

The Relationship Between Luxury Consumption and Savings Behavior of the Middle Class

Christina Lynn, CFP® AFC® CDFA®



Level of Experience

- First-time user
- Third year Ph.D. student in personal financial planning
- Intend to use the CE for my dataset in my dissertation

Research Question

How do non-essential expenditures (luxury consumption) effect the savings behavior of the middle class?



Variables

Predictor Variable = **luxury expenditure**

(luxury expenditure/total outlays)

- Apparel
- Eating out
- Entertainment
- Furniture
- Personal care
- Home outlays
- Pets & toys
- Telephones & televisions
- Transportation expenditures
- Trips

Dependent Variable = **savings behavior**

After tax income – total outlays

Control Variables

- Income
- Demographic

CE Variable Names

Luxury expenditure:

Total outlays `etotacx4 + etotapx4`

Apparel `apparcq + apparpq`

Eating out `fdawaycq + fdawaypq`

Entertainment `entertcq + entertpq`

Furniture `furntrcq + furntrpq`

Personal care `perscacq + perscapq`

Home outlays `esheltrc + esheltrp`

Pets & Toys `pettoycq + pettoypq`

Telephones & Televisions `telephcq + telephpq + tvrdiocq + tvrdiopq`

Transportation expenditures `etranptc + etranptp`

Trips `ttotalc + ttotalp`

Savings: After tax income – total outlays

After tax income `rowmean(finatxe1
finatxe2 finatxe3 finatxe4 finatxe5)`

Total outlays `etotacx4 + etotapx4`

Before tax income `rowmean(fincbtx1
fincbtx2 fincbtx3 fincbtx4 fincbtx5)`

Where did I find these variables?

Interview Survey

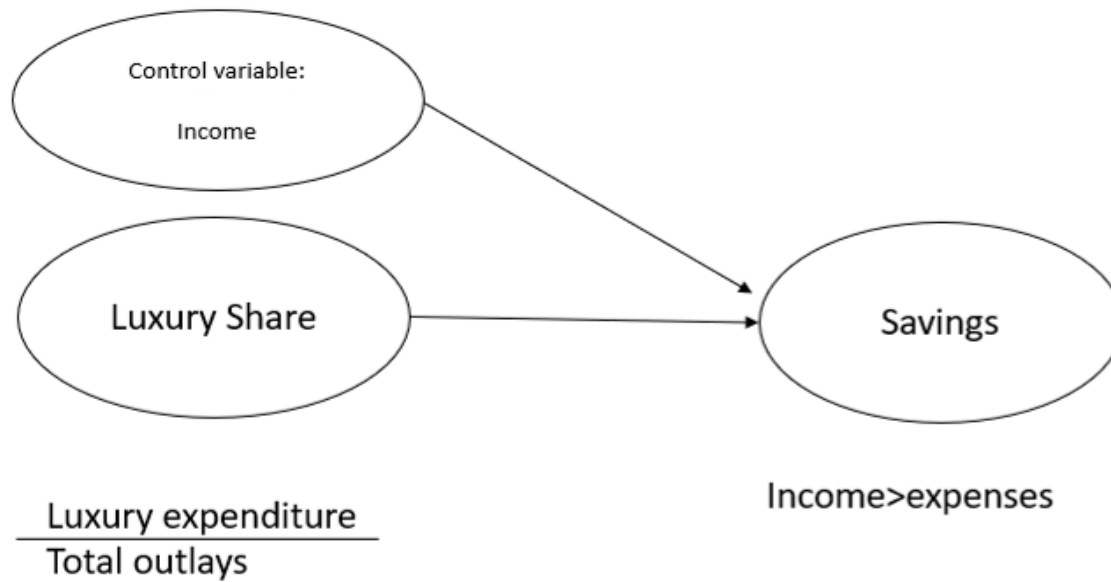
Quarters/years examined: 2019Q1-2019Q4

Data files used: FMLY (fmli191x, fmli192, fmli193, fmli194)

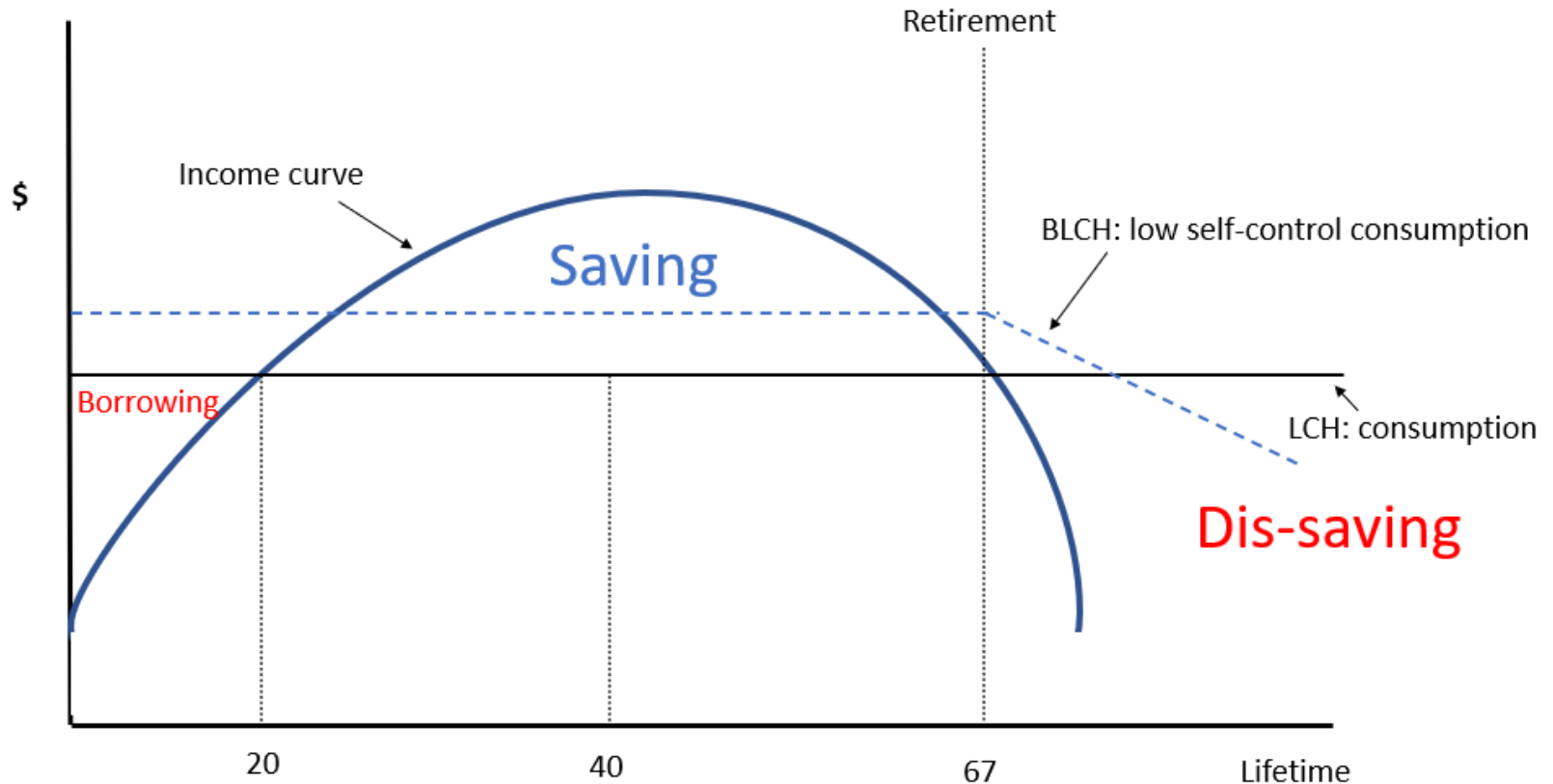
- Using STATA, I “appended” these 4 files.

Method Used

OLS regression



Theoretical Framework: Behavioral Life Cycle Hypothesis



Hypotheses

H₁: Self-control levels are associated with savings behavior.

(high self-control = low luxury share of total expenditures)

H₂: The impact of self-control (luxury share) on savings behavior varies according to a respondent's life cycle stage.

H₃: Household equivalent income is positively associated with savings behavior.

Problems I encountered

5th file for 2019: **fml201**.

- Problem: I didn't understand its purpose and if I needed to include it in order to obtain a year's worth of data. Intuitively I thought because it was included in the 2019 fml files, that it was a necessary element to complete a year's worth of data.
- Solution: Dr. Paulin pointed clarified what this was used for. I ended up excluding it from my sample.

Problems I encountered...continued

EXPN19 file **xpb19**

- Problem: I wanted to combine this file with my **fml**i files. The **xpb19** file contains data on **salon** services that I am interested in. Thus far, I have been unsuccessful in combining the **xpb19** and the **fml**i files, due to my lack of coding skills. Regardless of my limited coding skills, it seems puzzling that the PUMD files are not easily compatible.
 - An example of the problem I encountered in the attempt to combine the **xpb** with **fml**i files was receiving errors on merging incompatible storage types (string vs numeric).
- Solution: FMYL file has **perscare** summary variable the includes the salon services, but I just can't isolate the **salon** expenses.

Problems I encountered...continued

MEMB files (memi191x memi192 memi193, memi194)

- Problem: I wanted to use the **FINDRETX** variable as my dependent variable. This survey question is “Amount of money placed in a self-employment retirement plan in past year for all CU members”.
 - When attempting to merge the **memi** files, I was having trouble reconciling multiple **NEWID** tags for all the files.
 - Solution: My workaround for this was to operationalize my dependent variable (savings) in a different manner (which is less preferable but will suffice).

Problems I encountered...continued

EXP19 file (sub19)

- Problem: I was interested in including subscription expenditures in my analysis. **Sub19** is open to the public, but the column I needed to see (“**subdesc**”) didn’t include the description column. It appeared that this column data was not open to the public. Without this column data, I couldn’t determine the difference between subscription services on golf course dues and TV services.
- Solution: Not being able to determine between the types of subscription services was not helpful to me so I did not include it in my analysis.

Problems I encountered...continued

Cosmetic procedures

- Problem: I was interested in data on cosmetic procedures (e.g. plastic surgery, cosmetic fillers, etc.) but realized that the CE data was not suitable this type of analysis.
- Solution: relied on other luxury expenditures instead.

Problems I encountered...continued

“e” variables **ETOTALIX4** (total outlays) to **TOTEX4IQ** (total expenditures)

- Problem: I didn't see information in the Getting Started Guide on outlay variables, with an “e” at the front of the variable in the code dictionary.
- Solution: I relied on help from a CE expert, Dr. Paulin.
- Request: it would be helpful to have a clear written explanation of this in the Getting Started Guide, or somewhere else.

Problems I encountered...continued

Variables ending in **cq, pq, x, m**

- Problem:
 - It was challenging to understand how to handle CQ PQ variables. When I searched for these terms in the Getting Started Guide, there were no results.
 - I had difficulty understanding the nuanced difference between variables ending in cq or pq, and those ending in x or m (for example: **fpripenm** and **fpripenx**)
- Solution: I relied on help from a CE expert: Dr. Paulin, who explained this to me. I still need to better understand the variables ending in m and x.

Problems I encountered...continued

Time periods: Monthly vs Quarterly vs Annual

- Problem: after I append the 4 2019 fmli files, is my data monthly, quarterly, or annual?
 - **Annual?**
 - “The CE sample is designed to be representative of the entire **annual** U.S. population in the collection of each **quarter**. Thus, the weight (FINLWT21) needs to be divided by 4 to adjust for this fact. Without this adjustment the population in the denominator would be 4 times as large as the U.S. population. For example for an annual estimate (4 quarters) QNUM is 4.”
 - **Quarterly?**
 - the CE survey asks questions like
 - “Did (you/you or any members of your household) receive any reduced or free rent for this unit as a form of **pay since the first of (reference month)?** “
 - **Monthly?**
 - the CE survey asks questions like:
 - “How much was (your/your household's) payment on this (mortgage/lump sum home equity loan) **in (reference month)?**”
 - “What was the total amount paid in finance, late charges, and interest for all student loans **in the last month?**”
 - “How much were you billed for in **(last month)?** Do not include any unpaid charges from a previous billing period?”
- Solution: I have postponed utilizing the variables I don't understand until I am further along in the process.

Problems I encountered...continued

Cashcoiq

- Problem: I found this variable in the code dictionary, which applicable to my project. It was described as “Cash contributions” in the code dictionary, but I could not ascertain to what type of account. I did not see this clarified in the code dictionary.
 - This is an example of the challenge I encountered when comparing variables in the code dictionary to the survey questionnaire – they don’t appear to always line up in a way for a new user to understand.
- Solution: I did not include this variable in my study.

Problems I encountered...continued

Weights

- Problem: I had trouble understanding how to properly use weights.
 - FINLWT21
 - wtrep01 - wtrep 44
 - QNUM: "adjusts the weights from annual to quarterly. The CE sample is designed to be representative of the entire annual U.S. population in the collection of each quarter. Thus, the weight (FINLWT21) needs to be divided by 4 to adjust for this fact. Without this adjustment the population in the denominator would be 4 times as large as the U.S. population. For example for an annual estimate (4 quarters) QNUM is 4"
- Problem: I still do not fully understand this and how to properly use weights in my analysis.

Problems I encountered...continued

STATA

- Problem: I was using STATA and realized quickly that the CE does not offer as much support for STATA as they do for SAS.
 - Example: I'm looking for a better way to handle multiple imputation in STATA (STATA has different coding than SAS for this). There is SAS code for this, but I'm still trying to figure it out for STATA

```
egen aftertaxinc = rowmean(finatxe1 finatxe2 finatxe3 finatxe4 finatxe5)
```

- Solution: For now, I decided to continue using STATA, because that is the preferred coding platform of my university, but will switch to SAS if I am unable to perform suitable analyses. I hired a STATA tutor to help me with the coding.

Problems I encountered...continued

Employment

- Problem: How to control for full-time work.
 - Full time or part time or self-employed is not a question asked in FMLI. Any suggestions for how to measure this?
- Solution:
 - Used **incweek1** where 50 weeks worked per year = FT status
 - Incweek1 = "Number of weeks worked by reference person full or part time in last 12 months, including paid vacation and paid sick leave"

gen fulltime = 0

Replace fulltime = 1 if **incweek1** >=50

Requests

- Can you create a video explaining how to navigate some of these questions for beginners? For example: how to merge PUMD files (**fqli** and **memi**) in STATA.
- Explain how to use stub files:
 - Example: I wanted to know more details about **telephiq** variable in the **fqli** file (in the code dictionary, it just describes it as “telephone services this quarter part 1” – I was interested in whether that included cost of phone purchase or lease, unlimited data coverage, etc). Dr. Jimmy Choi was nice enough to direct me to using the stub files. I am not sure I found the right thing, but I didn’t understand how to use the file I found.

Ideas for improvement

- Relocate outdated variables to a different (new) tab in the code dictionary. Having only the latest variables listed would be more stream-lined and user friendly for new users.
- It looks like there are duplicates in the code dictionary of some variables, such as **age_ref** and **womensixiq** and **cashcoiq**. Remove duplicates to avoid confusion for new users, or clearly differentiate similar-looking variables. If there are not duplicates (maybe I'm just not seeing the distinction?), can you make the distinction clearer?
- Providing more detail in the code book of the survey question wording, and what the variable looks like in the data set (these did not always seem to match). Alternatively, list directly on the survey questionnaire the variable code and file it is housed in.
 - Maybe the stub file helps with this? Or there is another recourse available?
- **Pettoyiq** may be worth breaking apart in FMLI. I think expenditures on pets is a cultural trend to examine, and it would be easier to do so if it had its own variable.

Results

```
. reg savings apparel fdaway entertainment perscare tvphone petstoys trips furniture home
> t female i.age_cat i.education i.famsize i.race incweek1 if middle_3_quintiles == 1
```

Source	SS	df	MS	Number of obs	=	13,013
Model	1.4114e+12	28	5.0406e+10	F(28, 12984)	=	131.64
Residual	4.9717e+12	12,984	382911634	Prob > F	=	0.0000
				R-squared	=	0.2211
				Adj R-squared	=	0.2194
Total	6.3831e+12	13,012	490554917	Root MSE	=	19568

savings	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
apparel	-.693908	.4346405	-1.60	0.110	-1.545867	.1580512
fdaway	2.033305	.2771528	7.34	0.000	1.490045	2.576565
entertainment	-.9204673	.1577021	-5.84	0.000	-1.229587	-.611348
perscare	5.76344	1.350177	4.27	0.000	3.116894	8.409986
tvphone	3.848888	.4635574	8.30	0.000	2.940247	4.757528
petstoys	2.086434	.5000286	4.17	0.000	1.106304	3.066563
trips	-.4956498	.1478135	-3.35	0.001	-.785386	-.2059137
furniture	-1.598764	.3306644	-4.84	0.000	-2.246914	-.950613
home	-.0656214	.0647736	-1.01	0.311	-.1925871	.0613443
transport	-.5241463	.0670237	-7.82	0.000	-.6555225	-.39277

maritalstat						
widowed	-2948.957	746.9159	-3.95	0.000	-4413.022	-1484.892
divorced or sepearated	-3716.926	551.0837	-6.74	0.000	-4797.13	-2636.721
never married	-4701.353	557.9505	-8.43	0.000	-5795.017	-3607.688
female	-2579.138	358.6942	-7.19	0.000	-3282.231	-1876.045
age_cat						
30 - 40	3026.082	661.5	4.57	0.000	1729.445	4322.719
40 - 50	2090.155	694.6298	3.01	0.003	728.5782	3451.731
50 - 60	4001.464	695.6842	5.75	0.000	2637.821	5365.107
60 - 70	1767.788	710.4887	2.49	0.013	375.1256	3160.45
Age above 70	218.8974	791.5455	0.28	0.782	-1332.648	1770.443
education						
hsdiploma	3535.098	900.5291	3.93	0.000	1769.929	5300.268
somecollege	6608.634	870.4148	7.59	0.000	4902.494	8314.775
bachelors	11961.94	892.2281	13.41	0.000	10213.04	13710.83
advanceddegree	13488.86	946.0134	14.26	0.000	11634.54	15343.19
famsize						
2-3	9096.925	497.5599	18.28	0.000	8121.635	10072.22
4-5	14995.1	674.068	22.25	0.000	13673.83	16316.37
6plus	18366.84	1043.104	17.61	0.000	16322.21	20411.48
race						
nonwhite	-1949.195	449.4057	-4.34	0.000	-2830.096	-1068.294
incweek1	221.1118	9.343717	23.66	0.000	202.7967	239.4268
_cons	19454.6	1215.199	16.01	0.000	17072.63	21836.56

Limitations

- I'm new to this data set; there is a lot I don't know.
- CE dataset: lack of a direct measure of self-control, and no financial attitude questions.
- Endogeneity regarding savings behavior must also be considered.

Additional slides in case there are questions on variable coding:

Marital Status

```
gen maritalstat = 0
replace maritalstat =1 if marital1 == "1"
replace maritalstat =2 if marital1 == "2"
replace maritalstat =3 if marital1 == "3" marital1 == "4"
replace maritalstat =4 if marital1 == "5"
label define maritallabel 1 "married" 2 "widowed" 3
"divorced or sepearated" 4 "never married"
label values maritalstat maritallabel
```

Education

```
gen education = 0
replace education =1 if high_edu == "0" | high_edu == "10" |
high_edu == "11"
replace education =2 if high_edu == "12"
replace education =3 if high_edu == "13" | high_edu == "14"
replace education =4 if high_edu == "15"
replace education =5 if high_edu == "16"
label define edlabel 1 "lessthanhs" 2 "hsdiploma" 3
"somecollege" 4 "bachelors" 5 "advanceddegree"
label values education edlabel
```

Employment

```
gen fulltime = 0
replace fulltime = 1 if incweek1 >=50
```

Age

```
gen age_cat = .
replace age_cat =1 if age_ref < 30
replace age_cat =2 if age_ref >=30 & age_ref<40
replace age_cat =3 if age_ref >=40 & age_ref<50
replace age_cat =4 if age_ref >=50 & age_ref<60
replace age_cat =5 if age_ref >=60 & age_ref<70
replace age_cat =6 if age_ref >=70
label define agelabel 1 "Age up to 30" 2 "30 - 40" 3
"40 - 50" 4 "50 - 60" 5 "60 - 70" 6 "Age above 70"
label values age_cat agelabel
```

Gender

```
gen female = 0
replace female = 1 if sex_ref == "2"
```

Family size

```
gen famsize = .
replace famsize =1 if fam_size == 1
replace famsize =2 if fam_size >=2 & fam_size <=3
replace famsize =3 if fam_size >=4 & fam_size <=5
replace famsize =4 if fam_size >=6
label define famlabel 1 "1" 2 "2-3" 3 "4-5" 4 "6plus"
label values famsize famlabel
```

Employment

```
gen fulltime = 0  
replace fulltime = 1 if incweek1 >=50
```

Race

```
gen race = 0  
replace race =1 if ref_race =="1" | race2 =="1"  
replace race =2 if ref_race >"1" | race2 >"1"  
label define racelabel 1 "white" 2 "nonwhite"  
label values race racelabel
```

d