Economic Models, Spring 2020
Dr. Eric Van Dusen
Alan Liang
Umar Maniku

Lab 11: Consumer Spending and Economic Stimulus Payments

In [1]:
```
import warnings
warnings.simplefilter("ignore")
import statsmodels.api as sm
import numpy as np
import pandas as pd
import statsmodels.sandbox.regression.gmm as sm_gmm

def get_dummies(tbl, col, drop = True, drop_first = True):
    """Creates dummy variables for a column of a table"""
    values = np.unique(tbl[col])
    if drop_first:
        values = values[1:]
    for val in values:
        encoding = tbl[col].apply(lambda s: int(s == val))
        tbl[col + "_" + str(val)] = encoding
    if drop:
        tbl = tbl.drop(columns = col)
    return tbl
```

Welcome to lab 11!

In this lab, you will investigate how economic stimulus payments in the form of tax rebates affect household consumption. This lab is based on *Household Expenditure and the Income Tax Rebates of 2001* by David S. Johnson, Jonathan A Parker, and Nicholas S Souleles (which we will refer to as the JPS study for short).

In 2001, the Bush administration passed the *Economic Growth and Tax Relief Reconciliation Act of 2001*, which mainly reduced the rates of individual income taxes. In addition, the bill initiated a series of one-time rebates for all taxpayers that filed a tax return for 2000. The payment of these rebates were broadly announced, so that most households were aware of an incoming stimulus payment (much similar to the recent stimulus check). The rebate was as follows:

- Up to a maximum of $300 for single filers with no dependents
- Up to a maximum of $500 for single parents
- Up to a maximum of $600 for married couples

We are interested in determining how individuals altered their consumption patterns due to the economic stimulus payments, and by extent see if the permanent income hypothesis holds. The permanent income hypothesis states that consumers attempt to smooth their consumption across their life time, so that "changes in permanent income, rather than changes in temporary income, are what drive the changes in a consumer's consumption patterns." Intuitively, if the permanent income hypothesis were to hold, we would expect households to smooth out the spending of the rebate even before the rebate arrived. Thus, their consunmption would not change much between periods since they had known in the previous period that they would be receiving a sizable increase in income in the near future.

Notably, these stimulus payments were assigned to households at random periods in time, which allows us to better conclude a causal effect of a one time cash payment on changes in consumption.

To determine the true causal effect, we will use least squares regression. JPS propose the following regression for any household relating the stimulus payment to change in consumption:

$$C_{t+1} - C_t = \sum_s \beta_{0,s} \text{month}_s + \beta_1 \text{age} + \beta_2 \Delta \text{ children} + \beta_3 \Delta \text{ adults} + \beta_4 \text{Stimulus Payment}_{t+1} +$$

Here, we control for seasonal effects by creating dummy variables for each period (measured in months), and also control for changes in the number of children and adults in a household. Let's consider a few scenarios in context of the regression to gain some intuition:

- If a household received a stimulus in period $t+1$, then the change in consumption ($C_{t+1} - C_t$) due to the rebate should be captured by β_4 if we have sufficiently controlled for all potential factors of change in consumption between the 2 periods.
- If a household did not receive a stimulus payment in $t+1$, then the stimulus payment will be 0. Thus, the change in consumption will only be explained by our control variables: age, changes in family members, and seasonal variation.

Let's read in the table. The columns labels are:

Label	Description
`newid`	household ID
`year_month`	month when data was collected
`dcf`	change in food expenditures
`dcs`	change in strictly nondurable expenditures
`dcn`	change in nondurable expenditures
`dlcf`	change in log food expenditures
`dlcs`	change in log strictly nondurable expenditures
`dlcn`	change in log nondurable expenditures
`dnumadult`	change in number of adults
`dnumkids`	change in number of kids
`age`	average age of head & spouse (if exists)
`taxreb`	total rebates received in reference period
`ltaxreb`	rebates received in prior reference period (-1)
`l2taxreb`	rebates received in twice prior reference period (-2)

In [2]:
```
rebates = pd.read_csv("JPS.csv")
rebates.head()
```

Out[2]:

	newid	year_month	dcf	dcs	dcn	dlcf	dlcs	dlcn	dnumadult	dnumkids
0	113314	200103	281	343.0	352.0	0.618805	0.502042	0.284406	0	0
1	113314	200106	-129	-176.0	427.0	-0.238032	-0.226313	0.262581	0	0
2	113314	200109	-90	-42.0	169.0	-0.207639	-0.062520	0.087462	0	0
3	113318	200103	131	820.0	1241.0	0.067395	0.267209	0.344537	0	0
4	113318	200106	302	3147.0	3256.0	0.139978	0.641809	0.567968	0	0

One very important thing to note is that the unit of observation is not per household, but rather per time period per household. If a household were observed at 3 different time periods, then they would make up 3 rows and hence "contribute 3 times to the regression". This kind of set up is most oftne referred to as a *panel data study*.

Let's visualize the data. Below is a household that received a stimulus payment in August 2001.

In [3]:
```
rebates[rebates["newid"] == 116249]
```

Out[3]:

	newid	year_month	dcf	dcs	dcn	dlcf	dlcs	dlcn	dnumadult	dnu
1869	116249	200105	-469	-582.0	-465.0	-0.274464	-0.206017	-0.145046	0	
1870	116249	200108	-1226	-1359.0	-1494.0	-1.746342	-0.763995	-0.696173	0	
1871	116249	200111	555	646.0	471.0	1.145132	0.435119	0.275487	0	

Thus for the data point in which $t+1$ refers to August (so that t refers to May), $\text{Stimulus Payment}_{t+1} = 120$. For the data point in which $t+1$ refers to November (so that t refers to August), $\text{Stimulus Payment}_{t+1} = 0$ and $\text{Stimulus Payment}_{t} = 120$.

In general, we will use `taxreb` as the $\text{Stimulus Payment}_{t+1}$ variable.

Part 1 - OLS on Rebate as a Dollar Value

Let's try to recreate JPS' regression. We have selected the relevant columns for the independent variables:

In [4]:
```
X_q1 = rebates[["year_month", "dnumadult", "dnumkids", "age", "taxreb"]]
X_q1.head()
```

Out[4]:

	year_month	dnumadult	dnumkids	age	taxreb
0	200103	0	0	85.0	0
1	200106	0	0	85.0	0
2	200109	0	0	85.0	0
3	200103	0	0	51.0	0
4	200106	0	0	51.0	0

Question 1.1: Create dummy variables to represent the different months. Augment the `X_q1` table with dummy variables for `year_month` , and assign it to `X_q1_dummies` .

In [ ]:
```
X_q1_dummies = ...
X_q1_dummies.head()
```

In [5]:
```
## Solution ##
X_q1_dummies = get_dummies(X_q1, "year_month")
X_q1_dummies.head()
```

Out[5]:

	dnumadult	dnumkids	age	taxreb	year_month_200103	year_month_200104	year_month_200105
0	0	0	85.0	0	1	0	0
1	0	0	85.0	0	0	0	0
2	0	0	85.0	0	0	0	0
3	0	0	51.0	0	1	0	0
4	0	0	51.0	0	0	0	0

Question 1.2: Conduct an OLS regression of change in food consumption using `statsmodels` replicating JPS' setup. Interpret the coefficient on `taxreb` .

In [ ]:
```
q1_2_X = ...
q1_2_y = ...
model_q1_2 = sm.OLS(..., ...).fit()
model_q1_2.summary()
```

In [6]:
```python
## Solution ##
q1_2_X = X_q1_dummies
q1_2_y = rebates["dcf"]
model_q1_2 = sm.OLS(q1_2_y, q1_2_X).fit()
model_q1_2.summary()
```

Out[6]:

OLS Regression Results

Dep. Variable:	dcf	**R-squared (uncentered):**	0.006
Model:	OLS	**Adj. R-squared (uncentered):**	0.005
Method:	Least Squares	**F-statistic:**	4.383
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	2.06e-10
Time:	13:40:15	**Log-Likelihood:**	-1.2371e+05
No. Observations:	14960	**AIC:**	2.475e+05
Df Residuals:	14940	**BIC:**	2.476e+05
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	155.6932	30.700	5.071	0.000	95.517	215.869
dnumkids	53.1387	39.832	1.334	0.182	-24.938	131.215
age	0.6227	0.463	1.344	0.179	-0.285	1.531
taxreb	0.1044	0.050	2.093	0.036	0.007	0.202
year_month_200103	-5.5566	55.166	-0.101	0.920	-113.688	102.575
year_month_200104	-5.6264	52.532	-0.107	0.915	-108.595	97.343
year_month_200105	-50.4889	51.347	-0.983	0.325	-151.136	50.158
year_month_200106	-28.7730	39.826	-0.722	0.470	-106.836	49.290
year_month_200107	3.7976	40.657	0.093	0.926	-75.895	83.490
year_month_200108	30.2027	40.155	0.752	0.452	-48.505	108.911
year_month_200109	23.5979	35.794	0.659	0.510	-46.563	93.758
year_month_200110	15.3978	38.232	0.403	0.687	-59.541	90.337
year_month_200111	-113.1151	38.094	-2.969	0.003	-187.784	-38.446
year_month_200112	-158.7294	36.394	-4.361	0.000	-230.065	-87.393
year_month_200201	-100.7233	35.564	-2.832	0.005	-170.434	-31.013
year_month_200202	-74.0048	35.287	-2.097	0.036	-143.171	-4.838
year_month_200203	-10.9745	35.476	-0.309	0.757	-80.511	58.562
year_month_200204	-50.6713	41.458	-1.222	0.222	-131.934	30.591
year_month_200205	-18.3621	40.709	-0.451	0.652	-98.156	61.432
year_month_200206	-12.0670	41.915	-0.288	0.773	-94.225	70.091

Omnibus:	5669.566	**Durbin-Watson:**	2.568
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3743853.964
Skew:	0.232	**Prob(JB):**	0.00
Kurtosis:	80.498	**Cond. No.**	2.65e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.65e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Question 1.3: Conduct an OLS regression of change in consumption for strictly non-durable goods using `statsmodels` replicating JPS' setup. How does the coefficient on `taxreb` compare with that for 1.2?

In [ ]:
```
q1_3_X = ...
q1_3_y = ...
model_q1_3 = sm.OLS(..., ...).fit()
model_q1_3.summary()
```

In [7]:
```
## Solution ##
q1_3_X = q1_2_X
q1_3_y = rebates["dcs"]
model_q1_3 = sm.OLS(q1_3_y, q1_3_X).fit()
model_q1_3.summary()
```

Out[7]:

OLS Regression Results

Dep. Variable:	dcs	**R-squared (uncentered):**	0.007
Model:	OLS	**Adj. R-squared (uncentered):**	0.006
Method:	Least Squares	**F-statistic:**	5.475
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	2.79e-14
Time:	13:40:18	**Log-Likelihood:**	-1.3226e+05
No. Observations:	14960	**AIC:**	2.646e+05
Df Residuals:	14940	**BIC:**	2.647e+05
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	353.0230	54.334	6.497	0.000	246.521	459.525
dnumkids	100.8134	70.497	1.430	0.153	-37.368	238.995
age	0.6068	0.820	0.740	0.459	-1.000	2.214
taxreb	0.2501	0.088	2.833	0.005	0.077	0.423
year_month_200103	86.3582	97.634	0.885	0.376	-105.016	277.733
year_month_200104	200.3884	92.973	2.155	0.031	18.151	382.626
year_month_200105	99.3944	90.876	1.094	0.274	-78.734	277.523
year_month_200106	71.6821	70.485	1.017	0.309	-66.477	209.841
year_month_200107	-41.7049	71.956	-0.580	0.562	-182.748	99.338
year_month_200108	58.1288	71.067	0.818	0.413	-81.171	197.428
year_month_200109	35.5114	63.349	0.561	0.575	-88.661	159.683
year_month_200110	26.8340	67.664	0.397	0.692	-105.795	159.463
year_month_200111	-175.0873	67.420	-2.597	0.009	-307.239	-42.936
year_month_200112	-218.6684	64.411	-3.395	0.001	-344.921	-92.416
year_month_200201	-128.8856	62.943	-2.048	0.041	-252.261	-5.510
year_month_200202	-35.7558	62.452	-0.573	0.567	-158.169	86.657
year_month_200203	35.7564	62.786	0.569	0.569	-87.312	158.825
year_month_200204	52.3590	73.373	0.714	0.475	-91.462	196.180
year_month_200205	37.7762	72.048	0.524	0.600	-103.446	178.999
year_month_200206	71.8750	74.182	0.969	0.333	-73.530	217.280

Omnibus:	14750.825	**Durbin-Watson:**	2.437
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10235888.527
Skew:	3.927	**Prob(JB):**	0.00
Kurtosis:	130.904	**Cond. No.**	2.65e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.65e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Part 2 - OLS on Rebate as a Binary Value

For the second part, we will treat the variable Stimulus Payment as a binary variable, in which $\text{Stimulus Payment}_{t+1} = 1$ if the household received a stimulus payment in period $t + 1$, and $\text{Stimulus Payment}_{t+1} = 0$ if not.

In [9]:
```
X_q2 = rebates[["year_month", "dnumadult", "dnumkids", "age", "taxreb"]]
X_q2.head()
```

Out[9]:

	year_month	dnumadult	dnumkids	age	taxreb
0	200103	0	0	85.0	0
1	200106	0	0	85.0	0
2	200109	0	0	85.0	0
3	200103	0	0	51.0	0
4	200106	0	0	51.0	0

Question 2.1: Create a binary variable to represent whether a stimulus payment was received and add it to `X_q2` as a column called `itaxreb`. Make sure to drop `taxreb`.

In [ ]:
```
X_q2_1 = X_q2
X_q2_1["itaxreb"] = ...
X_q2_1 = ...
X_q2_1.head()
```

In [11]:
```
## Solution ##
X_q2_1 = X_q2
X_q2_1["itaxreb"] = (X_q2["taxreb"] > 0).astype(int)
X_q2_1 = X_q2_1.drop(columns = "taxreb")
X_q2_1.head()
```

Out[11]:

	year_month	dnumadult	dnumkids	age	itaxreb
0	200103	0	0	85.0	0
1	200106	0	0	85.0	0
2	200109	0	0	85.0	0
3	200103	0	0	51.0	0
4	200106	0	0	51.0	0

Question 2.2: Similar to 1.1, create dummy variables to represent the different months. Augment the `X_q2_1` table with dummy variables for `year_month`, and assign it to `X_q2_dummies`.

In [ ]:
```
X_q2_dummies = ...
X_q2_dummies.head()
```

In [12]:
```
## Solution ##
X_q2_dummies = get_dummies(X_q2_1, "year_month")
X_q2_dummies.head()
```

Out[12]:

	dnumadult	dnumkids	age	itaxreb	year_month_200103	year_month_200104	year_month_200105
0	0	0	85.0	0	1	0	0
1	0	0	85.0	0	0	0	0
2	0	0	85.0	0	0	0	0
3	0	0	51.0	0	1	0	0
4	0	0	51.0	0	0	0	0

Question 2.3: Conduct an OLS regression of change in food consumption using `statsmodels` replicating JPS' setup. Interpret the coefficient on `itaxreb`, and compare this with your results in 1.2.

In [ ]:
```
q2_3_X = ...
q2_3_y = ...
model_q2_3 = sm.OLS(..., ...).fit()
model_q2_3.summary()
```

In [13]:
```
## Solution ##
q2_3_X = X_q2_dummies
q2_3_y = rebates["dcf"]
model_q2_3 = sm.OLS(q2_3_y, q2_3_X).fit()
model_q2_3.summary()
```

Out[13]:

OLS Regression Results

Dep. Variable:	dcf	**R-squared (uncentered):**	0.006
Model:	OLS	**Adj. R-squared (uncentered):**	0.004
Method:	Least Squares	**F-statistic:**	4.328
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	3.18e-10
Time:	13:41:08	**Log-Likelihood:**	-1.2372e+05
No. Observations:	14960	**AIC:**	2.475e+05
Df Residuals:	14940	**BIC:**	2.476e+05
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	156.4078	30.698	5.095	0.000	96.237	216.579
dnumkids	53.3483	39.834	1.339	0.181	-24.732	131.428
age	0.6069	0.463	1.311	0.190	-0.301	1.514
itaxreb	48.2649	26.576	1.816	0.069	-3.828	100.358
year_month_200103	-4.7893	55.163	-0.087	0.931	-112.915	103.337
year_month_200104	-4.8074	52.528	-0.092	0.927	-107.769	98.154
year_month_200105	-49.6792	51.343	-0.968	0.333	-150.319	50.960
year_month_200106	-27.9922	39.820	-0.703	0.482	-106.044	50.060
year_month_200107	4.5889	40.652	0.113	0.910	-75.093	84.271
year_month_200108	30.8643	40.162	0.768	0.442	-47.858	109.587
year_month_200109	24.5977	36.013	0.683	0.495	-45.992	95.187
year_month_200110	17.8256	38.651	0.461	0.645	-57.935	93.586
year_month_200111	-110.9549	38.465	-2.885	0.004	-186.351	-35.558
year_month_200112	-157.5472	36.633	-4.301	0.000	-229.353	-85.742
year_month_200201	-99.9125	35.557	-2.810	0.005	-169.609	-30.216
year_month_200202	-73.2007	35.280	-2.075	0.038	-142.354	-4.048
year_month_200203	-10.1801	35.469	-0.287	0.774	-79.703	59.343
year_month_200204	-49.8565	41.452	-1.203	0.229	-131.108	31.395
year_month_200205	-17.5604	40.703	-0.431	0.666	-97.343	62.223
year_month_200206	-11.2663	41.909	-0.269	0.788	-93.413	70.881

Omnibus:	5671.471	**Durbin-Watson:**	2.568
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3744917.927
Skew:	0.233	**Prob(JB):**	0.00
Kurtosis:	80.509	**Cond. No.**	691.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Question 2.4: Conduct a simliar OLS regression of change in strictly non-durable consumption. How does the coefficient on `itaxreb` compare with your results in 1.3?

In [ ]:
```
q2_4_X = ...
q2_4_y = ...
model_q2_4 = sm.OLS(..., ...).fit()
model_q2_4.summary()
```

In [14]:
```
## Solution ##

q2_4_X = q2_3_X
q2_4_y = rebates["dcs"]
model_q2_4 = sm.OLS(q2_4_y, q2_4_X).fit()
model_q2_4.summary()
```

Out[14]:

OLS Regression Results

Dep. Variable:	dcs	**R-squared (uncentered):**	0.007
Model:	OLS	**Adj. R-squared (uncentered):**	0.006
Method:	Least Squares	**F-statistic:**	5.307
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	1.13e-13
Time:	13:41:13	**Log-Likelihood:**	-1.3226e+05
No. Observations:	14960	**AIC:**	2.646e+05
Df Residuals:	14940	**BIC:**	2.647e+05
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	354.8399	54.334	6.531	0.000	248.339	461.341
dnumkids	101.2076	70.505	1.435	0.151	-36.991	239.406
age	0.5567	0.820	0.679	0.497	-1.050	2.163
itaxreb	101.8670	47.039	2.166	0.030	9.665	194.069
year_month_200103	88.7778	97.636	0.909	0.363	-102.601	280.157
year_month_200104	202.9728	92.973	2.183	0.029	20.735	385.211
year_month_200105	101.9516	90.876	1.122	0.262	-76.176	280.080
year_month_200106	74.1539	70.480	1.052	0.293	-63.996	212.304
year_month_200107	-39.1984	71.952	-0.545	0.586	-180.233	101.836
year_month_200108	61.2649	71.085	0.862	0.389	-78.071	200.601
year_month_200109	43.0129	63.741	0.675	0.500	-81.928	167.954
year_month_200110	41.1439	68.411	0.601	0.548	-92.950	175.238
year_month_200111	-162.0852	68.082	-2.381	0.017	-295.534	-28.636
year_month_200112	-210.3061	64.839	-3.243	0.001	-337.399	-83.213
year_month_200201	-126.3338	62.935	-2.007	0.045	-249.694	-2.974
year_month_200202	-33.2261	62.444	-0.532	0.595	-155.624	89.172
year_month_200203	38.2612	62.779	0.609	0.542	-84.793	161.315
year_month_200204	54.9261	73.369	0.749	0.454	-88.885	198.738
year_month_200205	40.3025	72.043	0.559	0.576	-100.911	181.516

year_month_200206	74.4065	74.178	1.003	0.316	-70.991	219.804

Omnibus:	14768.797	**Durbin-Watson:**	2.437
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10258341.760
Skew:	3.936	**Prob(JB):**	0.00
Kurtosis:	131.044	**Cond. No.**	691.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Question 2.5: What are the differences in consumption changes between food and strictly non-durables?

Type your answer here, replacing this text

Solution: From parts 1 and 2, we observe that households typically on food a bit less than half of that of strictly non-durables. This makes sense, since food is a subset of strictly non-durables.

Question 2.6: Looking at non-durables is more relevant in the context of the permanent income hypothesis. Strictly non-durable goods like food do not last between time periods, so that households consume the good in 1 period only. Thus, we can attribute a change in non-durable consumption to consumption that was actually carried out in the corresponding period.

What would we expect β_4 to be if the permanent income hypothesis were to hold? Is this result in line with your OLS results from parts 1 and 2?

Type your answer here, replacing this text

Solution: If the permanent income hypothesis were true, we would expect a stimulus payment to not increase consumption between these 2 periods, i.e. $\beta_4 = 0$. This is because in the previous period t, households anticipate a change in incomes in the near future at $t + 1$, causing them to increase their consumption in t. Since we are statistically confident that β_4 is not 0, we can reject the permanent income hypothesis.

Part 3 - Instrumental Variables

One concern from the regerssion in part 1 is that there may be confounding variables between one's rebate amount and the change in consumption; for example, the size of a household will affect how much a household receives in the rebate and also how much in consumption changes across periods. JPS address this by conducting an instrumental variable regression to better determine causality. They use the binary variable on whether one received a rebate as an instrument for the rebate amount, and construct a 2 stage least squares regression. Note that all control variables are added in both the first stage and structural models.

For this part, you do not have to understand how the IV regression is conducted in python. You simply have to interpret the results from constructing this model.

Question 3.1: Does the instrument satisfy the conditions of *exogeneity* and *relevance*?

Type your answer here, replacing this text

Solution: The instrument is exogenous of all potentially confounding variables since the receipt of a rebate was based on one's social security number and thus essentially randomly timed. The instrument is clearly relevant as whether one receives a rebate is positively correlated with how much one receives on the rebate.

Question 3.2: We have constructed a 2 stage least squares model below for the change in food consumption. Do the results differ much from that of part 1?

In [15]:
```
Z_q3 = X_q2_dummies
X_q3 = X_q1_dummies
model_q3_2 = sm_gmm.IV2SLS(rebates['dcf'], X_q3, instrument = Z_q3).fit()
model_q3_2.summary()
```

Out[15]:

IV2SLS Regression Results

Dep. Variable:	dcf	**R-squared:**	0.006
Model:	IV2SLS	**Adj. R-squared:**	0.005
Method:	Two Stage	**F-statistic:**	nan
	Least Squares	**Prob (F-statistic):**	nan
Date:	Thu, 21 May 2020		
Time:	13:42:46		
No. Observations:	14960		
Df Residuals:	14940		
Df Model:	20		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	155.7284	30.701	5.072	0.000	95.550	215.907
dnumkids	53.1333	39.832	1.334	0.182	-24.943	131.210
age	0.6208	0.463	1.340	0.180	-0.288	1.529
taxreb	0.1010	0.056	1.816	0.069	-0.008	0.210
year_month_200103	-5.4654	55.170	-0.099	0.921	-113.605	102.674
year_month_200104	-5.5289	52.537	-0.105	0.916	-108.507	97.450
year_month_200105	-50.3922	51.352	-0.981	0.326	-151.049	50.264
year_month_200106	-28.6791	39.831	-0.720	0.472	-106.754	49.396
year_month_200107	3.8929	40.663	0.096	0.924	-75.811	83.597
year_month_200108	30.4010	40.180	0.757	0.449	-48.357	109.159
year_month_200109	24.2125	36.070	0.671	0.502	-46.489	94.914
year_month_200110	16.4450	38.978	0.422	0.673	-59.957	92.847
year_month_200111	-112.1526	38.728	-2.896	0.004	-188.064	-36.242
year_month_200112	-158.0605	36.715	-4.305	0.000	-230.027	-86.094
year_month_200201	-100.6275	35.571	-2.829	0.005	-170.351	-30.904
year_month_200202	-73.9098	35.294	-2.094	0.036	-143.090	-4.730
year_month_200203	-10.8800	35.482	-0.307	0.759	-80.430	58.670
year_month_200204	-50.5747	41.464	-1.220	0.223	-131.849	30.699
year_month_200205	-18.2670	40.715	-0.449	0.654	-98.073	61.539
year_month_200206	-11.9710	41.920	-0.286	0.775	-94.140	70.198

Omnibus:	5670.022	**Durbin-Watson:**	2.568
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3744309.268
Skew:	0.232	**Prob(JB):**	0.00
Kurtosis:	80.503	**Cond. No.**	2.65e+03

Question 3.3: We have constructed a 2 stage least squares model below for the change in strictly non-durable consumption. Do the results differ much from that of part 1?

In [16]:
```
model_q3_3 = sm_gmm.IV2SLS(rebates['dcs'], X_q3, instrument = Z_q3).fit()
model_q3_3.summary()
```

Out[16]:

IV2SLS Regression Results

Dep. Variable:	dcs	**R-squared:**	0.007
Model:	IV2SLS	**Adj. R-squared:**	0.006
Method:	Two Stage	**F-statistic:**	nan
	Least Squares	**Prob (F-statistic):**	nan
Date:	Thu, 21 May 2020		
Time:	13:42:48		
No. Observations:	14960		
Df Residuals:	14940		
Df Model:	20		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	353.4061	54.336	6.504	0.000	246.900	459.912
dnumkids	100.7537	70.497	1.429	0.153	-37.429	238.936
age	0.5862	0.820	0.715	0.475	-1.022	2.194
taxreb	0.2131	0.098	2.166	0.030	0.020	0.406
year_month_200103	87.3509	97.641	0.895	0.371	-104.038	278.740
year_month_200104	201.4500	92.982	2.167	0.030	19.195	383.705
year_month_200105	100.4468	90.885	1.105	0.269	-77.699	278.593
year_month_200106	72.7041	70.495	1.031	0.302	-65.476	210.884
year_month_200107	-40.6672	71.967	-0.565	0.572	-181.731	100.397
year_month_200108	60.2870	71.113	0.848	0.397	-79.103	199.677
year_month_200109	42.1999	63.838	0.661	0.509	-82.931	167.331
year_month_200110	38.2301	68.985	0.554	0.579	-96.990	173.450
year_month_200111	-164.6132	68.542	-2.402	0.016	-298.964	-30.263
year_month_200112	-211.3893	64.980	-3.253	0.001	-338.758	-84.020
year_month_200201	-127.8429	62.955	-2.031	0.042	-251.243	-4.443
year_month_200202	-34.7227	62.464	-0.556	0.578	-157.160	87.714
year_month_200203	36.7838	62.798	0.586	0.558	-86.308	159.876
year_month_200204	53.4103	73.384	0.728	0.467	-90.432	197.252
year_month_200205	38.8111	72.058	0.539	0.590	-102.432	180.054
year_month_200206	72.9190	74.192	0.983	0.326	-72.507	218.345

Omnibus:	14757.884	**Durbin-Watson:**	2.437

Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10247084.549
Skew:	3.931	**Prob(JB):**	0.00
Kurtosis:	130.974	**Cond. No.**	2.65e+03

Optional Part 4 - Incorporating Previous Period Payments into the OLS

One potential issue with the JPS study is that households who received a period in period t but not in $t+1$ will have the same $\text{Stimulus Payment}_{t+1}$ value as households who did not receive a stimulus payment at all. This may not be the case - intuitively we might expect consumption to decrease if a payment was issued in the previous period but not the current period. Instead, we will implement an added variable on whether a household received stimulus payment in period t to better control for the causal effect of a stimulus payment on changes in consumption.

We will augment the JPS setup with an added variable $\text{Stimulus Payment}_t$:

$$C_{t+1} - C_t = \sum_s \beta_{0,s}\,\text{month}_s + \beta_1\,\text{age} + \beta_2 \Delta\text{ children} + \beta_3 \Delta\text{ adults} + \beta_4 \text{Stimulus Payment}_{t+1} +$$

Thus:

- If a household received a stimulus in period $t+1$ only, then the change in consumption ($C_{t+1} - C_t$) due to the rebate should be captured by β_4 if we have sufficiently controlled for all potential factors of change between the 2 periods.
- If a household did not receive a stimulus payment in $t+1$ or t, then both stimulus payment variables will be 0. Thus, the change in consumption will only be explained by our control variables: age, changes in family members, and seasonal variation.
- If a household received a stimulus in period t only, then the change in consumption ($C_{t+1} - C_t$) due to the rebate should be captured by β_5 if we have sufficiently controlled for all potential factors of change between the 2 periods.

Notably, interpreting the coefficient β_5 for $\text{Stimulus Payment}_t$ will allow us to determine how much consumption will change in the period after a household receives the stimulus check.

The columns `ltaxreb` reflect the stimulus payment received in the previous period.

In [17]:
```
X_q4 = rebates[["year_month", "dnumadult", "dnumkids", "age", "taxreb", "lt
X_q4.head()
```

Out[17]:

	year_month	dnumadult	dnumkids	age	taxreb	ltaxreb
0	200103	0	0	85.0	0	0.0
1	200106	0	0	85.0	0	0.0
2	200109	0	0	85.0	0	0.0
3	200103	0	0	51.0	0	0.0
4	200106	0	0	51.0	0	0.0

Question 4.1: Conduct a new regression of change in food consumption using the new regression model proposed above. Interpret β_4 and β_5. Do your results differ much from that of part 1?

In [ ]:
```
q4_1_X = ...
q4_1_y = ...
model_q4_1 = ...
...
```

In [18]:
```
## Solution ##
q4_1_X = get_dummies(X_q4, "year_month")
q4_1_y = rebates["dcf"]
model_q4_1 = sm.OLS(q4_1_y, q4_1_X).fit()
model_q4_1.summary()
```

Out[18]:

OLS Regression Results

Dep. Variable:	dcf	**R-squared (uncentered):**	0.006
Model:	OLS	**Adj. R-squared (uncentered):**	0.004
Method:	Least Squares	**F-statistic:**	4.182
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	4.13e-10
Time:	13:42:58	**Log-Likelihood:**	-1.2371e+05
No. Observations:	14960	**AIC:**	2.475e+05
Df Residuals:	14939	**BIC:**	2.476e+05
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	155.5102	30.704	5.065	0.000	95.326	215.694
dnumkids	52.9902	39.835	1.330	0.183	-25.091	131.072
age	0.6104	0.464	1.315	0.189	-0.299	1.520
taxreb	0.1039	0.050	2.084	0.037	0.006	0.202
ltaxreb	-0.0214	0.051	-0.421	0.673	-0.121	0.078
year_month_200103	-4.9604	55.185	-0.090	0.928	-113.130	103.210
year_month_200104	-4.9892	52.555	-0.095	0.924	-108.004	98.025
year_month_200105	-49.8546	51.371	-0.970	0.332	-150.548	50.839
year_month_200106	-28.1522	39.854	-0.706	0.480	-106.271	49.967
year_month_200107	4.4274	40.686	0.109	0.913	-75.321	84.176
year_month_200108	30.8372	40.184	0.767	0.443	-47.928	109.603
year_month_200109	24.2828	35.832	0.678	0.498	-45.952	94.517
year_month_200110	16.1355	38.273	0.422	0.673	-58.884	91.155
year_month_200111	-111.7899	38.225	-2.925	0.003	-186.715	-36.865
year_month_200112	-155.0726	37.415	-4.145	0.000	-228.410	-81.735
year_month_200201	-94.3690	38.630	-2.443	0.015	-170.089	-18.649
year_month_200202	-67.9538	38.098	-1.784	0.074	-142.630	6.722
year_month_200203	-6.9524	36.738	-0.189	0.850	-78.964	65.059
year_month_200204	-50.0431	41.486	-1.206	0.228	-131.360	31.274
year_month_200205	-17.7434	40.736	-0.436	0.663	-97.592	62.105

year_month_200206	-11.4359	41.943	-0.273	0.785	-93.648	70.777

Omnibus:	5673.117	**Durbin-Watson:**	2.568
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3748573.516
Skew:	0.234	**Prob(JB):**	0.00
Kurtosis:	80.547	**Cond. No.**	2.73e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.73e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Question 4.2: Conduct a similar regression as 4.1 on change in strictly non-durable consumption. Interpret β_4 and β_5. Do your results differ much from that of part 1?

In [ ]:
```
q4_2_X = ...
q4_2_y = ...
model_q4_2 = ...
...
```

In [19]:

```
## Solution ##
q4_2_X = get_dummies(X_q4, "year_month")
q4_2_y = rebates["dcs"]
model_q4_2 = sm.OLS(q4_2_y, q4_2_X).fit()
model_q4_2.summary()
```

Out[19]:

OLS Regression Results

Dep. Variable:	dcs	**R-squared (uncentered):**	0.007
Model:	OLS	**Adj. R-squared (uncentered):**	0.006
Method:	Least Squares	**F-statistic:**	5.375
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	1.64e-14
Time:	13:43:06	**Log-Likelihood:**	-1.3225e+05
No. Observations:	14960	**AIC:**	2.645e+05
Df Residuals:	14939	**BIC:**	2.647e+05
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	351.6118	54.335	6.471	0.000	245.108	458.116
dnumkids	99.6685	70.494	1.414	0.157	-38.508	237.845
age	0.5119	0.821	0.623	0.533	-1.098	2.122
taxreb	0.2467	0.088	2.795	0.005	0.074	0.420
ltaxreb	-0.1652	0.090	-1.836	0.066	-0.342	0.011
year_month_200103	90.9539	97.658	0.931	0.352	-100.468	282.376
year_month_200104	205.3009	93.004	2.207	0.027	23.002	387.600
year_month_200105	104.2837	90.908	1.147	0.251	-73.907	282.474
year_month_200106	76.4679	70.527	1.084	0.278	-61.774	214.710
year_month_200107	-36.8500	71.999	-0.512	0.609	-177.977	104.277
year_month_200108	63.0199	71.111	0.886	0.376	-76.367	202.406
year_month_200109	40.7907	63.409	0.643	0.520	-83.499	165.081
year_month_200110	32.5202	67.729	0.480	0.631	-100.238	165.278
year_month_200111	-164.8717	67.644	-2.437	0.015	-297.462	-32.281
year_month_200112	-190.4803	66.211	-2.877	0.004	-320.262	-60.699
year_month_200201	-79.9038	68.361	-1.169	0.242	-213.900	54.093
year_month_200202	10.8876	67.419	0.161	0.872	-121.262	143.038
year_month_200203	66.7602	65.014	1.027	0.304	-60.674	194.195
year_month_200204	57.2010	73.415	0.779	0.436	-86.701	201.103
year_month_200205	42.5454	72.089	0.590	0.555	-98.758	183.848

year_month_200206	76.7396	74.223	1.034	0.301	-68.747	222.226

Omnibus:	14763.087	**Durbin-Watson:**	2.437
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10229816.482
Skew:	3.934	**Prob(JB):**	0.00
Kurtosis:	130.865	**Cond. No.**	2.73e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.73e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Question 4.3: We have repeated the regression of change in food consumption, but with the stimulus payment as a binary variable, like in part 2. Interpret β_4 and β_5. Do your results differ much from that of part 2?

In [20]:
```
X_q4_3 = X_q4
X_q4_3["itaxreb"] = (X_q4["taxreb"] > 0).astype(int)
X_q4_3["iltaxreb"] = (X_q4["ltaxreb"] > 0).astype(int)
X_q4_3.head()
```

Out[20]:

	year_month	dnumadult	dnumkids	age	taxreb	ltaxreb	year_month_200103	year_month_20010
0	200103	0	0	85.0	0	0.0	1	C
1	200106	0	0	85.0	0	0.0	0	C
2	200109	0	0	85.0	0	0.0	0	C
3	200103	0	0	51.0	0	0.0	1	C
4	200106	0	0	51.0	0	0.0	0	C

5 rows × 24 columns

In [21]:
```
q4_3_X = get_dummies(X_q4_3, "year_month")
q4_3_y = rebates["dcf"]
model_q4_3 = sm.OLS(q4_3_y, q4_3_X).fit()
model_q4_3.summary()
```

Out[21]:

OLS Regression Results

Dep. Variable:	dcf	**R-squared (uncentered):**	0.006
Model:	OLS	**Adj. R-squared (uncentered):**	0.004
Method:	Least Squares	**F-statistic:**	3.825
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	1.68e-09
Time:	13:43:24	**Log-Likelihood:**	-1.2371e+05
No. Observations:	14960	**AIC:**	2.475e+05
Df Residuals:	14937	**BIC:**	2.476e+05
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	155.4279	30.712	5.061	0.000	95.229	215.627
dnumkids	53.0338	39.839	1.331	0.183	-25.056	131.124
age	0.6142	0.464	1.323	0.186	-0.296	1.524
taxreb	0.1167	0.113	1.034	0.301	-0.105	0.338
ltaxreb	0.0160	0.115	0.139	0.889	-0.209	0.241
year_month_200103	-5.1419	55.191	-0.093	0.926	-113.322	103.039
year_month_200104	-5.1837	52.561	-0.099	0.921	-108.209	97.842
year_month_200105	-50.0473	51.376	-0.974	0.330	-150.751	50.657
year_month_200106	-28.3393	39.859	-0.711	0.477	-106.468	49.790
year_month_200107	4.2369	40.691	0.104	0.917	-75.523	83.996
year_month_200108	30.7986	40.199	0.766	0.444	-47.996	109.594
year_month_200109	24.7044	36.061	0.685	0.493	-45.979	95.388
year_month_200110	16.8597	38.731	0.435	0.663	-59.058	92.778
year_month_200111	-110.7440	38.634	-2.867	0.004	-186.470	-35.018
year_month_200112	-153.3293	37.794	-4.057	0.000	-227.410	-79.248
year_month_200201	-92.4552	38.980	-2.372	0.018	-168.860	-16.050
year_month_200202	-66.0671	38.443	-1.719	0.086	-141.420	9.286
year_month_200203	-5.3404	37.004	-0.144	0.885	-77.873	67.192
year_month_200204	-50.2357	41.491	-1.211	0.226	-131.564	31.092
year_month_200205	-17.9330	40.742	-0.440	0.660	-97.792	61.926
year_month_200206	-11.6272	41.948	-0.277	0.782	-93.850	70.596

itaxreb	-7.8537	60.147	-0.131	0.896	-125.750	110.043
iltaxreb	-22.0997	60.494	-0.365	0.715	-140.675	96.475

Omnibus:	5673.466	**Durbin-Watson:**	2.568
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3746101.951
Skew:	0.235	**Prob(JB):**	0.00
Kurtosis:	80.521	**Cond. No.**	2.74e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.74e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Question 4.4: Lastly we have also repeated the regression of change in consumption of strictly non-durables on `Stimulus Payment` as a binary variable. Interpret β_4 and β_5. Do your results differ much from that of part 2?

In [22]:
```
q4_4_X = q4_3_X
q4_4_y = rebates["dcs"]
model_q4_4 = sm.OLS(q4_4_y, q4_4_X).fit()
model_q4_4.summary()
```

Out[22]:

OLS Regression Results

Dep. Variable:	dcs	**R-squared (uncentered):**	0.008
Model:	OLS	**Adj. R-squared (uncentered):**	0.006
Method:	Least Squares	**F-statistic:**	4.969
Date:	Thu, 21 May 2020	**Prob (F-statistic):**	5.06e-14
Time:	13:43:27	**Log-Likelihood:**	-1.3225e+05
No. Observations:	14960	**AIC:**	2.646e+05
Df Residuals:	14937	**BIC:**	2.647e+05
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
dnumadult	350.6928	54.347	6.453	0.000	244.167	457.219
dnumkids	99.5177	70.499	1.412	0.158	-38.668	237.704
age	0.5295	0.822	0.644	0.519	-1.081	2.140
taxreb	0.3975	0.200	1.990	0.047	0.006	0.789
ltaxreb	-0.0233	0.203	-0.115	0.909	-0.421	0.375
year_month_200103	90.1086	97.664	0.923	0.356	-101.324	281.541
year_month_200104	204.3969	93.010	2.198	0.028	22.086	386.707
year_month_200105	103.3911	90.914	1.137	0.255	-74.811	281.593
year_month_200106	75.6083	70.534	1.072	0.284	-62.646	213.863
year_month_200107	-37.7231	72.006	-0.524	0.600	-178.863	103.417
year_month_200108	63.7727	71.135	0.897	0.370	-75.660	203.206
year_month_200109	46.4465	63.812	0.728	0.467	-78.632	171.525
year_month_200110	41.2486	68.537	0.602	0.547	-93.093	175.590
year_month_200111	-155.2748	68.365	-2.271	0.023	-289.278	-21.272
year_month_200112	-179.3889	66.879	-2.682	0.007	-310.480	-48.298
year_month_200201	-72.4692	68.977	-1.051	0.293	-207.673	62.735
year_month_200202	18.2045	68.027	0.268	0.789	-115.137	151.547
year_month_200203	72.9335	65.481	1.114	0.265	-55.417	201.284
year_month_200204	56.3013	73.421	0.767	0.443	-87.614	200.216
year_month_200205	41.6603	72.095	0.578	0.563	-99.655	182.976
year_month_200206	75.8569	74.230	1.022	0.307	-69.642	221.356

itaxreb	-90.5998	106.435	-0.851	0.395	-299.225	118.026
iltaxreb	-84.4608	107.047	-0.789	0.430	-294.287	125.365

Omnibus:	14756.315	**Durbin-Watson:**	2.437
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	10222619.640
Skew:	3.931	**Prob(JB):**	0.00
Kurtosis:	130.821	**Cond. No.**	2.74e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.74e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

An Afterword

Johnson, Parker, and Souleles' results from their paper can be seen below. Part 1 corresponds to the leftmost set of regressions, part 2 to matches the second set from the left, and part 3 is the right-most set.

Table 2: The contemporaneous response of expenditures to the tax rebate

Dependent Variable:	ΔC Dollar change in			ΔC Dollar change in			ΔlnC Percent change in			ΔC Dollar change in		
	Food	Non-durable goods (strict)	Non-durable goods	Food	Non-durable goods (strict)	Non-durable goods	Food	Non-durable goods (strict)	Non-durable goods	Food	Non-durable goods (strict)	Non-durable goods
Estimation method:	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS	2SLS
Rebate	0.109 (0.056)	0.239 (0.115)	0.373 (0.135)							0.108 (0.058)	0.202 (0.112)	0.375 (0.136)
I(Rebate>0)				51.5 (27.6)	96.2 (53.6)	178.8 (65.0)	2.72 (1.36)	1.76 (1.05)	3.16 (1.02)			
Age	0.570 (0.320)	0.449 (0.550)	1.165 (0.673)	0.552 (0.318)	0.391 (0.548)	1.106 (0.670)	0.035 (0.020)	0.005 (0.016)	0.023 (0.015)	0.569 (0.320)	0.424 (0.549)	1.166 (0.671)
Change in adults	130.3 (57.8)	285.8 (90.0)	415.8 (102.8)	131.1 (57.8)	287.7 (90.2)	418.6 (102.9)	6.16 (2.08)	6.22 (1.58)	7.55 (1.50)	130.3 (57.7)	286.2 (90.0)	415.7 (102.7)
Change in children	73.7 (45.3)	98.3 (82.4)	178.4 (98.3)	74.0 (45.3)	98.7 (82.5)	179.2 (98.3)	3.99 (2.36)	3.73 (1.66)	4.59 (1.66)	73.7 (45.3)	98.3 (82.5)	178.4 (98.3)
N	13,066	13,066	13,066	13,066	13,066	13,066	13,007	13,066	13,066	13,066	13,066	13,066

Notes: All regressions include a full set of month dummies. Reported standard errors are adjusted for arbitrary within-household correlations and heteroskedasticity. The third triplet of three columns is multiplied by 100 so as to report a percent change. The last three columns report results from two-stage least squares regressions where *I(Rebate>0)* with the other regressors are used as instruments for *Rebate.*

In this lab, you conducted 4 types of regressions to 'pin down' the causal effect of rebates on changes in non-durable consumption. We made multiple design choices in each model and could have made many other adjustments as well. For example, we could have determined changes in log consumption, controlled for more variables, or only regressed on households that received a stimulus payment.

As you can see, doing econometrics often relies on many value judgments. Each subtle decision you make on your data or model may lead to large changes in your regression outcomes, and could be the difference between statistical significance and insignificance.

Something to keep in mind as we conduct many models to try out different adjustments is to be aware of potential p-hacking. With a p-value of 5%, we would expect to see statistically significant results even if the null hypothesis were true 1 in 20 times. Thus, even if the null hypothesis were true, a model may produce statistically significant results if we ran enough variations of the model.

In [ ]:

```
```