Discretionary spending by high-income households and the marketing of ultra-premium credit cards

All code and text AI created with prompts generated by Chris Blanchard (chris@chrisblanchard.org)

H_1 (Research Hypothesis)

Increases in discretionary spending by high-income households (top 10%) are positively associated with the intensity of marketing campaigns for ultra-premium credit cards (e.g., Amex Platinum, Chase Sapphire Reserve), such that periods of elevated affluent spending predict subsequent increases in issuer advertising activity.

This formulation does three things:

1. **Specifies the independent variable**: discretionary spending by wealthy households.
2. **Specifies the dependent variable**: marketing intensity for premium cards.
3. **States the expected direction**: positive association, with predictive timing implied.

Would you like me to also draft the **null hypothesis (H_0)** and a **testable operationalization** (including measurable proxies and time-lag structure)? Or even a **full hypothesis set** for a paper (H_1, H_2 for lead-lag, H_3 for interaction with seasonality)?

> What kind of correlation might we run to confirm that increased spending by wealthy households is driving the revamp and heavy marketing of high-end credit cards?

Great question. If we want to **test whether increased spending by wealthy households is driving the revamp and heavy marketing of ultra-premium cards**, we'll need to set up a design that can distinguish simple co-movement from directional influence.

Below is a practical playbook—variables, methods, and code templates—to move from "it looks like a lot of ads" to statistically credible evidence.

1) Define the Variables

A. Affluent spending (the potential driver)

- *Primary metric*: Monthly or quarterly **consumer spending by income cohort** (top 10%, top 5%, top 1%).
- *Practical sources*:
    - Bank/processor aggregates by income cohort (e.g., bank institute research).
    - **BLS Consumer Expenditure Survey (CES)** microdata (build cohorts and time series).
    - Card network or issuer transaction aggregates (if accessible).
- *Transformations*: Deflate by CPI; seasonal adjustment; log transformation for stability.

B. Premium card marketing intensity (the outcome)

- *Direct measures*: **Ad spend** (TV/digital/out-of-home), **GRPs/impressions**, **share of voice** by issuer (Amex, Chase, Capital One, Citi).
- *Proxies when spend is hard to get*:
    - **Google Trends** indices for brand+card names (e.g., “Amex Platinum”, “Chase Sapphire Reserve”).
    - Count of **campaign launches/press releases/creative refreshes** per month.
    - **Paid social ad volume** from public ad libraries (Meta, YouTube) aggregated to time series.

C. Control variables (to avoid spurious correlation)

- Macro: CPI (inflation), unemployment, Fed funds rate, consumer confidence, air travel volumes (TSA throughput), hotel RevPAR.
- Card-specific: lounge access expansions, annual fee changes, interchange policy shocks, merchant fee news.
- Seasonality: holiday/month fixed effects, travel season dummies.

2) Analytical Approaches (from simplest to stronger inference)

(i) Pearson/Spearman Correlation (baseline)

- **Ask**: Do affluent spending and marketing intensity rise/fall together?
- **Do**: Compute correlations at 0-lag and with **lead/lag variants** (marketing lagged on spending and vice versa).

(ii) Cross-Correlation Function (CCF)

- **Ask**: Does *affluent spending* tend to move **before** *marketing*, or the other way around?
- **Do**: Compute CCF across ±12 months of lags to detect lead-lag structure.
- **Interpretation**: Peaks at positive lags (spending leads marketing) support your hypothesis; peaks at negative lags suggest marketing might be *driving* spending or that campaigns are anticipatory.

(iii) Granger Causality (time-series causality test)

- **Ask**: Do past values of affluent spending **help predict** marketing intensity beyond marketing's own history?
- **Do**: VAR/Granger tests with appropriate lag length, stationarity checks (ADF tests), and differencing if needed.
- **Caveat**: Granger causality indicates predictive direction, not structural causality.

(iv) Panel Fixed-Effects Regression (issuer × time)

- **Ask**: Across issuers, does higher affluent spending coincide with **higher marketing intensity**, controlling for issuer-level unobservables?
- **Do**: Monthly panel (Amex, Chase, Capital One, Citi) with issuer fixed effects, time fixed effects, macro controls.
- **Spec**:\ [\text{Marketing}*{i,t} = \alpha_i + \gamma_t + \beta \cdot \text{AffluentSpend}*t + \mathbf{\delta}'\mathbf{X}*{t} + \varepsilon*{i,t}]

(v) Event Study / Difference-in-Differences (DiD)

- **Ask**: Around identifiable **revamp dates** (product refresh/fee changes), do issuers that refresh see **disproportionate increases** in marketing relative to those that don't, **conditional on affluent spending**?
- **Do**:
  - Treat refresh events as "shocks."
  - Compare treated vs. control issuers before/after.
  - Interact with affluent spending to test whether revamps are **more intense when affluent spending is high**.

(vi) Instrumental Variables (IV) / 2SLS (for reverse causality)

- **Problem**: Marketing might also affect affluent spending (or both respond to a third factor).

- **Instrument idea** (must be exogenous to affluent spending but shifts marketing costs):
  - **Ad price shocks** (CPMs) by channel due to inventory constraints (sports events, elections).
  - **Regulatory or platform policy changes** that alter ad delivery or targeting rules.
- **Goal**: Isolate variation in marketing intensity unrelated to demand, then test the **affluent spending → marketing** channel.

3) Practical Steps & Diagnostics

1. **Stationarity & transformations**: Run ADF tests; difference or log-difference non-stationary series.
2. **Lag selection**: Use information criteria (AIC/BIC) to set Granger/VAR lags.
3. **Seasonality**: Include month dummies; consider STL/Seasonal-ARIMA to remove seasonality.
4. **Robustness**:
   a. Replace affluent spending with **alternative proxies** (e.g., top 5%, luxury category spend).
   b. Swap marketing measure (spend vs. impressions vs. SOV).
   c. **Rolling windows** (e.g., 24-month) to see if relationships strengthen during certain cycles (holidays, travel surges).
5. **Directionality**: Prefer CCF and Granger first; then **DiD** around refresh events; consider **IV** if you have credible instruments.

4) Example Code Templates (drop-in with your data)

Replace affluent_spend.csv and marketing_intensity.csv with your files (monthly, same date index).

```
# === Baseline correlation & rolling windows ===
import pandas as pd

aff = pd.readcsv('affluentspend.csv', parsedates=['date']).setindex('date')
mkt = pd.readcsv('marketingintensity.csv', parsedates=['date']).setindex('date')

# Align and transform (log to stabilize variance)
df = pd.DataFrame({
    'logaffluent': (aff['value']).apply(lambda x: pd.NA if x<=0 else
x).dropna().apply(pd.Series).squeeze().apply(lambda x: pd.np.log(x)),
```

```
    'logmarketing': (mkt['value']).apply(lambda x: pd.NA if x<=0 else
x).dropna().apply(pd.Series).squeeze().apply(lambda x: pd.np.log(x))
}).dropna()

# Simple Pearson
pearson = df['logaffluent'].corr(df['logmarketing'])
print("Pearson correlation:", pearson)

# Rolling 12-month correlation
rollingcorr = df['logaffluent'].rolling(12).corr(df['logmarketing'])
print(rollingcorr.tail())
# === Cross-correlation function (which leads?) ===
import numpy as np

def crosscorr(x, y, maxlag=12):
    x = (x - x.mean()) / x.std()
    y = (y - y.mean()) / y.std()
    corrs = {}
    for lag in range(-maxlag, maxlag+1):
        if lag < 0:
            corrs[lag] = np.corrcoef(x[-lag:], y[:len(y)+lag])[0,1]
        else:
            corrs[lag] = np.corrcoef(x[:len(x)-lag], y[lag:])[0,1]
    return pd.Series(corrs)

ccf = crosscorr(df['logaffluent'], df['logmarketing'], maxlag=12)
print(ccf)
# Interpret: positive lag peak => affluent spending leads marketing by that many months.
# === Granger causality ===
import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests

# Build a 2-column DataFrame: [marketing, affluent]
gcdf = pd.concat([df['logmarketing'], df['logaffluent']], axis=1).dropna()
# Test whether affluent 'Granger-causes' marketing
res = grangercausalitytests(gcdf, maxlag=12, verbose=True)
# === Panel fixed effects (issuer x time), if you have per-issuer marketing ===
```

```
# Columns: date, issuer, marketingvalue; plus an external series affluentvalue by date
panel = pd.readcsv('issuermarketingpanel.csv', parsedates=['date'])
panel = panel.merge(aff.resetindex().rename(columns={'value':'affluentvalue'}), on='date', how='left')

# Fixed effects via dummy variables (simple approach)
panel['month'] = panel['date'].dt.toperiod('M').astype(str)
X = pd.getdummies(panel[['issuer','month']], dropfirst=True)
X['affluentvalue'] = panel['affluentvalue']
y = panel['marketingvalue']

import statsmodels.api as sm
model = sm.OLS(y, sm.addconstant(X)).fit(covtype='HC3')  # robust SEs
print(model.summary())
# === Event study / DiD ===
# Suppose you have a 'revampdate' per issuer and 'treated' dummy.
# Create 'post' dummy = 1 after revampdate for treated issuer.
panel['post'] = (panel['date'] >= panel['revampdate']) & (panel['treated']==1)
panel['did'] = panel['treated'] panel['post'].astype(int)

# Interaction with affluent spending (test amplification)
X = pd.getdummies(panel[['issuer','month']], dropfirst=True)
X['treated'] = panel['treated'].astype(int)
X['post'] = panel['post'].astype(int)
X['did'] = panel['did'].astype(int)
X['affluent'] = panel['affluentvalue']
X['didaffluent'] = panel['did'] panel['affluent']

y = panel['marketingvalue']
model = sm.OLS(y, sm.addconstant(X)).fit(covtype='HC3')
print(model.summary())
# Key coefficient: did_affluent (revamp effect that scales with affluent spending)
```

5) Interpreting Results

- **Strong positive CCF peaks at positive lags** (e.g., affluent spending leading marketing by 3–6 months) support your theory that issuers respond to high-end demand.

- **Granger causality significance** from affluent → marketing strengthens the case for predictive direction.
- **DiD interaction** (did_affluent) positive and significant implies **revamps are more aggressive when affluent spending is high**.
- **IV estimates** mitigate reverse causality; if the IV-driven effect remains positive, your causal story is more credible.

6) Common Pitfalls to Avoid

- **Seasonality** (holidays/travel): always include month fixed effects or seasonal adjustment.
- **Simultaneity**: marketing can also stimulate spend; consider **IV** or **structural VAR**.
- **Measurement error**: impressions ≠ spend; mix multiple marketing proxies.
- **Overfitting**: keep models parsimonious; validate on hold-out periods.

If you can share (even partially) the datasets you have—affluent spend by cohort and any marketing intensity series—I can **run the CCF and Granger tests for you** and visualize the lead-lag structure.

Would you prefer we start with **Google Trends (as a proxy)** plus an **income-cohort spending series** to get an initial read, then refine with ad-spend data once we secure it?