Implications of Coarse Data Allocation Methods for Flood Mitigation Analysis

James Howard, Kore Federal

2013 Joint Statistical Meetings, Montréal, Canada

Introduction

Efforts to perform fine-grained analysis are often hampered by data provided by government agencies that do not reflect appropriate granularity. Coarse-grained government data may reflect the data collection methods, strategies, reflect the reality of what the data represents, or be intentionally introduced (Heitjan & Rubin, 1991).For example, the United States Flood Mitigation Assistance (FMA) grant program makes grants to both state and local governments (King, 2005). By employing data on the FMA program, this analysis examines allocation strategies for coarse data. Between 1996 and 2010, approximately one-quarter of FMA grants were given to state governments with the remaining three-quarters given to local governments. Performing a local-level analysis of the impact of these grants requires an allocation method that fairly reflects the local impact of statewide grants. This analysis considers several allocation strategies and how these strategies affect the implementation and interpretation of statistical models for public policymaking.

Data

This analysis uses information from the Federal Emergency Management Agency (FEMA) on the financial history of the National Flood Insurance Program (NFIP) and the FMA programs for the period from 1996 through 2010. The NFIP datasets include premiums paid by policyholders, the aggregate coverage levels, the count of policies in force for each year, the number of claims made against the NFIP, the amount of claims paid for structural damage, the amount of claims paid for building contents, and the amount of claims paid for Increased Cost Compliance (ICC) (Fraser, Doyle, & Young, 2006). The FMA dataset includes includes each grant made under the FMA, Severe Repetitive Loss (SRL), and Repetitive Flood Claims (RFC) programs. For each grant, the data includes the state, county, subgrantee (an agency receiving the grant), a grant program identifier, the year, and the amount.

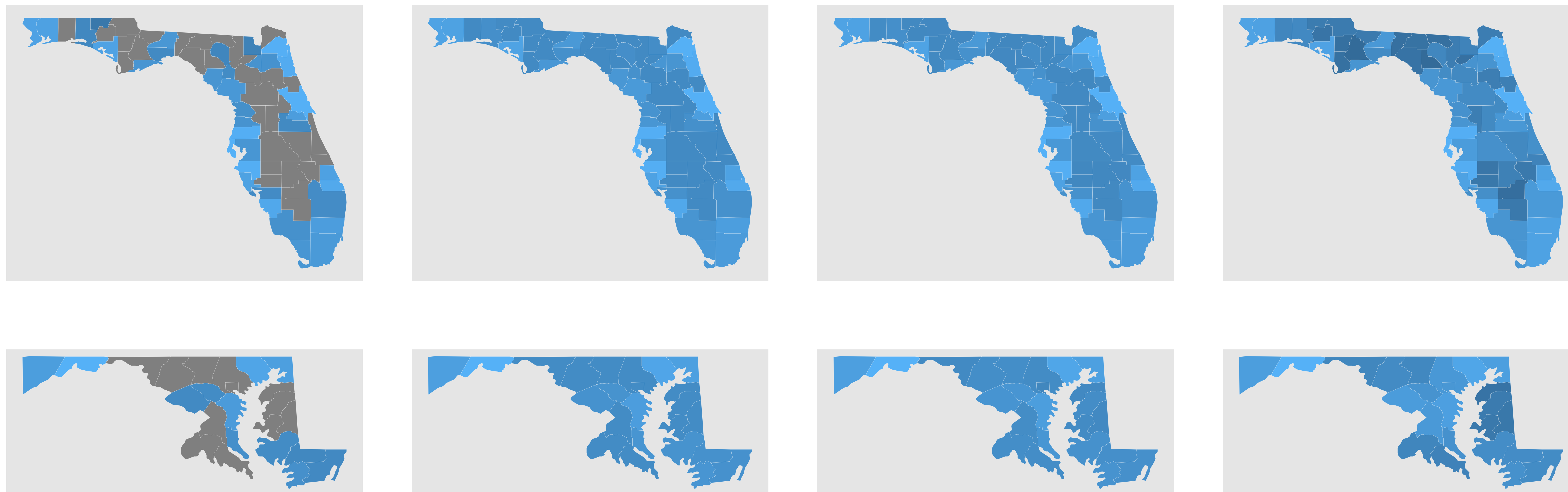
Methods

FEMA readily makes available county-level data on the NFIP's finances and using this data to understand the FMA program is a natural choice. But when assessing the FMA program for its effectiveness or to understands how grants are made by FEMA, grants issued to state-level agencies can cause analytical problems. One approach to manage these grants is to allocate them across the state. This would allow impacts to be measured in a coherent way. This analysis uses four distinct allocation strategies based on demographic data and considers the results of each. These methods are:

1. No allocation: Silently drop the state-wide grants. This is considered the base case.
2. Equally by County: Evenly allocate the grants to each county.
3. By Median Income: Proportionally distribute the grants to each county by the median income. This is designed to proxy a distribution by wealth.
4. By Population: Proportionally distribute the grants to each county by population.

Florida and Maryland Counties by FMA Grant Allocation Strategy

No Allocation (1) | **Equally by County (2)** | **By Median Income (3)** | **By Population (4)**



Results

This analysis used a panel linear model to test the impacts of different allocation strategies on the linear model. The model is a simple non-lagged model such that,

$$grants = pif + avgclaims + premiums,$$

where *grants* is the amount of grants in a given year, *pif* is the number of policies in force, *avgclaim* is the average claim size, and *premiums* is the amount of premiums paid to the NFIP. This model attempts to track whether the current-year activities of the NFIP have any bearing on how much is given to a local jurisdiction in FMA grants. This model is estimated four times to capture the differences between the impacts on *grants* at different allocation strategies. This estimation presumes a random effects model since the variance from county to county is generally random. However, this may not necessarily be the case.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-3084.27	-3075.65	-3084.27	-3444.79
	(1991.06)	(1991.03)	(1991.06)	(1994.91)
pif	-12.30^{***}	-12.31^{***}	-12.30^{***}	-12.39^{***}
	(0.32)	(0.32)	(0.32)	(0.33)
avgclaim	0.22^{*}	0.22^{*}	0.22^{*}	0.22^{*}
	(0.09)	(0.09)	(0.09)	(0.09)
premiums	0.04^{***}	0.04^{***}	0.04^{***}	0.05^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.09	0.09	0.09	0.09
Adj. R^2	0.09	0.09	0.09	0.09
Num. obs.	40627	40627	40627	40627

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Discussion

From a purely analytical standpoint, allocation strategy has little impact on on regression results. In the case of *pif*, the impacts are statistically significant but also consistent to three significant digits across all four models. The impacts for both *avgclaim* and *premiums* respond similarly. Remarkably, the allocation strategy that disposes of more than a quarter of the data is approximately the same as the the demographically-allocated grants. This implies allocation strategies, provided they are done in a reasonable and repeatable manner, are largely inconsequential to the final regression estimates.

Allocation strategy is not meaningless. As is evident in the maps of Florida and Maryland, there is substantial difference in the final amounts of grants awarded to a county depending on the allocation strategy employed. If allocating equally or by median income, the maps show little difference. But allocation by population shows a different grant pattern. This is shown in the estimated y-intercepts, which varies for the population-based allocation method from the other approaches. This suggests that a principal driver of the grants given to a county is not one of the three variables estimated in the regression and the robustness of the regression method is on display.

Conclusions

This analysis used multiple allocation strategies with FMA grant data to determine the effects of those allocation strategies on statistical analysis. There is little impact on regression coefficients for any variable, despite the differences shown on the sample maps. However, there is a dramatic change in the y-intercept. This points to other potential regressors that could have control over the outcome variable that are not included in the model. There are implications for this across multiple policy domains (*e.g.*, Janssen and Sklar (1998)).

References

Fraser, J. C., Doyle, M. W., & Young, H. (2006, July). Creating effective flood mitigation policies. *Eos, Transactions, American Geophysical Union, 87*(27), 265–270.

Heitjan, D. F. & Rubin, D. B. (1991, December). Ignorability and coarse data. *Annals of Statistics, 19*(4), 2244–2253.

Janssen, G. & Sklar, M. (1998, September 16). *Geographic allocation of state level nonroad engine population data to the county level.* (No. NR-014). United States Environmental Protection Agency. Washington.

King, R. O. (2005, June 30). *Federal flood insurance: the repetitive loss problem.* Congressional Research Services. Washington, D.C.

KoreFederal