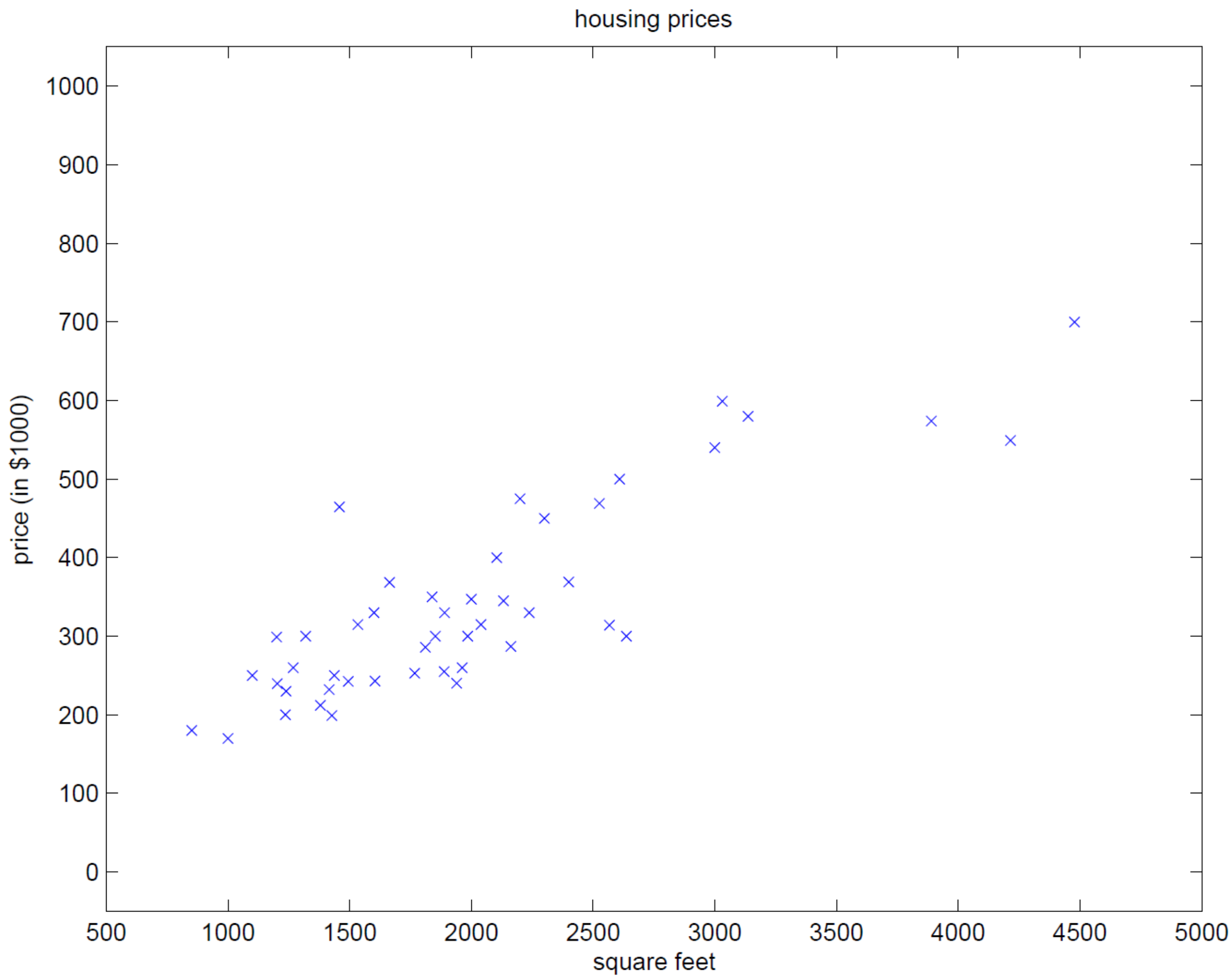
LEC003 Demand Forecasting

VG441 SS2020

Cong Shi
Industrial & Operations Engineering
University of Michigan

Linear Regression

- “Best fitting line”



Linear Regression

- (Linear) Hypothesis Function:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

- Minimizing the least-squares cost:

$$J\left(\theta_{0\ldots n}\right) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Matrix Derivation

- Turn everything into matrix notation

$$h_\theta(x) = \theta^T x \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_n \end{pmatrix} \in \mathbb{R}^{n+1} \qquad x = \begin{pmatrix} x_0 \triangleq 1 \\ x_1 \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^{n+1}$$

- Design matrix

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m-1)} \\ x^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & \dots & \dots & x_n^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(m)} & \dots & \dots & \dots & x_n^{(m)} \end{bmatrix}$$

Normal Equation

- Cost function

$$J(\theta) = \frac{1}{2m}(X\theta - y)^T(X\theta - y)$$

- Throwing out the constant

$$J(\theta) = \theta^T X^T X\theta - 2(X\theta)^T y + y^T y$$

- The “famous” normal equation

$$\frac{\partial J}{\partial \theta} = \underline{2X^T X\theta} - 2X^T y = 0$$

$$X^T X\theta = X^T y$$

$$\theta = (X^T X)^{-1} X^T y$$

$$\begin{cases} \frac{\partial x^T B}{\partial x} = B \\ \frac{\partial x^T b}{\partial x} = b \\ \frac{\partial x^T x}{\partial x} = 2x \end{cases}$$

$$\begin{cases} \frac{\partial x^T B x}{\partial x} \\ = 2Bx \end{cases}$$

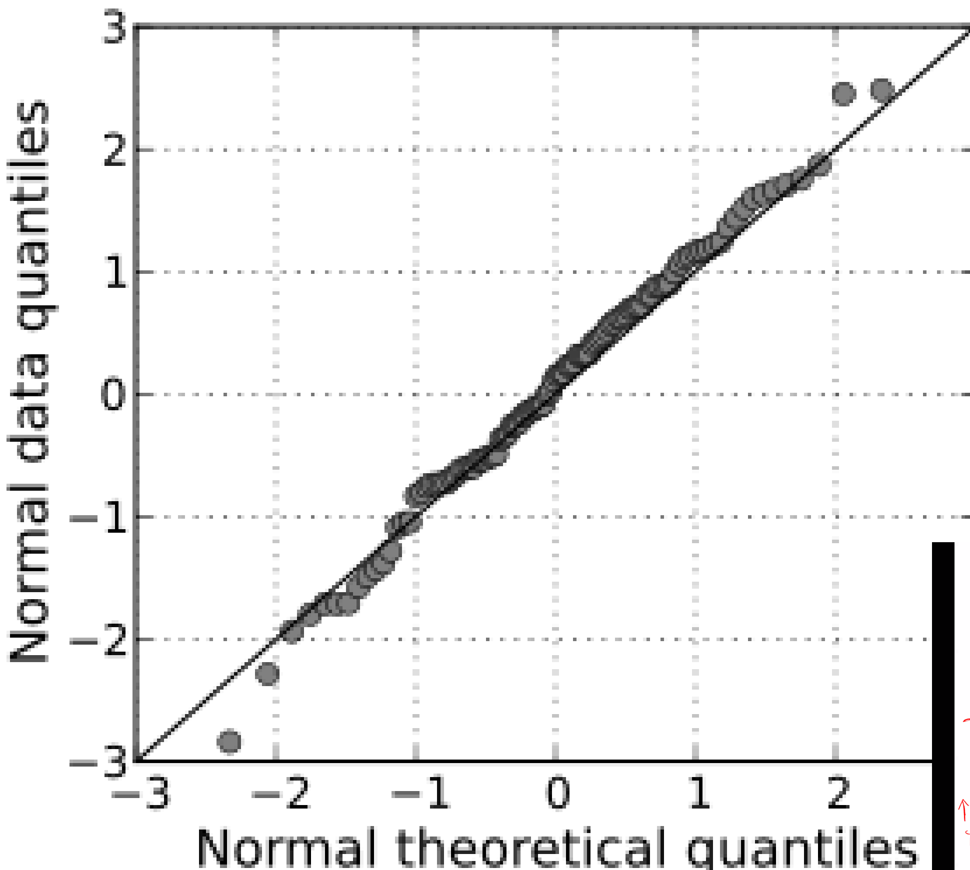
Symmetric

$B = B^T$

Residuals

- The residual should be normally distributed

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$



Ho: R^2 not sig.

$$F = \frac{[SS(mean) - SS(fit)]/(\#param - 1)}{SS(fit)/(n - \#param)}$$

variance that can be explained by features.

variance not explained by features.

F large → small p-value → ~~more lik~~ → more likely to reject Ho

lin. reg. Two most important questions

(1) R^2 r-squared $\in [0, 1]$

(2) Is R^2 significant? $SS(fit) = a_1^2 + a_2^2 + a_3^2 + a_4^2$

R^2: ↑ knowledge in SCM

$y - (\theta_0 + \theta_1 x)$

$var(fit) = \frac{SS(fit)}{n}$

$var(mean) = \frac{SS(mean)}{n}$ (without considering x)

x explains R^2 amount of variation

x: time spent on VG441

$$R^2 = \frac{var(mean) - var(fit)}{var(mean)}$$

Probabilistic Interpretation

- Deriving the log-likilihood function

$$\epsilon^{(i)} = y^{(i)} - \theta^T x^{(i)} \sim \mathcal{N}\left(0, \sigma^2\right)$$

$$p\left(\epsilon^{(i)}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(\epsilon^{(i)}\right)^2}{2\sigma^2}\right)$$

$$p\left(y^{(i)} | x^{(i)}; \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

Probabilistic Interpretation

- Deriving the log-likilihood function

$$L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y}|X;\theta)$$

$$L(\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)};\theta\right)$$
$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

Probabilistic Interpretation

- Maximum Likelihood Estimator (M

$$
\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right) \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right) \\
&= m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^{m} \left(y^{(i)} - \theta^T x^{(i)}\right)^2
\end{aligned}
$$

Out[12]:

OLS Regression Results

Dep. Variable:	HOUSING PRICE	R-squared:	0.959
Model:	OLS	Adj. R-squared:	0.958
Method:	Least Squares	F-statistic:	891.1
Date:	Tue, 12 May 2020	Prob (F-statistic):	0.00
Time:	23:16:28	Log-Likelihood:	-1523.8
No. Observations:	506	AIC:	3074.
Df Residuals:	493	BIC:	3129.
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
CRIM	-0.0916	0.034	-2.675	0.008	-0.159	-0.024
ZN	0.0487	0.014	3.379	0.001	0.020	0.077
INDUS	-0.0038	0.064	-0.059	0.953	-0.130	0.123
CHAS	2.8564	0.904	3.160	0.002	1.080	4.633
NOX	-2.8808	3.359	-0.858	0.392	-9.481	3.720
RM	5.9252	0.309	19.168	0.000	5.318	6.533
AGE	-0.0072	0.014	-0.523	0.601	-0.034	0.020

96%

t-test

x_1

θ_1

{crim rate does not matter}

$H_0: \theta_1 = 0$

$H_A: \theta_1 \neq 0$

P-value small

→ reject H_0

Python Time!

- from sklearn import linear_model
- from statsmodels.api import OLS

ARMA(p,q)

- AR(p)

$$X_t = \mu + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

- MA(q)

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

- ARMA(p,q)

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

ARIMA(p,d,q)

- d is the degree of difference, e.g., d =1

$$Y_t = X_t - X_{t-1}$$

$$Y_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

- How about d = 2?

```
df = pd.read_csv('PowerUsage.csv', names=['value'], header=0)
df.head()
```

Out[2]:

	value
0	88
1	84
2	85
3	85
4	84

In [3]:

```
from statsmodels.tsa.stattools import adfuller
from numpy import log
result = adfuller(df.value.dropna())
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -2.464240
p-value: 0.124419
```

hypo test for stationarity

Augmented Dickey-Fuller test.

H_0: t.s. not stationary

H_A: t.s stationary

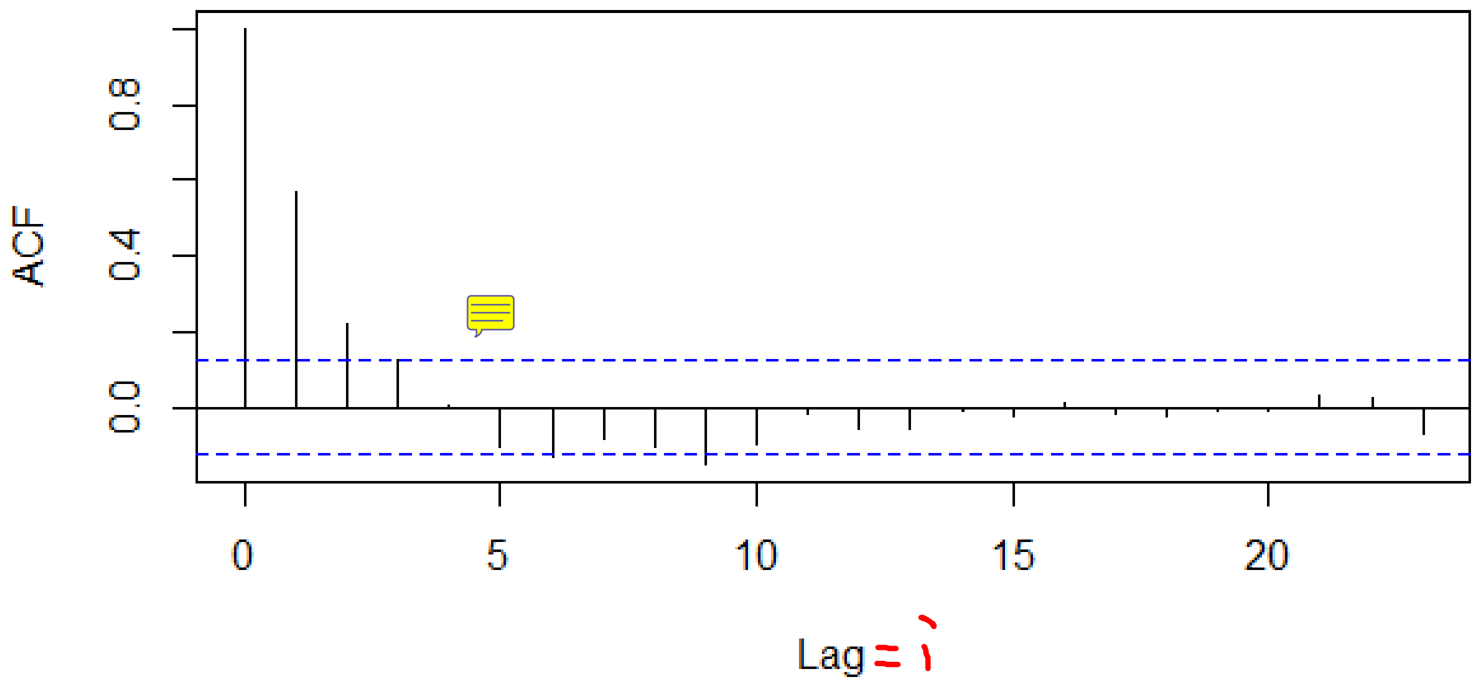
large

fail to reject H_0

In [ ]:

```
# Original Series
fig, axes = plt.subplots(3, 2, sharex=True)
axes[0, 0].plot(df.value); axes[0, 0].set_title('Original Series')
plot_acf(df.value, ax=axes[0, 1])
```

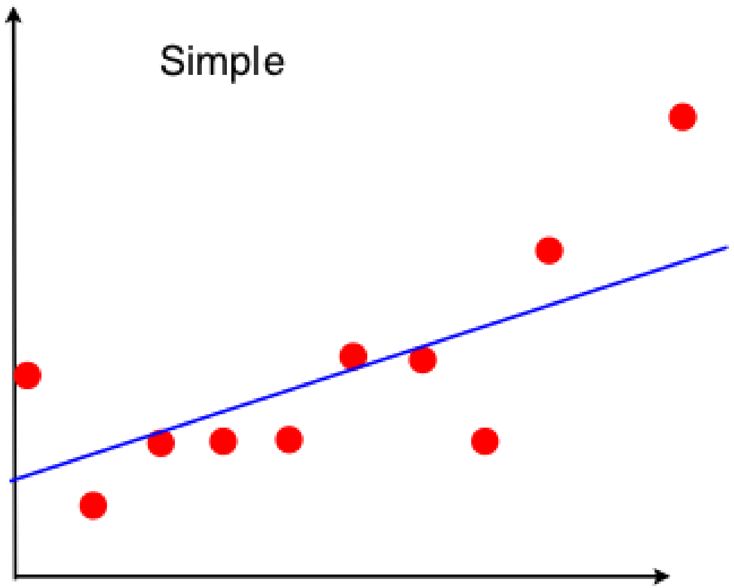
ACF/PACF

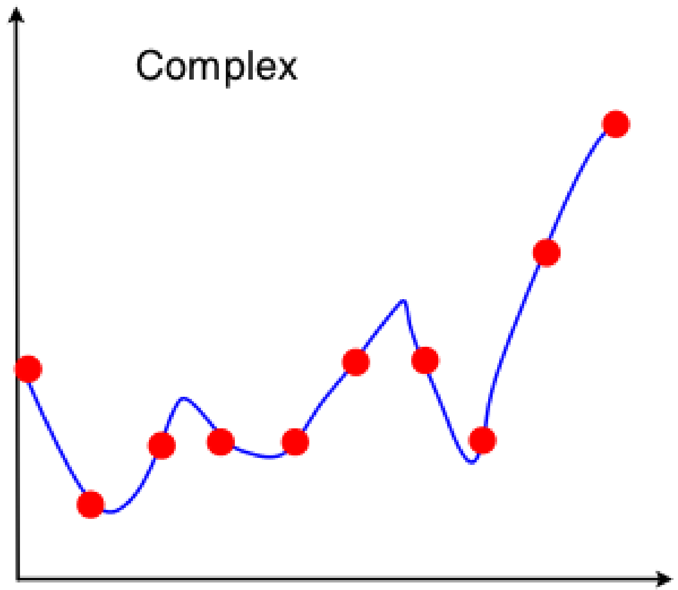


Python Time!

- statsmodels.tsa.arima_model

Bias-Variance Tradeoff





Kaggle Competition

All Competitions

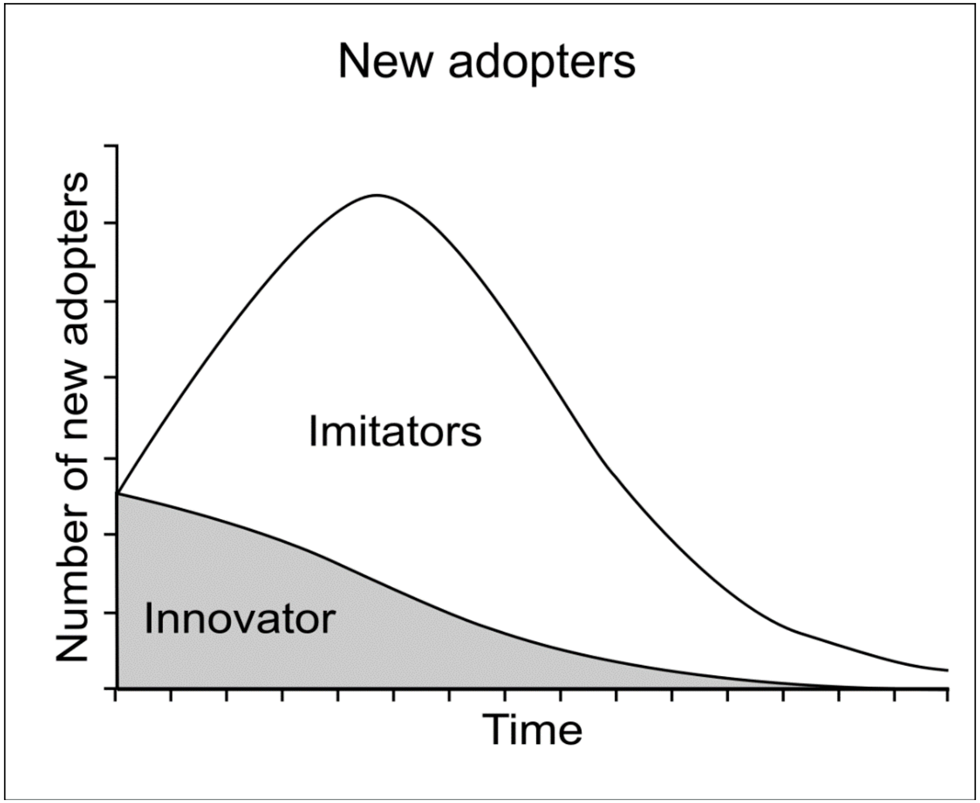
Active | Completed | InClass

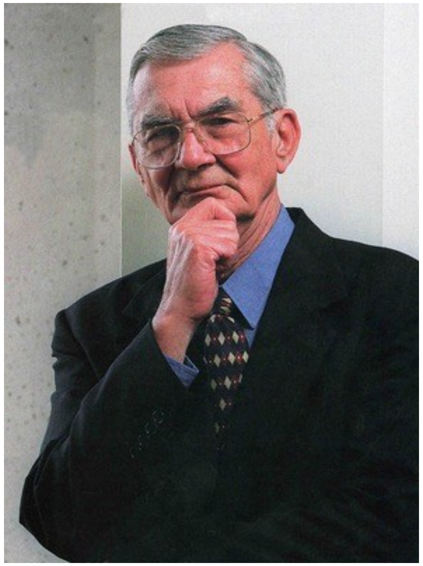
All Categories ▾ | Default Sort ▾

Jigsaw Multilingual Toxic Comment Classification Use TPUs to identify toxicity comments across multiple languages Featured • a month to go • Code Competition • 862 Teams	$50,000
M5 Forecasting - Accuracy Estimate the unit sales of Walmart retail goods Featured • 2 months to go • 3589 Teams	$50,000
M5 Forecasting - Uncertainty Estimate the uncertainty distribution of Walmart unit sales. Featured • 2 months to go • 389 Teams	$50,000
University of Liverpool - Ion Switching Identify the number of channels open at each time point Research • 16 days to go • 2333 Teams	$25,000
TReNDS Neuroimaging Multiscanner normative age and assessments prediction with brain function, structure, and connectivity Research • 2 months to go • 275 Teams	$25,000
ALASKA2 Image Steganalysis Detect secret data hidden within digital images Research • 2 months to go • 237 Teams	$25,000
Prostate cANcer graDe Assessment (PANDA) Challenge Prostate cancer diagnosis using the Gleason grading system Featured • 2 months to go • Code Competition • 309 Teams	$25,000

Bass Diffusion Model

- How about projecting sales of new products?





Frank Bass

Bass Diffusion Model

Cumulative purchase probability of a random customer $F(t)$

Purchase probability at time t $f(t) = F'(t)$

The rate of purchase at time t (given no purchase so far)

$$\frac{f(t)}{1 - F(t)} = p + qF(t)$$

Coefficient of innovation

Coefficient of imitation

Bass Solution

$$\frac{dF/dt}{1-F} = p + qF$$

$$\frac{dF}{dt} = p + (q-p)F - qF^2$$

$$\int \frac{1}{p + (q-p)F - qF^2} dF = \int dt$$

$$\begin{aligned}
\frac{1}{(p+qF)(1-F)} &= \frac{A}{p+qF} + \frac{B}{1-F} \\
&= \frac{A - AF + pB + qFB}{(p+qF)(1-F)} \\
&= \frac{A + pB + F(qB - A)}{(p+qF)(1-F)}
\end{aligned}$$

$$A = q/(p+q)$$
$$B = 1/(p+q)$$

Bass Solution

$$\int \frac{1}{(p+qF)(1-F)} dF = \int dt$$

$$\int \left(\frac{A}{p+qF} + \frac{B}{1-F} \right) dF = t + c_1$$

$$\int \left(\frac{q/(p+q)}{p+qF} + \frac{1/(p+q)}{1-F} \right) dF = t + c_1$$

$$\frac{1}{p+q} \ln(p+qF) - \frac{1}{p+q} \ln(1-F) = t + c_1$$

$$\frac{\ln(p+qF) - \ln(1-F)}{p+q} = t + c_1$$

Boundary Condition

$$t = 0 \quad \Rightarrow \quad F(0) = 0$$

$$t = 0 \quad \Rightarrow \quad c_1 = \frac{\ln p}{p+q}$$

Bass Solution

$$F(t) = \frac{p\left(e^{(p+q)t} - 1\right)}{pe^{(p+q)t} + q}$$

Calibration

- Sales in any period are $s(t) = mf(t)$
- Cumulative sales up to time t are $S(t) = mF(t)$

$$\frac{s(t)/m}{1 - S(t)/m} = p + qS(t)/m$$

$$s(t) = [p + qS(t)/m][m - S(t)]$$

$$s(t) = \beta_0 + \beta_1 S(t) + \beta_2 S(t)^2 \quad (BASS)$$
$$\beta_0 = pm$$
$$\beta_1 = q - p$$
$$\beta_2 = -q/m$$

$$m = \frac{-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_0\beta_2}}{2\beta_1}$$

$$p = \frac{\beta_0}{m}; \quad q = -m\beta_2$$

Conduct a linear regression!

Python Time!

- from sklearn import linear_model
- from statsmodels.api import OLS

Discrete Choice Model

- Multinomial Logit Model

$$U_{ni} = V_{ni} + \epsilon_{ni}$$

Gumbel

$$f(x) = e^{-x} e^{-e^{-x}}$$

$$F(x) = e^{-e^{-x}}$$

$$\begin{aligned} P_{ni} &= \mathbb{P}\left(U_{ni} > U_{nj} \quad \forall j \neq i\right) \\ &= \mathbb{P}\left(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj} \quad \forall j \neq i\right) \end{aligned}$$

Discrete Choice Model

- Multinomial Logit Model

Gumbel $f(x) = e^{-x} e^{-e^{-x}}$

$F(x) = e^{-e^{-x}}$

$$
\begin{aligned}
P_{ni} &= \int \left(P_{ni} | \epsilon_{ni}\right) f\left(\epsilon_{ni}\right) d\epsilon_{ni} \\
&= \int \left(P_{ni} | \epsilon_{ni}\right) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} d\epsilon_{ni} \\
&= \int \left(\prod_{j \neq i} e^{-e^{-\left(\epsilon_{ni} + V_{ni} - V_{nj}\right)}} \right) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} d\epsilon_{ni}
\end{aligned}
$$

$$\Longrightarrow \qquad P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$