

Микроэконометрика

Тема 1. Типы ошибок спецификации модели

Specification errors types: Omission of a relevant variables, Inclusion of an unnecessary variables, Adopting the wrong functional form, Errors of measurement, Incorrect specification of the stochastic error term

Эндогенность из-за пропуска существенной переменной: True regression: $y_i = \beta_1 + \beta_2 \cdot x_i + \beta_3 \cdot w_i + \varepsilon_i, \beta_3 \neq 0$

Estimated regression: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$

В этом случае МНК-оценка будет, вообще говоря, смещена, что само по себе является достаточно серьезной проблемой. На самом деле, в большинстве ситуаций (кроме одного частного случая) она будет еще и несостоятельной. То есть возникшую проблему нельзя будет компенсировать использованием сколь угодно большого массива данных.

Докажем это:

$$\hat{\beta}_2 = \frac{\widehat{\text{cov}}(x, y)}{\widehat{\text{var}}(x)} \rightarrow \frac{\text{cov}(x_i, y_i)}{\text{var}(x_i)} = \frac{\text{cov}(x_i, \beta_1 + \beta_2 \cdot x_i + \beta_3 \cdot w_i + \varepsilon_i)}{\text{var}(x_i)} = \beta_2 + \beta_3 \frac{\text{cov}(x_i, w_i)}{\text{var}(x_i)}$$

Смещение из-за пропущенных переменных:

- Пропущенная переменная должна: (1) влиять на Y, и (2) быть коррелированной с одной или более включенных в регрессию независимых переменных,
- Смещение сохраняется даже в больших выборках – МНК-оценка является несостоятельной.

пропущенных переменных; контрольных переменных (переменные, не изменяющиеся во времени);

Излишние переменные: Ухудшение точности оценок (увеличение оценок дисперсий) при включении в модель излишних переменных.

Неправильная функциональная форма модели: Рассмотрение линейной модели вместо нелинейной эквивалентно пропуску существенной переменной и приводит к таким же последствиям. логарифм можно аппроксимировать квадратичной функцией или полиномом более высокого порядка.).

Проверка гипотезы о группе излишних переменных.

F-статистика позволяет проверять одновременно все ограничения совместной гипотезы Для частного случая

$$H_0 : \beta_1 = \beta_{1,0} \text{ и } \beta_2 = \beta_{2,0}$$

в регрессии с двумя регрессорами:

$$F = \frac{1}{2} \left(\frac{t_1^2 + t_2^2 - \hat{\rho}_{t_1, t_2} t_1 t_2}{1 - \hat{\rho}_{t_1, t_2}^2} \right)$$

где $\hat{\rho}_{t_1, t_2}$ - оценка коэффициента корреляции между t_1 и t_2 .

* статистика корректирует наличие корреляции между t_1 и t_2

Для случая гомоскедастичных ошибок F-статистику можно вычислить проще:

Тогда

$$F = \frac{(R_{UR}^2 - R_R^2) / q}{(1 - R_{UR}^2) / (n - k_{UR} - 1)}$$

где $R_R^2 - R^2$ в регрессии в ограничениями $R_{UR}^2 - R^2$ в регрессии без ограничений q - число ограничений в нулевой гипотезе k_{UR} - число объясняющих переменных (без константы) в модели без ограничений.

RESET тест Рамсея (Ramsey's RESET test) для проверки гипотезы о существовании упущенных переменных.

$$y_i = \beta_0 + \beta_1 + \beta_1 x_{1i} + \dots + \beta_{k-1} x_{k-1,i} + \epsilon_i \text{ (restricted model)}$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_{k-1} x_{k-1,i} + \gamma_1 \hat{y}_i^2 + \dots + \gamma_q \hat{y}_i^{(q+1)} + \epsilon_i \text{ (unrestricted model)}$$

$$H_0 : \gamma_1 = \dots = \gamma_q = 0$$

$$H_a : \text{at least one } \gamma_j \neq 0, j = 1, \dots, q$$

$$F = \frac{(RSS_R - RSS_{UR}) / q}{RSS_{UR} / (n - k - q)} = \frac{(R_{UR}^2 - R_R^2) / q}{(1 - R_{UR}^2) / (n - k - q)} \sim F(q, n - k - q)$$

k - number of estimated coefficients in a model we are conducting specification test for, q - number of included powers of \hat{y} .

If $F_{obs} > F_{1-\alpha, q, n-k-q}$ null is rejected, model suffers from misspecification, and, otherwise, if $F_{obs} \leq F_{1-\alpha, q, n-k-q}$, there is **not enough information to reject null**.

Тема 2. Прогнозирование по регрессионной модели

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

$$\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1}$$

Такой прогноз будет несмещенным и эффективным (т.е. будет характеризоваться минимальной ожидаемой квадратичной ошибкой прогноза).

$$E(\hat{y}_{n+1}) = E(\hat{\beta}_1 + \hat{\beta}_2 x_{n+1}) = E(\hat{\beta}_1) + E(\hat{\beta}_2) x_{n+1} = \beta_1 + \beta_2 x_{n+1}$$

$$\text{Кроме того: } E(y_{n+1}) = E(\beta_1 + \beta_2 x_{n+1} + \varepsilon_{n+1}) = \beta_1 + \beta_2 x_{n+1} + E(\varepsilon_{n+1}) = \beta_1 + \beta_2 x_{n+1}$$

$$\text{Следовательно, } E(y_{n+1}) = E(\hat{y}_{n+1}).$$

Кроме самого прогноза нас интересует его точность:

$$\begin{aligned} E(\hat{y}_{n+1} - y_{n+1})^2 &= E\left(\left(\hat{\beta}_1 - \beta_1\right) + \left(\hat{\beta}_2 - \beta_2\right) x_{n+1} - \varepsilon_{n+1}\right)^2 = \\ &= \text{var}(\hat{\beta}_1) + x_{n+1}^2 \text{var}(\hat{\beta}_2) + \sigma^2 + 2x_{n+1} \text{cov}(\hat{\beta}_1, \hat{\beta}_2) - 0 - 0 = \\ &= \frac{\sigma^2}{n} \cdot \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2} + x_{n+1}^2 \frac{\sigma^2}{\sum (x_i - \bar{x})^2} + \sigma^2 - 2x_{n+1} \bar{x} \cdot \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \end{aligned}$$

$$\text{Стандартная ошибка прогноза: } \delta = \sqrt{S^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}$$

$$\text{Доверительные интервалы для прогнозных значений: } (\hat{y}_{n+1} - \delta \cdot t_{n-2}^\alpha, \hat{y}_{n+1} + \delta \cdot t_{n-2}^\alpha)$$

Тема 3. Метод максимального правдоподобия. Тесты Вальда, отношения правдоподобия, множителей Лагранжа

Метод максимального правдоподобия.

Evaluating the joint density at the observed data sample $\mathbf{y} = (y_1, y_2, \dots, y_n)$ gives a real-valued function,

$$L_n(\theta) = L_n(\theta; \mathbf{y}) = f_n(\mathbf{y}; \theta)$$

which is called the likelihood function. For independent and identically distributed random variables, $f_n(\mathbf{y}; \theta)$ will be the product of univariate density functions.

The goal of maximum likelihood estimation is to find the values of the model parameters that maximize the likelihood function over the parameter space, that is

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta; \mathbf{y})$$

Свойства оценок метода максимального правдоподобия.

As the sample size increases to infinity, sequences of maximum likelihood estimators have these properties:

- Consistency: the sequence of MLEs converges in probability to the value being estimated.
- Functional Invariance: If $\hat{\theta}$ is the maximum likelihood estimator for θ , and if $g(\theta)$ is any transformation of θ , then the maximum likelihood estimator for $\alpha = g(\theta)$ is $\hat{\alpha} = g(\hat{\theta})$.
- Efficiency, i.e. it achieves the Cramér-Rao lower bound when the sample size tends to infinity. This means that no consistent estimator has lower asymptotic mean squared error than the MLE (or other estimators attaining this bound), which also means that MLE has asymptotic normality.
- Second-order efficiency after correction for bias.

Связь между МНК и ММП.

Однако точка максимума $\ln [L(\tilde{S}_t, \beta_0, \dots, \beta_n)]$ совпадает с точкой минимума функции $Q(\tilde{S}_t, \beta_0, \beta_1, \dots, \beta_n)$, оптимизируемой в методе МНК, что позволяет сделать вывод о эквивалентности ММП и МНК

Проверка гипотез с помощью теста Вальда:

Wald tests are based on the same principle as t tests in that they evaluate whether the discrepancy between the maximum likelihood estimate θ and the hypothetical value θ_0 is significant, taking account of the variance in the estimate. The test statistic for the null hypothesis $H_0 : \hat{\theta} - \theta_0 = 0$ is:

$$\frac{(\hat{\theta} - \theta_0)^2}{\hat{\sigma}_{\hat{\theta}}^2}$$

where $\hat{\sigma}_{\hat{\theta}}^2$ is the estimate of the variance of θ evaluated at the maximum likelihood value. $\hat{\sigma}_{\hat{\theta}}^2$ can be estimated in various ways that are asymptotically equivalent if the likelihood function has been specified correctly. A

common estimator is that obtained as minus the inverse of the second differential of the log-likelihood function evaluated at the maximum likelihood estimate. Under the null hypothesis that the restriction is valid, the test statistic has a chi-squared distribution with one degree of freedom. When there are multiple restrictions, the test statistic becomes more complex and the number of degrees of freedom is equal to the number of restrictions.

$$W_T = \frac{[\hat{\theta} - \theta_0]^2}{1/I_n(\hat{\theta})} = I_n(\hat{\theta}) [\hat{\theta} - \theta_0]^2$$

The Wald test (also called the Wald Chi-Squared Test) is a way to find out if explanatory variables in a model are significant. "Significant" means that they add something to the model; variables that add nothing can be deleted without affecting the model in any meaningful way. The test can be used for a multitude of different models including those with binary variables or continuous variables.

If the Wald test shows that the parameters for certain explanatory variables are zero, you can remove the variables from the model.

If the test shows the parameters are not zero, you should include the variables in the model.

The Wald test is usually talked about in terms of chi-squared, because the sampling distribution (as n approaches infinity) is usually known. This variant of the test is sometimes called the Wald Chi-Squared Test to differentiate it from the Wald Log-Linear Chi-Square Test, which is a non-parametric variant based on the log odds ratios.

Тест отношения правдоподобия

A likelihood ratio test compares the value of the likelihood function at $\theta = \hat{\theta}$ with its value at $\theta = \theta_0$. In view of the definition of $\hat{\theta}$, $L(\hat{\theta}) \geq L(\theta_0)$ for all θ_0 . However, if the null hypothesis is true, the ratio $L(\hat{\theta})/L(\theta_0)$ should not be significantly greater than 1. As a consequence, the logarithm of the ratio:

$$\log \left(\frac{L(\hat{\theta})}{L(\theta_0)} \right) = \log L(\hat{\theta}) - \log L(\theta_0)$$

should not be significantly different from zero. In that it involves a comparison of the measures of goodness of fit for unrestricted and restricted versions of the model, the LR test is similar to an F test. Under the null hypothesis, it can be shown that in large samples the test statistic:

$$LR = 2 \left(\log L(\hat{\theta}) - \log L(\theta_0) \right)$$

has a chi-squared distribution with one degree of freedom. If there are multiple parameters of interest, and multiple restrictions, the number of degrees of freedom is equal to the number of restrictions.

Тест множителей Лагранжа.

Wald tests are based on the same principle as t tests in that they evaluate whether the discrepancy between the maximum likelihood estimate $\hat{\theta}$ and the hypothetical value θ_0 is significant, taking account of the variance in the estimate. The test statistic for the null hypothesis $H_0 : \hat{\theta} - \theta_0 = 0$ is:

$$\frac{(\hat{\theta} - \theta_0)^2}{\hat{\sigma}_{\hat{\theta}}^2}$$

where $\hat{\sigma}_{\hat{\theta}}^2$ is the estimate of the variance of θ evaluated at the maximum likelihood value. $\hat{\sigma}_{\hat{\theta}}^2$ can be estimated in various ways that are asymptotically equivalent if the likelihood function has been specified correctly. A common estimator is that obtained as minus the inverse of the second differential of the log-likelihood function evaluated at the maximum likelihood estimate. Under the null hypothesis that the restriction is valid, the test statistic has a chi-squared distribution with one degree of freedom. When there are multiple restrictions, the test statistic becomes more complex and the number of degrees of freedom is equal to the number of restrictions.

If the null hypothesis is true, the likelihood ratio test, the Wald test, and the Score test are asymptotically equivalent tests of hypotheses. When testing nested models, the statistics for each test then converge to a Chi-squared distribution with degrees of freedom equal to the difference in degrees of freedom in the two models. If the null hypothesis is not true, however, the statistics converge to a noncentral chi-squared distribution with possibly different noncentrality parameters.

Тема 4. Гетероскедастичность

Случайная ошибка u_i называется гомоскедастичной, если условная дисперсия u_i относительно X_i постоянна для $i = 1, 2, \dots, n$ (т.е. $\text{var}(u_i | X_i = x) = \text{const}$, $i = 1, \dots, n$). В частности, условная дисперсия u_i относительно X_i не зависит от X_i .

Экономические причины гетероскедастичности:

- Неоднородность исследуемых объектов (например, если исследуется зависимость спроса от дохода потребителя, то обнаруживается, что чем больше доход, тем больше индивидуальное значение спроса колеблется относительно ожидаемого значения)
- Характер наблюдений (например, временной ряд)

Последствия гетероскедастичности:

- МНК оценки коэффициентов – несмещенные, состоятельные и асимптотически нормальные как для в случае гомоскедастичности, так и в случае гетероскедастичности ошибок
- МНК оценки эффективны только в случае гомоскедастичности ошибок

Методы борьбы с гетероскедастичностью:

- Моделировать (ВМНК)
- Устойчивые к гетероскедастичности стандартные ошибки Эйкера-Хьюбера-Уайта

Устойчивые к гетероскедастичности стандартные ошибки Эйкера-Хьюбера-Уайта (состоятельные при выполнении трех предположений МНК):

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n \hat{H}_i^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n \hat{H}_i^2 \right]^2}$$

где $\hat{H}_i = 1 - \left[\bar{X} / \frac{1}{n} \sum_{i=1}^n X_i^2 \right] X_i$

Робастная (устойчивая) к гетероскедастичности оценка ковариационной матрицы - Вместо $\widehat{\text{Var}}(\hat{\beta} | X) = \frac{RSS}{n-k} (X'X)^{-1}$ использовать $\widehat{\text{Var}}_{HC}(\hat{\beta} | X) = (X'X)^{-1} X' \hat{\Omega} X (X'X)^{-1}$ - Уайт, 1980, НС0: $\hat{\Omega} = \text{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2)$ - Современный вариант $\hat{\Omega} = \text{diag}\left(\frac{\varepsilon_1^2}{(1-h_{11})^2}, \dots, \frac{\varepsilon_n^2}{(1-h_{nn})^2}\right)$

Взвешенный метод наименьших квадратов:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Пусть

$$\text{var}(u_i | X_i) = \lambda h(X_i)$$

где λ - некоторая константа, $h(X_i)$ - известная функция. Рассмотрим преобразованные переменные:

$$\tilde{Y}_i = \frac{Y_i}{\sqrt{h(X_i)}}; \tilde{X}_{0i} = \frac{1}{\sqrt{h(X_i)}}; \tilde{X}_{1i} = \frac{X_i}{\sqrt{h(X_i)}}; \text{ и } \tilde{u}_i = \frac{u_i}{\sqrt{h(X_i)}}$$

Тогда (1) примет вид

$$\tilde{Y}_i = \beta_0 \tilde{X}_{0i} + \beta_1 \tilde{X}_{1i} + \tilde{u}_i$$

$\hat{\beta}_1$ - ВМНК-оценка коэффициента наклона в уравнении (1) и она BLUE

Заметим, что

$$\begin{aligned} \text{var}(\tilde{u}_i | X_i) &= \text{var}\left(\frac{u_i}{\sqrt{h(X_i)}} | X_i\right) = \\ &= \frac{\text{var}(u_i | X_i)}{h(X_i)} = \frac{\lambda h(X_i)}{h(X_i)} = \lambda = \text{const} \end{aligned}$$

НО! $h(X_i)$ - в реальной жизни неизвестна \rightarrow такая оценка называется недоступной ВМНК - оценкой
Доступный ВМНК - алгоритм:

1. Оцените регрессию (1) на при помощи МНК и вычислите остатки $\hat{u}_i, i = 1, \dots, n$
2. Оцените предполагаемую модель функции условной дисперсии $\text{var}(u_i | X_i)$ (оцениваем \hat{u}_i^2 на набор переменных, от которых она зависит).
3. Используйте оцененную функцию для расчета предсказанных значений функции условной дисперсии $\widehat{\text{var}}(u_i | X_i)$
4. Умножьте зависимую переменную и регрессоры (в том числе и свободный член) на величину, обратную квадратному корню из оцененной функции условной дисперсии (см. (2)).
5. Оцените коэффициенты взвешенной регрессии (3) при помощи МНК; полученные оценки являются ВМНК-оценками (1).

ВМНК или устойчивые к гетероскедастичности стандартные ошибки:

1. ВМНК:
 - «+» более эффективные, чем МНК, оценки коэффициентов с использованием сырых (не скорректированных) регрессоров;

- «-» необходимо знать функцию условной дисперсии и оценки ее параметров, что может быть затруднительно на практике;
- «-» если функциональная форма дисперсии неверна, то ВМНК с.о. регрессии являются неверными в том смысле, что они приводят к неправильным статистическим выводам;

2. Устойчивые к гетероскедастичности с.о.:

- «+» асимптотически верные выводы, даже если вид функции условной дисперсии неизвестен;
- «+» легко вычисляются;
- «-» не эффективные по сравнению с ВМНК-оценкой (полученной на основе верной функции условной дисперсии), по крайней мере, асимптотически.

Обобщенный метод наименьших квадратов:

Generalized Least Square (GLS) So far, we have been dealing with heteroskedasticity under OLS framework. But if we knew the variance-covariance matrix of the error term, then we can make a heteroskedastic model into a homoskedastic model. As we defined before

$$E(uu') = \sigma^2 \Omega = \Sigma$$

Define further that

$$\Omega^{-1} = P'P$$

P is a "n x n" matrix Pre-multiply P on a regression model

$$Py = PX\beta + Pu$$

or

$$\tilde{y} = \tilde{X}\beta + \tilde{u}$$

In this model, the variance of \tilde{u} is

$$E(\tilde{u}\tilde{u}') = E(Puu'P') = PE(uu')P' = P\sigma^2\Omega P' = \sigma^2 P\Omega P' = \sigma^2 I$$

Note that $P\Omega P' = I$, because define $P\Omega P' = A$, then $P'P\Omega P' = P'A$. By the definition of P , $\Omega^{-1}\Omega P' = P'A$, thus $P' = P'A$. Therefore, A must be I .

Weighted Least Squares Estimation (WLS) Consider a general case of heteroskedasticity.

$$\text{Var}(u_i) = \sigma_t^2 = \sigma^2 \omega_i$$

Then,

$$E(uu') = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 \\ 0 & \omega_2 & 0 \\ 0 & 0 & \omega_n \end{bmatrix} = \sigma^2 \Omega, \text{ thus } \Omega^{-1} = \begin{bmatrix} \omega_1^{-1} & 0 & 0 \\ 0 & \omega_2^{-1} & 0 \\ 0 & 0 & \omega_n^{-1} \end{bmatrix}$$

Тестирование гетероскедастичности:

- Графические методы: график остатков модели \hat{u}_i , график стандартизованных остатков $c_i = \frac{\hat{u}_i}{s_{\hat{u}}}$ от
 - оцененных значений \hat{y}_i : позволяют выявить выбросы, возможную гетероскедастичность, неправильную спецификацию модели;
 - отдельных объясняющих переменных: позволяют выявить возможную нелинейность;
 - номера наблюдения (последовательные моменты времени): позволяют выявить изменение дисперсии ошибок во времени, пропуск переменных, автокоррелированность ошибок.
- Формальные процедуры: тесты Голдфелда-Квандта (Goldfeld-Quandt), Глейзера (Glejser), Бройша-Пагана (Breusch-Pagan).

Тесты Голдфелда-Квандта (Goldfeld-Quandt)

1. Упорядочиваем наблюдения по предполагаемому возрастанию дисперсий ошибок (по предполагаемой переменной);
2. Отбрасываем r центральных наблюдений;
3. Оцениваем выбранную модель отдельно на первом и последнем интервале;

4. Вычисляем $F^{act} = \frac{RSS_2}{RSS_1}$;

5. Гипотеза $H_0 : \text{var}(u_i | X_i) = \text{const}, i = 1, n$ (гомоскедастичность) отвергается, если $F^{act} > F_{1-\alpha}(\frac{n-r}{2} - 2; \frac{n-r}{2} - 2)$

Тест Глейзера (Glejser)

$$\text{var}(u_i | X_i) = \theta_0 + \theta_1 X_i^\gamma$$

1. Оцените регрессию (1) на при помощи МНК и вычислите остатки $\hat{u}_i, i = 1, \dots, n$

2. Оцените регрессию $|\hat{u}_i|$ на константу и X_i^γ , специфицировав γ . (Выбрав модель с самым высоким R^2 . Обычно $\gamma = 1, \gamma = -1, \gamma = 0.5$)

3. $H_0 : \theta_1 = 0, H_1 : \theta_1 \neq 0$

4. Отвергаем H_0 на 5% ур.зн., если $|t^{act}| > 1,96$

Тест Бройша-Пагана (Breusch-Pagan)

Тест Бреуша - Пагана Тестируемая гипотеза в данном тесте состоит в том, что гетероскедастичности в модели нет:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

Тест Бреуша - Пагана Тестируемая гипотеза в данном тесте состоит в том, что гетероскедастичности в модели нет:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

Процедура осуществления теста устроена так: сначала при помощи обычного МНК оценивается исходная модель (для которой мы хотим проверить отсутствие гетероскедастичности) и вычисляются соответствующие остатки e_i . Далее вычисляется вспомогательное значение $\tilde{\sigma}^2 = \frac{1}{n} \sum e_i^2$. После этого необходимо оценить вспомогательное уравнение, в котором справа стоят переменные, потенциально влияющие на дисперсию случайной ошибки:

$$\frac{e_i^2}{\tilde{\sigma}^2} = \gamma_0 + \gamma_1 z_i^{(1)} + \dots + \gamma_p z_i^{(p)} + u_i$$

Далее вычисляется расчетное значение тестовой статистики по формуле: $ESS/2$

Если верна нулевая гипотеза, то указанная статистика асимптотически имеет распределение Хи-квадрат с p степенями свободы. Поэтому, если расчетное значение больше критического значения, взятого из таблиц распределения χ^2 с p степенями свободы для выбранного исследователем уровня значимости, то следует отвергнуть нулевую гипотезу и заключить, что в данных есть гетероскедастичность (необходимые таблицы доступны, например, в приложении к гл. 6). В противном случае можно сделать вывод в пользу гомоскедастичности.

Тема 5. Стохастические регрессоры. Эндогенность. Инструментальные переменные

Предпосылки линейной модели со стохастическим регрессором (случай парной регрессии): 1. Модель линейна по параметрам и правильно специфицирована:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

2. Наблюдения $\{(x_i, y_i), i = 1, \dots, n\}$ независимы и одинаково распределены. 3. x_i и y_i имеют ненулевые конечные четвертые моменты распределения $E(x_i^4) < \infty$, $E(y_i^4) < \infty$. 4. Случайные ошибки имеют нулевое условное математическое ожидание при заданном x : $E(\varepsilon_i | x_i) = 0$.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \equiv \alpha + \beta_1 (X_i - \bar{X}) \quad i = 1, 2, \dots, n$$

Assumptions:

- $\{X_i, Y_i\} (i = 1, 2, \dots, n)$ are IID pairs, with $\text{cov}(X_i, Y_i) \geq 0$, $\text{var}(X_i) = \sigma_X^2 > 0$, $\text{var}(Y_i) = \sigma_Y^2 > 0 \forall i$
- $E(\varepsilon_i | X_h) = 0, \forall i$ and h
- $\text{var}(\varepsilon_i | X_h) = \sigma^2, \forall i$ and h and finite 4 th moments (no excess kurtosis)
- $\text{cov}(\varepsilon_i, \varepsilon_j | X_h) = 0, \forall i \neq j$, and for all h
- β_0, β_1 and σ^2 are constant parameters
- For the purpose of statistical inference in small samples we may invoke normally distributed disturbances: $\varepsilon_i \sim \text{IIN}(0, \sigma^2 | X_h)$.

Обобщение теоремы Гаусса-Маркова на случай стохастических регрессоров (без доказательства).

THE EXTENDED LEAST SQUARES ASSUMPTIONS IN THE MULTIPLE REGRESSION MODEL The linear regression model with multiple regressors is

$$Y_i = X_i' \beta + u_i, i = 1, \dots, n$$

The extended least squares assumptions are

1. $E(u_i | X_i) = 0$ (u_i has conditional mean zero);
2. $(X_i, Y_i), i = 1, \dots, n$ are independently and identically distributed (i.i.d.) draws from their joint distribution;
3. X_i and u_i have nonzero finite fourth moments;
4. X has full column rank (there is no perfect multicollinearity);
5. $\text{var}(u_i | X_i) = \sigma_u^2$ (homoskedasticity); and
6. the conditional distribution of u_i given X_i is normal (normal errors).

The Gauss-Markov conditions for multiple regression are

- (i) $E(\mathbf{U} | \mathbf{X}) = \mathbf{0}_m$
- (ii) $E(\mathbf{U}\mathbf{U}' | \mathbf{X}) = \sigma_u^2 \mathbf{I}_n$, and
- (iii) X has full column rank.

The Gauss-Markov conditions for multiple regression are (i) $E(\mathbf{U} | \mathbf{X}) = \mathbf{0}_m$ (ii) $E(\mathbf{U}\mathbf{U}' | \mathbf{X}) = \sigma_u^2 \mathbf{I}_n$, and (iii) X has full column rank.

Suppose that the Gauss-Markov conditions for multiple regression in Equation (18.38) hold. Then the OLS estimator $\tilde{\beta}$ is BLUE. That is, let $\tilde{\beta}$ be a linear conditionally unbiased estimator of β and let c be a nonrandom $k + 1$ dimensional vector. Then $\text{var}(c'\tilde{\beta} | X) \leq \text{var}(c'\beta | X)$ for every nonzero vector c . where the inequality holds with equality for all c only if $\tilde{\beta} = \hat{\beta}$.

Угрозы для внутренней обоснованности

1. пропущенные переменные,
2. неправильная спецификация функциональной формы регрессии,
3. неточное измерение независимых переменных («ошибки в переменных»),
4. отбор наблюдений,
5. одновременная причинность.

Напомним, что в предыдущей главе мы сформулировали два важных определения:

- Эндогенный регрессор - это регрессор, который коррелирован со случайными ошибками модели: $\text{cov}(x_i, \varepsilon_i) \neq 0$.

- Экзогенный регрессор - это регрессор, который не коррелирован со случайными ошибками модели: $\text{cov}(x_i, \varepsilon_i) = 0$.

Могут возникнуть следующие типичные ситуации:

1. Эндогенность регрессора из-за пропуска существенной переменной. В качестве важного частного случая тут также следует указать проблему эндогенности из-за самоотбора.

2. Эндогенность регрессора из-за выбора неверной функциональной формы связи.
3. Эндогенность регрессора из-за двусторонней причинно-следственной связи.
4. Эндогенность регрессора из-за ошибок измерения.

Проблема эндогенности.

Метод инструментальных переменных (instrumental variables, IV).

Представим, что в нашем распоряжении кроме информации о переменных x и y есть еще данные о третьей переменной z , которая удовлетворяет двум свойствам:

- релевантность (relevance): $\text{cov}(z_i, x_i) \neq 0$;
- экзогенность (exogeneity): $\text{cov}(z_i, \varepsilon_i) = 0$.

Первый шаг Оцениваем регрессию переменной x по переменной z :

$$\hat{x}_i = \hat{\theta}_1 + \hat{\theta}_2 z_i$$

Получаем прогнозные значения \hat{x}_i

Второй Шаг Оцениваем регрессию переменной y по этим предсказанным значениям \hat{x} :

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \hat{x}_i$$

Чтобы отличать оценку коэффициента, полученную таким способом, от обычной МНК-оценки, мы иногда будем использовать в обозначении и дополнительный индекс: $\hat{\beta}_2^{\text{TSLS}}$

$$\hat{\beta}_2^{\text{TSLS}} = \frac{\frac{1}{n} \sum (y_i - \bar{y})(z_i - \bar{z})}{\frac{1}{n} \sum (x_i - \bar{x})(z_i - \bar{z})} = \frac{\widehat{\text{cov}}(y, z)}{\widehat{\text{cov}}(x, z)}$$

Для случая гомоскедастичности корректная формула стандартной ошибки коэффициента при регрессоре имеет

$$\text{se}(\hat{\beta}_2) = \sqrt{\frac{S^2}{\sum (x_i - \bar{x})^2} \cdot \frac{1}{\text{corr}^2(x, z)}}, \quad S^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

Векторно-матричная форма записи

$$y_i = \beta_0 + \beta_1 \cdot x_i^{(1)} + \dots + \beta_p x_i^{(p)} + \beta_{p+1} w_i^{(1)} + \dots + \beta_{p+r} w_i^{(r)} + \varepsilon_i$$

$$y = X\beta + \varepsilon$$

На первом шаге двухшагового МНК мы оцениваем регрессию X по Z . Это можно записать так:

$$\hat{X} = Z\hat{\alpha}.$$

По формуле для МНК-оценки из §3.3 $\hat{\alpha} = (Z'Z)^{-1} Z'X$, следовательно:

$$\hat{X} = Z(Z'Z)^{-1} Z'X$$

На втором шаге двухшагового МНК мы оцениваем регрессию y по \hat{X} :

$$\hat{y} = \hat{X}\hat{\beta}^{\text{TSLS}}$$

Таким образом, окончательно имеем следующую формулу для 2МНК-оценки:

$$\hat{\beta}^{\text{TSLS}} = \left(X'Z(Z'Z)^{-1} Z'X \right)^{-1} X'Z(Z'Z)^{-1} Z'y.$$

Если число инструментов в точности совпадает с числом эндогенных регрессоров ($m = p$), то матрицы X и Z имеют одинаковую размерность. В этом случае выражение для оценки $\hat{\beta}^{\text{TSLS}}$ можно упростить:

$$\hat{\beta}^{\text{IV}} = (Z'X)^{-1} Z'y$$

Обобщенный метод моментов

http://www.eco.uc3m.es/~ricmora/mei/materials/Session_12_GMM.pdf

Тема 6. Системы одновременных уравнений

Структурная форма системы одновременных уравнений - в явном виде представлены взаимные связи между входящими в модель переменными

$$\begin{cases} C_t = \alpha + \beta Y_t + \epsilon_t \\ Y_t = C_t + I_t \end{cases}$$

Приведенная форма системы одновременных уравнений - выраженные из системы объясняемые переменные (взаимодействие между переменными представлено неявном, однако коэффициенты приведенной формы отражают итог взаимодействия этих переменных)

$$\begin{cases} C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + \frac{1}{1-\beta} \epsilon_t \\ Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{1}{1-\beta} \epsilon_t \end{cases}$$

Идентифицируемость

Структурное уравнение называется **идентифицируемым**, если его коэффициенты можно выразить через коэффициенты приведённой формы. Если это можно сделать единственным способом, то говорят о точной идентифицируемости, если несколькими способами — о **сверхидентифицируемости**. В противном случае оно называется неидентифицируемым. Сверхидентифицируемость фактически означает, что на коэффициенты приведённой формы наложены некоторые ограничения (сверхидентифицирующие).

Система частично идентифицируема, если коэффициенты одного из уравнений восстанавливаются по коэффициентам приведенной формы.

При рассмотрении условий идентифицируемости отдельных структурных уравнений, предполагается, что переменные, задействованные в системе, подразделяются на три типа:

- эндогенные переменные (эндогенная переменная, входящая в i -е уравнение системы, коррелирована с ошибкой в этом уравнении.) - экзогенные переменные (экзогенные переменные не коррелированы с ошибками во всех уравнениях системы для всех моментов времени.) - предопределенные переменные (значение предопределенной переменной, входящей в уравнение, не должно быть коррелированным со значениями ошибки в этом уравнении, соответствующими моментам $t, t + 1, \dots$

Предполагается, что:

- система состоит из g уравнений - g эндогенных переменных; экзогенных и предопределенных переменных; - каждое из g уравнений нормировано (normalized), так что коэффициент при одной из эндогенных переменных, входящих в уравнение, равен 1.

Условие порядка и условие ранга.

Условия идентифицируемости отдельного структурного уравнения

Поскольку $\Pi \Gamma = B$ и мы использовали это соотношение для восстановления коэффициентов структурных уравнений двух последних систем, вопрос об идентифицируемости структурной формы — это вопрос о возможности однозначного восстановления всех коэффициентов структурной формы, т.е. восстановления матриц Π и Γ , на основании матрицы $B = -1$.

В совокупности матрицы Π и Γ состоят из $g^2 + Kg$ элементов, тогда как в матрице B всего Kg элементов. Поэтому однозначное восстановление коэффициентов структурной формы по коэффициентам приведенной формы невозможно без использования дополнительной информации в виде невключения в отдельные уравнения тех или иных переменных, нормировки коэффициентов, линейных ограничений на параметры структуры.

Пусть матрица Γ размера $(g + 1) * g$ составлена из матриц Γ_i и таким образом, что матрица Γ располагается над матрицей

Коэффициенты при g эндогенных и предопределенных переменных в i -м структурном уравнении составляют i -й столбец Γ_i , матрицы A .

Существенным является то обстоятельство, что коэффициенты i -го структурного уравнения не могут быть восстановлены на основании коэффициентов приведенной формы, если в это уравнение входят все $g + 1$ переменные системы. Поэтому далее будем предполагать, что на элементы вектора Γ_i , помимо нормировочного накладываются еще и некоторые дополнительные однородные линейные ограничения в виде уравнений:

$$\Phi_i a_i = 0$$

где Φ_i — матрица размера $R_i * (g + 1)$, R_i — количество линейных ограничений.

Пусть Γ_i — матрица, получаемая из матрицы Γ вычеркиванием ее i -го столбца так что $\Gamma_i = [\Gamma : \Gamma_i]$.

Тогда $\text{rank}(\Phi_i) = \text{rank}(\Phi_i a_i : \Phi_i \Gamma_i) = \text{rank}(0 : \Phi_i \Gamma_i) = \text{rank}(\Phi_i \Gamma_i)$

Но матрица $\Phi_i \Gamma_i$ имеет размер $R_i * (g - 1)$, и, чтобы ее ранг был равен $(g - 1)$, во всяком случае необходимо выполнение следующего порядкового условия идентифицируемости

Необходимое условие идентифицируемости структурного уравнения (порядковое условие):

Количество переменных правой части уравнения должно быть не больше количества всех экзогенных переменных системы.

В канонической форме: количество исключенных из данного уравнения экзогенных переменных должно быть не меньше количества включенных эндогенных переменных уравнения минус единица.

$$R_i \geq g - 1$$

Достаточное условие идентифицируемости структурного уравнения (ранговое условие):

Ранг матрицы, составленной из коэффициентов (в других уравнениях) при переменных, отсутствующих в данном уравнении, не меньше общего числа эндогенных переменных системы минус единица:

$$\text{rank}(\Phi_i A) = g - 1$$

- $\text{rank}(\Phi_i A) < g - 1$ - уравнение неидентифицируемо - $\text{rank}(\Phi_i A) = g - 1$ и $R_i = g - 1$ уравнение идентифицируемо точно - $\text{rank}(\Phi_i A) = g - 1$ и $R_i > g - 1$ - уравнение сверхидентифицируемо

В ситуациях 2 и 3, если рассматривать задачу восстановления коэффициентов i -го структурного уравнения на основании МНК-оценок коэффициентов приведенной формы каждого отдельного уравнения приведенной системы, не учитывая ограничения на коэффициенты приведенной формы, то по $\hat{\Pi}$ в ситуации 2 структурного уравнения восстанавливаются однозначным образом, тогда как в ситуации 3 существует несколько вариантов восстановления.

Оценивание систем одновременных уравнений.

Идентифицируемость системы структурных уравнений в целом (на основании приведенной формы системы) означает идентифицируемость не только всех коэффициентов системы, но и ковариационной матрицы случайных ошибок, входящих в правые части уравнений системы.

При этом при восстановлении коэффициентов и ковариационной матрицы ошибок в структурной форме используются не только коэффициенты приведенной формы, но и ковариационная матрица ошибок в приведенной форме.

Пусть $E(u_t) = 0$, $\Sigma = (\text{Cov}(u_{it}, u_{tj})) = (\sigma_{ij})$ - диагональная матрица

Ошибки не коррелированы по времени, но для одного и того же момента ошибки в разных уравнениях могут быть коррелированными между собой:

Тогда $E(w_t) = 0$, $\Omega = (\text{Cov}(w_{it}, w_{tj})) = \text{Cov}(u_t \Gamma^{-1}) = (\Gamma^{-1})^T \Sigma (\Gamma^{-1})$

$$\Sigma = \Gamma^T \Omega \Gamma$$

Следовательно, если структурная система идентифицируема (коэффициенты структурной системы однозначно восстанавливаются по коэффициентам приведенной формы), то тогда, восстановив по коэффициентам приведенной формы матрицу Γ , можно, используя эти восстановленные коэффициенты и матрицу Ω , восстановить ковариационную матрицу Σ .

Если структурная форма не восстанавливается целиком, а возможно лишь восстановление некоторых ее уравнений, то тогда для полной идентификации i -го стохастического структурного уравнения надо восстановить все его коэффициенты и дисперсию случайной составляющей этого уравнения.

Рекурсивная СОУ

Рекурсивная система - благодаря последовательному определению эндогенных переменных в таких системах при переходе от уравнения к уравнению, в правых частях каждого из уравнений системы не оказываются переменных, значения которых коррелированы со значением ошибки в этом уравнении при одном и том же t .

Косвенный метод наименьших квадратов (ILS)

Если i -е стохастическое уравнение структурной формы идентифицируемо точно, то параметры этого уравнения (коэффициенты уравнения и дисперсия случайной ошибки) восстанавливаются по параметрам приведенной системы однозначно. Поэтому для оценивания параметров такого уравнения достаточно оценить коэффициенты каждого из уравнений приведенной формы с помощью МНК (отдельно для каждого уравнения) и получить оценку ковариационной матрицы Σ ошибок в приведенной форме, после чего воспользоваться соотношениями $\Gamma = \Sigma^{-1} \Omega \Gamma$ по $\hat{\Pi}, \hat{\Omega}$.

Полученные в результате оценки коэффициентов i -го стохастического уравнения структурной формы наследуют свойство состоятельности оценок приведенной формы. Однако они не наследуют таких свойств оценок приведенной формы, как несмещенность и эффективность, из-за того что получаются в результате нелинейных преобразований.

Двухшаговый метод наименьших квадратов (2SLS)

Рассмотрим систему g одновременных уравнений:

$$y_t \Gamma = x_t B + u_t$$

Считая, что первое уравнение нормировано на коэффициент при переменной y_{t1} , преобразуем уравнение к виду:

$$y_{t1}\Gamma = \alpha_{11}y_{t1}^* + \dots + \alpha_{g1}y_{tg}^* + \theta_{11}x_{t1}^* + \dots + \theta_{K1}x_{tK}^* + u_{t1}$$

Состоятельному оцениванию коэффициентов уравнения мешает эндогенность переменных $y_{t1}^*, \dots, y_{tg1}^*$.

1 шаг:

Производится оценивание уравнений регрессии каждой из этих переменных на все предопределенные переменные, включенные в систему.

$$Y_i = X\Pi_i + W_i$$

Оценив отдельные уравнения с помощью МНК, получим оценку \hat{Y}_i , матрицы \hat{X}_i и на ее основе — оценку матрицы \hat{Y}_i , в виде $\hat{Y}_i = \hat{X}_i$

Матрица \hat{Y}_i , содержит значения эндогенных переменных, включенных в правую часть i -го структурного уравнения, «очищенных» от корреляции с ошибкой в этом уравнении.

2 шаг:

Значения эндогенных переменных в первом уравнении $y_{t1}^*, \dots, y_{tg1}^*$ заменяются значениями $y_{t1}^*, \dots, y_{tg1}^*$. Полученное уравнение оценивается МНК.

Запишем уравнение в виде:

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = \hat{Y}_i\alpha_i + X_i\theta_i + \epsilon_i = \hat{Z}_i\delta_i + \epsilon_i$$

где $\epsilon_i = (Y_i - \hat{Y}_i)\alpha_i + u_i$

Вычислим МНК-оценки вектора δ_i :

$$\hat{\delta}_i^{2SLS} = (\hat{Z}_i^T \hat{Z}_i)^{-1} \hat{Z}_i^T y_i$$

Для состоятельности $\hat{\delta}_i^{2SLS}$ требуется еще, чтобы предельная матрица имела конечные элементы и была невырожденной, а для этого матрица \hat{Z}_i должна иметь полный столбцовый ранг.

$$p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \hat{Z}_i^T \hat{Z}_i \right) = Q_i$$

Для построения t -статистик и доверительных интервалов, значения стандартных ошибок коэффициентов должны быть скорректированы

Проверка адекватности

Применив 2SLS к i -му структурному уравнению, получим вектор остатков $\hat{u}_i^{2SLS} = y_i - \hat{y}_i^{2SLS}$, опираясь на который можно обычным образом проверять адекватность уравнения, используя критерии: - нормальности (Харке — Бера); - линейности (добавляя степени и перекрестные произведения предопределенных переменных) - гомоскедастичности (Уайта, Пагана — Холла для выявления гетероскедастичности в отдельном уравнении системы); - автокоррелированности остатков (Бройша — Годфри)

Тема 7. Бинарные объясняемые переменные. Логит и пробит модели

Классическая линейная модель:

$$y_i = \theta_1 x_{i1} + \dots + \theta_p x_{ip} + u_i$$

Попытка оценить такую модель МНК наталкивается на определенные трудности:

При обычном предположении:

$$E(y_i | x_i) = x_i^T \theta$$

В то же время, поскольку y_i принимает только значения — 0 и 1, ее условное математическое ожидание (при заданном значении x_i) равно:

$$E(y_i | x_i) = \{y_i = 1 | x_i\}$$

Таким образом

$$x_i^T \theta = \{y_i = 1 | x_i\}$$

- вероятность, а значит, она должна быть в рамках $[0,1]$

$$Var(\epsilon_i | x_i) = x_i^T \theta (1 - x_i^T \theta)$$

Также возникает проблема гетероскедастичности, осложненная еще и тем, что в выражения для дисперсий входит (неизвестный) вектор параметров θ .

Коэффициент θ практически всегда является неинтерпретируемым.

Логит-, пробит-, гомпит-модели

$$y_i = G(\theta_1 x_{i1} + \dots + \theta_p x_{ip}) + u_i = G(x_i^T \theta) + u_i$$

Предположим, что при фиксированных значениях объясняющих переменных, случайные ошибки статистически независимы, так что функция правдоподобия параметров имеет вид:

$$L(\theta|x) = \prod_i (G(x_i^T \theta))^{y_i} (1 - G(x_i^T \theta))^{1-y_i}$$

Максимизируя логарифмическую функцию правдоподобия, получаем оценки $\hat{\theta}$.

- Пробит-модель - функция стандартного нормального распределения

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

- Логит-модель - функция стандартного логистического распределения

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

- Гомпит-модель - функция стандартного распределения экстремальных значений (минимума) I типа (распределение Гомпертца)

$$G(z) = 1 - \exp(-e^z)$$

Интерпретация коэффициентов

Поскольку модели логит, пробит и гомпит являются нелинейными, оцененные коэффициенты в этих моделях имеют интерпретацию, отличающуюся от интерпретации коэффициентов в линейной модели.

Пусть k -я объясняющая переменная является непрерывной переменной. Тогда предельный эффект (marginal effect) этой переменной определяется как производная:

$$\frac{\partial P\{y_i = 1|x_i\}}{\partial x_{ik}} = \frac{\partial G(x_i^T \theta)}{\partial x_{ik}}$$

и в отличие от линейной модели этот эффект зависит от значений объясняющих переменных для i -го субъекта.

$$\Delta P\{y_i = 1|x_i\} = \frac{\partial P\{y_i = 1|x_i\}}{\partial x_{ik}} \Delta x_{ik} = \frac{\partial G(x_i^T \theta)}{\partial x_{ik}} \Delta x_{ik}$$

В случае когда сама объясняющая переменная - дамми-переменная, предельный эффект определяют просто как разность

$$\Delta P\{y_i = 1|x_i, d_i = 1\} - \Delta P\{y_i = 1|x_i, d_i = 0\}$$

Пусть — вероятность некоторого события. Отношение шансов $\frac{p}{1-p}$

Логарифм отношения шансов называют логитом: $\text{logit}() = \ln \frac{p}{1-p}$

Если $\text{logit}(p) > 0$, то больше шансов, что событие А произойдет. Если $\text{logit}(p) < 0$, то больше шансов, что событие А не произойдет.

Логит-модель линейна в отношении логита. Отсюда вытекает, что изменение значения k -й объясняющей переменной на величину Δx_{ik} приводит (при неизменных значениях остальных объясняющих переменных) к изменению значения логита на $\theta_k \Delta x_{ik}$

ROC –кривая.

<https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>

Тема 8. Модели панельных данных

Панельные данные (panel data) - данные о значениях переменных и для N субъектов в T последовательных моментов времени.

$$\{i, t; i = 1, \dots, N, t = 1, \dots, T\}$$

В общем случае является вектором конечной размерности, и наиболее общей формой линейной модели наблюдений для такой ситуации являлась бы спецификация

$$y_{it} = x_{it}^u + \theta_{it} + u_{it}$$

где θ_{it} измеряет частное влияние x_{it} в момент t для субъекта i .

Модели сквозной регрессии.

Модель пула ($\theta_{it} = \theta$) :

$$y_{it} = x_{it}^T \theta + u_{it}$$

в которой предполагается, что

$$u_{it} \sim i.i.d, E(x_{it}u_{js}) = 0 \text{ for } \forall i, j = 1, \dots, N, t, s = 1, \dots, T$$

так что является экзогенной переменной.

В этом случае имеем дело с обычной линейной регрессией с NT наблюдениями, удовлетворяющей предположениям классической нормальной линейной модели.

Эффективные оценки параметров модели пула:

Для получения эффективных оценок вектора коэффициентов в достаточно использовать обычный метод наименьших квадратов (OLS). При соответствующих предположениях о поведении значений объясняющих переменных, когда $N \rightarrow \infty$ или $T \rightarrow \infty$, эта оценка является также состоятельной оценкой этого вектора.

Модели с фиксированными эффектами.

Обратимся теперь к методам анализа панельных данных, предназначенным в основном для анализа данных, в которых количество субъектов исследования N велико, а количество наблюдений над каждым субъектом мало.

Модели со скалярной объясняющей переменной

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}$$

равносильна

$$y_{it} = \sum \alpha_i d_{ij} + \beta x_{it} + u_{it}$$

где $d_{ij} = 1$, если $j = i$

Здесь α_i - неизвестные фиксированные эффекты.

$$u_{it} \sim N(0, \sigma_u^2), E(x_{it}u_{js}) = 0 \quad \forall i, j = 1, \dots, N, t, s = 1, \dots, T$$

так что x является экзогенной переменной.

Модели с фиксированными эффектами наиболее подходят для случаев, когда субъектами исследования выступают страны, крупные компании или предприятия. Эти эффекты по существу, отражают наличие у субъектов индивидуальных характеристик, не изменяющихся со временем в процессе наблюдений, которые трудно или даже невозможно наблюдать или измерить. Если значения таких характеристик не наблюдаются, то эти характеристики невозможно непосредственно включить в правые части уравнений регрессии в качестве объясняющих переменных, тогда чтобы исключить смещение оценок из-за «пропущенных переменных», в правые части уравнений вместо значений ненаблюдаемых индивидуальных характеристик как раз и вводятся ненаблюдаемые эффекты α_i .

FE-оценка (within-estimator)

Оценки для фиксированных эффектов вычисляются по формуле:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta} \bar{x}_i$$

OLS-оценка имеет вид:

$$\hat{\beta}_{CV} = \frac{\sum (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum (x_{it} - \bar{x}_i)^2}$$
$$Var(\hat{\beta}_{CV}) = \frac{\sigma_u^2}{\sum (x_{it} - \bar{x}_i)^2}$$

Оценка имеет одно и то же значение как в рамках статистической модели с дамми-переменными

$$y_{it} = \sum \alpha_i d_{ij} + \beta x_{it} + u_{it}$$

и в рамках модели отклонения от групповых средних

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)$$

Свойства получаемых оценок

При сделанных предположениях $\hat{\beta}_{CV}$ является BLUE.

- $\hat{\beta}_{CV}$ является состоятельной оценкой и когда $N \rightarrow \infty$, и когда $T \rightarrow \infty$ (если нас интересует только состоятельность $\hat{\beta}_{CV}$, но не ее эффективность, то для этого не требуется строгая экзогенность x ; в этом случае достаточно экзогенности в рамках каждого отдельного субъекта исследования)

- $\hat{\alpha}_i$ состоятельна только тогда, когда $T \rightarrow \infty$, т.к. оценивание каждого α_i производится фактически лишь по наблюдениям, так что при фиксированном с ростом N происходит лишь увеличение количества параметров α_i , что не приводит к повышению точности оценивания каждого конкретного α_i

Модели со случайными эффектами.

Модель со случайными эффектами (variance components model)

Модель

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}$$

равносильна

$$y_{it} = \mu + \alpha_i + \beta x_{it} + u_{it}$$

где $\sum \alpha_i = 0$, α_i - дифференциальные эффекты.

В ряде ситуаций субъекты могут рассматриваться как случайная выборка из некоторой более широкой ГС, и исследователя интересуют обезличенные субъекты, имеющие заданные характеристики. В таких ситуациях предполагается, что α_i - случайные величины или случайные эффекты, которые не интерпретируются и не подлежат оцениванию. Вместо этого оцениваются параметры распределения случайных величин α_i :

$$y_{it} = \mu + \beta x_{it} + (\alpha_i + u_{it}) = \mu + \beta x_{it} + v_{it}$$

где ошибка $v_{it} = \alpha_i + u_{it}$ состоит из двух компонент.

Случайные эффекты α_i отражают наличие у субъектов исследования индивидуальных характеристик, не изменяющихся со временем в процессе наблюдений, которые невозможно измерить. Однако теперь их значения встраиваются в состав случайной ошибки.

Предположим:

$$u_{it} \sim N(0, \sigma_u^2), E(x_{it}u_{js}) = 0 \quad \forall i, j = 1, \dots, N, t, s = 1, \dots, T$$

а также:

$$E(\alpha_i) = 0$$

$$E(\alpha_i \alpha_j) = \begin{cases} 0 & \text{для } i \neq j \\ \sigma_{ij}^2 & \text{для } i = j \end{cases}$$

$E(x_{it} \alpha_j) = 0$, так в RE-модели x экзогенна

Если $E(u_{it} \alpha_i) = 0$, то условная относительно x_{it} дисперсия случайной величины v_{it} складывается из двух некоррелированных компонент (variance components):

$$Var(v_{it}|x_{it}) = \sigma_{\alpha}^2 + \sigma_u^2$$

В векторной форме эта модель имеет вид:

$$y_i = [ex_i]\delta + v_i$$

Заметим, что случайные величины $v_{it}v_{is}$ коррелированы, даже если не коррелированы ошибки u_{it} :

$$V = E(v_i v_i^T) = \sigma_u^2 I_T + \sigma_{\alpha}^2 e e^T$$

При этом выполняется предположение равной коррелированности в модели компонент дисперсии.

Оценка RE-модели

В RE - модели оценка OLS-оценка остается несмещенной и состоятельной оценкой для β , но перестает быть эффективной, поскольку не учитывает коррелированность v_{it} во времени для субъекта i .

$$\hat{\beta}_{CV} = \frac{\sum (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum (x_{it} - \bar{x}_i)^2}$$

GLS-оценка, учитывающая эту коррелированность, будет более эффективной:

$$\hat{\beta}_{GLS} = \frac{\sum (x_{it}^* - \bar{x}_i^*)(y_{it}^* - \bar{y}_i^*)}{\sum (x_{it}^* - \bar{x}_i^*)^2} = w\hat{\beta}_b + (1 - w)\hat{\beta}_{CV}$$

$$\text{где } \hat{\beta}_b = \frac{\sum (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum (\bar{x}_i - \bar{x})^2}$$

$\hat{\beta}_b$ — «межгрупповая» оценка (between estimator), соответствующая регрессии средних значений \bar{y}_i на константу и средние значения \bar{x}_i («модель для групповых средних», игнорирующая внутригрупповую изменчивость):

$$\bar{y}_i = \mu + \beta\bar{x}_i + \bar{v}_i$$

Таким образом, $\hat{\beta}_{GLS}$ RE-модели учитывает и внутригрупповую, и межгрупповую изменчивость. Она является взвешенным средним between-оценки (учитывающей только межгрупповую изменчивость) и «внутригрупповой» оценки (учитывающей только внутригрупповую изменчивость).

Свойства оценок

При сделанных предположениях обе оценки — $\hat{\beta}_b, \hat{\beta}_{CV}$ — состоятельны, следовательно, состоятельна и сама $w\hat{\beta}_{GLS}$

- Если $w \rightarrow \infty$, то $w \rightarrow 0$ и $\hat{\beta}_{GLS} \rightarrow \infty\hat{\beta}_{CV}$, так что при больших w оценки, получаемые в рамках моделей фиксированных и случайных эффектов, эквивалентны.

- Если $\sigma_\alpha^2 \rightarrow 0$, GLS-оценка переходит в OLS-оценку, т.е. в пределе нет никаких эффектов.

GLS -оценка эффективнее, чем $\hat{\beta}_{CV}$, так как она использует информацию как о внутригрупповой изменчивости, так и о межгрупповой изменчивости.

Чтобы получить FGLS-оценку, надо подставить подходящие оценки дисперсий.

GLS-оценка является линейной комбинацией «within»-оценки и «between»-оценки. Эта линейная комбинация оптимальна. Поэтому, например, оценка также являющаяся линейной комбинацией этих двух оценок состоятельна, но менее эффективна.

Тесты Бройша-Пагана и Хаусмана для выбора между моделями.

- FE: получаемые выводы — условные по отношению к значениям эффектов α_i , в выборке. Это соответствует ситуациям, когда эти значения нельзя рассматривать как случайную выборку из некоторой более широкой совокупности (популяции). - RE: получаемые выводы — безусловные относительно совокупности всех эффектов α_i . Исследователя не интересуют конкретные субъекты в выборке — для него это обезличенные субъекты, выбранные случайным образом из более широкой совокупности.

Критерий Бройша — Пагана для индивидуальных эффектов

Проверки в рамках RE-модели гипотезы сведения к модели пула:

$$H_0 : \sigma_\alpha^2 = 0$$

Статистика критерия Бройша — Пагана:

$$BP = \frac{NT}{2(T-1)} \left(\frac{\sum (\sum \hat{u}_{it})^2}{\sum \sum \hat{u}_{it}^2} - 1 \right)^2 \sim \chi_1^2$$

Критерии спецификации

Речь здесь идет о том, совпадает или нет условное распределение α_i при заданном i с безусловным распределением α_i . Если не совпадает — наилучшей оценкой является $\hat{\beta}_{CV}$, если совпадает — наилучшей оценкой является $\hat{\beta}_{GLS}$

Критерий 1. Используя формулировку Мундлака, проверяем гипотезу

$$H_0 : a = 0$$

против $H_1 : a \neq 0$.

Критерий 2. Критерий Хаусмана:

$$H_0 : E(\alpha_i | x_{it}) = 0$$

против $H_1 : E(\alpha_i | x_{it}) \neq 0$ Идея критерия 2 основывается на следующих фактах: - при гипотезе H_0 и $\hat{\beta}_{GLS}$, соответствующая RE-модели, и $\hat{\beta}_{CV}$, соответствующая FE-модели, состоятельны; - при гипотезе H_1 $\hat{\beta}_{GLS}$ несостоятельна, а $\hat{\beta}_{CV}$ состоятельна.

Соответственно если гипотеза H_0 верна, то между оценками не должно наблюдаться систематического расхождения, и эта гипотеза должна отвергаться при слишком больших (в сравнении со стандартной ошибкой этой разности) по абсолютной величине значениях разности этих оценок.

Статистика критерия Хаусмана:

$$m = \frac{\hat{q}^2}{\hat{Var}(\hat{q})} \sim \chi_1^2$$

$$\hat{q} = \hat{\beta}_{CV} - \hat{\beta}_{GLS}$$

Если выполнены предположения модели со случайными эффектами, то все четыре оценки состоятельны (если, конечно, объясняющие переменные не коррелированы с ошибкой), и при этом RE-оценка имеет наибольшую эффективность. Если, однако, индивидуальные эффекты α_i коррелированы хотя бы с одной из объясняющих переменных, то состоятельной остается только FE-оценка. Поэтому встает вопрос о проверке гипотезы о том, что модель является RE- моделью. Для этого можно сравнивать FE и RE оценки, используя критерий Хаусмана.

Анализ и моделирование временных рядов

Тема 1.

Говорят, что наблюдаемый временной ряд является реализацией стохастического процесса, или временной ряд порождается стохастическим процессом. Часто имеет смысл трактовать последовательность $\{X_t, t \in [1, T]\}$ как подпоследовательность бесконечной последовательности $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ и именно эту последовательность назвать стохастическим процессом, порождающим наблюдаемые данные. Как правило, в дальнейшем, если не оговорено обратное, в курсе под стохастическим процессом понимается бесконечная последовательность $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$. Если t пробегает непрерывный отрезок времени, а иногда аналитически удобнее работать с непрерывным временем, то $X(\omega, t)$ называют случайным процессом с непрерывным временем.

Слабо и сильно стационарные случайные процессы.

Строгая стационарность, или стационарность в узком смысле. Мы будем называть случайный процесс строго стационарным, если сдвиг во времени не меняет ни одну из функций плотности распределения. Это значит, что если ко всем моментам времени прибавить некоторую (целочисленную) величину, то сама функция плотности не изменится, $f_n(x_t, \dots, x_{t_n}) = f_n(x_{t+\Delta}, \dots, x_{t_n+\Delta})$ для всех n , моментов времени t_1, \dots, t_n и целочисленных Δ .

1-ое следствие: если процесс строго стационарный, то его математическое ожидание не зависит от времени. Обратите внимание, что если свойства стационарности нет, то в различные моменты времени может быть разное по величине математическое ожидание.

2-ое следствие: дисперсия строго стационарного процесса в каждый момент времени одинакова.

3-е следствие: автоковариационная функция стационарного временного ряда зависит только от разности моментов времени $(t_1 - t_2)$.

Слабая стационарность, или стационарность в широком смысле. Если случайный процесс таков, что у него математическое ожидание и дисперсия существуют и не зависят от времени, а автокорреляционная (автоковариационная) функция зависит только от разности значений $(t_1 - t_2)$, то такой процесс мы назовем стационарным в широком смысле, или слабо стационарным. Существует еще одно название процесса с такими свойствами - стационарный в ковариациях случайный процесс. Всякий строго стационарный процесс является слабо стационарным, обратное, вообще говоря, не верно, например третий центральный момент может зависеть от времени. Не зря эти термины подобраны так, а не наоборот. Для очень важного класса нормальных процессов два эти определения стационарности эквивалентны. Мы назовем случайный процесс нормальным или гауссовым, если все его многомерные функции распределения нормальны. Поскольку многомерные функции распределения нормального вектора любого порядка полностью определяются вектором математических ожиданий и ковариационной матрицей, то для гауссова случайного процесса из слабой стационарности следует сильная стационарность.

Теорема Вольда.

Декомпозиция или теорема Вольда (Wold decomposition). Мы только сформулируем этот результат и не будем его доказывать. Вольд доказал, что чисто недетерминированный стационарный в широком смысле случайный процесс может быть представлен в следующем виде: $X_t - \mu = \sum_{\tau=0}^{\infty} \psi_{\tau} \cdot \varepsilon_{t-\tau}$, где μ_t - математическое ожидание этого процесса, а ε_j - белый шум с конечными математическим ожиданием и дисперсией. То есть всякий слабо стационарный процесс представляется в виде линейной комбинации белых шумов, с разными весовыми коэффициентами. Так как это выражение линейное, его часто называют линейным фильтром. Как бы белый шум пропустили через линейный фильтр. И это очень удобно, это значит, что без потери общности мы можем ограничиться удобным линейным представлением и все про стационарные процессы изучить.

Разумеется, для того, чтобы это выражение имело смысл, надо потребовать выполнения условия сходимости, причем сходимости по вероятности, потому что суммируются случайные величины. Это условие очень простое, мы его выпишем и больше к нему возвращаться не будем: $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Поскольку реализации белого шума не наблюдаем, весовые коэффициенты определены с точностью до множителя. Поэтому договорились без потери общности считать $\psi_0 = 1$. Чем больше весовой коэффициент ψ_{τ} , тем больше влияние случайного возмущения в момент $t - \tau$ на текущий момент t . Разумеется, бесконечное число слагаемых порождает технические проблемы. К счастью, оказалось, что во многих случаях достаточно рассматривать не общее представление Вольда, а его частные случаи, когда число слагаемых конечно.

Оператор лага

Для изучения свойств временных рядов удобно использовать оператор сдвига L . Определим $LX_t = X_{t-1}$, то есть действие оператора сдвига на временной ряд дает значение временного ряда в предыдущий момент времени. Естественным образом последовательное применение оператора сдвига p раз дает значение временного ряда в момент времени на p периодов ранее, что позволяет ввести степень оператора сдвига: $L^p X_t = L(L \dots) X_t = X_{t-p}$. Иногда удобно использовать нулевую степень оператора лага: $L^0 X_t = X_t$, играющую роль единичного оператора. В дальнейшем мы не будем специально оговаривать разницу между умножением на единицу, единичным оператором и нулевой степенью оператора сдвига,

используя по обстоятельствам тот из них, который удобнее. Также естественным образом вводится умножение оператора сдвига на число и операции сложения и умножения степеней этого оператора.

Тема 2.

Модели скользящего среднего $MA(q)$.

В прошлой лекции мы остановились на рассмотрении процессов скользящего среднего, которые мы определили как процессы, для которых сумма ряда в разложении Вольда простирается не до бесконечности, а до некоторого конечного целого числа q . Для обозначения коэффициентов конечного ряда $MA(q)$ мы будем использовать иную букву β : $X_t = \sum_{\tau=0}^q \beta_\tau \cdot \varepsilon_{t-\tau}$. Это означает, что ψ_i до $i = q$ включительно равны β_i , а остальные ψ_i равны нулю. Тогда выражение для ковариационной функции этого процесса принимает вид: $\text{Cov}(X_t, X_{t+\tau}) = \gamma(\tau) = \sigma_\varepsilon^2 \sum_{i=0}^{q-\tau} \beta_i \cdot \beta_{i+\tau}$, где $\tau \leq q$, а $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_t)$. Другими словами, между значениями временного ряда, достаточно далеко отстоящими друг от друга, отсутствует корреляционная связь. Из полученных выражений для ковариаций и дисперсии непосредственно следует выражение для автокорреляционной функции: $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$.

Условие обратимости.

Таким образом, получим выражение следующего вида: $(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_k L^k + \dots) X_t$, где коэффициенты α_i сложным образом будут выражены через характеристические корни π . Если это все выполнено, то ε выражено через текущие и предыдущие значения X .

Проведенные преобразования справедливы, если каждая дробь типа $\frac{A_i}{1-\pi_i L}$ может трактоваться как сумма бесконечно убывающей геометрической прогрессии, то есть мы возвращаемся к условию, что все корни по модулю строго меньше единицы: $|\pi_i| < 1$.

Это условие называется условием обратимости (invertability) процесса MA . При этом выражение для ε_t будет бесконечным, но сходящимся операторным полиномом, действующим на X_τ . Это значит, что ε_t будет выражено через уходящие в бесконечно далекое прошлое значения X . Можно показать, что условие обратимости сохраняет свой вид, а именно: $|\pi_i| < 1$ и для случая комплексных и/или кратных корней.

Модели авторегрессии $AR(p)$.

Если значение случайного процесса определяется линейной комбинацией конечного числа его предыдущих значений и добавлением белого шума, то такой процесс называется процессом авторегрессии (autoregression) порядка p , и его общее уравнение имеет вид: $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t$, где ε_t — белый шум. В этих терминах процесс, обратный к $MA(q)$, может быть обозначен как $AR(\infty)$. Однако у нас нет гарантии, что при любых коэффициентах $\alpha_1, \alpha_2, \dots, \alpha_p$ этот процесс будет стационарным. Для того чтобы он был стационарным, нужно, чтобы он был представим в виде разложения Вольда, чтобы его можно было перевести в $MA(q)$ представление, имеющее смысл.

Условие стационарности.

В результате условие, что все корни характеристического уравнения для процесса $AR(p)$ лежат внутри единичного круга, является необходимым и достаточным для того, чтобы ряд был стационарным.

Тут есть некая асимметрия. Процессы $AR(p)$ и $MA(q)$ в чем-то похожи. Но процесс $MA(q)$ всегда стационарен, а условия обратимости обеспечивают его некоторое полезное свойство, не затрагивая фундаментального свойства стационарности. Для $AR(p)$ это условие очень жесткое: либо он стационарен и сводится к $MA(\infty)$, либо он вообще не стационарен, и тогда он выпадает (пока) из нашего рассмотрения.

Уравнения Юла-Уокера.

Перейдем к автокорреляционной функции: $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$. Иногда ее обозначают асф. Получаем более удобное соотношение: $\rho(1) = \alpha_1 + \alpha_2 \rho(1)$. Для упрощения записи будем использовать индекс вместо аргумента у автоковариационной и автокорреляционной функций. Отсюда следует, что $\rho(1) = \rho_1 = \frac{\alpha_1}{1-\alpha_2}$. В результате аналогичных действий получаем: $\rho_2 = \rho(2) = \alpha_1 \rho_1 + \alpha_2 = \frac{\alpha_1^2}{1-\alpha_2} + \alpha_2$. При умножении на X_{t-3} получаем: $\rho_3 = \alpha_1 \rho_2 + \alpha_2 \rho_1$. Если так делать дальше, то получим:

$\rho_n = \alpha_1 \rho_{n-1} + \alpha_2 \rho_{n-2}$. Это то, что мы отметили в прошлой лекции в общем виде, начиная с номера большего, чем p , значения автоковариационной функции подчиняются разностному уравнению. Совокупность таких разностных уравнений для значений автокорреляционной функции при различных n носит название уравнений Юла-Уокера (Yule-Walker equations). Решение уравнения в случае разных по величине корней имеет следующий вид: $\rho_n = C_1 (\pi_1)^n + C_2 (\pi_2)^n$, где π_1 и π_2 — характеристические корни. Для совпадающих корней решение имеет вид:

$$\rho_n = C_1 (\pi_1)^n + C_2 n (\pi_1)^n$$

Исходя из условий стационарности, корни по модулю меньше 1, решение разностного уравнения устойчиво, поэтому это выражение затухающее.

Автокорреляционная и частная автокорреляционная функции.

Поэтому хотелось бы получить характеристику, качественно ведущую себя сходным образом, указывающую на точный порядок p авторегрессионного уравнения.

Такой характеристикой является частная автокорреляционная функция (partial autocorrelation function). Стандартная английская аббревиатура - расf. В определение автокорреляционной функции $\rho(\tau) = \frac{1}{\text{Var}(X_t)} E \{(X_t - \mu)(X_{t-\tau} - \mu)\}$ входит ковариация между значениями процесса, отстоящими на τ шагов по времени друг от друга.

Коэффициент корреляции при элиминировании промежуточных значений называется частным коэффициентом корреляции. В курсе эконометрики подобная конструкция встречалась при рассмотрении излишних и не включенных в модель переменных. Рассматривалось влияние последней добавленной переменной с учетом того, что все предыдущие уже включены.

Обозначим через $\varphi_{kk} - k$ -е значение частной автокорреляционной функции. По определению φ_{kk} — это коэффициент корреляции между X_{t-k} и X_t (или между x_{t-k} и x_t) за вычетом той части x_t , которая линейно объяснена промежуточными

лагами $x_{t-1}, x_{t-2}, \dots, x_{t-k-1}$. Эта объясняющая линейная комбинация называется оптимальным линейным предиктором, если она реализует минимальную среднеквадратичную ошибку предсказания (прогноза) x_t . Более строго, нас интересует коэффициент корреляции

$$\text{Corr}(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k-1}, x_{t-k})$$

где $\varphi_{k1}, \varphi_{k2}, \dots, \varphi_{kk-1}$ — коэффициенты линейной комбинации, обеспечивающей минимальную среднеквадратичную ошибку предсказания

$$\min E \left\{ (x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k-1})^2 \right\}$$

Тема 3.

Модели авторегрессии-скользящего среднего ARMA(p,q).

Теперь делаем следующий шаг. Переходим к комбинации двух различных типов процессов AR(p) и MA(q). Рассмотрим общий процесс авторегрессии-скользящего среднего, который носит название: ARMA - авторегрессия-скользящее среднее. Мы назовем так процесс следующего вида: $\alpha_p(L)X_t = \beta_q(L)\varepsilon_t$ или в развернутом виде: $X_t = \alpha_1X_{t-1} + \dots + \alpha_pX_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \dots + \beta_q\varepsilon_{t-q}$

Выведем условия стационарности процесса ARMA. Мы уже установили, что если все корни полинома $\alpha_p(L)$ по модулю меньше единицы, то существует обратный оператор, и мы можем записать: $X_t = [\alpha_p(L)]^{-1} \beta_q(L)\varepsilon_t$. Обратный оператор можно разложить в сумму элементарных дробей, каждую представить как бесконечно убывающую геометрическую прогрессию, то есть бесконечный операторный полином. При умножении на конечный полином мы получим вновь бесконечный полином. Выражение имеет смысл, если все характеристические корни полинома $\alpha_p(L)$ по модулю меньше единицы. А тогда полученное выражение является разложением Вольда, и процесс стационарен. То есть вся эта цепочка приводит к тому, что стационарность процесса ARMA определяется только его AR-частью. Поэтому условия те же самые, что у процесса AR. Процесс ARMA стационарен, если корни характеристического уравнения AR-части по модулю меньше единицы. Точно так же, условия обратимости процесса, то есть возможность выразить ε_t через X_t , то есть существование выражения: $\varepsilon_t = [\beta_q(L)]^{-1} \alpha_p(L)X_t$, полностью определяются условиями обратимости MA-части. Если MA-часть обратима, то и весь процесс обратим.

Итак, если процесс ARMA стационарный, то он имеет обязательно MA(∞) представление, как любой стационарный процесс по теореме Вольда, но он имеет и конечное представление ARMA(p,q). Он может иметь еще и бесконечное AR(∞) представление. То есть мы видим, что ARMA(p,q) может быть очень удобным представлением одного и того же процесса. Поэтому, если можно «согнать» процесс в ARMA, то он определяется всего $p + q$ параметрами.

Очевидно, что математическое ожидание стационарного процесса ARMA равно нулю: $E\{X_t\} = 0$.

Оценивание коэффициентов авторегрессионных моделей.

Оценивание коэффициентов моделей скользящего среднего.

Оценивание коэффициентов процессов ARMA(p).

(Канторович 5)

Качество подгонки моделей временных рядов.

(Канторович 5)

Информационные критерии Акаике (AIC) и Шварца (BIC).

“Портмونت”-статистика.

Тема 4.

Подход Бокса-Дженкинса к идентификации моделей стационарных временных рядов.

Прогнозирование в модели Бокса-Дженкинса.

Тренд и сезонность в модели Бокса-Дженкинса.

Коэффициент множественной детерминации в моделях временных рядов.

(Лекция 7)

Тема 5.

Нестационарные временные ряды.

Случайное блуждание.

Ряды с нестационарной дисперсией. Нестационарное среднее.

Процессы, приводимые к стационарным, выделением тренда (TSP) и взятием последовательных разностей (DSP).

(Лекция 7)

Модели ARIMA (p,1, q).

Подход Бокса-Дженкинса к определению степени интеграции временного ряда.