Term Paper Checklist … just a start

Replication of the one reported regression of interest:

- n obs right? correct LHS and RHS vars?
- Do your summary stats match theirs? (Do you have the right dataset?)
- R-squared, adj R-squared, MSE, coeffs and t-stats right?
- You should be close, if not right on … but close is good enough (recall NSRate analysis)

Where are the bodies buried?

- Your going-in assumption should be that the analysis is rife with errors and skullduggery… and you will almost always be wrong! (Remember: someone with an agenda ran hundreds of regressions before they settled on the ones they are presenting.)
- Is this a favorite coefficient model? If it is, then you have a head start in the investigation.
- Low R-squared usually means there's lots of room for improvement… but not always!
- Are there any suspicious RHS coeffs? (variables that have no right to be in the model; highly statistically significant with the *wrong* sign; peculiar RHS variable constructions (e.g. $X1=\sin(z^2.379)$; X2 lagged 12 days; X3 is an avg over the last 7 wks; X4 on full moons)?
- What obvious RHS vars didn't make it into the final model?
- Look at the discussion: Do they discuss heteroskedasticity or multicollinearity? What did they do about these? Do they run weighted regressions… and do the weights make sense?
- And if it's a time series model, did they worry about things like serial correlation?
- Look at all reported regression results, not just the one of interest. How do things move around with the different variations? Which obvious regression results were not reported?
- Look for the footnotes saying things like "we dropped the ABC observations because of course… ." or "it's obvious that the right way to handle this is to …" … Of course not! … obvious not!... and beware of any claims of "sophisticated analysis"

Improvements to consider

- You know about omitted variable bias… find those omitted variables and see how they impact the model.
- Grab more data: other countries, states, years, etc.
- Grab similar RHS variables from other sources: e.g. try the World Bank's X metric instead of the IMF's; grab Freedom House's corruption index instead of transparency.org's
- Run VIF to see how much multicollinearity you have? And if it's there, do something about it (drop vars to assess impacts; get more data) ... or maybe not!
- If there are obvious F/Chow tests to run, do those as well and see what you conclude. Relatedly: Run the regression on different subsets of data to see how robust the model is.
- If data is over time, make sure that dollar denominated vars are in real $s
- The heteroskedasticity correction is easy (, robust) so why not do it?
- Functional forms: RHS vars: try X^2 , X^3, $\ln(X)$ etc ; interactions… for LHS, try $\ln(Y)$ (but be careful as it's a challenge to compare models with different dependent variables)

- And dummies, dummies, … dummies: try out lots of intercept and slope dummies to see what they missed… especially useful if it's a favorite coefficient model. Remember that dummies just capture what the rest of the model missed; so figure out what you missed and build a better model.

Your writeup

- A paragraph or two describing the paper and regression of interest. (What's the question being addressed? How do they address that question (what data and techniques)? What do they conclude?)
- Snapshots/copies of their regression results and your replication.
  - You might separately do the same for summary stats for their and your datasets.
- A paragraph or two comparing the two regression results (and summary stats?)… focusing both on what matched and what didn't, and explaining why you weren't able to match exactly.
- Your improvements (at least several pages, and often more) … yes that should be plural. (These should be your improvements, not someone else's!)
  - Use scatterplots, eststo and esttab to tell the story… yes, econometrics is storytelling!
  - What did you consider? What worked? What didn't work? What did you learn?
  - If you had more time and money, what would you do next?