Finance PhD Student Tutorial: Producing Tables Workflow

Kyle Zimmerschied[1]

[1]University of Arkansas

April 29, 2026

Today's Agenda

Focus: How do I create clean, professional tables?

1. **Opening Discussion and Reflection (10 min):** What does my current workflow look like to produce tables?
2. **Overview (20 min):** What does the flow of tables in a high-quality paper normally look like?
3. **Practical (20 min):** How do I make tables in Stata?

How Do You Currently Build Tables?

Reflect on Your Experience

1. What does my current workflow look like to produce tables?
   - What programming language(s) do I use to produce these and how are these integrated to my paper?
   - How challenging is it for me to update tables in my paper?

Discussion time: 10 minutes

How Strong Papers Usually Organize Tables

- In the research process, brainstorming, producing, and placing tables should precede writing
- Papers typically follow similar flows of results and writing
  - It can be helpful to read adjacent papers to mirror its flow of results and writing
  - It's not ideal to surprise your reader
- Increasingly, many papers are reducing the number of tables to instead visually show results in figures
  - We'll discuss figure construction next week

Typical Order of Tables in a Paper

- Most papers follow a fairly predictable table flow:

1. **Summary Tables:** Describe the sample, key variables, and comparison of interest
2. **Context/Motivation Tables:** Motivate the identification strategy or empirical design
3. **Main Results Tables:** Present the core empirical finding of the paper
4. **Robustness Tables:** Show the result is not driven by alternative specifications or explanations
5. **Heterogeneity Tables:** Show how the effect varies across groups or settings
6. **Mechanism Tables:** Provide evidence for why the effect occurs
7. **Real Effects/Implications Tables:** Show why the main result matters economically

Example of Table Flows

1. Causal, corporate finance example: Bernstein (2015)
2. Data, innovative question example: Falato *et al.* (2022)
3. Data, methodological innovation: Anarkulova *et al.* (2022)

- Each paper has a different focus and style depending on where the innovation of the paper lies
- Your paper should be structured likewise:
  1. Emphasizing results that are focal to developing your main argument
  2. Placing tables in accordance with their value and flow of writing
- Table placement and writing flow should align well

Regression Tables: Clear vs. Unclear

Good Table

	(1) Base	(2) +Ctrls	(3) +FE
Immigration Inflow	-0.021***	-0.018***	-0.015**
	(0.005)	(0.006)	(0.007)
Income Growth		0.011*	0.009
Population		-0.003	-0.002
County FE	No	No	Yes
Obs.	3142	3142	3142
R^2	0.12	0.18	0.29

Why it works

- Clear variables
- Columns build logically
- Main coefficient visible

Bad Table

	(1)	(2)	(3)	(4)
x1	-0.021***	-0.012	-0.019	-0.014
x2	0.031	0.014	0.022	0.017
var_3	-0.002	-0.001	0.003	0.001
pop_lag4	0.014	0.010	0.013	0.012
zscore	0.006	0.005	0.002	0.003
ctrl_set	Yes	Yes	Yes	Yes
Obs	4000	3500	3200	3142

Problems

- Unclear variables
- Controls unexplained
- No logical progression

Key Lesson

A good table should answer in 10 seconds: *What is the main variable? What controls are included? What changes across columns?*

- The use of headers and multiple panels can also be beneficial

A Clean Workflow for Producing Tables

- It's beneficial to have a clean, professional system for building tables
  - Default choice for many scholars is Stata but alternative frameworks exist in R, Python, SAS
- Many scholars use LaTex to streamline table production and placement
  - Outputting to Word and making edits can be quite time consuming
- Some fixed costs to learning LaTex, but it's a worthwhile investment for the long-run time it will save you
  - Online LaTex table generator can help build non-traditional tables
- We'll do a live demo in 2 weeks with LaTex to input tables and figures we've built

Building Tables (Stata + LaTex)

1. Run estimation in Stata
2. Store results with eststo
3. Export table fragment with esttab
4. Insert into LaTex

- This process enables clean, reproducible tables which be easily updated and shared

Creating a Summary Statistics Table in Stata

```
global xlist1 final_mat_yield_ta_rf_w years_to_maturity ///
    amount_final_mat_mil composite_amount_usd_millions ///
    callable_issue insured_flag negotiated_bid_flag ///
    revenue_type_flag tax_exempt_flag ratings_comb ///
    sinking_fund_flag refinancing_refunding_flag
{
    qui estpost su $xlist1 if longest_bond == 1 & in_sample == 1, detail
    est store A
    esttab A using "$main_output_path/TAB_SUMMARY_BOND.tex", replace ///
        b(a2) collabels(\multicolumn{1}{c}{{N}} ///
        \multicolumn{1}{c}{{Mean}} ///
        \multicolumn{1}{c}{{SD}} ///
        \multicolumn{1}{c}{{p1}} ///
        \multicolumn{1}{c}{{p25}} ///
        \multicolumn{1}{c}{{Median}} ///
        \multicolumn{1}{c}{{p75}} ///
        \multicolumn{1}{c}{{p99}}) ///
        cells("count mean(fmt(%10.2fc)) sd(fmt(%10.2fc)) ///
        p1(fmt(%10.2fc)) p25(fmt(%10.2fc)) ///
        p50(fmt(%10.2fc)) p75(fmt(%10.2fc)) ///
        p99(fmt(%10.2fc))") ///
        label booktabs nonumber noobs fragment ///
        nonotes nomtitles nolines posthead(\midrule)
}
```

Stata Example: Build Regression Table

```
* (A) OLS Baseline
reghdfe final_mat_yield_ta_rf_w $endog_x ///
    if in_sample == 1, absorb($fe_state_year) cluster(state_fips)
eststo A

* (B) + Bond Controls
reghdfe final_mat_yield_ta_rf_w $endog_x $bond_controls ///
    if in_sample == 1, absorb($fe_state_year ratings_comb) cluster(state_fips)
eststo B

* (C) + County Controls
reghdfe final_mat_yield_ta_rf_w $endog_x $county_controls ///
    if in_sample == 1, absorb($fe_state_year) cluster(state_fips)
eststo C

* Export to LaTex
esttab A B C using "$main_output_path/IV_YIELD.tex", booktabs fragment brackets replace ///
    se(3) b(3) label nobaselevels order() drop($bond_controls) star(* 0.1 ** 0.05 *** 0.01) ///
    nonotes nomtitles nolines posthead(\midrule) stats(N r2_a state_fe year county_fe ///
    bond_controls county_controls ymean,  fmt(0 2 0 0 0 0 0 2) ///
    labels("Observations" "Adjusted \( R^2 \)" "State F.E." "Year F.E." "County F.E." ///
    "Bond Controls" "County Controls" "Y-mean")) mgroups("Yield Spread", pattern(1 0 0) ///
            prefix(\multicolumn{@span}{c}{) suffix(}) ///
            span erepeat(\cmidrule(lr){@span}))
```

Before Next Week (1 Hour)

- Reflect (30 minutes):
  1. How can my current workflow be improved?
  2. Does my placement and focus on tables typically align well with my writing?
- Explore (30 minutes):
  1. Find one table with an excellent table flow in your area of interest
  2. R for Data Science Chapter 1
  3. R Graph Gallery
  4. Data Visualization in Stata

References

Anarkulova, Aizhan, Cederburg, Scott, & O'Doherty, Michael S. 2022. Stocks for the long run? Evidence from a broad sample of developed markets. *Journal of Financial Economics*, **143**(1), 409–433.

Bernstein, Shai. 2015. Does going public affect innovation? *The Journal of finance*, **70**(4), 1365–1403.

Falato, Antonio, Kim, Hyunseob, & von Wachter, Till M. 2022. *Shareholder power and the decline of labor*. Tech. rept. National Bureau of Economic Research.