Эконометрика.

Цель курса.

Цель курса – дать понятие об основных методах современной эконометрики в строгом математическом изложении. Предполагается, что слушатель по окончанию курса будет обладать двумя компетенциями: (i) уметь самостоятельно проводить анализ экономических данных в эмпирическом исследовании, (ii) обладать достаточным арсеналом методов и понятий, чтобы понимать эконометрический анализ в экономических статьях.

Организация курса.

В курсе предполагается интенсивная лекционная нагрузка в соотношении 2:1 к семинарским занятиям. Задания на самостоятельную и домашнюю работу будут включать в себя как аналитические задачи, так и задачи на компьютере на языке с открытым кодом R. Задачи будут как теоретическими, так и практическими с использованием реальных данных.

Основная литература.

1. Hansen B.E. Econometrics. University of Wisconsin, Department of Economics, 2016. http://www.ssc.wisc.edu/∼bhansen/econometrics

2. Anatolyev S. Intermediate and Advanced Econometrics: Problems and Solutions. 3rd edition, New Economical School, 2009.

3. Анатольев С.А. Курс лекций по эконометрике для продолжающих. М.: Российская экономическая школа, 2002.

4. Анатольев С.А. Курс лекций по эконометрике для подготовленных. М.: Российская Экономическая Школа, 2006.

5. Hayashi F. Econometrics, Princeton University Press, 2000.

6. Baltagi B. Econometric Analysis of Panel Data, John Wiley & Sons, 2008.

Программа курса

1. Основные понятия.

   - Условное распределение и условное математическое ожидание.
   - Закон повторных математических ожиданий.
   - Регрессия среднего, медианная и квантильная регрессии.
   - Наилучший линейный предиктор. Линейная проекция.
   - Случайные и неслучайные выборки. Принцип аналогий.

2. Асимптотические методы в эконометрике.
    - Точные и асимптотические подходы. Их преимущества и недостатки.
    - Закон больших чисел, центральная предельная теорема, дельта-метод, теорема о непрерывном отображении.
    - Асимптотические доверительные интервалы.
    - Асимптотические методы проверки гипотез о случайных выборках.
    - Временные ряды: стационарность, эргодичность, эргодическая теорема, центральная предельная теорема, долгосрочная дисперсия и ее оценка.
3. Бутстрап.
    - Бутстрап для случайных выборок. Эмпирическая функция распределения, аппроксимация распределения статистик с помощью бутстрапа.
    - Бутстраповские доверительные интервалы. Бутстраповская корректировка смещения оценок. Тестирование гипотез с помощью бутстрапа.
    - Бутстрап для временных рядов.
4. Линейная регрессия.
    - Метод наименьших квадратов: обычный и обобщенный.
    - Асимптотические свойства МНК-оценок. Проверка гипотез об МНК-оценках.
5. Нелинейная регрессия.
    - Нелинейный метод наименьших квадратов: обычный и обобщенный.
    - Асимптотические свойства НМНК-оценок.
    - Модель бинарного выбора и ее НМНК-оценка.
    - Проверка гипотез об НМНК-оценках в случае, когда часть параметров не идентифицируется при нулевой гипотезе.
6. Непараметрическая регрессия.
    - Непараметрическая регрессии: дискретные и непрерывные регрессоры.
    - Непараметрическая ядерная оценка и ее асимптотические свойства.
    - Методы выбора ширины окна в ядерных оценках.
    - Многомерная непараметрическая ядерная оценка.
    - Неядерные непараметрические методы.
7. Экстремальные оценки и метод максимума правдоподобия (ММП).

- Экстремальные оценки и их асимптотические свойства.
- Функция правдоподобия и основное информационное неравенство.
- ММП-оценка и НМНК-оценка как экстремальные оценки.
- Асимптотические свойства ММП-оценок.
- Условный, совместный и маржинальный ММП.
- Тестирование гипотез о ММП-оценках: тесты Вальда, множителей Лагранжа и отношения правдоподобия.
- ММП-оценивание моделей бинарного выбора.
- ММП-оценивание моделей временных рядов.

8. Метод моментов.

    - Моментные ограничения и моментные функции. Точная идентификация и сверхидентификация.
    - Классический и обобщенный метод моментов (ОММ).
    - Асимптотические свойства ОММ-оценок. Эффективность ОММ-оценок.
    - Тесты Вальда, множителей Лагранжа и разности расстояний.
    - J-тест о сверхидентифицирущих ограничениях.
    - Метод инструментальных переменных.
    - Обобщенный метод моментов для временных рядов. Примеры применения.
    - Бутстрапирование ОММ-оценок.
    - Тест Хаусмана о спецификации модели.

9. Панельные данные.

    - Оценка панельной регрессии методом наименьших квадратов.
    - Асимптотические свойства МНК-оценок.
    - Случайные и фиксированные эффекты. Тест Хаусмана.
    - Динамическая панельная регрессия и ее оценка методом инструментальных переменных.
    - Панельные модели бинарного выбора: логит-модель с фиксированными эффектами и пробит-модель со случайными эффектами. Их оценка.

Типовые задачи к курсу.

Далее $\alpha, \beta, \sigma, \rho, \ldots$ – параметры статистической модели; $\widehat{\alpha}_n, \widehat{\beta}_n, \widehat{\sigma}_n, \widehat{\rho}_n, \ldots$ – оценки параметров, построенные по выборке или временному ряду длины n; $x, y, z, e, \ldots$ – случайные величины или вектора, имеющие все необходимые моменты; $f(y|x)$ – функция плотности условного распределения y при фиксированном x; $\mathrm{Med}(y|x)$ – условная медиана; $\mathrm{E}(y|x)$ – условное среднее; $\mathrm{Var}(y|x)$ – условная дисперсия.

Тема «Основные понятия».

1. Пусть распределение (x, y) является смесью распределений $\mathcal{N}((0,0)', I_2)$ с весом $1/2$ и $\mathcal{N}((1,1)', I_2)$ с весом $1/2$, где I_2 есть 2×2 единичная матрица. Найти $\mathrm{E}(y|x)$, $\mathrm{Var}(y|x)$ и $f(y|x)$.

2. Верно ли, что $\mathrm{E}(y|x) = \mathrm{E}(y|x,z)$ п.н., если (i) y и z независимы, (ii) (x, y) и z независимы?

3. Найти функцию f, реализующую минимум выражения $\mathrm{E}|y - f(x)|$.

4. Установить, что минимум выражений $\mathrm{E}|y - (ax + b)|^2$ и $\mathrm{E}|\mathrm{E}(y|x) - (ax + b)|^2$ достигается при одних и тех же $a, b \in \mathbb{R}$.

5. Пусть $\{(x_i, y_i)\}_{i=1}^n$ – случайная выборка. Из нее выбрали все элементы (x_i, y_i), у которых $y_i > 0$, получив новый набор данных $\{(x_j^*, y_j^*)\}_{j=1}^m$, где m – число элементов, у которых $y_i > 0$. Показать, что $\{(x_j^*, y_j^*)\}_{j=1}^m$ – случайная выборка.

Тема «Асимптотические методы в эконометрике».

6. Пусть $\{x_i\}_{i=1}^n$ – случайная выборка. Показать, что (i) $\sqrt{n}(\overline{x}_n - \alpha) \sim \mathcal{N}(0, \sigma^2)$ при любом $n \geqslant 1$, если $x_1 \sim \mathcal{N}(\alpha, \sigma^2)$; (ii) $\sqrt{n}(\overline{x}_n - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ при $n \to \infty$, если $\mathrm{E}x_1 = \alpha, \mathrm{Var}(x_1) = \sigma^2$.

7. Пусть $\sqrt{n}(\widehat{\rho}_n - \rho) \xrightarrow{d} \mathcal{N}((0,0)', I_2)$. Найти асимптотическое распределение $\|\widehat{\rho}_n\|^2$ при всех значениях $\rho \in \mathbb{R}^2$.

8. В условиях предыдущей задачи построить 5% асимптотический доверительный интервал для $\|\rho\|^2$.

9. В условиях предыдущей задачи проверить гипотезу $H_0 : \|\rho\|^2 = 0$.

10. Пусть $(e_t)_{t\in\mathbb{Z}}$ – н.о.р. случайные величины с $\mathrm{E}e_t = 0, \mathrm{E}e_t^2 = 1$. Для каких $\beta_1, \beta_2 \in \mathbb{R}$ существует единственное стационарное решение реккурентного уравнения $x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + e_t$, $t \in \mathbb{Z}$? Будет ли это решение эргодичным? Найти $\sigma > 0$, для которого $\sqrt{T}\overline{x}_T \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ при $T \to \infty$. Построить состоятельную оценку σ^2.

Тема «Бутстрап».

11. Пусть $(x_1, x_2, x_3) = (1, 1, 2)$. Построить эмпирическую функцию распределения по $\{x_i\}_{i=1}^3$. Найти бутстраповское распределение статистики $T = \overline{x}_3 - \mathrm{E}x$. Сколько значений с ненулевой вероятностью принимает бутстраповский аналог статистики T?

12. Пусть $\rho = \mathrm{Corr}(x, y)$ и $\widehat{\rho}$ – выборочный коэффициент корреляции, отвечающий случайной выборке $\{(x_i, y_i)\}_{i=1}^n$. Скорректировать смещение оценки $\widehat{\rho}$ с помощью бутстрапа. Построить бутстраповский доверительный интервал для ρ. Проверить гипотезу $H_0 : \rho = 0$ против $H_1 : \rho \neq 0$ с помощью бутстрапа.

13. В условиях предыдущей задачи сделать все то же самое, предполагая, что $\{(x_t, y_t)\}_{t\geqslant 1}$ – стационарный эргодичный временной ряд.

Тема «Линейная регрессия».

14. Пусть в популяции $y = \alpha + \beta x + e$, где $\mathrm{E}(e|x) = 0$ и $\mathrm{E}(e^2|x) = \sigma^2$. Имеется две случайные выборки $\{(x_i^{(1)}, y_i^{(1)})\}_{i=1}^n$ и $\{(x_i^{(2)}, y_i^{(2)})\}_{i=1}^n$ из подпопуляций с $y \geqslant 0$ и $y < 0$ соответственно. Как оценить (α, β) наиболее эффективным образом с помощью взвешенного метода наименьших квадратов по выборке $\{(x_i, y_i)\}_{i=1}^{2n} = \bigcup_{j=1}^2 \{(x_i^{(j)}, y_i^{(j)})\}_{i=1}^n$?

15. Пусть y^* – зависимая переменная, а x^* – объясняющий фактор. Наблюдается зашумленное значение (x, y) вектора (x^*, y^*), т.е. $x = x^* + u$, $y = y^* + v$, где u, v –

ошибки измерения. Пусть $y^* = \beta x^*$ для некоторого $\beta \in \mathbb{R}$, а x^*, u, v – независимые нормальные величины со средним нуль. Имеется случайная выборка $\{(x_i, y_i)\}_{i=1}^n$.

(i) Найти предел $\widehat{\beta} - \beta$ при $n \to \infty$, где $\widehat{\beta}$ – МНК-оценка в регрессии y на x.

(ii) Показать, что, вообще говоря, β невозможно оценить по данным $\{(x_i, y_i)\}_{i=1}^n$.

(iii) Проверить гипотезу $H_0 : \beta = 0$ против альтернативы $H_1 : \beta > 0$ с помощью асимптотических методов и с помощью бутстрапа наиболее эффективным образом.

Тема «Нелинейная регрессия».

16. Пусть $y \in \{0, 1\}$ и $\mathrm{P}(y = 1|x) = p(x, \beta)$ для некоторой гладкой функции p и $\beta \in \mathbb{R}$. Как оценить β наиболее эффективным образом с помощью взвешенного нелинейного метода наименьших квадратов?

17. Рассмотрим модель $\mathrm{E}[c_t|y_{t-1}, y_{t-2}, \ldots] = \alpha + \beta I(y_{t-1} > \Delta) + \delta y_{t-1}^{\gamma}$, где c_t – потребление агента в момент t, а y_t – его доход в момент t. Пары $\{(c_t, y_t)\}_{t=1}^T$ имеют непрерывное распределение, стационарны и эргодичны по t (при $T \to \infty$). Параметр Δ представляет собой нормальный уровень дохода и является известным.

(i) Опишите не менее 3х ситуаций, когда параметры модели не идентифицируются.

(ii) Как численно посчитать НМНК-оценки параметров модели на компьютере с помощью метода концентрации?

(iii) Проверить на 5% уровне значимости гипотезу $H_0 : \gamma = 1$ против $H_1 : \gamma < 1$ с помощью асимптотических методов и с помощью бутстрапа.

(iv) Некто проводит линейную регрессию c_t на $(1, y_{t-1}, y_{t-2})$, получая соответствующие оценки коэффициентов $(\widehat{\alpha}, \widehat{\beta}_1, \widehat{\beta}_2)$. Какие величины состоятельно оценивают $(\widehat{\alpha}, \widehat{\beta}_1, \widehat{\beta}_2)$?

(v) Рассмотрим гипотезу $H_0 : \delta = 0$ против $H_1 : \delta \neq 0$. При H_0 параметр γ не идентифицируется. Как проверить гипотезу H_0 на 5% уровне значимости?

Тема «Непараметрическая регрессия».

18. Пусть $f(x) = \mathrm{E}(y|x)$ и $x \in \{0, 1, 2\}$. Оцените вектор параметров $(f(k))_{k=0}^2$ и выпишите асимптотическое распределение соответствующей оценки.

19. Имеются реальные данные $\{r_t\}_{t=1}^T$ по доходностям 3-месячных казначейских облигаций США. По этим данным требуется на компьютере в пакете R выполнить следующие задания, где в качестве оценок следует брать оценки Надарая-Ватсона с ядром Епанечникова.

(i) Оцените непараметрически плотность r_t. Сделайте это для 3х значений ширины окна, которые на глаз приводят к излишнему, недостаточному и оптимальному сглаживанию. Постройте графики полученных оценок, отложив на оси абсцисс значения доходностей r_t.

(ii) Оцените непараметрически условное среднее $\mathrm{E}(\Delta r_t|r_{t-1})$ с $\Delta r_t = r_t - r_{t-1}$. Сделайте это для 3х значений ширины окна по аналогии с пунктом (i). Постройте графики полученных оценок на плоскости $(r_{t-1}, \Delta r_t)$, отметив на этой плоскости изначальные данные. Постройте аналогичный график для ширины окна, выбранного по правилу Сильвермана.

(iii) Оцените непараметрически условную дисперсию $\mathrm{Var}(\Delta r_t|r_{t-1})$. Сделайте это для 3х значений ширины окна по аналогии с пунктом (i). Постройте графики полученных оценок.

(iv) В пункте (ii) для оптимального значения ширины окна вычислите и нарисуйте асимптотические границы 5% доверительного интервала вокруг оценки регрессионной функции $f(r) = \mathrm{E}(\Delta r_t | r_{t-1} = r)$.

Тема «Экстремальные оценки и метод максимального правдоподобия».

20. Пусть $\{r_t\}_{t \geqslant 1}$ – доходности некоторых финансовых активов. Предполагается, что r_t стационарно и эргодично во времени. Рассмотрим модель $r_t = \alpha + \beta r_{t-1} + e_t$ при $\mathrm{E}(\exp\{\theta e_t\} | I_{t-1}) = 1$ для некоторого известного $\theta \in \mathbb{R}$, где $I_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$. Пусть также $(\widehat{\alpha}_T, \widehat{\beta}_T)$ – экстремальные оценки параметров (α, β), для которых

$$(\widehat{\alpha}_T, \widehat{\beta}_T) = \arg\min_{a,b} \sum_{t=1}^{T} L_\theta(r_t - a - b r_{t-1}),$$

где $L_\theta(u) = \exp\{\theta u\} - \theta u - 1$ – линейно-экспоненциальная функция потерь.

(i) Доказать, что $\widehat{\alpha}_T, \widehat{\beta}_T$ являются состоятельными оценками параметров α, β.

(ii) Посчитать асимптотическое распределение и стандартные ошибки $(\widehat{\alpha}_T, \widehat{\beta}_T)$.

(iii) Проверить гипотезу $H_0 : \beta = 0$ против $H_1 : \beta \neq 0$ на 5% уровне значимости с помощью бутстрапа.

(iv) Некто оценивает параметры α, β с помощью обычного метода наименьших квадратов. Получатся ли таким методом состоятельные оценки для α и β?

Предположим далее, что $\exp\{\theta r_t\}$ имеет экспоненциальное распределение условно на I_{t-1}.

(v) Выпишите экстремальную задачу, которая возникает в методе максимального правдоподобия, и вычислите асимптотическое распределение ММП-оценок.

(vi) Использую ММП-оценки, проверьте гипотезу $H_0 : \beta = 0$ против $H_0 : \beta \neq 0$ с помощью тестов Вальда, отношения правдоподобия и множителей Лагранжа.

(vii) Допустим, что параметр θ изначально неизвестен. Можно ли получить состоятельные оценки параметров (α, β, θ), решая экстремальную задачу из пункта (v), в которой оптимизация ведется в том числе и по θ?

Тема «Метод моментов».

21. Дана система уравнений

$$\begin{cases} q = \alpha(p + w) + u, & \text{уравнение спроса на рыбу,} \\ q = \beta_1 p + \beta_2 z + v, & \text{уравнение предложения рыбы на рынке,} \end{cases}$$

где q – количество проданной рыбы, p – равновесная цена рыбы, z – погода в море ($z = 0$ плохая/$z = 1$ хорошая), w – налог, который взимает администрация рынка за каждую купленную рыбу, u, v – какие-то остатки (все другие факторы). Пусть $\mathrm{E}zu = \mathrm{E}zv = \mathrm{E}wu = \mathrm{E}wv = 0$. Пусть также имеется случайная выборка $\{(p_i, q_i, w_i, z_i)\}_{i=1}^n$ за n периодов.

(i) Какие параметры модели можно состоятельно оценить, если z ненаблюдаемо? Как это сделать?

(ii) Можно ли оценить параметр α с помощью МНК? Тот же вопрос для β_1, β_2.

(iii) Как оценить параметры модели наиболее эффективным образом с помощью обобщенного метода моментов.

(iv) Предполагая, что $\mathrm{E}[uv|z,w]=0$, $\mathrm{E}[u^2|z,w]=\sigma_u^2$, $\mathrm{E}[v^2|z,w]=\sigma_v^2$, выписать явную формулу оценок из пункта (iii).

(v) Выписать асимптотические распределения полученных оценок.

(vi) Как проверить гипотезу $H_0:\beta_2=0$ против $H_1:\beta_2>0$ на 5% уровне значимости с помощью асимптотических методов? С помощью бутстрапа?

(v) Проверить спецификацию модели с помощью J-теста. Какие моментные условия в действительности проверяет J-тест, когда $\mathrm{E}[uv|z,w]=0$, $\mathrm{E}[u^2|z,w]=\sigma_u^2$, $\mathrm{E}[v^2|z,w]=\sigma_v^2$?

Тема «Панельные данные».

22. Имеются данные за 10 периодов $t=1,\dots,10$ по избыточным доходностям рыночного портфеля r_t и по избыточным доходностям $\{r_{it}\}_{i=1}^n$ для n случайно выбранных акций рынка США.

(i) Опишите подлежащую популяцию и объясните, почему величины r_t^m могут считаться неслучайными в данной постановке.

Рассмотрим модель вида $r_{it}=\alpha_i+\beta_i r_t+u_{it}$, где $\{u_{it}\}$ – н.о.р. случайные величины по (i,t) с $\mathrm{E}u_{it}=0$ и $\mathrm{E}u_{it}^2=\sigma^2$. Пусть вначале $\alpha_i\equiv\alpha$.

(ii) Предложите какое-нибудь преобразование уравнений модели, которое уберет шумовые величины β_i и позволит оценить α состоятельно.

(iii) Оцените α наиболее эффективным образом, использую преобразованные уравнения из пункта (ii). Приведите вашу оценку $\widehat{\alpha}$ к наиболее компактному виду.

(iv) Является ли $\widehat{\alpha}$ несмещенной оценкой? Состоятельной оценкой?

(v) Выпишите асимптотическую дисперсию $\widehat{\alpha}$ и отвечающую ей стандартную ошибку.

(vi) Проверьте на 5% уровне значимости гипотезу $H_0:\alpha=0$ против $H_1:\alpha\neq 0$.

(vii) Найдите наилучшую несмещенную линейную (по r_{it}) оценку α на основе первоначальных уравнений модели.

Пусть теперь $\{\alpha_i\}_{i=1}^n$ – н.о.р. случайные величины.

(viii) Какую величину состоятельно оценивает $\widehat{\alpha}$ из пункта (iii)?

(ix) Какую гипотезу H_0 в действительности проверяет ваш тест из пункта (vi)? На каком уровне значимости?