Sensitivity of the results in "Economic Crises, Civilian Mobilization, and Repression in Developing States" to clustering at higher levels of analysis using more constricting visualization techniques

In February, 2024, Professor Dave Armstrong approached Bumba Mukherjee and myself with an interesting observation on our article, "Economic Crises, Civilian Mobilization, and Repression in Developing States" (Koren and Mukherjee, 2022): when he plotted the effect of commodity price shocks on the predicted change in (log) number of civilian deaths across the range of concentrated urban infrastructure, *adjusted for clustering at the country level*, he noticed that the 95% CIs intersect with the X-axis across the entire range of the plot. This was different than the substantive figures we presented in our study, where the 95% CIs were separate from the X-axis across almost the entire range. The primary explanation for this difference is that the R package (R was used in all our statistical analyses in this study) used for plotting the figures, "interplot," did not allow us to use objects from R packages that allow for effective clustering, including "lmtest" and "lfe." This forced us to take a different approach and rely on the "cluster()" option(from the "survival" package) to adjust the estimates for country level heterogeneities (we choose this higher level of heterogeneity despite the fact that our cross-sectional unit, which is – by far – the most oft used for clustering standard errors in studies of similar phenomena at similar levels of resolution – the 0.5 degree grid cell – as I discuss below). Even though it is sometimes used in conflict and violence research, this approach provide inferior clustering estimates compared with the two aforementioned packages, which rely on the sandwich estimator. Considering this concern, we made sure to illustrate the findings' robustness to this decision, as follows:

1. **Ensuring the results hold when clustering by country is done using the "lmtest" and "lfe" packages:** before reporting the "cluster()" models, we estimated the same specifications when clustering country level heterogeneities using the "lmtest" package for the year fixed effects only and country + year fixed effects models, and the "lfe" package for the 0.5 degree grid cell + year models. A minimalist illustration of this exercise is reported in Table 1, which shows that – even when clustering using these more constricting packages – the interaction term remains positive and statistically significant to the $p < .05$ level across all three models. Considering the results held to a constricting threshold ($p < .05$), and the fact that (at the time) the most effective way available to us to plot these interactive effects was using a package that couldn't incorporate clustering (i.e., using the "interplot" package), we made a choice to report plots that relied on models with the "cluster()" option.

2. **When clustering at the cross-sectional unit level (0.5 degree grid cell) the**

plots remain largely unchanged: Additionally, as mentioned above, in order to provide a higher threshold of significance, we made the decision to cluster our standard errors at the country rather than at our cross-sectional unit's level, the 0.5 degree grid cell. This was an unusual decision – in the majority of conflict and violence studies, clustering of standard errors is almost always done at the cross-sectional unit level (e.g., O'Loughlin et al., 2012; O'Loughlin, Linke and Witmer, 2014; King and Roberts, 2015; Esarey and Menger, 2019). The reason was that we believed that country-level heterogeneities might matter considering price shocks could induce relevant effects in different locations within the country. However, this decision also increased the number of units in each cluster significantly, in effect, from several dozens to several hundreds or thousands, greatly increasing the risk of type II error (i.e., falsely failing to reject the null hypothesis). For illustration, I plot a similar figure to the one plotted by Professor Armstrong below, this time clustering standard errors at the more standard cross-sectional unit (0.5 degree grid cell) rather than higher (country) level in Figure 1. As can be seen, in this case, the 95% CIs remain relatively tight around the estimates, and closer to the plot that does not use standard error clustering.

These two issues notwithstanding, considering this – how can clustered interaction results yield a statistically significant ($p < .05$) coefficient and yet, when plotted, do not show robustness across the moderator's range – is a generally interesting question , and in the full interest of transparency, I asked Professor Armstrong if he would agree to write a short description of the issue and present his findings. This discussion is provided below.

Table 1: Civilian Mobilization and Killings by Government Forces, 1998-2007 (Baseline Models)

	Year	**Country and year**	**GID and year**
Concentrated Urban Infrastructure$_{it-1}$	0.011	0.023*	0.004
	(0.011)	(0.013)	(0.015)
Price Shock$_{jt-1}$	−0.014	−0.014	−0.013
	(0.010)	(0.010)	(0.010)
Concentrated Urban Infrastructure$_{it-1}$ × *Price Shock*$_{jt-1}$	0.031**	0.027**	0.020**
	(0.015)	(0.013)	(0.010)
Constant	0.024	0.007	
	(0.016)	(0.013)	
FEs	Year	Country +year	Grid cell + year
Observations	77,042	77,042	77,042
R^2	0.002	0.070	0.190
Adjusted R^2	0.002	0.068	0.128

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Variable coefficients are reported with standard errors clustered by country in parentheses. Fixed effects by year and geospatial level are not reported.

[1] Natural log

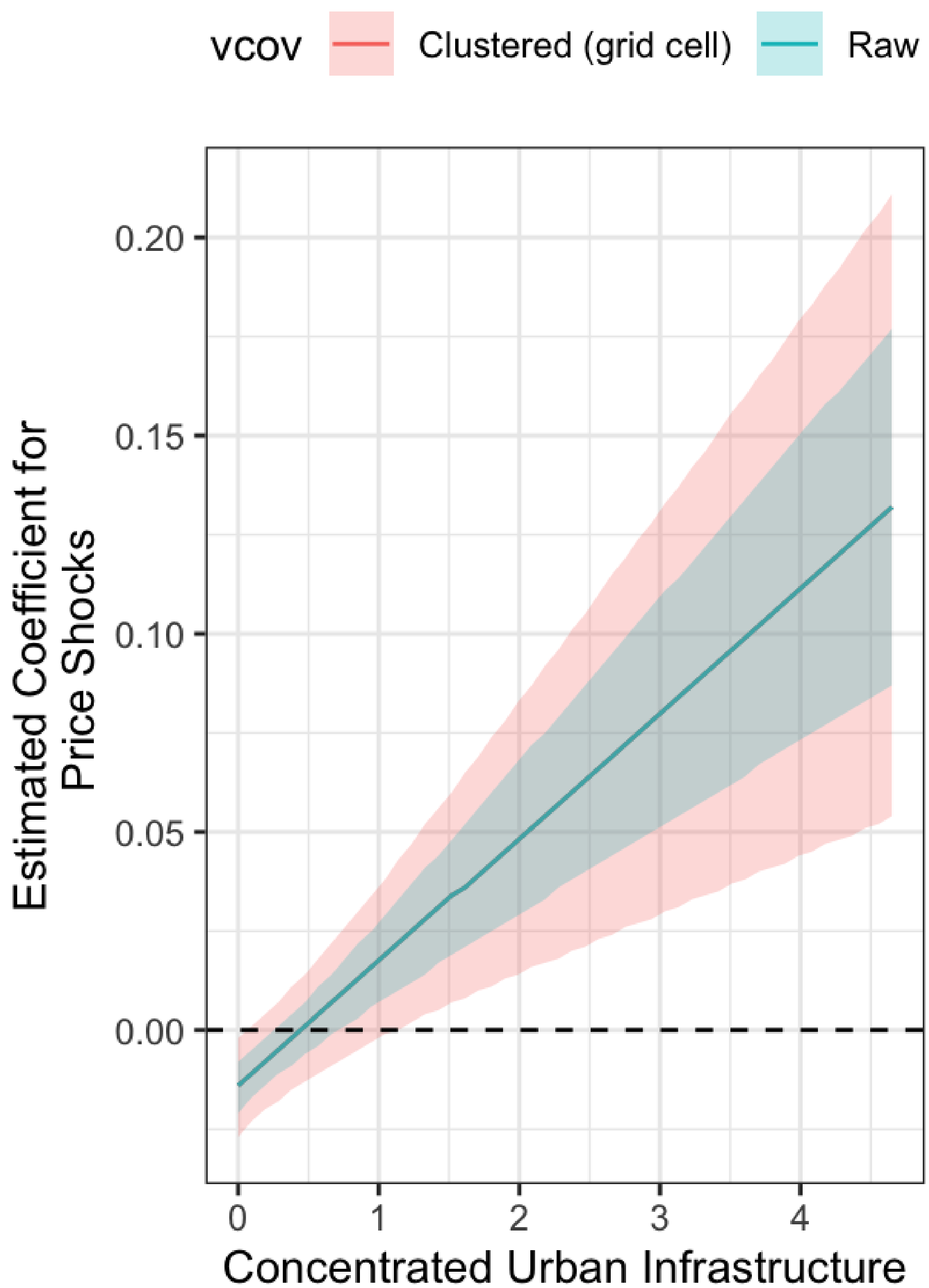


Figure 1: Estimated Effect of Price Shocks on Urban Development PC's Propensity of Regime Violence

References

Esarey, Justin and Andrew Menger. 2019. "Practical and effective approaches to dealing with clustered data." *Political Science Research and Methods* 7(3):541–559.

King, Gary and Margaret E Roberts. 2015. "How robust standard errors expose methodological problems they do not fix, and what to do about it." *Political Analysis* 23(2):159–179.

Koren, Ore and Bumba Mukherjee. 2022. "Economic crises, civilian mobilization, and repression in developing states." *Conflict Management and Peace Science* 39(5):520–541.

O'Loughlin, John, Andrew M Linke and Frank DW Witmer. 2014. "Modeling and data choices sway conclusions about climate-conflict links." *Proceedings of the National Academy of Sciences* 111(6):2054–2055.

O'Loughlin, John, Frank DW Witmer, Andrew M Linke, Arlene Laing, Andrew Gettelman and Jimy Dudhia. 2012. "Climate variability and conflict risk in East Africa, 1990–2009." *Proceedings of the National Academy of Sciences* 109(45):18344–18349.

Interaction Analysis

2024-02-14

I think we're looking at the interaction in two different ways and coming up with two different answers about its meaning. First, let's say that the model is:

$$y_{ig} = b_1 s_{ig} + b_2 u_{ig} + b_3 s_{ig} u_{ig} + \mathbf{aX} + e_{ig}$$

where s_{ig} is the shock for observation i in grid g and u_{ig} is urban concentration for observation i in group g. We know that:

$$\frac{\partial y}{\partial s} = b_1 + b_3 u$$

So, we can show that for any two values of u, say u_0 and $u_0 + \delta$, the difference between the two estimates is:

$$\begin{aligned}(b_1 + b_3 u_0) - (b_1 + b_3(u_0 + \delta)) &= 0\\ b_1 + b_3 u_0 - b_1 - b_3 u_0 - b_3\delta &= 0\\ -b_3\delta &= 0\\ b_3 &= \frac{0}{-\delta}\\ b_3 &= 0\end{aligned}$$

So, as you rightly note, if the interaction coefficient is different from zero, that means that for any value of $\delta \neq 0$ and any value of u_0, we would find a significant difference between

$$b_1 + b_3 u_0$$

and

$$b_1 + b_3(u_0 + \delta)$$

. It would seem that this corroborates your hypothesis and, in fact, even with the clustered covariance matrices, this seems to hold across models.

I was looking at this a different way. Rather than wondering whether any two coefficients are significantly different from each other, which ones are significantly different from zero. To figure this out, we would simply have to calculate the t-statistic for the conditional effect:

$$t = \frac{b_1 + b_3 u}{\sqrt{V(b_1) + u^2 V(b_2) + 2uV(b_1, b_2)}}$$

Let's estimate your model:

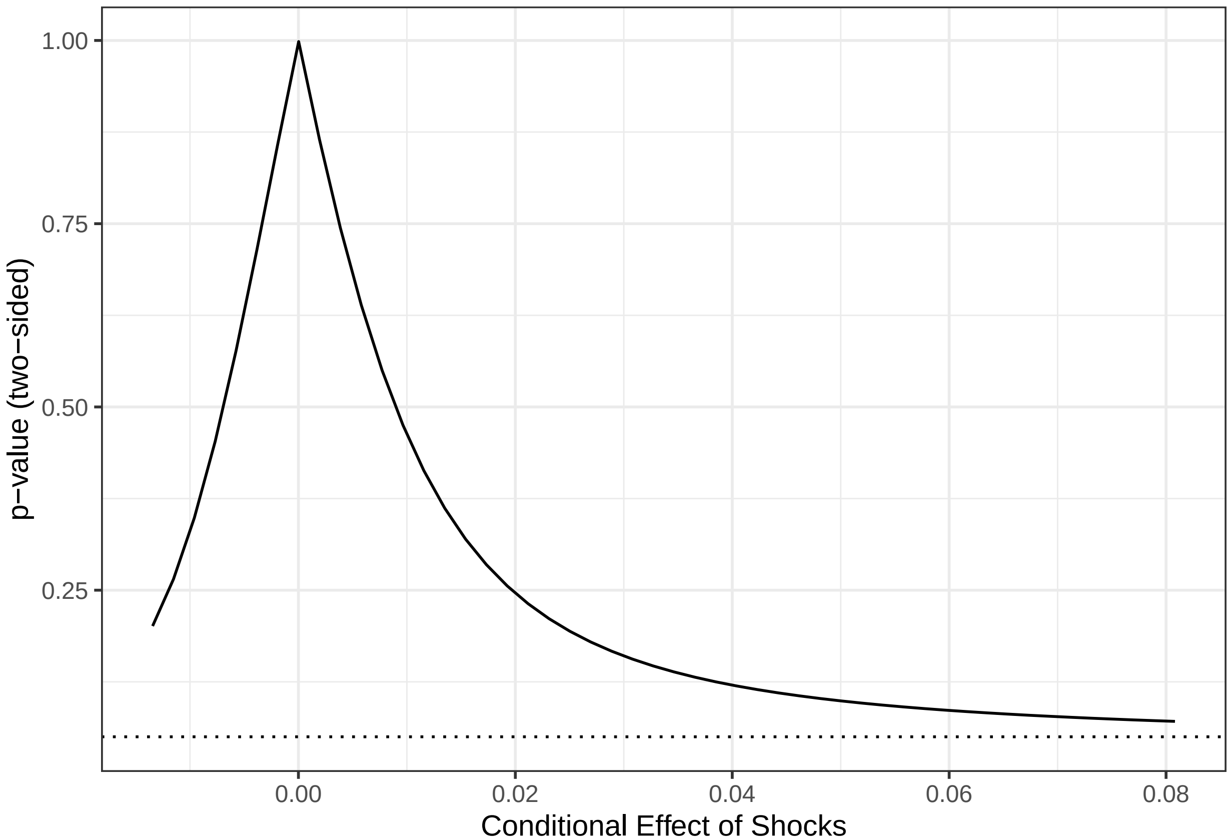
```
lm.1.g <- felm(loggov_deaths ~ lagpshock_sd*lagnlight_pc|
                 gid+year|
                 0|ccode, data=urb.dat)
summary(lm.1.g)
```

```
##
## Call:
##    felm(formula = loggov_deaths ~ lagpshock_sd * lagnlight_pc |      gid + year | 0 | ccode, data = 
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1445 -0.0023  0.0005  0.0026 10.3860
##
## Coefficients:
##                            Estimate Cluster s.e. t value Pr(>|t|)
## lagpshock_sd1             -0.013446     0.010451  -1.287   0.2011
## lagnlight_pc               0.003539     0.014537   0.243   0.8082
## lagpshock_sd1:lagnlight_pc  0.020272     0.009781   2.073   0.0407 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1728 on 71520 degrees of freedom
## Multiple R-squared(full model): 0.1903   Adjusted R-squared: 0.1278
## Multiple R-squared(proj model): 0.0004086   Adjusted R-squared: -0.07676
## F-statistic(full model, *iid*):3.045 on 5521 and 71520 DF, p-value: < 2.2e-16
## F-statistic(proj model): 1.816 on 3 and 105 DF, p-value: 0.1488
```

Now, let's calculate the t-statistic for lots of different values of u between its minimum (0) and maximum (4.65)

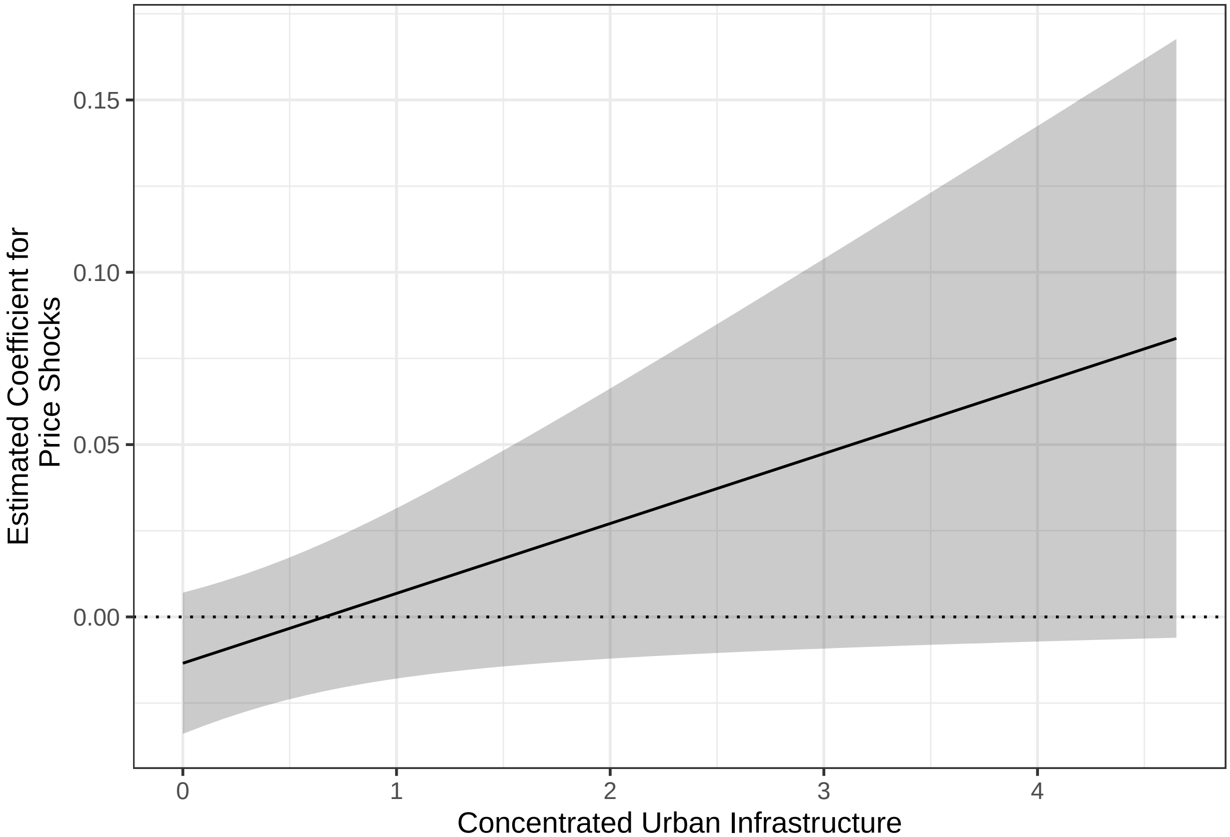
```r
u <- seq(0, 4.65, length=50)
b <- coef(lm.1.g)
V <- vcov(lm.1.g)

## numerator of t-stat
num <- b[1] + b[3]*u

## denominator of t-stat
denom <- sqrt(V[1,1] + u^2*V[3,3] + 2*u*V[1,3])

## t-statistic
t_stat <- num/denom

## two-sided p-value
pvals <- 2*pt(abs(t_stat), df = lm.1.g$df.residual, lower.tail=FALSE)
```

```r
plot1 <- tibble(
  effect = num,
  pval = pvals
)
ggplot(plot1, aes(x=effect, y=pval)) +
  geom_hline(yintercept=.05, linetype=3) +
  geom_line() +
  theme_bw() +
  labs(x="Conditional Effect of Shocks", y="p-value (two-sided)")
```



As you can see, even though the conditional effects are different from each other, none of them is significantly different from zero. This is corroborated by the effect plot, if we make it with the clustered covariance matrix.

```
library(DAMisc)
plot2 <- tibble(x = u,
                eff = num,
                lwr = eff - 1.96*denom,
                upr = eff + 1.96*denom)

ggplot(plot2, aes(x=x, y=eff, ymin = lwr, ymax=upr)) +
  geom_ribbon(alpha=.25) +
  geom_line() +
  geom_hline(yintercept=0, linetype=3) +
  theme_bw() +
  labs(x="Concentrated Urban Infrastructure", y="Estimated Coefficient for\n Price Shocks")
```



This shows that the effects are never significantly different from zero. This is consistent with an interesting article by Gelman and Stern (2006) - *The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant"*