первое ПФМ и получаем:

y1=сигма10+сигма11\*x1+сигма12\*(y2-сигма20-сигма21\*x1)/сигма22. После преобразований получаем:

y1=[сигма10-сигма12(сигма20/сигма22)]+[сигма11-сигма12(сигма21/сигма22)]\*x1

[сигма10-сигма12(сигма20/сигма22)]=c1

[сигма11-сигма12(сигма21/сигма22)]=сигма12

[сигма11-сигма12(сигма21/сигма22)]=a11

Аналогично получаем y2, и соответственно коэф-ты при втором уравнении.

### 22. Двухшаговый метод наименьших квадратов.

Если система сверхидентифицируема, то КосвМНК не исполь-з. т.к. не дает однозначных оценок для параметров структурной модели. Тогда ДМНК: на основе приведенной формы модели получить для

сверхидентифицируемого уравнения теор значения эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений можно исп-е обычный МНК к структурной форме сверхид уравнения. Т.е. дважды используется МНК

Rt=g20+g21\*Rt+g22\*Yt+g23\*Yt-1+Vt

Qt=σ10+σ11\*Rt+σ12\*Yt+σ13\*Yt-1+Ut

2. Для определения к-тов просто идем ур-ий – косвенный метод

3. Для определения к-тов сверхид уравнений: в каждом ур-нии выявляем эндогенные переменные, нах-ся в правой части уравнения и заменяем теоретическим значением, вычисленным по приведенной форме. Так со всеми эндогенными переменными правой части.

4. Компануем все переменные, включая экзогенные, входящие в правую часть - они будут векторами иксов. Переменная в левой части = Y. Применяем МНК для определения a1, a2, a3.

Qt=a1+a2\*Pt+a3\*Pt+ε1t сверх

Pt=b1+b2\*Qt+b3\*Yt+b4\*Yt-1+ε2t просто

Выражаем: Pt=(Qt-σ10-σ2\*Yt-σ13\*Yt-1-Ut)/σ11

### 23. Предпосылки применения метода наименьших квадратов (МНК).

При оценке параметров Ур-я регрессии - МНК. при этом делаются предпосылки отн-но случ-стей «Е». В модели y=a+b1\*x1+b2\*x2+...+bp\*xr+ε ε= ненаблюдаемая величина. Делается предположение о поведении остатков е-это независимые случ-е величины и их ср знач = 0 они имеют одинаковую дисперсию и подчиняются норм-му распределению. Оценки параметров регрессии при МНК должны отв-ть след треб-м: дб несмещенными (матем ожидание остатков =0), дб эффективными (хар-ся наименьшей дисперсией), дб состоятельными (увеличение их точности с увеличением выборки). МНК строит оценки регрессии на основе минимизации суммы квадратов остатков. Важно исслед поведение остаточных величин регрессии е. Условия, необх для получения несмеш, эфф, сост оценок = предпосылки МНК, соблюдение которых желательно для получения достоверного результата регрессии.

Исследование остатков еі предп-т проверку наличия след предпосылок МНК:

-случ-й характер остатков

-нулевая средняя величина остатков, не зависящая от xi

-гомоскедастичность - те дисперсия каждого отклонения еі одинакова для всех знач x

-отсутствие автокорреляции остатков, те распределены независимо др от друга

-остатки подчиняются нормальному распределению.

Если не исполн-ся хотя бы 1 предп-ка, то нужно корректировать модель.

1 предпосылка: строим график зависимости остатков еі от теор знач результативного признака. Если на графике горизонт-я полоса, то ост-ки = случайные величины и МНК оправдан.

2 означает, что ∑(y-уеог)=0. Это выполняемо для линейных моделей, затем строится график анализ-но. График зависимости сл остатков от факторов x. Если горизонтальная полоса – независимы, модель адекватна.

Для 5 позволяет проводить проверку параметров регрессии и корреляции с помощью критерия F.

3 Если это условие не собл-ся то имеет место гетероскедастичность (дисперсия остатков растет по мере увелич x, дисперсия остатков достиг-т макс величины при ср знач перем x и уменьш при мин и макс знач). Ведет к смещенности оценок.

Чтобы оценить нарушение гомоскед-ти можно сделать параметрический тест, шага:

1)упорядочение n наблюдений по мере возрастания переменной x

2)Исключение из рассм-я С центральных наблюдений, при этом (n-С):2>r, где r - число оцениваемых параметров

3)Разд-е совокуп-ти из (n-С) набл-й на 2 группы (с малым и большим знач фактора x) и определение по каждой из групп уравнений регрессии

4)Опр-е остаточной суммы квадратов для 1-й (S1) и 2-й (S2) групп и нахождение их отношения: R=S1/S2.

Чем больше величина R превышает табл знач F-критерия тем больше нарушена предпосылка о равенстве дисперсий остаточных величин.

4. автокорреляция = наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений.

К корреляции между еі и еj, где еj –остатки предыдущих наблюдений j=i-1 может быть определен как refej=cov(ei,ej)/σei\*σej

Т.е. по обычной формуле линейного коэффициента корреляции. Если он окажется существенно отличен от нуля, то остатки автокоррелированы и функция плотности вероятности зависит от житой точки наблюдения и от распределения остатков в других точках наблюдения.

Отсутствие автокорреляции остаточных величин обеспечивает состоятельность и эфф-ть оценок коэф-тов регрессии.

24. Коэффициент ранговой корреляции Спирмена (КРКС).

Иногда нужно установить связь не только между 2мя колич. Переменными, но между ординальными (порядковыми) переменными – качество жилья, оценки экз. Тогда объекты анализа

ранжируют по степени выраженности измеряемых переменных. Каждому объекту присваивается № (ранг). Напр, объекту с наименьшим проявлением признака – ранг 1, следующему – ранг 2. Если объекты ранжированы по 2м признакам, то можно оценить тесноту связи, основываясь на рангах (тесноту ранговой корреляции).

КРКС находится по формуле  $r=1-\frac{6\sum i \cdot i^2}{n(n^2-1)}$   
 Где  $i_i, s_i$  - ранги I-го объекта по переменным X и Y; n-число пар наблюдений.

Если ранги всех объектов равны ( $i_i=s_i, i=1,2,\dots,n$ ), то  $r=1$ . Т.е. при полной прямой связи  $r=1$ . При полной обратной связи, когда ранги объектов по двум переменным расположены в обратном порядке, можно показать, что  $r=-1$ . Во всех остальных случаях  $|r|<1$ .

При ранжировании иногда невозможно найти существенные различия между объектами по величине проявления рассматриваемого признака: объекты оказываются связанными. Им присписывают одинаковые средние ранги так, чтобы сумма всех рангов оставалась такой же, как при отсутствии связанных рангов. Например, если 4 объекта оказались равнозначными в отношении рассматриваемого признака и невозможно определить, какие из 4х рангов (4,5,6,7) присписать этим объектам, то каждому присписывается средний ранг, равный  $(4+5+6+7)/4=5,5$ . В модификациях формула на связанные ранги вводится поправки. При проверке значимости р исходят из того, что в случае справедливости нулевой гипотезы об отсутствии корреляционной связи между переменными при  $n>10$  статистика  $t=(r \cdot \sqrt{\text{корень из } (n-2)})/(\text{корень из } (1-r^2))$  имеет t-распределение Стьюдента с (n-2) степенями свободы. Поэтому р значим на уровне  $\alpha$  при числе степеней свободы (n-2)

Ранговый коэфф корреляции r может быть использован и для оценки тесноты связи между обычными колич переменными. Достоинство r здесь - его нахождение не требует нормального распределения переменных, линейной связи между ними. Но нужно учитывать, что при переходе от первоначальных значений переменных к их рангам происходит определенная потеря информации. Чем теснее связь, тем чем меньше корр зависимость между переменными отличается от линейной, тем ближе КРКС r к коэффциенту парной корреляции g.

## 25. Тест Голдфелда-Квандта о наличии гетероскедастичности.

Гетероскедастичность – это свойство стохастической компоненты регрессии, выражающееся в том, что их дисперсия (а, следовательно, разброс значений ошибок) не является постоянной. Следствие гетероскедастичности – найденные оценки коэффциентов регрессии больше не представляют собой наилучшие оценки с наименьшей дисперсией. В результате значение критерия проверки значимости коэффциентов будет искажено: при отрицательном смещении дисперсий он будет неосновательно завышен, при положительном – занижен, что повлечет за собой неверные выводы о значимости оценок коэффциентов регрессии.

Тест Голдфелда-Квандта применяют, если есть предположение о прямой зависимости дисперсии ошибки от величины некоторой независимой переменной модели. Для этого надо действовать по следующему алгоритму:

- 1) все наблюдения упорядочиваются по величине независимой переменной, относительно которой есть подозрение на гетероскедастичность;
- 2) остатки в этой упорядоченной совокупности делят на две равных группы, при чем находящиеся посредине между ними d наблюдений исключаются из рассмотрения (d обычно равно около четверти от общего количества наблюдений);
- 3) рассчитывается две независимых регрессии по первой и второй группе, количество наблюдений в которых составляет  $n/2-d/2$  (при этом должно быть  $n/2-d/2 > k+1$ , где k – число независимых переменных), и находятся соответствующие остатки для первой и для второй регрессии  $e_1$  и  $e_2$ ;
- 4) если предположение о прямой зависимости дисперсии ошибки от величины данной независимой переменной верно, то в первой группе сумма квадратов остатков (а значит и их дисперсия) будет меньше, чем во второй; затем рассчитывают критерий Голдфелда-Квандта: в случае предположения прямой пропорциональности между величиной дисперсии отклонений и значением независимой переменной сумму квадратов остатков во второй группе делят на сумму квадратов остатков в первой. Рассчитанный критерий имеет F-распределение с  $(n/2-d/2-k)$  и  $(n/2-d/2-k)$  степенями свободы. В случае обратной пропорциональности дисперсии отклонений значению независимой переменной сумму квадратов остатков в первой группе делят на сумму квадратов остатков во второй, распределение критерия также имеет вид F-распределения с теми же степенями свободы.

В случае, когда в модели имеет место гетероскедастичность остатков, требуется изучить взаимосвязь между значениями остатков и переменными модели, после чего скорректировать регрессионную модель таким образом, чтобы она учитывала эту взаимосвязь.

## 26. Модели с распределёнными лагами. Модель Койка.

Можно выделить два основных типа динамических эконометрических моделей. К моделям первого типа относятся модели авторегрессии или модели с распределённым лагом, в

которых значения переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель. Модели второго типа учитывают динамическую информацию в неявном виде. В эти модели включены переменные, характеризующие ожидаемый или желаемый уровень результата, или одного из факторов в момент времени t.

При исследовании экономических процессов нередко приходится моделировать ситуации, когда значение результирующего признака в текущий момент времени t формируется под воздействием ряда факторов, действовавших в прошлые моменты времени t-1, t-2, ..., t-l. Величину, характеризующую запаздывание в воздействии фактора на результат, называют в эконометрике лагом, а временные ряды самих факторных переменных, сдвинутые на один или более моментов времени, — лаговыми переменными.

Эконометрическое моделирование осуществляется с применением моделей, содержащих не только текущие, но и лаговые значения факторных переменных. Эти модели называются моделями с распределённым лагом. Модель вида  $y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \dots + b_k \cdot x_{t-k} + \epsilon_t$  является примером модели с распределённым лагом.

Наряду с лаговыми значениями независимых, или факторных, переменных на величину зависимой переменной текущего периода могут оказывать влияние ее значения в прошлые моменты или периоды времени. Эти процессы обычно описывают с помощью моделей регрессии, содержащих в качестве факторов лаговые значения зависимой переменной, которые называются моделями авторегрессии. Модель вида  $y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \dots + c_k \cdot y_{t-k} + \epsilon_t$  относится к моделям авторегрессии.

Очевидно, что параметры такой модели обычным МНК или с помощью иных стандартных статистических методов определить нельзя. Впервые подход к оценке параметров моделей с распределённым лагом был предложен Л.М. Койком. Койк предположил, что существует некоторый постоянный темп  $\lambda$  ( $0 < \lambda < 1$ ) уменьшения во времени лаговых воздействий фактора на результат. Если, например, в период t результат изменялся под воздействием изменения фактора в этот же период времени на  $\Delta y$  ед., то под воздействием изменения фактора, имевшего место в период (t-1), результат изменится на  $\Delta y \cdot \lambda$  ед.; в период (t-2) — на  $\Delta y \cdot \lambda^2$  ед., и т. д. В более общем виде можно записать:  $\Delta y_t = \Delta y \cdot \lambda^j$ ,  $j=0,1,2,\dots,0 < \lambda < 1$

Ограничение на значения  $\lambda > 0$  обеспечивает одинаковые знаки для всех коэффциентов  $b_j > 0$ , а ограничение  $\lambda < 1$  означает, что с увеличением лага значения параметров модели убывают в геометрической прогрессии. Чем ближе  $\lambda$  к 0, тем выше темп снижения воздействия фактора на результат во времени и тем большая доля воздействия на результат приходится на текущие значения фактора  $x_t$ .

Выразим все коэффциенты  $b_j$  в модели через  $b_0$  и  $\lambda$ . Тогда для периода (t-1) модель можно записать следующим образом:  $y_{t-1} = a + b_0 \cdot x_{t-1} + b_1 \cdot \lambda \cdot x_{t-2} + b_2 \cdot \lambda^2 \cdot x_{t-3} + \dots + \lambda^k \cdot \epsilon_{t-1}$

Умножим обе части модели на  $\lambda$ . Преобразования приводят, нас к получению модели Койка:  $y_t = a \cdot (1 - \lambda) + b_0 \cdot \lambda \cdot x_t + (1 - \lambda) \cdot y_{t-1} + \lambda \cdot \epsilon_t$ , где  $\epsilon_t = \epsilon_{t-1} \cdot \lambda$ . Полученная модель есть модель двухфакторной линейной регрессии (точнее — авторегрессии). Определив ее параметры, мы найдем  $\lambda$  и оценки параметров  $a$  и  $b_0$  исходной модели. Далее с помощью соотношений несложно определить параметры,  $b_1, b_2, \dots$  модели. Отметим, что применение обычного МНК к оценке параметров модели приведет к получению смещенных оценок ее параметров ввиду наличия в этой модели в качестве фактора лаговой результирующей переменной  $y_{t-1}$ .

Описанный выше алгоритм получил название преобразования Койка. Это преобразование позволяет перейти от модели с бесконечными распределёнными лагами к модели авторегрессии, содержащей две независимые переменные  $x_t$  и  $y_{t-1}$

## 27. Модели Ш. Алмон.

Модели с распределённым лагом. Общий вид:

$$Y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \dots + b_p \cdot x_{t-p} + \epsilon_t$$

Эта модель показывает, что если в некоторый момент времени t происходит изменение независимой переменной x, то это изменение будет влиять на значение переменной y в течение l следующих моментов времени.

Лаги, структуру которых можно описать с помощью полиномов, называют лагами Алмон. Предположим, что в модели — полиномиальная структура лага, т.е. зависимость коэффциентов регрессии  $b_i$  от величины лага описывается полиномом k-той степени.

Формально модель зависимости коэффциентов  $b_j$  от величины лага j в форме полинома можно записать так:  
 Для полинома 1й степени  $b_j = c_0 + c_1 j$   
 Для полинома 2й степени  $b_j = c_0 + c_1 j + c_2 j^2$   
 Для полинома 3й степени  $b_j = c_0 + c_1 j + c_2 j^2 + c_3 j^3$

В наиболее общем виде для полинома k-й степени имеем:  
 $b_j = c_0 + c_1 j + c_2 j^2 + \dots + c_k j^k$   
 Тогда каждый из коэффциентов модели можно выразить: (1)  
 $b_0 = c_0$ ;  $b_1 = c_0 + c_1 + \dots + c_k$ ;  $b_2 = c_0 + 2c_1 + 4c_2 + \dots + 2^k c_k$ ;  
 $b_3 = c_0 + 3c_1 + 9c_2 + \dots + 3^k c_k$ ;  
 $\dots$ ;  $b_l = c_0 + l c_1 + l^2 c_2 + \dots + l^k c_k$

Подставляем в общий вид модели найденные соотношения, перегруппировываем слагаемые и обозначаем слагаемые в скобках перед  $C_i$  как новые переменные: (2)

$$z_0 = x_t + x_{t-1} + x_{t-2} + \dots + x_{t-l} = \sum_{j=0}^l x_{t-j}$$

$$z_1 = x_{t-1} + 2x_{t-2} + 3x_{t-3} + \dots + l x_{t-l} = \sum_{j=1}^l j x_{t-j}$$

$$z_2 = x_{t-1} + 4x_{t-2} + 9x_{t-3} + \dots + l^2 x_{t-l} = \sum_{j=1}^l j^2 x_{t-j}$$

$$\dots$$

$$z_k = x_{t-1} + 2^k x_{t-2} + 3^k x_{t-3} + \dots + l^k x_{t-l} = \sum_{j=1}^l j^k x_{t-j}$$

Перепишем модель с учетом полученных соотношений: (3)  
 $y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \dots + c_k z_k + \epsilon_t$   
*Процедура применения для расчетов параметров модели с распределённым лагом:*

1. Определяется максимальная величина лага
2. Определяется степень полинома k, описывающего структуру лага
3. по соотношениям (2) рассчитываются значения переменных
4. определяются параметры уравнения линейной регрессии (3)
5. по (1) рассчитываются параметры исходной модели с распределённым лагом.

Проблемы:  
 1. величина лага l должна быть известна заранее. Если выбрать меньший лаг, чем реальный, то в модели регрессии не будет учтен фактор, оказывающий значительное влияние на результат. Если выбрать больший, то в модель включится статистически незначимый фактор. Величина лага определяется измерением тесноты связи между результатом и лаговым значением фактора.

2. Необходимо установить степень полинома k. Она должна быть на единицу больше числа экстремумов в структуре лага. Определяется сравнением моделей, построенных для различных значений k, выбирается наилучшая модель.

3. переменные зет будут коррелировать между собой, когда наблюдается высокая связь между самими исходными переменными.

Преимущества:  
 1. универсален, может быть применен для моделирования процессов, которые характеризуются разнообразными структурами лагов.

2. При относительно небольшом кол-ве переменных в (3) (обычно выбирают k=2 или k=3), которое не приводит к потере значительного числа степеней свободы, с помощью метода Алмон можно построить модели с распределённым лагом любой длины.