Workshop - Regression-Based Classification

Does `statsmodels` marginal effect use the average of covariates or the average predicted values?

- Use the class data.
- Show your work.

Load the necessary packages and data:

Favorite Food: Cantonese Dimsum Hobbies: Basketball, Barista

In [2]:
```
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
```

In [3]:
```
df = pd.read_pickle('/Users/liaohaitao/Desktop/ECON 490/Lecture 2.2/class_data.pkl')
df.head()
```

Out[3]:

fips	year	GeoName	pct_d_rgdp	urate_bin	pos_net_jobs	emp_estabs	estabs_entry_rate	estabs_exit_rate	pop	pop_pct_black	pop_pct_hisp
1001	2002	Autauga, AL	3.202147	lower	1	12.531208	11.268	9.256	45909.0	17.386569	1.611884
	2003	Autauga, AL	1.434404	lower	1	12.598415	10.603	9.940	46800.0	17.493590	1.692308
	2004	Autauga, AL	15.061365	lower	1	12.780078	11.140	8.519	48366.0	17.584667	1.796717
	2005	Autauga, AL	0.333105	higher	1	12.856784	11.735	8.673	49676.0	17.612127	1.986875
	2006	Autauga, AL	7.440034	higher	1	12.832506	10.645	8.766	51328.0	17.898613	2.032029

Fit a logistic regression using either `sm.Logit()` or `smf.logit()`.

In [6]:
```
fit_logit = smf.logit(data = df, formula = 'pos_net_jobs ~ pct_d_rgdp + estabs_exit_rate').fit()
```

```
Optimization terminated successfully.
         Current function value: 0.670572
         Iterations 5
```

Get the marginal effects (`.get_margeff()`). Print the summary (`.summary()`).

In [7]:
```
fit_logit.get_margeff().summary()
```

Out[7]:

Logit Marginal Effects

Dep. Variable:	pos_net_jobs
Method:	dydx
At:	overall

	dy/dx	std err	z	P>|z|	[0.025	0.975]
pct_d_rgdp	0.0056	0.000	21.167	0.000	0.005	0.006
estabs_exit_rate	-0.0278	0.001	-32.051	0.000	-0.029	-0.026

Covariate Averages

$$\frac{\partial p(x_i)}{\partial \beta_1} \approx \frac{e^{\hat{\beta}_0+\bar{x}\hat{\beta}_1+\bar{x}\hat{\beta}_2}}{(1+e^{\hat{\beta}_0+\bar{x}\hat{\beta}_1+\bar{x}\hat{\beta}_2})^2}\hat{\beta}$$

In [8]:
```
beta = fit_logit.params
avgs = np.array([1., np.mean(df.pct_d_rgdp), np.mean(df.estabs_exit_rate)])
```

In [9]:
```
( np.exp(sum(beta*avgs)))/((1 + np.exp(beta*avgs))**2)*beta
```

Out[9]:
```
Intercept           0.080768
pct_d_rgdp          0.007228
estabs_exit_rate   -0.081482
dtype: float64
```

Predicted values Averages

$$\frac{\partial p(x_i)}{\partial \beta_1} \approx \frac{1}{n} \sum_{i=1}^{n} \frac{e^{\hat{y}_i}}{1 + e^{\hat{y}_i}} \hat{\beta}$$

In [10]:
```
yhat = fit_logit.fittedvalues
```

In [12]:
```
np.mean((np.exp(yhat))/((1 + np.exp(yhat)))**2)*beta
```

Out[12]:
```
Intercept           0.296016
pct_d_rgdp          0.005610
estabs_exit_rate   -0.027778
dtype: float64
```

Interpretaton

Interpret the marginal effect on one feature.

An increase in the pct_d_rgdp by one point is associated with an increase in the probability of positive net job creation rate of 0.00561. An increase in the establishment exit rate by one percentage point is associated with an decrease in the probability of positive net job creation rate of 0.02778.

In [ ]:
```

```