Основы эконометрики

Рысьмятова Анастасия

ВМК МГУ 317 группа

22.10.2014

Содержание

Что такое эконометрика

Эконометрика

Наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и статистических методов и моделей

Основные задачи эконометрики

- Обнаружение и анализ статистических закономерностей в экономике;
- Построение на базе выявленных эмпирических экономических зависимостей эконометрических моделей.

Эконометрические методы

- Регрессионный анализ
- Панельный анализ
- Анализ временных рядов

Содержание

Основные понятия

Временной ряд

Совокупность наблюдений экономической величины в различные моменты времени

Будем рассматривать временной ряд как выборку X_t

Как использовать данную конструкцию

Наиболее полная характеристика процесса является совместная функция распределения, или плотностей распределения. Поэтому, чтобы задать все вероятностые свойства этой конструкции, нам нужна совокупность функций распределения. А именно одномерная функция распределения, двумерная, и.т.д. $f_1(x_{t_1}), f_2(x_{t_1}, x_{t_2}), f_3(x_{t_1}, x_{t_2}, x_{t_3})$

Основные понятия

Строгая стационарность

Будем называть случайный процесс строго стационарным, если сдвиг во времени не меняет ни одну из функций плотности распределения

Следствия:

- В строго стационарном процессе математическое ожидание не зависит от времени
- В строго стационарном процессе дисперсия не зависит от времени
- Автоковариационная и автокорреляционные функции зависят только от разности моментов времени

Основные понятия

Слабая стационарность

Будем называть случайный процесс слабо стационарным, если:

- его математическое ожидание и дисперсия существуют и не зависят от времени
- автоковариационная функция зависит только от разности значений t_1-t_2

Примеры процессов

Белый шум

Процесс ε_t такой, что $E(\varepsilon_t) = 0$; $Var(\varepsilon_t) = 0$; $Cov(\varepsilon_i, \varepsilon_j) = 0$ при i$\neq$j
если добавить, что ε_i распределены нормально то будем называть это гауссовским белым шумом и писать $\varepsilon_e \backsim WN(0, \sigma^2)$
этот процесс слабо стационарен

Процесс случайного блуждения

Процесс задается следующм образом: $X_t = X_{t-1} + \varepsilon_t$,
где ε_t белый шум
данный процесс является нестационарным

Замечание

$$X_t = X_{t-1} + \varepsilon_t$$

$$\Delta\ X_t = \varepsilon_t$$

Пусть $z_t = \Delta\ X_t$, расмотрим z_t как новый временной ряд,

$$\Rightarrow z_t = \varepsilon_t$$

$\Rightarrow$ получили стационарный ряд

Теорема Вольда

Теорема

Недетерминированный слабо стационарный случайный процесс может быть представлен в следующем виде:
$X_t - \mu_t = \sum_{\tau=0}^{\infty} \Psi_\tau \cdot \varepsilon_{t-\tau}$
Где μ_t - математическое ожидание, ε_j - белый шум.

То есть всякий слабо стационарный процесс может быть представлен в виде линейной комбинации белых шумов с разными весовыми коффициентами.

Содержание

Процесс скользящего среднего

Процесс скользящего среднего

Стохастический процесс называется процессом скользящего среднего порядка q, если в разложении Вольда присутствуют только q слагаемых

MA(q): $X_t - \mu_t = \sum_{\tau=0}^{q} \Psi_\tau \cdot \varepsilon_{t-\tau}$

Где μ_t - математическое ожидание, ε_j - белый шум.

Оператор сдвига

Определим оператор сдвига

$LX_t = X_{t-1}$

$L^p X_t = L(L(L...))X_t$

Запишем процесс скользящего среднего через оператор сдвига

$X_t = \beta_q(L)\varepsilon_t$, где

$\beta_q(L) = \beta_0 L^0 + \beta_1 L^1 + \beta_2 L^2 + ... + \beta_q L^q$ - операторный полином

Процесс авторегрессии

Авторегрессия порядка p

AR(p): $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t$
где ε_t - белый шум
Запишем в эквивалентном виде
$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - ... - \alpha_p X_{t-p} = \varepsilon_t$
$\alpha(L) X_t = \varepsilon_t$

Процесс авторегрессии

Процессы AR(p) и MA(q) чем-то похожи
AR(p): $\alpha_p(L)X_t = \varepsilon_t$
MA(q): $\beta_q(L)\varepsilon_t = X_t$
MA(q) - всегда стационарен
AR(p) - стационарен тогда и только тогда, если все корни его характеристического уравнения лежат внутри еденичного круга.
Причем, если он стационарен его можно свести к MA(∞)

Свойства стационарных процессов AR

Математическое ожидание $E(X_t) = 0$
Это следует из теоремы Вольда: $X_t - \mu_t = \sum_{\tau=0}^{\infty} \Psi_\tau \cdot \varepsilon_{t-\tau}$
Автоковариационная функция
при $\tau > p$
$\gamma(\tau) = \alpha_1\gamma(\tau - 1) + ... + \alpha_p\gamma(\tau - p)$
Запишем авторегрессию порядка р:
$X_t = \alpha_1 X_{t-1} + ... + \alpha_p Xt - 1 + \varepsilon_t | * X_{t-\tau}$
$\gamma(\tau) = E(X_t X_{t-\tau}) = \alpha_1\gamma(\tau - 1) + ... + \alpha_p\gamma(\tau - p) + \underbrace{E(X_{t-\tau}\varepsilon_t)}$

Свойства стационарных процессов AR

Автоковариационная функция

при $\tau < p$

Рассмотрим частный случай AR(1)

$X_t = \alpha\ X_{t-1} + \varepsilon_t | * X_{t-\tau}$

$\gamma(1) = \alpha\gamma(0)$

$\gamma(2) = \alpha\gamma(1) = \alpha^2\gamma(0)$

...

$\gamma(\tau) = \alpha^\tau\gamma(0)$

Автокорреляция $\rho(\tau) = \alpha^\tau$

Автокорреляция для AR(1)

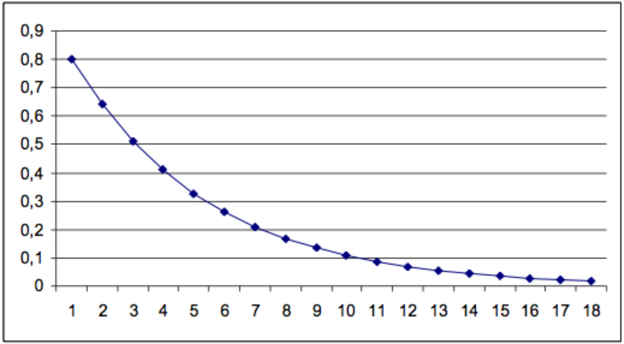


Рис. 1. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном 0,8

Автокорреляция для AR(1)

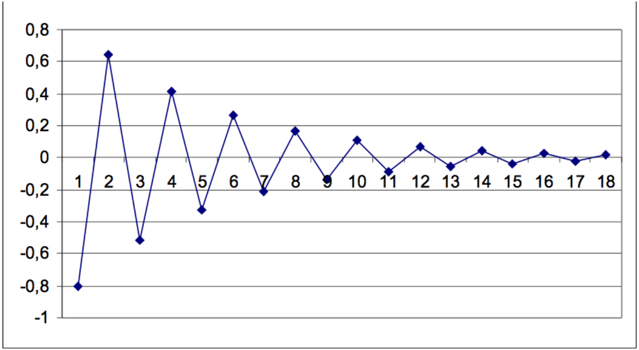


Рис. 2. Значения автокорреляционной функции процесса AR(1) при значении коэффициента равном −0,8

Свойства стационарных процессов

Автоковариационная функция для AR(2)

$X_t = \alpha_1 \; X_{t-1} + \alpha_2 \; X_{t-2} + \varepsilon_t | * X_{t-\tau}$

$E(X_{t-1}X_t) = \gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(-1)$

Автокорреляция

$\rho(\tau) = \gamma(\tau)/\gamma(0)$

$\rho_n = \alpha_1\rho_{n-1} + \alpha_2\rho_{n-2}$

Данное уравнение имеет решение:

$\rho_n = C_1(\pi_1)^n + C_2(\pi_2)^n$, где π_1, π_2 - характеристические корни из условий стационарности корни по модулю меньше еденицы $\Rightarrow$ выражение затухающее

Свойства стационарных процессов

Общий вывод

Частная автокорреляционная функция авторегрессионного процесса AR(p), равна 0 для $\tau > p$ и, вообще говоря, не равна 0 при $\tau \leq p$

Комбинация процессов ARMA

ARMA

ARMA(p,q)- авторегрессия-скользящее среднее
Комбинация процессов AR(p) и MA(q)
Уравнение:

$$\alpha_p(L)X_t = \beta_q(L)\varepsilon_t$$

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q}$$

Условие стационарности ARMA

Мы уже установили, что если все корни полинома $\alpha_p(L)$ по модулю меньше 1 то $\exists$ обратный оператор и
$X_t = [\alpha_p(L)]^{-1}\beta_q(L)\varepsilon_t$
Тогда X_t можно представить в разложение Вольда.
Значит процесс стационарный.
Значит стационарность процесса определяется только его AR-частью

Свойства стационарного процесса ARMA

- Математическое ожидание $EX_t = 0$
- Автокорреляционная функция "затухает"в том же смысле что и для процесса AR

Приведение к стационарности

Вспомним про нестационарный ряд случайного блуждания
Уравнение $X_t = X_{t-1} + \varepsilon_t$
Можно привести к виду $\Delta\ X_t = z_t = \varepsilon_t$

Приведение к стационарности

Уравнение
$X_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t$
Возьмем разность
$\Delta X_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t - \alpha - \beta(t-1) - \gamma(t-1)^2 - \varepsilon_{t-1} =$
$\beta + 2\gamma t - \gamma + (\varepsilon_t - \varepsilon_{t-1})$
Проведем взятие второй разности
$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = 2\gamma + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t+2})$
Получили стационарный процесс, который является MA(2)

Содержание

Класс ARIMA

ARIMA(p,d,q)

Класс нестационарных рядов, которые взятием d последовательных разностей можно привести к стационарному виду ARMA(p,q).
В операорном виде ARIMA(p,d,q) записывается как
$\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t$

Класс ARIMA

Этапы построения модели ARIMA Бокса и Дженкинса

- Установить порядок интеграции d и добиться стационарности ряда, взяв достаточное количество разностей.
- Исходя из поведения автокорреляционной и частной автоковариационной функции установить значение p, q
- Оценивание коэффициентов $\alpha_1..\alpha_p, \beta_1...\beta_q$
- Построение модели
- Использование модели для прогнозирования

Содержание

Пример реальной задачи

Рассмотрим пример анализа временного ряда с помощью *python* и модуля *statsmodels*.

Загрузка данных

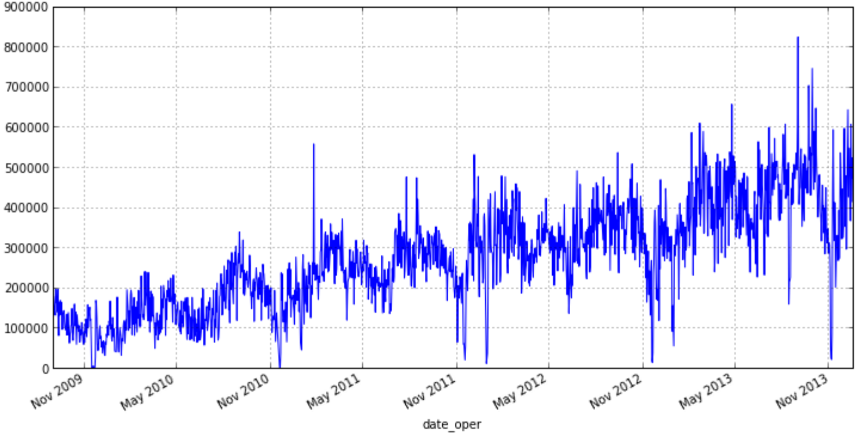
Для начала загрузим данные и посмотрим на них:

```
from pandas import read_csv, DataFrame
import statsmodels.api as sm
from statsmodels.iolib.table import SimpleTable
from sklearn.metrics import r2_score
import ml_metrics as metrics
In [2]:
dataset = read_csv('tovar_moving.csv',';', index_col=['date_oper'], parse_dates=['date_oper'], dayfirst=True)
dataset.head()
```

	Otgruzka	**priemka**
date_oper		
2009-09-01	179667	276712
2009-09-02	177670	164999
2009-09-03	152112	189181
2009-09-04	142938	254581
2009-09-05	130741	192486

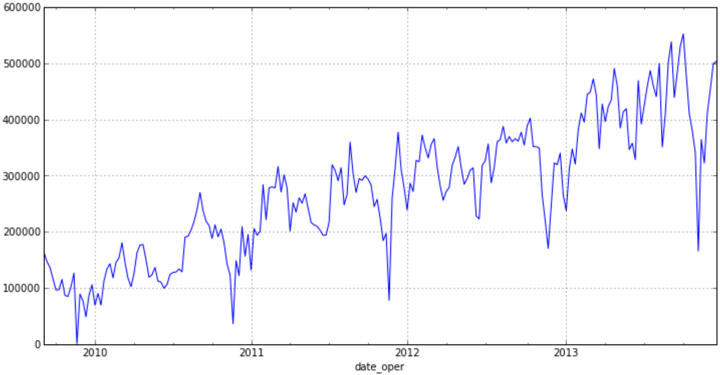
Анализ временного ряда

Посмортим график нашего ряда:



Анализ временного ряда

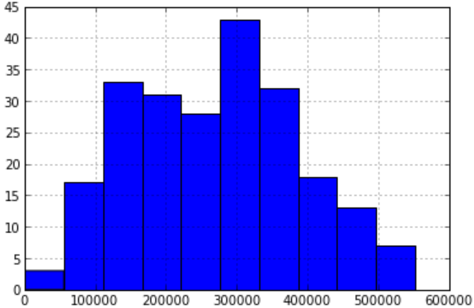
Уменьшим колебания нашего ряда



Как можно заметить, новый график не имеет ярких выбросов и имеет ярко выраженный тренд. Из это можно сделать вывод о том, что ряд не является стационарным

Анализ временного ряда

Построим гистограмму



Как можно заметить из гистограммы, ряд у нас более менее однородный и имеет относительно небольшой разброс

Анализ временного ряда

Проведем тест Харки — Бера для определения номарльности распределения, чтобы подтвердить предположение об однородности. Для этого в существует функция jarque_bera() которая возвращает значения данной статистики
$p - value = 0.0606724103$
Нулевая гипотеза о нормальности распределения отвергается с малой вероятностью , значит наш ряд имеет нормального распределения.

Анализ временного ряда

Для проверки проверки стационарности давайте проведем обобщенный тест Дикки-Фуллера на наличие единичных корней. Для этого в модуле statsmodels есть функция adfuller():

```
test = sm.tsa.adfuller(otg)
print 'adf: ', test[0]
print 'p-value: ', test[1]
print'Critical values: ', test[4]
if test[0]> test[4]['5%']:
    print 'есть единичные корни, ряд не стационарен'
else:
    print 'единичных корней нет, ряд стационарен'
```

adf: -1.38835541357

p-value: 0.58784577297

Critical values: {'5%': -2.8753374677799957, '1%': -3.4617274344627398, '10%': -2.5741240890815571}

есть единичные корни, ряд не стационарен

Анализ временного ряда

Определим порядок интегрированного ряда для нашего ряда:
otg1diff = otg.diff(periods=1).dropna()

```
test = sm.tsa.adfuller(otg1diff)
print 'adf: ', test[0]
print 'p-value: ', test[1]
print'Critical values: ', test[4]
if test[0]> test[4]['5%']:
    print 'есть единичные корни, ряд не стационарен'
else:
    print 'единичных корней нет, ряд стационарен'
```

adf: -5.95204224907

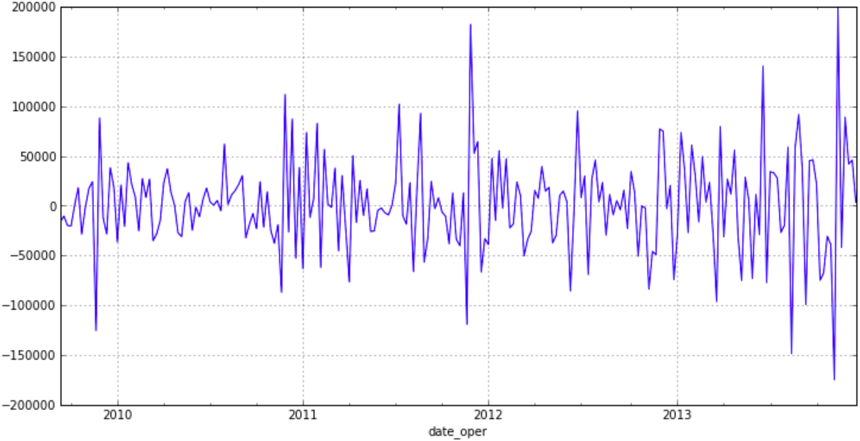
p-value: 2.13583392404e-07

Critical values: {'5%': -2.8755379867788462, '1%': -3.4621857592784546, '10%': -2.574231080806213}

единичных корней нет, ряд стационарен

Анализ временного ряда

Построим график:

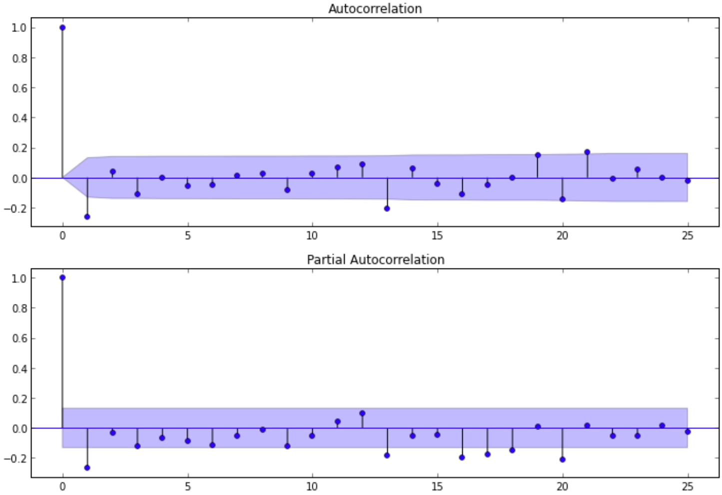


Построение модели временного ряда

Чтобы построить соответствующие коррелограммы, в пакете statsmodels имеются следующие функции: plot_acf() и plot_pacf(). Они выводят графики ACF и PACF, у которых по оси X откладываются номера лагов, а по оси Y значения соответствующих функций. Нужно отметить, что количество лагов в функциях и определяет число значимых коэффициентов. Итак, наши функции выглядят так:

```
ig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig = sm.graphics.tsa.plot_acf(otg1diff.values.squeeze(), lags=25, ax=ax1)
ax2 = fig.add_subplot(212)
fig = sm.graphics.tsa.plot_pacf(otg1diff, lags=25, ax=ax2)
```

Построение модели временного ряда



Построение модели временного ряда

Итак, когда известны все параметры можно построить модель

```
src_data_model = otg[:'2013-05-26']
model = sm.tsa.ARIMA(src_data_model, order=(1,1,1), freq='W').fit(full_output=False, disp=0)
```

Среднеквадратичное отклонение нашей модели: 80919.057367642512
Средняя абсолютная ошибка прогноза: 63092.763277651895
Осталось нарисовать наш прогноз на графике

Построение модели временного ряда