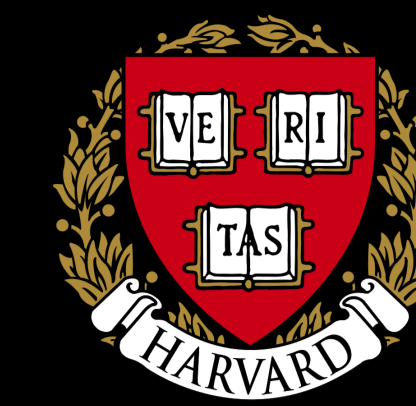


Where's Waldo: Searching for the Hidden Variable of Corruption

Jennifer Pan[†] Chiara Superti[‡]

Department of Government, Harvard University, Cambridge, MA



SUMMARY

We provide a methodological approach to uncover the “true” value of corruption. By using substantive knowledge about the data generation process (DGP) of corruption measures and their biases, we identify “filter variables” that mitigate non-classical measurement error through an instrumental variables approach.

MOTIVATION

Corruption is an important topic in social science, but the difficulty of measuring corruption presents a significant barrier to research. This measurement problem has led to misleading results and has forced scholars to change their quantity of interest, e.g., to study the “perception of corruption” rather than “corruption.”

OUR METHOD

NOTATION

$$y = \alpha + \beta x^* + \epsilon_i$$

x^* is the latent, unobserved variable (“true” value of corruption)

$$x_g = x^* + \eta_g$$

x_g is the observed variable (corruption measure)

η_g is the measurement error in x_g not independent of x^*

OUR APPROACH

1. Identify 2 measures of corruption with different DGP x_1, x_2
2. Identify and apply “filter” variables z_1, z_2 , that give rise to the error in x_1, x_2 . Using substantive knowledge, identify z_g a $n \times k$ matrix where n is the number of observations and k is the number of covariates that lead to measurement error

$$x^* = x_g - z_g \quad (1)$$

$$z_g = \eta_g$$

Regress corruption measures on on the filter variable to obtain x''_g , the residuals from Equation 2

$$x_g = \theta_g z_g + x'_g \quad (2)$$

3. Use residuals in instrumental variables approach to obtain an estimator with reduced bias.

$$\begin{aligned} x''_1 &= \gamma x''_2 + v \\ y &= \beta_{iv} \hat{x}''_1 + \epsilon \end{aligned} \quad (3)$$

β_{iv} is consistent and unbiased when $z_g = \eta_g$ and $\hat{\theta}_g = 1$

$$plim \hat{\beta}_{iv} = \frac{cov(\beta x^* + \epsilon_i, x^* + \eta_2 - \hat{\theta}_2 z_2)}{cov(x^* + \eta_1 - \hat{\theta}_1 z_1, x^* + \eta_2 - \hat{\theta}_2 z_2)} \quad (4)$$

SIMULATION

We conduct 2 simulations to test the validity of our approach:

1. Comparison of correlation between residuals (x''_1, x''_2) versus correlation between original measures (x_1, x_2)
2. Improvement in coefficient estimate toward true value x^*

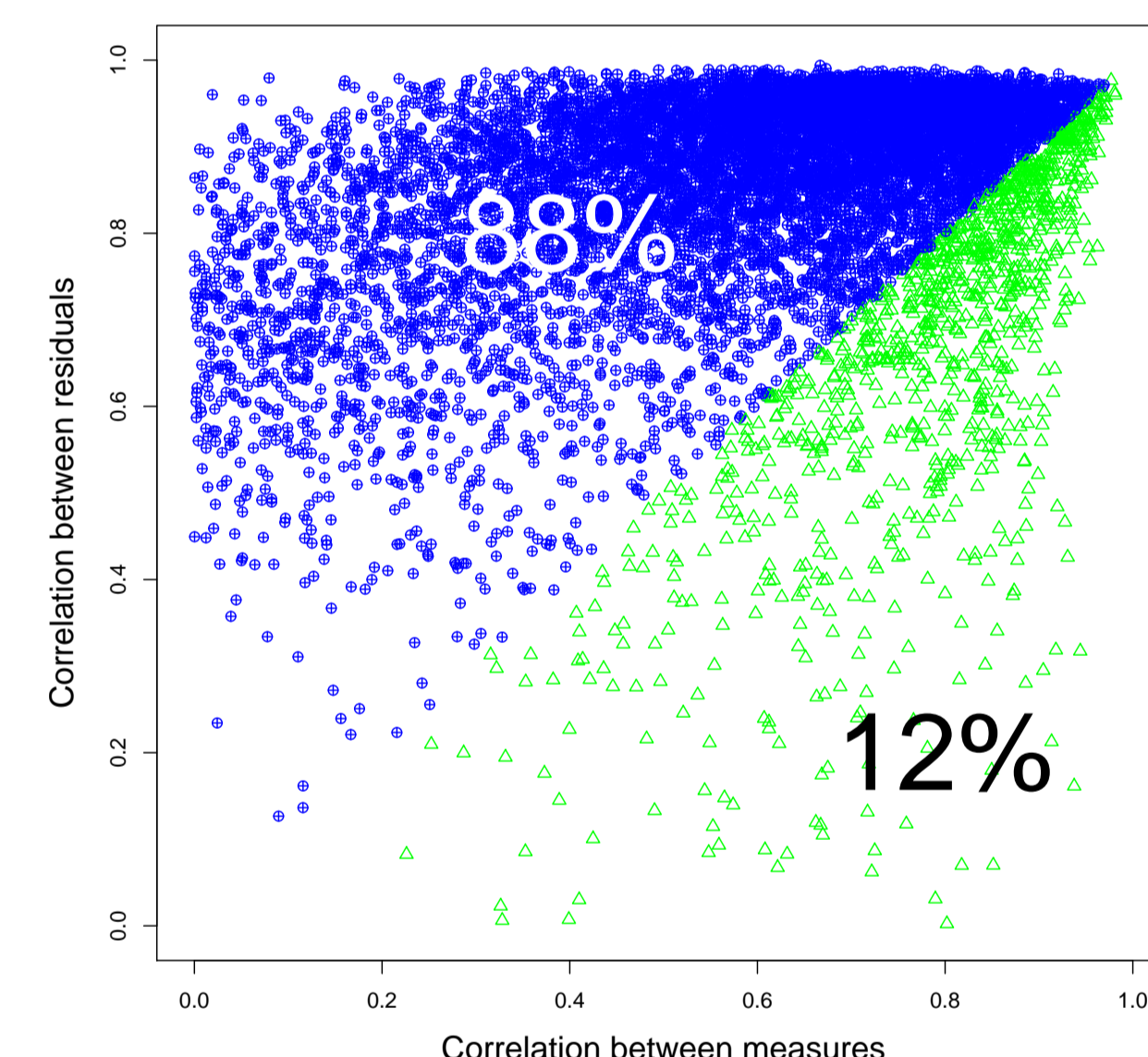
54,813 ($n=35$) simulated datasets with randomly generated values for $x^*, z_1, z_2, \eta_1, \eta_2$ that exhibit a wide range of correlations. The simulations confirm that the only necessary assumption (which is a key rule in choosing z_g) is:

$$cor(x^*, z_g) < cor(\eta_g, z_g)$$

SIMULATION 1: COMPARISON OF CORRELATIONS

In 88% of the simulated datasets, the correlation between residuals is higher than the correlation between original measures. We would expect the improvement to be greater if z_g were selected based on substantive knowledge.

Figure: Correlation of Residuals x''_g versus Correlation of Original Measures x_g



SIMULATION 2: IMPROVEMENT IN COEFFICIENT ESTIMATE

In 64-71% of datasets, the coefficient estimate move closer to the true coefficient estimate (based on x^*)

Figure: Improvement in Coefficient Estimate Compared to x_1

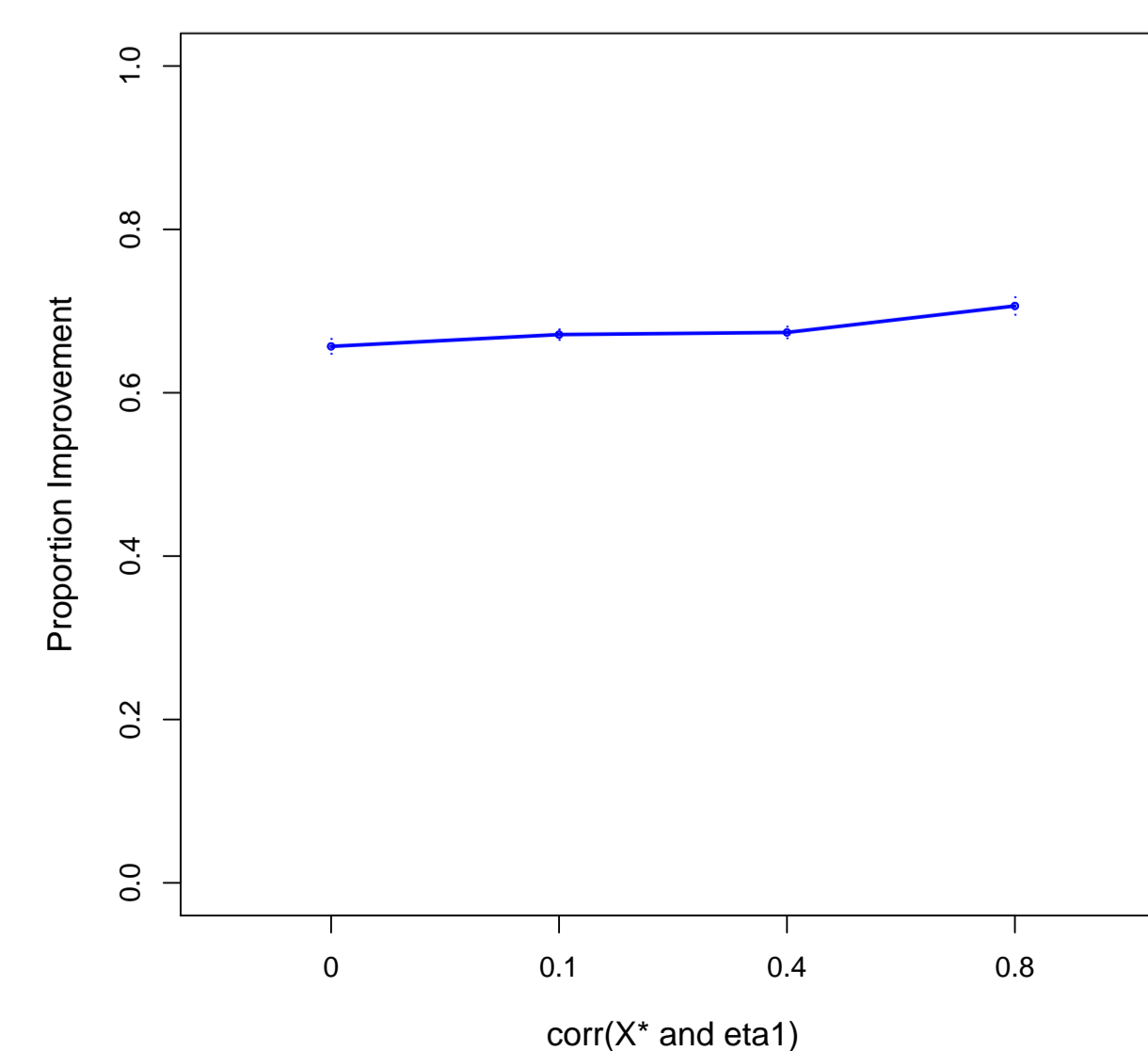
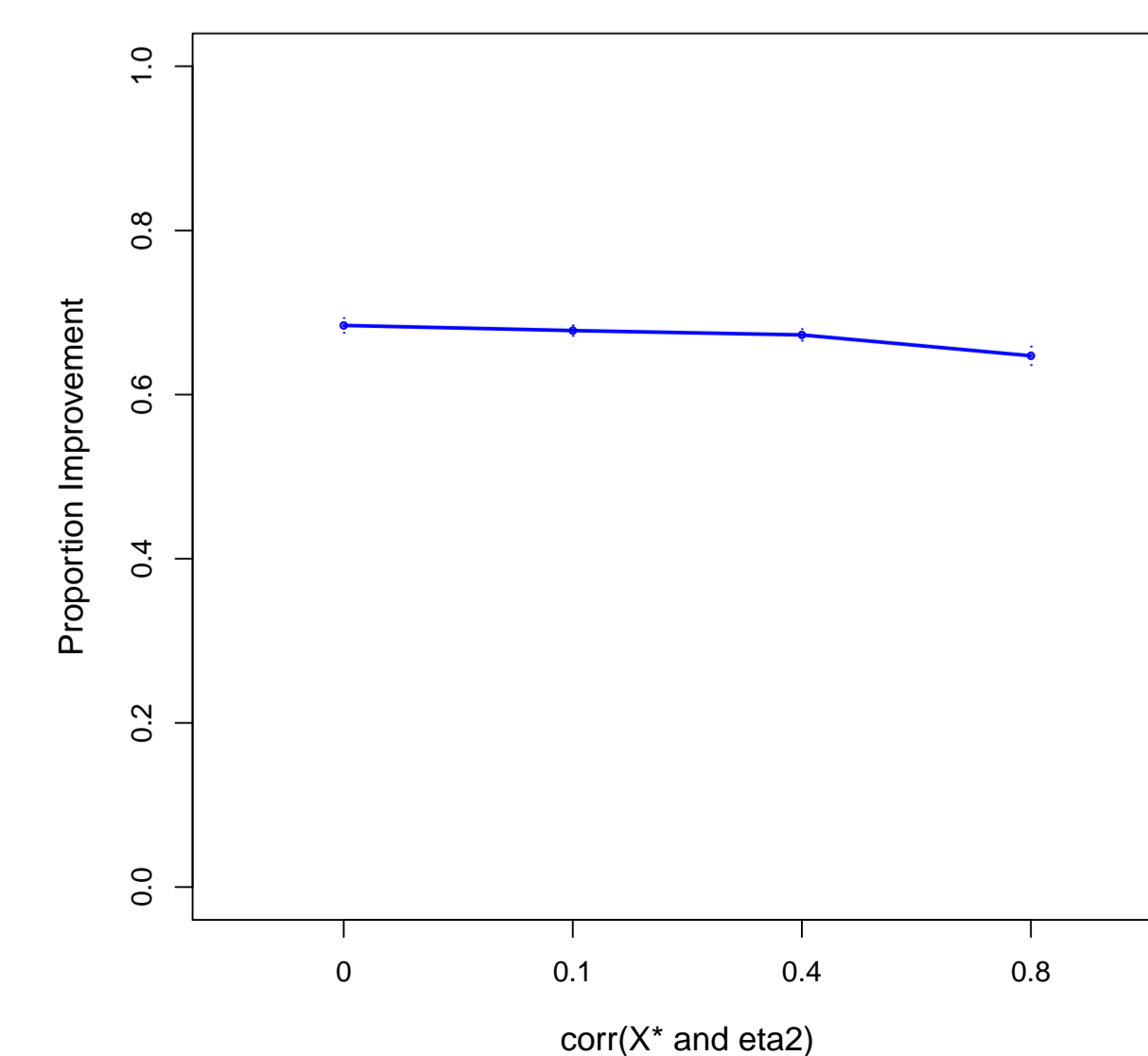


Figure: Improvement in Coefficient Estimate Compared to x_2



APPLICATION

We apply our method in an example of the effect of corruption on economic development:

$$GDP = \beta_0 + \beta_1 \text{Corruption} + \beta_2 \text{Democracy} + \epsilon$$

Two measures of corruption

- x_1 Experience-based measure of corruption from World Economic Forum *Global Competitiveness Report*
- x_2 Government-recorded measure of corruption from the United Nations Office on Drugs and Crime

Identify and apply filter variables

- z_1 Sources of bias in perception of local business people: historic trends in economic development, long-term changes in democracy, and ethnic fractionalization
- z_2 Sources of bias from legitimacy and effectiveness of criminal justice system: country conviction rate

Correlation of measures improves 50% after filtering.

Original Measure

| | Estimate | Std. Error |
|-------------|------------|------------|
| (Intercept) | 8.0827*** | 0.1312 |
| x_1 | -0.6471*** | 0.0698 |
| Democracy | 0.1257*** | 0.0195 |

More corruption predicts a decrease in economic development

Our Approach

| | Estimate | Std. Error |
|---------------|------------|------------|
| (Intercept) | 7.45196*** | 0.55812 |
| \hat{x}''_1 | -1.82503 | 1.88483 |
| Democracy | 0.24485** | 0.09166 |

Corruption has no effect on the level of economic development

Conclusion

Applying our method changes the substantive conclusion because the original experience-based measure is biased by heuristics and perceptions of respondents that have more to do with factors like ethnic fractionalization than with corruption. It is these biases rather than corruption that has an effect on economic development, and thus, when these biases are controlled for through our method, the effect of original corruption measure goes away.

[†] <http://www.people.fas.harvard.edu/jjpan/>

[‡] <http://www.iq.harvard.edu/people/chiara-superti>