In Class Exercise 06

Priyanka Nayyar, Shruthi Khurana

2025-11-04

EXECUTIVE SUMMARY

Instagram Complaint Engagement In a study of 1,000 real-world Instagram complaint posts, we used two-stage least squares (2SLS) to isolate the causal impact of post format and self-presentation on engagement, addressing endogeneity arising from unobserved user traits like personality type, storytelling ability, and emotional investment that influence both content choices and audience response.

Leveraging pre-existing behavioral habits—proportion of past photography-related posts as an instrument for video use, and proportion of past self-posts for featuring oneself—we satisfied both IV conditions: strong relevance (F-statistics of 155 and 358, $p < 2\times10\text{-}16$). The Wu-Hausman test confirmed endogeneity ($p < 1e\text{-}6$), validating the need for IV over OLS. Results show featuring the victim drives broad engagement, adding 33.5 likes and 5.7 comments ($p < 0.001$), while videos reduce likes by 7.3 ($p = 0.005$) but boost comments by 8.6 ($p < 0.001$)—a trade-off of virality for depth. Instagram should encourage authentic, victim-centered video complaints to maximize issue visibility and discussion, even if it means fewer likes, as deeper interaction better signals urgency to brands and accelerates resolution.

1. Why are the two X variables endogenous in a model for engagement? The two X variables we have here are Video & FeatureSelf are endogenous because the same unseen qualities (like charisma or confidence) that make someone pick those formats also make their posts more engaging — so we can't tell what's cause and what's just personality.

In our observation, Endogeneity is strongly present ($p < 1e\text{-}6$ for both likes and comments), meaning a standard regression would give us biased, misleading results. That's why we need instrumental variables.

2. Why are the two proposed instruments good? The two proposed instruments are:

1. Photography — the proportion of a user's past posts with a camera emoji or photography hashtag (instrument for video choice)
2. SelfPct — the proportion of a user's past posts that feature themselves (instrument for self-feature choice)

Our instruments — the share of a user's past posts with photography hashtags or camera emojis, and the share featuring themselves — work because they're rooted in habit, not the heat of the moment. Someone who regularly posts about photography is far more likely to feel comfortable shooting and uploading a video complaint; someone who routinely posts selfies will naturally do it again when upset. These are pre-existing behaviors, formed long before the current grievance, making them powerful predictors of format choice. And crucially, they don't directly influence how many likes or comments the complaint gets — only through the channel of choosing video or self-inclusion. The F-statistics back this up: 155 for video, 358 for self-feature. With exactly two instruments for two endogenous variables and strong theoretical exclusion, these satisfy the textbook IV conditions and give us clean, causal estimates.

3. Which choices (featuring one's self and posting a video instead of a picture) affect engagement? Featuring yourself is a winner across the board: it causally adds 33.5 likes and 5.7 comments (both $p < 0.001$),

showing that authenticity and emotional connection resonate powerfully with audiences. Video, however, pans out differently: it actually reduces likes by about 7.3 ($p = 0.005$ and significant) — likely because video thumbnails are less scroll-friendly — but increases comments by 8.6 ($p < 0.001$), suggesting videos spark conversation, debate, or support. In short: self-features drive broader engagement; videos trade likes for depth. Follower count remains the biggest lever (100+ likes per 1,000 followers), but for complaint posts, victim-centered videos are the smart play — they may not go viral in likes, but they generate the kind of discussion that gets corporate attention. Instagram should encourage this format to surface real issues faster.

```
load("06_Social_IV.Rdata")

# Loading the required packages
library(ivreg)  # for ivreg()
library(AER)  # for diagnostic tests (optional)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
## Registered S3 methods overwritten by 'AER':
##   method              from
##   print.ivreg         ivreg
##   print.summary.ivreg ivreg
##   summary.ivreg       ivreg
##   vcov.ivreg          ivreg
##   bread.ivreg         ivreg
##   estfun.ivreg        ivreg
##   hatvalues.ivreg     ivreg
##   predict.ivreg       ivreg
##   anova.ivreg         ivreg
##   terms.ivreg         ivreg
##   model.matrix.ivreg  ivreg
##   update.ivreg        ivreg
```

```
##
## Attaching package: 'AER'
```

```
## The following objects are masked from 'package:ivreg':
##
##     ivreg, ivreg.fit
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# MODEL 1: LIKES (2SLS IV Regression)
iv_likes <- ivreg(IGLikes ~ LengthChar + Followers + Video +
    FeatureSelf | LengthChar + Followers + Photography + SelfPct,
    data = IGdata  # Use IGdata, not complaints
)
summary(iv_likes, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = IGLikes ~ LengthChar + Followers + Video + FeatureSelf |
##     LengthChar + Followers + Photography + SelfPct, data = IGdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -56.9920 -12.0585   0.2401  11.3731  54.5992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.578877   2.860386   4.398 1.21e-05 ***
## LengthChar   0.032827   0.022190   1.479   0.1394
## Followers    0.103785   0.002966  34.993  < 2e-16 ***
## Video       -7.262961   2.587628  -2.807   0.0051 **
## FeatureSelf 33.450827   1.905693  17.553  < 2e-16 ***
##
## Diagnostic tests:
##                                 df1 df2 statistic  p-value
## Weak instruments (Video)          2 995    155.09  < 2e-16
## Weak instruments (FeatureSelf)    2 995    358.19  < 2e-16
## Wu-Hausman                        2 993     15.54 2.26e-07
## Sargan                            0  NA        NA       NA
##
```

```
## Weak instruments (Video)       ***
## Weak instruments (FeatureSelf) ***
## Wu-Hausman                     ***
## Sargan
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.14 on 995 degrees of freedom
## Multiple R-Squared: 0.6629,  Adjusted R-squared: 0.6615
## Wald test: 404.5 on 4 and 995 DF,  p-value: < 2.2e-16
```

```r
# MODEL 2: COMMENTS (2SLS IV Regression)
iv_comments <- ivreg(IGComments ~ LengthChar + Followers + Video +
    FeatureSelf | LengthChar + Followers + Photography + SelfPct,
    data = IGdata  # Use IGdata here too
)
summary(iv_comments, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = IGComments ~ LengthChar + Followers + Video +
##     FeatureSelf | LengthChar + Followers + Photography + SelfPct,
##     data = IGdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -31.76605  -8.44516  -0.01403   7.80890  30.34749
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.595004   1.809603   0.881  0.37831
## LengthChar  0.040637   0.014038   2.895  0.00388 **
## Followers   0.033371   0.001876  17.785  < 2e-16 ***
## Video       8.621388   1.637045   5.266 1.70e-07 ***
## FeatureSelf 5.742178   1.205623   4.763 2.19e-06 ***
##
## Diagnostic tests:
##                                 df1 df2 statistic  p-value
## Weak instruments (Video)          2 995    155.09  < 2e-16
## Weak instruments (FeatureSelf)    2 995    358.19  < 2e-16
## Wu-Hausman                        2 993     24.98 2.61e-11
## Sargan                            0  NA        NA       NA
##
## Weak instruments (Video)       ***
## Weak instruments (FeatureSelf) ***
## Wu-Hausman                     ***
## Sargan
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.48 on 995 degrees of freedom
## Multiple R-Squared: 0.4334,  Adjusted R-squared: 0.4311
```

```
## Wald test: 93.64 on 4 and 995 DF,  p-value: < 2.2e-16
```

Summary - For Likes, featuring oneself in a post has a strong positive impact, while posting a video slightly decreases likes after correcting for endogeneity. n contrast, for Comments, both videos and self-featuring posts significantly increase engagement, indicating that dynamic or personal content stimulates more active audience interaction and conversation. Across both models, follower count consistently drives higher engagement, while caption length exerts a modest but positive effect. Diagnostic tests confirm that endogeneity is present and that the chosen instruments are valid and strong, reinforcing the reliability of these insights. Overall, these results suggest that creators seeking visibility through likes may benefit from well-composed image posts, whereas those aiming for deeper engagement and dialogue should focus on video and self-featured content.