TO: Mark Hand

FROM: Bill Spelman

ABOUT: Improving efficiency (and a few other things)

DATE: 5 March 2019

(**Emphasis** and*comments*Mark’s)

Your paper is interesting. It’s on a great subject with potentially important policy implications. Oddly, I’d find myself siding with Clarence Thomas if I found your results convincing. If campaign limits don’t reduce corruption or the public appearance thereof, there’s really no justification for them. Let $10,000 contributions bloom.

The big problem (as you quite properly mention) is that absence of evidence isn’t the same thing as evidence of absence. It’s hard to prove a null finding, particularly when the data and model are as messy as these are. On the other hand, a nonscientist (like, for example, Justice Thomas) might interpret your *caveats* as mere *pro forma* caution or, worse, as an attempt to reduce the damage on the part of a namby-pamby liberal partisan. Thus he might interpret your paper as the very evidence of absence you’re pretty sure you’re not providing.

I can think of two partial solutions to this (basically political) problem. **First, you could reframe the analysis from significance-seeking to fact-finding**. There *is* a relationship between campaign contribution limits and corruption.[[1]](#footnote-1) You’re currently using significance tests in an attempt to prove that it’s negative: *t* < -1.96 means “more limits, less corruption.” Your *t* values are nowhere near that low, so you can’t make that case. But what *can* you say? Suppose you **shift to confidence intervals**. In this case, the issue changes to “How big an effect, positive or negative, could contribution limits have on corruption?” Now that currently insignificant coefficient on the individual limits dummy in Table 4 starts to look pretty large. Individual limits might increase the CRI by (-.56 + 1.96 × .55 = +.518) × .19 = +.098, but it might *decrease* the CRI by (-.56 – 1.96 × .55 = -1.638) × .19 = -.311. That second effect is a little nonsensical, since mean CRI = .28. But it’s conceivable given your regression results that adopting individual limits reduces corruption references in the news media (or at least the AP) to nearly zero. No, you can’t prove it, but yes, it’s no less likely than the positive confidence limit.

*-.56 (the mean)*

*+ 1.96 × .55 (the z-score for a 95% confidence interval times the standard error)*

*= 0.518) × .19 (what is this 0.19? maybe standard deviation of CRI?)*

*= +.098*

More generally, a focus on confidence intervals shifts the frame from what you can prove to what’s consistent with your data. It also demonstrates a lot more clearly what you can’t expect to show, and why: Your standard errors are too big for this to be a very powerful test of anything. No reflection on you. It’s the data that suck.

This leads to the second solution. Your analysis isn’t wrong, but it *is* inefficient. That is, your confidence intervals are wide because your standard errors are large. If there’s anything you can do to reduce the size of those standard errors, your confidence intervals may be small enough to clarify the results.

There’s a bunch of things you can do to improve efficiency. (I had exactly this problem in a paper I finished last night, so I got to be rather an expert on the subject.)

1. All of your key independent variables in Table 1 have standard deviations larger than the means. That means they’ve got long right-hand tails – a few cases are much, much larger than the average. Under pretty general conditions, taking logs will make the distribution a lot more symmetrical. Your coefficients won’t have to stretch so far to accommodate the largest values, and your standard errors will go down.

*Done*

1. To a lesser extent, the same is true of your dependent variables (Table 2). (How did the max for a proportion get to be 2.16, anyway? But I digress.) Taking logs of the CCI, and either logs or (if that max value is a coding error or misprint) logits of the CRI would likewise prevent your coefficients from having to stretch so far, and your standard errors would probably descend from the stratosphere.

This means you should probably take logs of all the control variables, too, but I’m not so worried about them.

*Done – and correct!*

1. You (quite properly) tested fixed state and year effects, but I’m not certain you need them. The simplest approach is to **conduct and report an *F* test on each**. I suspect from the findings on Tables 3 and 4 that the *F* value is significant on year, but not on state. That shouldn’t be too surprising, given that most of the control variables (e.g., population, real income, urbanization, citizen ideology) are likely to differ among states but not so much over time. If state isn’t close to significant, that’s another 50 degrees of freedom you can use to reduce your standard errors, and a substantial increase in adjusted *R*2.

*Done – and correct!*

1. A state panel makes sense, but all states are not equal. You’ve got a lot more information about corruption in California (with 40 million population and some really good newspapers) than in Wyoming (with 600,000 people, including tourists, and a lonely AP outpost in Cheyenne). If nothing else, you can expect a lot more random sampling error in Wyoming than in California. The simplest solution is to **weight your cases by the square root of population**. If random error is an important reason for your large standard errors (I bet it is), then weighting the cases should reduce them considerably. The Stata code is

gen rpop = sqrt(population)

reg Y X [aw = rpop]

*Population is already logged. ¿?*

1. Your first differences model, though a good idea in theory, may be overdoing it. I suspect there’s a high positive correlation between *Yt* and *Yt*-1 for both CCI and CRI. All by itself, positive serial correlation causes you to understate your standard errors. If the correlation is high enough (ρ > +.8 or so), then first differences are justified and your standard errors are about right. But if the correlation is relatively low (say, ρ = .2 or .3), then taking first differences will induce a high *negative* serial correlation in the differenced dependent variable. (You’re overcompensating by nuking a fairly minor problem.) This means your standard errors will be systematically too *high*, and your results are actually a lot better than they appear to be. Given the rather absurd standard errors you report on Table 5, I suspect this may have been the case.

There are a lot of technical fixes for this, but if you rely on a simple procedure you won’t go far wrong:

1. Check and report the serial correlation in your dependent variable. Easiest way is to run the regression *Yt* = α + ρ *Yt*-1.
2. If ρ > .8, difference your data and stop. Your standard errors were right all along.
3. If ρ < .3, don’t difference your data. You get a lot more information from levels than you do from differences, and it’s unrealistic to expect a sensible result in differences.
4. If .3 < ρ < .8, you can difference the data if you like, but use Arellano (“cluster robust”) standard errors. In Stata, the necessary code is

reg Y X,vce(cluster(state)).

You can look up the details, but basically this will clean your standard errors of both serial correlation and (bonus!) heteroscedasticity. Probably, they’ll be smaller, maybe by a lot.

If you want to get a little radical, you can also try feasible generalized least squares (FGLS). This is a “partial difference” procedure. In Stata, the code is

prais Y X, robust.

A lot of stuff can go wrong with FGLS and you should test a few other things first to be sure it’s valid for your data, but for some data sets it’s the most efficient way to deal with serial correlation. See me if you want to know more.

While I’m on a techie rant, allow me to point out one more thing. The scale of your variables is all over the place. This is why you’ve got significant findings for the citizen ideology measure (but the coefficients are zero to two decimal places), while your coefficient on government wages is a few thousand times bigger but not significant at all. Coefficients like these are hard to parse. The two obvious fixes (pick one) are to use a log-log model (for which all of your coefficients are expressed as *elasticities*) or to report standardized regression coefficients (βstd = β × *Ss*/*Sy*). Political scientists tend to use standardized coefficients, but they’ll forgive your elasticities if a log-log model makes sense.

This is a good start, but if you want to see it in print I think you’ll need to address the framing and efficiency issues. Fortunately, they won’t take too long to fix.

1. If a butterfly flapped its wings in Rio, Garret Graves will kick poor people’s ass in DC. Actually he’ll do that, anyway, but it’s still true that everything is related to everything else. [↑](#footnote-ref-1)