Case Study Design and Analysis as a Complementary Empirical Strategy to Econometric Analysis in the Study of Public Agencies: Deploying Mutually Supportive Mixed Methods

Dan Honig

March 29, 2018*

There is little methodological guidance regarding how to best integrate qualitative observational case study data and quantitative large-N observational data in the study of public agencies. This paper examines the role of case studies in mixed methods work via an inferential problem of relevance to the study of public agencies for which mutually supportive mixed methods may be particularly helpful: simultaneously estimating the effect of a slow-moving or time-invariant variable while controlling for unit level fixed effects. The paper's central points are illustrated using mixed method data on foreign aid agency management practice and performance.

* Assistant Professor of International Development, Johns Hopkins School of Advanced International Studies. Many thanks to James Bisbee, Peter Hall, Rich Nielsen, Tom Pavone, David Steinberg, and Michael Woolcock for their helpful comments, and to Anthony Bertelli, Peter Hall, Michael Woolcock, and others for their encouragement to take up this project. I can be reached at dhonig@jhu.edu.

1. Introduction

The literature on case selection and methods is increasingly complex, as befits a maturing methodological sub-field. As Pavone (2017) notes, a recent synthesis of case study selection methods derived "no less than five distinct 'types' (representative, anomalous, most-similar, crucial, and most-different) and eighteen 'subtypes' of cases, each with its own logic of case selection." (Pavone 2017, 2 describing Gerring and Cojocaru 2015).[1] This paper is not meant to add to this complexity, but rather to focus on a specific application of qualitative case studies: as complementary to quantitative methods in making a causal argument.

Qualitative methodologists (e.g. Brady and Collier 2004; George and Bennett 2005; Levy 2008) have argued that qualitative scholarship has moved beyond the "single logic of inference" made famous by KKV (King Keohane and Verba 1994). That said, mixed methods scholars who wish to appeal to a broad range of scholars (including those for whom qualitative methods are less familiar) are often are called upon in practice to wrestle with the logic of positivist quantitative analysis. Scholars steeped in quantitative epistemology may also wish to incorporate cases into their work, and struggle with how to do so.

This article is thus oriented towards scholars whose epistemic compasses, as Mahoney (2010) puts it in describing the appeal of seminal pieces by Lieberman (2005) and Gerring (2007), "share with KKV a statistically oriented approach to social science." This article focuses on how to employ case studies in ways complementary with the causal logic of mainstream quantitative work, with particular attention to bolstering causal inference in contexts where collinearity between terms bounds the usefulness of econometric analysis.

Case study empirics in a mixed methods context are often usually conceived as providing more fine-grained interrogation of claims made in quantitative work. In his book <u>Delegation in the Regulatory State</u>, Gilardi conducts a primarily quantitative inquiry, but concludes by examining the establishment of the German energy regulator as an example of an "interesting case." Drawing on Lieberman (2005), Gilardi argues that qualitative cases can be chosen based on the conclusions of the quantitative analysis. (Gilardi 2008) For Gilardi, it appears qualitative and quantitative analyses are conceived of as complementary, but in a particular sense: they are layered upon one another (with appropriate linkages between the qualitative and quantitative "layers," of course) to present a more complete picture than would otherwise be available. The qualitative data does not, in this illustrative case, actually do any of the primary hypothesis testing.

There are, however, a variety of situations in which econometric analysis is possible, but is incomplete. In these situations, qualitative exploration is not merely about "adding layers." it is also about "filling holes," in what I term a mutually supportive mixed methods strategy. I define a mutually supportive mixed methods approach as one where the design and/or

[1] This itself is arguably an elaboration of Seawright and Gerring 2008's typology, which includes typical, diverse, extreme, deviant, influential, most similar, and most different cases.

analysis techniques employ both qualitative and quantitative empirics through a single logic of inference. Mutually supportive mixed methods has much in common with Seawright's (2016) "integrative multi-method research", which focuses on research design (with special attention to optimal case selection). Seawright describes his book as "the first systematic guide to designing multi-method research", suggesting the relatively uncharted terrain of methodological integration.

One way in which mutually supportive mixed methods differ from the logic of nested analysis (e.g. Lieberman 2005) is that the "small N" qualitative work is not merely a fine-grained look at mechanisms from "large N" analysis, but rather the small- and large-N analyses can be conceived of as mutually supportive hypothesis testing. The focus on variation in independent variables for case selection also contrasts with nested analysis in allowing simultaneous design of qualitative and quantitative empirical strategies, rather than conceiving of the former as endogenous to the results of the latter. In this sense, the logic of mutually supportive mixed methods has more in common with Humpreys and Jacobs (2015), as it conceives of qualitative and quantitative evidence as complementary data whose optimal proportion depends on the nature of the data environment and inferential challenge. [2]

This article is situated in a broader stream of methodological literature that collectively suggests there are alternatives to consigning research ideas for which the best possible quantitative empirical strategies are less than fully satisfactory to the rubbish bin. This article explores the integration of quantitative and qualitative data in the context of an inferential challenge of potentially broad applicability to the study of public agencies, collinearity between slow-moving or time invariant features of agencies, contexts, and/or measurement strategies and unit-level fixed effects.

This article first introduces the motivating problem of how to account for fixed or slow-moving features of analytic units in multi-unit studies via a general model (Section 2). The paper then instantiates the challenge and discusses solutions in the context of aid agencies and the success of foreign assistance efforts, drawing on my (2018, 2019) work on same (Section 3). In this work the quantitative analysis fails not only to elucidate causal mechanisms but also to estimate a critical substantive relationship of interest. The paper then draws broader methodological lessons (Section 4) and illustrates the potentially broad applicability of the proposed method using published work from the Journal of Public Administration Research and Theory (Section 5) before turning to a broader discussion and conclusion (Section 6).

2. Time-invariant or Slow-Moving Features of Agencies and Contexts: A Source of Holes in the Public Management Landscape

One central concern of public management scholarship is the relationship between management practices and outcomes such as employee motivation or organizational

[2] This in turn has echoes of Adcock and Collier 2001 on shared measurement validity standards for qualitative and quantitative research.

performance. These relationships are typically contingent. An identical feature – e.g. a performance management system – may have very different associations with performance in different organizations. O'Toole and Meier (2014) introduce a simple general model of context's interaction with agency management on performance of the form

$$O = \beta_1 M + \beta_2 C + \beta_3 MC + \beta_4 X + \varepsilon.$$

This paper adapts the form of model slightly, taking O to be organizational performance, M management practice of a given agency, C specific features of the broader context in which these actions take place or a given agency sits, and X a vector of control variables.[3]

The logic of this model is simple and general: actions and features of agencies have an impact on performance. So, too, do features of the context in which agencies operate. However, context and agency features interact. As a result, there is no single strictly dominant agency design or management practice, with the right "tools for the job" of delivering public value a function of the nature of the job itself and where the job is located.

A quantitative, positivist scholar will at this juncture likely be thinking that variation within M and C will be necessary to properly estimate the full model in a context where the researcher wishes to give the analysis a causal flavor. This variation might be cross-sectional or time series. Better yet, we could examine both cross-sectional and time-series variation, which would strengthen causal claims regarding how a change in M, or a change in C, affects O. Using panel data to estimate the model for agency i in context j would yield a model of the general form

$$O_{i,j,t} = \beta_1 M_{i,t} + \beta_2 C_{j,t} + \beta_3 MC_{i,j,t} + \beta_4 X + \varepsilon.$$

Using panel (cross sectional time series) data introduces the question of fixed effects. It is highly unlikely that the vector of controls *X* will fully capture all of the ways features of agency *i* or context *j* might impact performance *O*. To ensure time-invariant features of *i* and *j* are not introducing omitted variable bias, fixed effects at the *i* and *j* levels are appropriate; to ensure common temporal shocks (e.g. a global recession) are not contributing bias, time fixed effects may also be useful. This jointly yields:

$$O_{i,j,t} = \beta_1 M_{i,t} + \beta_2 C_{j,t} + \beta_3 MC_{i,j,t} + \beta_4 X + \text{Fixed Effects}_i + \text{Fixed Effects}_j + \text{Fixed Effects}_t + \varepsilon.$$

Critical to straightforward estimation of this model is that *M* and *C* are time variant; if $M_{i,t}$ or $C_{j,t}$ have no temporal variation they will be collinear with fixed effects at the $_i$ or $_j$ level.

[3]This differs slightly from the original. In O'Toole and Meier's original formulation, C is a vector describing the context in whole; in this paper C is a specific contextual feature. Similarly M is in this paper a specific management practice, rather than a vector of managerial actions. This adaptation also reverses β_3 and β_4 relative to the original, to put the terms of primary focus ($\beta_{1,2,3}$) in front. O and X are unchanged from the original, as is the functional form.

This will make either β_1 or β_2 (or both, where neither M or C have temporal variation) unestimable.[4]

The simplest solution is to ensure that the features of *M* and *C* that are estimated have inter-temporal variation. This is often the case: if *M* is a performance management system that has been introduced in the middle of the time period the data covers, then fixed effects at agency_i level will simply (and appropriately) allow a straightforward intra-organizational comparison of what occurred before and after the performance management system's introduction. Similarly, if the theoretically interesting feature of *C* is the party to lead the national government, fixed effects at country_j level will allow comparison of the time under each party's leadership. If the researcher's interest is in simply controlling for time invariant features, these features will be absorbed by fixed effects.

However, there are many features of agencies and context that may be important to model but are unlikely to be time variant. On the agency level, features might include the formal structure of the agency; the year of founding of the agency; whether the agency is an independent agency; whether the head of an agency is a cabinet member; the year of an agency's founding; the existence in the agency of a particular feature (e.g. an internal audit function); the organization's level of centralization; etc. On the country level (as a particular instantiation of context) features might include whether the state is unitary or there are shared powers; the degree of federalism in the country; the legal tradition of the state; the state's status as a developed or developing country; etc. There are also the myriad contexts in which a given feature of agencies (e.g., their level of professionalism) or contexts (e.g., societal social capital) are time variant but collected by survey. Creating a panel using surveys will require multiple administrations, and may not be feasible due to financial, logistical, or other constraints.

To estimate a full model when *M* is time varying and *C* is time invariant for a given agency_i requires estimating not just the coefficient β_3 on the interaction term *MC* but also β_1 and β_2. The same is true for a case where M is time invariant (either in practice or in estimation) and C is time variant. The β cannot be estimated for the term which is time invariant, be it C or M. Thus an econometric analysis can estimate marginal, but is not conventionally seen as being able to estimate aggregate substantive, effects.[5]

---

[4] One solution here would be to employ a multilevel random effects model; however, there will be many situations where multilevel models either cannot be estimated or are inappropriate (e.g. because the need for fixed effects to control for what would otherwise be a source of bias is clear, as in this paper's example).

[5] Plumper and Troeger (2007) put forward a three-stage procedure for, as they put it, "the estimation of time-invariant and rarely changing variables in panel data models with unit effects". The qualitative empirical strategy described below can be read as a complement to their vector decomposition model, with the degree to which we ought update our priors (in a Bayesian sense) given addition weight by the combination of qualitative and vector decomposition empirical strategies.

The following section discusses this problem and solutions in the context of a particular application: aid agencies' management practices and their performance.

3. Aid Agencies' Management Practices, Project Performance, and Fixed Effects

In a recent book (Honig 2018) and a related paper (Honig 2019), I investigate the causal effect of greater or lesser field agent control (M) on performance (O). These studies examine the practices of foreign aid agencies (e.g. the US Agency for International Development and the World Bank) that give field agents greater (or lesser) control over the design, revision, and day-to-day management of foreign aid interventions.

An Econometric Hole: Missing Quantities of Interest Problem

The argument in Honig 2018 and 2019 explicitly conceives of greater or lesser field agency control as an agency management practice (M) that interacts with features of the broader context (C). I hypothesize that the returns to giving foreign aid's street-level bureaucrats (Lipsky 1980) greater control will have increasing returns to performance when recipient country environments are more unpredictable. Uncodifiable information about context (tacit knowledge in the sense of Polanyi 1966, or soft information in the sense of Stein 2002) to which only field agents have access will be in higher demand in more unpredictable environments, and is more likely to be gathered and used (Aghion Tirole 1997) when field agents have greater control.

To test this theory, I assembled a database of over 14,000 discrete development projects from nine aid agencies. These projects have Likert-type outcome scores assigned by the aid agencies themselves, allowing multivariate (OLS and ordered logit) regression models to be fit to the data. Collectively the projects span over forty years and 178 recipient countries. I was thus able to exploit the intertemporal nature of the data to control for time, agency, and recipient country fixed effects. This ensures that fixed features of agencies, common temporal shocks, and fixed features of recipient countries are not biasing the results. In the terms of the econometric model outlined above, my model can be described as:

$$\text{Project Success}_{i,j,t} = \beta_1\text{M(Field Agent Control)}_{i,t} + \beta_2\text{C(Country Unpredictability)}_{j,t} + \beta_3\text{MC}_{i,j,t} + \beta_4 X + \text{Agency Fixed Effects}_i + \text{Country Fixed Effects}_j + \text{Time Fixed Effects}_t + \varepsilon.$$

The measure of C, Country Unpredictability, draws from the country-level panel data in the Polity IV State Fragility Index. (Center for System Peace 2014). The measure of field agent control M, however, is a time-invariant survey measure.[6]

The quantity of primary interest – management practice – is collinear to agency fixed effects. But agency fixed effects are critical for this analysis. As Honig 2018 notes, there is no reason to believe that each agency assesses project success using a parallel scale; for

[6] While the measure is in fact from multiple waves of the Paris Declaration Monitoring Surveys (OECD 2012), these surveys are quite proximate in time (2004, 2007, 2010); Honig conceives of them as multiple measures of the same construct rather than a true time series.

example, that a given agency's rating of four on a six-point scale is equivalent to another. This research thus faces the challenge described in section 2 above. I can, and do, fit models that generate estimates of β_2 and β_3. However, these models cannot estimate β_1, given the collinearity of the measure of *M* to agency fixed effects.

The analysis finds that the coefficient on β_3 is positive and statistically significant, suggesting that there are increasing returns to field agent control as a given aid recipient country becomes more unpredictable. While this marginal effect may be of interest, it leaves open the question of whether greater field agent control is actually *better* for any given project. Figures 1, 2, and 3 illustrate this problem graphically.[7]

The marginal effects of differential levels of field agent control can be estimated using the sum of β_2 and β_3; we can observe the differential slopes of project performance by level of *C*, recipient country environmental unpredictability. But I cannot estimate the relative *levels* of the marginal effects plot lines, as indicated by the vertical black arrow superimposed onto the marginal effects plot (which is meant to indicate that the relative vertical positions of the lines are not estimated; that e.g. the line that appears to be on the bottom may in fact be above the line that appears above it).

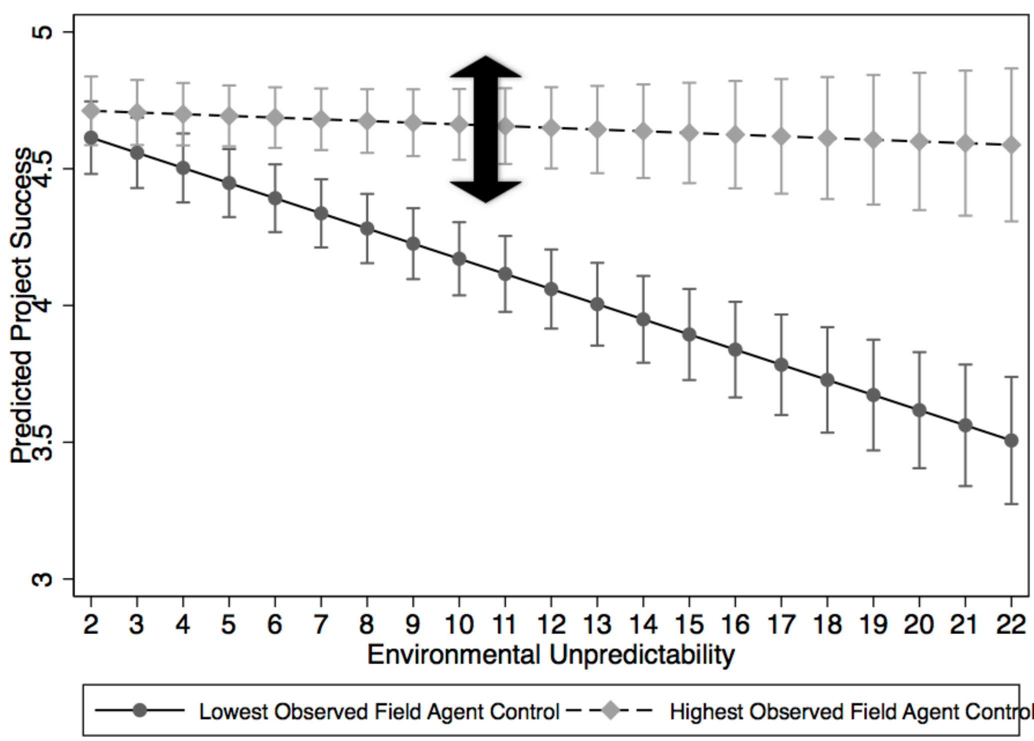
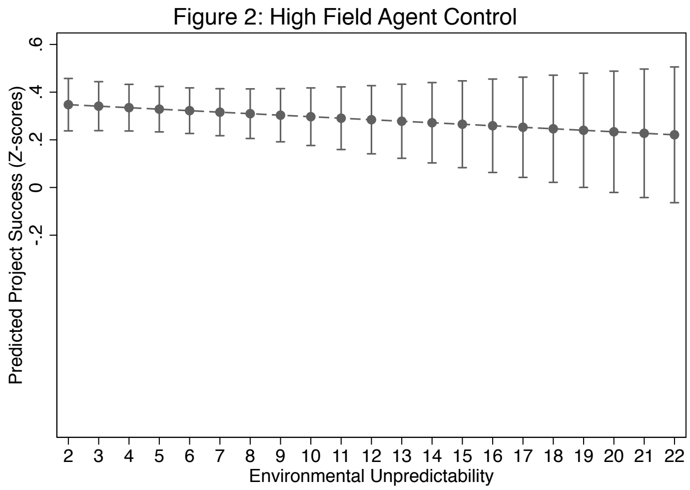
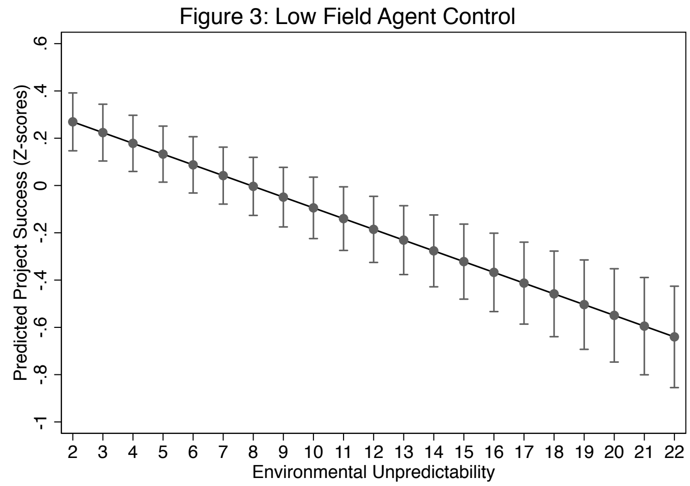


Figure 1: Graphical Estimate of the Interaction of Environmental Unpredictability and Field Agent Control in the Absence of β_1

[7] Adapted from Honig 2019, figure 2.





Figures 2 and 3: Figure 1 disaggregated for High Field Agent Control (Left, Figure 2) and Low Field Agent Control (Right, Figure 3)

The information in figure 1 is re-expressed in figures 2 and 3. Figures 2 and 3 express the same two lines as present in figure 1, but using a normalized agency-specific z-score as the outcome variable. The analysis demonstrates that the slope of the two lines (for high levels of field agent control in figure 2, and low levels of field agent control in figure 3) differs. But as they are both plotted against normalized values of each agency's own predicted success, there is no way to know whether e.g. a z-score of .5 in figure 2 is in actual, real-world terms of project success higher or lower than e.g. a z-score of .2 in figure 3.

As a result of these econometric limitations, this analysis does little to clarify the substantive effect of field agent control. Figure 4 demonstrates the stylized possibilities by holding the slope of the estimates for both low and high field agent control constant, but arbitrarily varying the intercept of the high field agent control line in the plot.

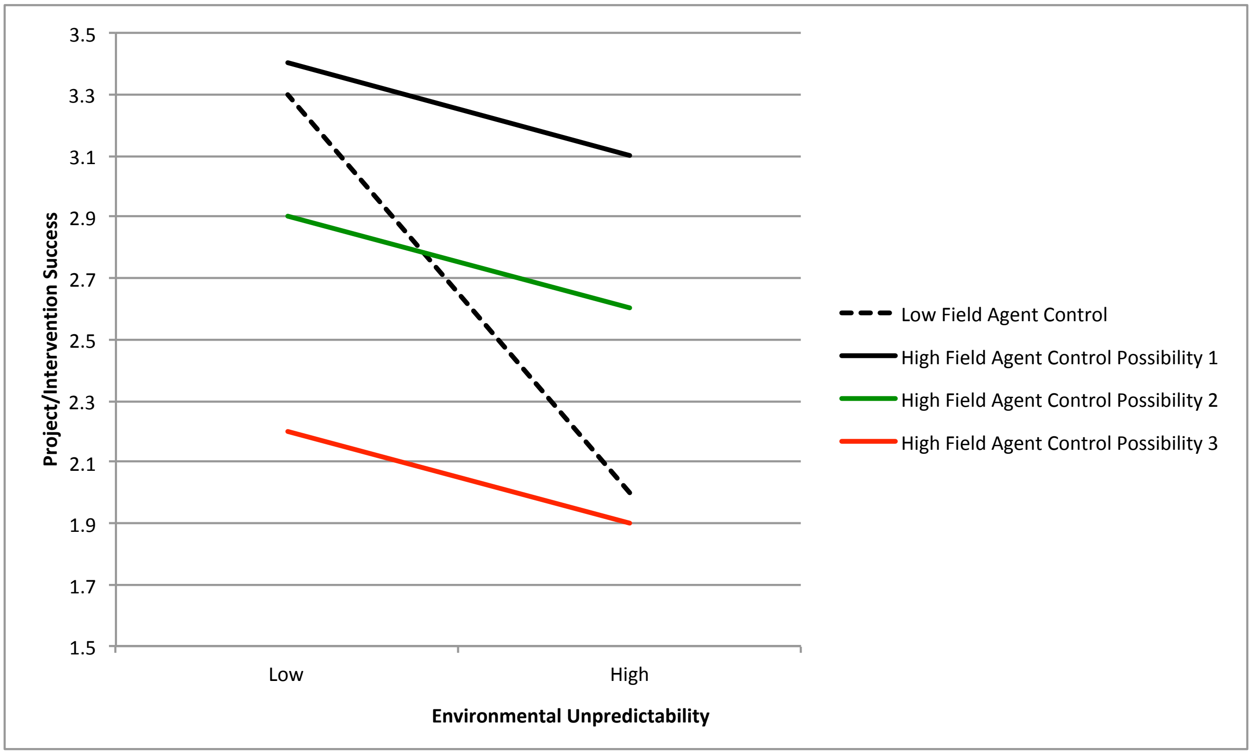


Figure 4: Greater or Less Field Agent Control May Strictly Dominate, or There may be a Conditional Net Relationship (We Cannot Know Based on the Quantitative Empirics)

It is possible that higher levels of field agent control strictly dominate lower levels (Possibility 1 of Figure 4); that greater levels of field agent control are associated with better aid project performance in all environments, with increasing returns in more unpredictable environments. It is also possible that lower levels of field agent control strictly dominate higher levels (Possibility 3 of Figure 4): that even as lower levels of control are associated with declining of project effectiveness in contexts of greater environmental unpredictability, this decline still leaves less field agent control the more effective strategy in even the most unpredictable contexts. It is also possible that the lines cross (Possibility 2 of Figure 4): that lower levels of field agent control are superior in more predictable environments, and higher levels of field agent control are superior in less predictable environments. The three stylized possibilities depicted in Figure 4 are identical in the slopes they give the "high level of field agent control" and "low level of field agent control" organizations, implying the same β_3 in the econometric model. But this leaves open the central question: what is the right management strategy *M*, and does the "right-ness" of this strategy depend on context *C*? Which strategy is better? If the answer is conditional, where exactly the optimal strategy "flips" cannot be estimated econometrically.

Mutually Supportive Mixed Methods Part 1: Choosing Case Studies to Complement Econometric Weaknesses

I turn to case studies to complement the econometric analysis. While I describe the case studies as exploring mechanisms in a way consistent with mainstream process tracing (Bennett and Checkel 2015; Blatter and Blume 2008; Hall 2013), Honig 2018 and 2019 make clear that a critical role of the case studies is to allow a direct comparison of relative project success of different levels of field agent control *(M).* The case studies are used to directly observe the slow-moving variable (agencies' level of field agent control *M)* that was collinear to fixed effects in the large-N analysis, thus precluding an estimate of β_1.

As such, the case studies were chosen along what might be best described as a "similar enough" case selection strategy (Nielsen 2016). This is a cousin of the "most similar" selection strategy (e.g. Seawright and Gerring 2008, though the concept dates at least to Mill 1843) which attempts to hold constant all factors other than level of field agent control, but recognizes that not all information is knowable and estimable to truly choose the absolute maximally similar cases. I examine pairs of cases from two organizations where the quantitative measure used for *M* varies substantially in level of field agent control – the United States Agency for International Development (USAID) and UK Department for International Development (DFID). I choose cases where USAID and DFID attempt to accomplish similar goals in the same country over the same time period.

As described above, however, a critical question remains: whether one management strategy *M* (more/less field agent control) clearly dominates the other. To investigate further, I examine pairs of USAID and DFID cases in countries of different levels of environmental unpredictability *C*. I examine four pairs of cases, or eight case studies in Honig 2018. Two pairs of cases occur in Liberia in the 2000s, a relatively high unpredictability environment; two pairs occur in South Africa in the 2000s, a relatively low unpredictability environment.

I also vary one additional feature of context *C* in case selection: the degree to which the task domain of the aid project is tractable to external measurement, or the project's external verifiability. This is a feature of context that cannot be accurately measured econometrically. Thus while project external verifiability features in the theory - I argue the less tractable a given project to performance measures, the greater the returns to field agent control – project verifiability plays only a limited role in the 2019 quantitative analysis.[8] A full schematic of the case selection strategy is presented in figure 5 below.[9]

[8] Honig (2019, chapter 6) provides some suggestive quantitative analysis by sector, but makes clear that sectors are largely poor proxies for external verifiability.
[9] Adapted from Honig 2018, Figure 1.2.

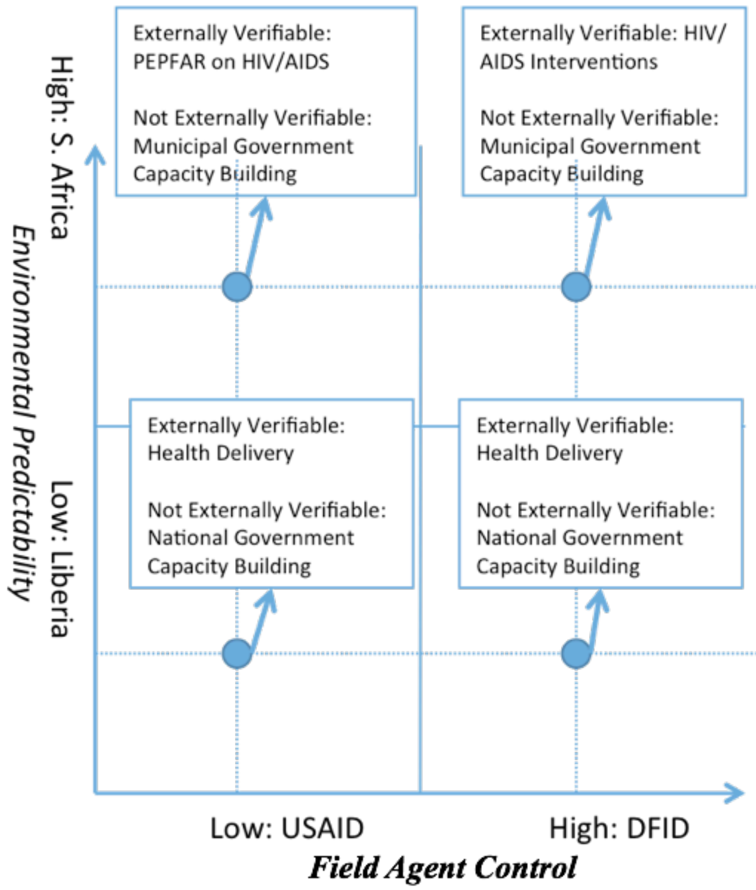


Figure 5: Schematic of Selected Cases

These case selection methods have something in common with mainstream nested analysis (Lieberman 2005). Case selection is informed by quantitative parameters. However, unlike in most applications of nested analysis, it is not the *results* of econometric analysis that inform case selection. Case selection is done to maximize variation in the independent variables *M* and *C*. Cases are chosen to maximize variation in the key independent variables so as to allow the qualitative analysis to complement econometric analysis, given the limitations of the econometric model. It is the level and variation of independent variable regressors, rather than the results of econometric analysis regressions, that inform case selection.

Mutually Supportive Mixed Methods Part 2: Conducting Case Analysis to Complement Econometric Limitations

Ensuring case studies complement econometric analysis does not end at the case selection stage. In any qualitative case study, researchers must choose to focus on specific elements of the case rather than others. When case studies are used to complement econometric analysis, the choice of how to construct cases can, and should, be informed by the limitations of the econometric analysis they are meant to complement.

Each case study in Honig 2018 explores the design, implementation, and revision of the development projects it examines. In each of the four case study pairs the success of USAID and DFID projects are directly compared as well. Process tracing links a given project's success to M and C. I also take pains to establish in each case pair the differing levels of field agent control M in the projects. I trace how M affects project success in each case, focusing on critical junctures in the design, implementation, and revision of projects. The case studies thus provide (qualitative) estimation of the net effect of M and C in interaction – that is, the qualitative equivalent of the sum of β_1, β_2, and β_3. The cases can provide suggestive, small-N evidence on whether high or low levels of M strictly dominate one another, or rather if more field agent control M is associated with greater project success in some contexts C, and lower levels of project success in other contexts C.

By way of brief illustration, one case pair examines USAID and DFID efforts to strengthen South African municipal fiscal management. Both USAID and DFID sought to strengthen the ability of selected South African local governments to manage their budgets and expenditures. The two foreign aid agencies managed their parallel projects very differently, and had very different results.

USAID's project aimed to help municipalities deliver public services more effectively by transferring knowledge to municipal staff. A training plan was centrally developed with modules including municipal accounting, billing systems, and debt management. On a designated day, a trainer would arrive and hold a session on a given topic. The trainings were easily monitored, and measurable. Whether the trainings were actually effective for the people in the seats, however, was less clear. A leader of the USAID project suggested indicators were chosen "because [they were] easier to count… but the number didn't tell about the impact." Another project leader suggested the USAID project "might have not made the most dent or impact." One of the trainers reported he didn't "think [the trainings] contributed much."

By contrast, DFID's project strategy centered around embedding advisors in local municipalities. The advisers resided in the communities for extended periods of time, building skills and systems on an ongoing basis. These advisers set the specific goals against which they reported. Where USAID's project focused on quantifiable metrics like "all staff trained," DFID's asked advisers to "implement their work plans and report on progress." Effectively, these advisers and their judgments led the project. DFID not only condoned this strategy; they explicitly designed it into the project. As full-time residents for the long term (two to three years), DFID project advisors were often – though not always - able to find a way to positively influence municipal systems. In interviews, both beneficiaries and project staff reported that advisors achieved some shifts in municipal practices. As one implementer put it, DFID's reporting was "more content-rich; it was not a numbers game."

The research directly compares project success in each case pair and the relationship between management practice M and project success in each case. I determine that DFID's project was by no means an overwhelming success. But it was substantially more successful than USAID's. USAID and DFID implemented programs with similar goals, through similar contracting structures. USAID's project had very little field agent control; DFID's had a great deal. This difference in field agent control is linked at length to field agent success. (Honig 2018)

Each individual case pair can say little about the overall context C. However, by comparing across case pairs, the qualitative empirics can provide leverage on environmental unpredictability C's role in mediating the relationship between field agent control M and success O. The case analysis ultimately concludes that in three of the four case pairs, greater field agent control is a factor in the relatively greater success of projects. In the case pair where theory predicted the returns to field agent control likely to be lowest – in the most stable environment (South Africa) with the most externally verifiable projects (delivery of drugs for HIV/AIDS) – greater field agent control is a factor in the relatively lower success of DFID's project as compared to USAID's. The case pairs thus collectively suggest that the relationship between M and O is indeed conditioned on environmental unpredictability C.

The case studies explore mechanisms, but with attention to the weaknesses of the econometric analysis. As the quantitative analysis does not allow direct estimation of comparative project success, given the agency-specific nature of the measurement regime. The qualitative cases establish these levels of comparative project success directly. In finding that in all four case pairs USAID and DFID differ on M as predicted (with DFID having much greater levels of field agent control than USAID), the cases also serve to reduce concerns that the agency-level measurement of management practice M in the quantitative analysis is introducing bias in being insufficiently sensitive to comparative intra-organizational variation in field agent control M. The clear links between management practice M and project success suggest that it is indeed field agent control M, and not some other agency-level features which co-vary with agent control, that plays a causal role in project success and failure.

4. The Connective Tissue of Mutually Supportive Mixed Methods: Using Parallel Quantitative and Qualitative Approaches

The discussion thus far has framed qualitative analysis as filling the inferential holes of the quantitative strategy. The inverse is also true of the Honig 2018 analysis: the quantitative analysis undergirds weaknesses in the qualitative case design. Indeed, section 3 could have proceeded in precisely the inverse way: by beginning with the qualitative logic of causal inference, suggesting its weaknesses, and then finding econometric complements. As I acknowledge, the qualitative cases are intentionally chosen to maximize variation. It is possible that the functional form of the relationship between field agent control and outcomes is not linear, but a purely qualitative examination might erroneously ignore, for example, a parabolic relationship due to not sampling the middle of the distribution. This is but the tip of the iceberg of potential problems with causal inference with which qualitative methodologists are rightly concerned. The cases could be outliers in a variety of unintended ways, and thus not provide accurate systematic data. The large-N analysis helps to "fill the holes" left by the qualitative analysis.

Put another way, the case studies might be conceived of by a primarily quantitative scholar as estimates of net marginal effects that allow simultaneous estimation of $\beta_{1,}$ $\beta_{2,}$ and $\beta_{3.}$ The case study pairs suggest that it is possibility 2 of Figure 4 above, not possibilities 1 or 3, that depicts the correct stylized relationship; that the best strategy M depends on C. This quantitative scholar might conclude that the case studies provide an important complementary source of evidence to the "primary" quantitative empirics.

A primarily qualitative scholar might begin by conceiving of the case studies as providing strong suggestive evidence that field agent control is an important component of development project outcomes in the case study projects, with the impacts of this management practice conditioned by the unpredictability of recipient country environments. A qualitative scholar might see the quantitative analysis as suggesting that case selection and analysis methods and features of the agencies chosen are not driving the findings, strengthening the claim of the qualitative cases to broader generalizability. This qualitative scholar might conclude that the large-N econometrics provide a complementary source of evidence to the "primary" qualitative empirics. Both the primarily quantitative and primarily qualitative scholars might agree that the existence of both sets of empirics allows us to update (in the Bayesian sense) our priors more firmly, even as they disagree on which component of the empirical strategy provides more useful information.

Inferential challenges are not method-specific; the limitations above are not quantitative limitations to be addressed with qualitative data, nor are they qualitative limitations to be addressed with quantitative data. To take the last item mentioned – estimating substantive significance – in the example in section 3, the quantitative data could not tell us whether field agent control was substantially driving differences in project outcomes, and so the qualitative case studies were designed to directly examine this link. But we could imagine a scenario where qualitative data illustrated one step in a causal chain, but not the outcome of ultimate interest. Perhaps qualitative case studies drawing on interviews conducted both before and after the introduction of performance evaluations in a given agency convincingly demonstrate that the introduction of new evaluation methods improves employee attitudes about their work. The researcher, however, cannot determine whether improvements in employee engagement lead to better client outcomes. In this case, large-N data linking evaluation system change-induced differences in employee engagement survey scores to client outcomes might "fill the hole" by estimating the size of the substantive effect as well as the proportion of the variance (the R^2) explained by the econometric model. Both qualitative and quantitative empirical strategies may yield unsatisfying answers as to substantive significance; both qualitative and quantitative strategies may be useful in complementing these weaknesses via carefully co-designed, mutually supportive mixed methods.

Whether one begins by examining the case studies or the large-N analysis, at the heart of the common inferential challenge is the need to leverage both within-and between-unit variation in making strong empirical claims. One way of describing the fixed-effects strategy in the quantitative analysis is as shifting the inquiry to within-unit variation. When I employ fixed effects at the recipient country$_j$ (context C) level, I shift the analysis from one that examines e.g. the differences between country A and country B, to focusing only on changes within country A, and within country B. When I employ fixed effects at the agency level, I shift the analysis to within-agency differences in performance. Taken together, the analysis thus looks at within-agency differences in project performance as within-recipient country unpredictability rises or falls. I use within-agency, within-recipient variation to make between-agency claims; see e.g. β_3 in the model above, or the differing slopes of figures 1, 2, 3, and 4.[10]

This empirical strategy has clear parallels in qualitative research design. Leveraging within-case differences in case pairs, I make causal claims about the relationship of variables examined in each case to the between-case differences (or in Mahoney 2007's framing, cross-case comparisons) in project outcomes. Figure 5's case design schematic, then, has intuitive parallels with the fixed effects models that underlie figure 1, 2, and 3's marginal effects plots. Using within-agency variation in success to illustrate between-agency differences in the relationship between management practices, the case study pairs and large-N analysis jointly provide the basis for a mutually supportive mixed methods conclusion. Namely, both top-down management control and greater field control have their appropriate environments, with unpredictability conditioning which is the superior strategy. This mutually supportive mixed methods conclusion is facilitated by the common inferential logic on which the quantitative and qualitative empirics rely: leveraging differential within-unit variation to make between-unit comparisons.

Designing qualitative and quantitative strategies so they rely on the same logic of inference but are mutually supportive requires co-design of quantitative and qualitative strategies. My choice of agencies that vary widely in level of M (field agent control) in countries that vary widely in level of C (country unpredictability) is informed by both a parallel logic of inference as the quantitative empirics and a recognition of the limitations of those empirics. The choice to examine pairs of cases from different agencies using a "most similar" case strategy is driven by the inferential logic of leveraging "within" variation to make "between" comparisons.

The parallel strategies of causal inference in the qualitative and quantitative analysis are the analytic connective tissue that allows the limitations of the qualitative analysis to be supported by those of the quantitative analysis, and those of the quantitative analysis to be supported by the qualitative analysis. To build this type of connective tissue requires researchers to abstract away from the particular context of their qualitative and quantitative strategies, to interrogate the logic of inference in each case, and ensure the quantitative and qualitative empirical strategies are mutually supportive of one another.

Mutually supportive mixed methods do not in some way allow any given method to transcend their limitations. The case studies are a small-N sample, subject to the standard limitations of small-N samples; a scholar who does not believe qualitative cases can provide useful information is unlikely to be convinced to additionally update their priors in

[10] To be clear, this is not a new or novel method, but rather perfectly conventional in panel data econometric analysis.

response to the case data. The quantitative data is observational, subject to the standard limitations of quantitative analysis. A scholar who believes we can learn little from large-N observational data is unlikely to be swayed by the analysis. But for, I believe, the great majority of scholars who would agree that large-N econometric analysis, process tracing (Bennett and Checkel 2015; Blatter and Blume 2008; Hall 2013), and controlled case comparisons (Mahoney 2007; Slater and Ziblatt 2013) can all provide useful evidence for updating one's priors, mutually supportive mixed methods can help develop a stronger combined picture than any one method can provide in isolation. The strength of mutually supportive mixed methods lies in the diversity of each method's weaknesses; a reason for being dubious about one part of the mutually supportive mixed methods analysis often does not apply to other components of the analysis.

5. Mutually Supportive Mixed Methods in the Study of Public Agencies: Imagining Case Study Complements to the Econometric Analysis in JPART's Highly Cited Articles Archive

This paper has focused on but one of the many possible inferential challenges researchers face. There are a variety of situations in which mutually supportive mixed methods might be helpful in addressing challenges, and Seawright (2016) attempts to provide some structure to the variety of general problems researchers may face. One challenge that scholars of public agencies face is the slow-moving nature of important empirical features. This section aims to briefly explore how mutually supportive mixed methods might be more generally used in the face of these challenges.

In the O'Toole and Meier (2014) article from which the general model introduced in section 2 is adopted, the authors provide a "Public Management Context Matrix" (Table 1). This matrix lists important causal variables that could be classified as part of political, environmental, and/or internal context. In the first, political, category are the slow-moving at best and time-invariant at worst "Separation of Powers" and "Federalism". The same can be said of all four environmental variables (Complexity, Turbulence, Munificence, and Social capital) and all three internal agency variables (Goals, Centralization, and Professionalization). This suggests the centrality of slow-moving or time-invariant variables to the study of public agencies.

To further demonstrate the broad potential applicability of case studies as a mutually supportive method in the field, I turn to articles included in the "Highly Cited Articles" section of the Journal of Public Administration Research and Theory.[11] Described on the journal's website as "a selection of five highly cited articles from recent years", these articles presumably combine methodological rigor with a focus on substantive topics of interest to the core of the journal's scholarly community. This implicitly high standard of rigor in these standout articles makes this sample a "high bar" test for the relevance of mutually supportive mixed methods; nonetheless I believe that use of case studies as a mutually supportive method to econometric analysis could in principle have further strengthened the empirics of some of these papers.

[11] Available at https://academic.oup.com/jpart/pages/Impact_Factor. Articles as of mid-March 2018.

Table 1: The Potential for Case Studies to Complement Existing Econometric Analyses in JPART's Public "Highly Cited Articles" Archive

Paper Author-Year	Short Title	Study Population	Key Finding	Slow moving or time-invariant features of theoretic interest to the author(s)	Case Studies Possibly Helpful?
Ennser-Jednastik 2016	The Politicization of Regulatory Agencies	100 Regulatory Agencies Across 16 West European Countries	Individuals with ties to a government party are more likely to be appointed as formal agency independence increases.	Legal independence of regulatory agencies	Yes
Ingold & Leifeld 2016	Structural and Institutional Determinants of Influence Reputation	Four policy networks/cases on climate, flood, telecom, and toxic chemicals in Sweden and Germany	Institutional roles and structural positions in the network impact the perceived power of actors	Dependent variable (survey measure of perceived importance), structural independent variable (one-off survey) in main (Table 2) analysis	Yes
Favero, Meier, and O'Toole 2016	Goals, Trust, Participation, and Feedback	1,100 New York City Schools	Internal management has a causal impact on school delivery of educational outcomes	School internal management practices, sampled with time-invariant survey of teachers	Yes
Jilke, Van Ryzen, and Van de Walle 2016	Responses to a Decline in Marketized Public Services	Randomized survey experiment using MTurk	Choice overload holds for electricity markets – more choice reduces welfare-enhancing switches	No – individual's choice preferences are arguably slow moving/changing, but the evolution of these preferences is not the paper's subject	No
Marvel 2016	Unconscious Bias in Citizens' Evaluations of Public Sector Performance	Randomized survey experiment using MTurk	Individuals' evaluations of government performance is weighed down by their unconscious views of the public sector	No – individuals' unconscious views of the public sector are likely slow moving, but the study is concerned with the consequence of these views rather than their evolution or alteration	No

This paper has argued that the collinearity of slow moving or time-invariant features sometimes has consequences for estimating substantive quantities of interest in the field. It also has argued that this collinearity has the potential to make attribution of a given effect to particular mechanisms more difficult. Table 1 reviews each of the five papers and summarizes the extent to which each has a slow moving of time invariant feature, and whether mutually supportive mixed methods (MSMM in the table) as outlined in this paper might have further strengthened the central claims of the paper, had it been employed. As each of the five papers is a quantitative econometric exploration of the topic, this section focuses on the ability of qualitative data to complement the econometrics.

Of the five studies included in Table 1, two (Jilke, Van Ryzen, and Van de Walle 2016; Marvel 2016) involve randomized survey experiments conducted through the online platform M-Turk. While we could imagine case studies, or for that matter observational large-N data, that might address the central concerns of both papers, there is no obvious way for qualitative data to integrate with the existing experiments. The remaining three non-experimental papers (Ennser-Jednastik 2016; Ingold & Leifeld 2016; Favero et. al. 2016) all involve survey and/or observational administrative data. In all three cases, I believe it is possible that carefully designed case studies fitting the same logic of causal inquiry as the existing econometric analysis might have further bolstered the central claim of this paper. This is because in each of these three articles at least one key variable is slow-moving or time-invariant.

In Ennser-Jednastik 2016's exploration of the impact of legal independence on the politicization of appointments to regulatory agencies the key independent variable - the Gilardi measure of formal independence used by the author - is time-invariant.[12] The author controls for many potential confounds (agency resources, agency age, rule of law, etc.), as well as country-level fixed effects, and models the data using mixed-effects models. However collinearity between legal independence and unit-level fixed effects precludes the inclusion of agency-level fixed effects, which would absorb all possible fixed features of agencies other than their legal independence. As such, qualitative case studies of agencies at the extremes of the key independent variable (that is, those with very low and very high levels of legal independence) might have allowed additional exploration of the link between legal independence and appointment decisions. This might provide additional confidence that it was the legal independence of the agencies rather than potential agency-level confounds that were driving results, as well as strengthening the claim that the theorized mechanism (appointment of co-partisans to ensure independent agencies carry out the party's desired policy) was in fact operative.

The data Ingold & Leifeld 2016 use in their primary analysis of the determinants of perceived influence is largely drawn from a time-invariant survey (though an additional analysis of the one case for which it is available exploits longitudinal data from two survey waves). The use of a single survey wave for both the dependent and a key independent variable increases the potential threat to validity of omitted variable bias. Their analysis is already at the case level. Qualitative case data about the process of how actors came to be perceived as influential would provide additional empirical leverage on their question of central interest, allowing estimates of how changes in institutional roles and/or structural positions led to changes in perceived influence. Qualitative case data might also link perceived influence to actual impact (if e.g. a figure influential on their measure clearly

[12] By coincidence, this measure is developed in Gilardi 2008, the book discussed in this paper's introduction.

carried the day at the end of a contentious policy debate), a link of the causal chain for which the authors currently rely on theory.

Favero et. al. 2016 go some way to trying to address the inferential challenges posed by their time-invariant measure of internal management practice and its relationship with student success. The authors address potential mis-measurement of their key dependent variable (via a "halo effect" where management practice is rated better by staff at higher performing schools independent of its "true" level). They also control for the previous year's performance in a given school, lessening the risk that omitted variables such as e.g. more qualified teachers (the composition of a school's teachers being unlikely to change radically from one year to the next) are leading both to better student performance and better internal management practices.

Favero et al. conclude there is "little doubt… that the positive effects of management in the New York City school system are real." While the existing econometric strategy certainly has its strengths, the paper's focus on causal inference suggests that the authors might themselves agree that were time series data available for management practices and student achievement, a model with school fixed effects that looked at how changes in management practice affect changes in performance would have been even more convincing. Put another way, given available panel the authors may well have chosen to use it rather than employ the cross-sectional approach to management practice employed in the article. If indeed temporal data would have been preferred but was unavailable, carefully selected case studies (e.g. of schools where student performance radically improved or fell) might have provided an additional temporal dimension to the analysis. This examination might have provided qualitative evidence of first differences; of the marginal impact on education provision provided by marginal changes in management practice. Such data would have strengthened even further the paper's central claim that management practices play a causal role in student achievement.

In three of the five papers in JPART's "most cited" archive – and all the non-experimental papers in the archive - case study empirics might have further strengthened the authors' quantitative empirical strategy in making the core claims pursued by the authors. This is not to suggest these papers are, at present, insufficiently rigorous; indeed, their inclusion in the JPART archive suggests much the opposite. Nor does it suggest the three papers for case studies might have served as a mutually supportive mixed method are weaker than the two for which it is not the case. It does suggest the frequency and centrality of slow-moving and time-invariant variables of significance to the field, and the the broad potential for case studies to serve as a mutually supportive mixed method in the field.

Further empirical exploration of these papers using case studies would certainly be costly in both time and money; many, perhaps even most, researchers would conclude the econometric rigor of each of these five papers sufficient. The claim here is not that all work in the field *must* include mutually supportive mixed methods; it is that empirical work *can* benefit from careful case study design which complements econometric analysis. This claim applies, at least in principle, to some of JPART's strongest recent papers as determined by the field via citations and the journal itself via inclusion in a special curated section of highly cited articles made freely available to the public as exemplars.

6. Discussion and Conclusion

Rare is the empirical study – qualitative or quantitative – that faces no inferential challenges. Challenges are a function of the data-generating environment, the tractability of the objects of inquiry to direct observation and/or manipulation, and the causal density of the context (Woolcock 2013), amongst myriad other factors. Perhaps the only broad, universal statement that can be made about the nature of inferential challenges is that, unhelpfully for prescribing particular responses, the inferential challenges of a given empirical strategy are deeply contextual and often affect only a very narrow slice of the empirical work with which the researcher is familiar.

What, then, is a researcher to do? The first step, perhaps, is to step away from considering the limitations of the data to a more abstract consideration of the flavor of empirical challenge with which the researcher is grappling. One way of identifying the inferential weaknesses of one's own empirical strategy is to borrow from the toolkit of experimental economists. Experimental economists sometimes speak of the "God experiment"; the experiment that would allow perfect identification and causal inference, were it possible to manipulate all the relevant features of the environment.

The often unexamined intuition behind the God experiment is the notion that we can learn about our own strategy's limitations, and work to address them as best as possible in the design phase, by considering the "breach" between the perfect and the possible. Step one of a general mutually supportive mixed methods design strategy, then, might be to consider what the perfect data (quantitative and/or qualitative) to test a given theory might be. Step two would be to think about what data is available, or what qualitative and quantitative empirical investigation might make available.

In some instances, researchers will determine that the best empirical strategy to address these challenges is entirely quantitative, or entirely qualitative. In many instance, however, both qualitative and quantitative empirics will be illuminating. One general example of a kind of instance where mutually supportive mixed methods are likely to be of use is when a key quantity of interest is slow moving or time-invariant, leading to collinearity in econometric analysis and thus difficulty in determining quantities of interest. These quantities of interest may relate to substantive significance of findings or to mechanisms of action; or, as in the case of section 3, may impact interpretation of both substantive significance and mechanisms. This collinearity is a problem of particular relevance to the econometric study of public agencies, given the slow-moving or time-invariant nature of both many measurement strategies (e.g. surveys) and many important features of both agencies themselves and their broader contexts.

Where the researcher determines that quantitative and qualitative case study methods might be mutually helpful in examining the question and quantities of interest, it is critical to integrate these two parts of the empirical strategy. Integration involves carefully thinking through the logic of causal inference of both parts of the holistic empirical strategy: determining where the holes in each part of the strategy lie, and determining how the other part of the empirical strategy might best fill the hole. Rare will be the researcher who can come to the best strategy on first draft. Iteration and collaboration are critical in crafting a well-integrated, mutually supportive mixed methods strategy. One implication of this approach, then, is to invest greater time at the outset in considering a range of qualitative and quantitative empirical approaches than may be conventional in many corners of the academy.

A variety of disciplines have wrestled with the integration of qualitative and quantitative methods. From biostatistics, Rosenbaum & Silber (2001) argue for what might be thought of as the inverse of the Lieberman (2015) nested analysis approach – using "thick description" to improve matching strategies for quantitative analysis.[13] In development economics, Blattman et al (2016) argue for a form of what I would term mutually supportive mixed methods, integrating qualitative and quantitative data to develop better informed survey results in a single logic of inference to better validate survey responses. Michael Woolcock, an international development scholar with roots in sociology, argues that mutually supportive mixed methods are critical to addressing internal and external validity concerns in understanding the effect of development projects. (Woolcock 2013, 2018)

While there are particular types of methodological difficulties more common in the study of public agencies, it is not that public administration and management are unique in benefiting from mutually supportive mixed methods. It is rather that public administration and management are not an exception from the general case, and the study of public agencies could benefit from greater use of mutually supportive mixed method empirical strategies. We have no idea how many studies have never been attempted, or placed in the proverbial file drawer at the concept stage, because scholars found econometric challenges insurmountable. As public administration moves towards large-N quantitative research (e.g. Boyne et. al 2005; Lynn, Heinrich and Hill 2001; Walker, Boyne, and Brewer 2010) while also taking seriously management context (e.g. Andrews, Beynon and McDermott 2016; Bullock, Stritch and Rainey 2015; Meier et al 2015; O'Toole and Meier 2014) the research design issues explored in this paper will become more common.

To the extent that the study of public management and public administration becomes more econometrically rigorous, the need to address the kinds of causal inference problems on which this paper focuses will likely become more acute. Econometric analysis is sometimes conceived of as coincident with rigorous. But as more and more public management scholars proceed in this direction, not just the strength and sophistication but also the potential limitations of econometric analysis may become more apparent. Econometric analysis is an incredibly powerful tool, but it is nonetheless a tool in a toolkit,

[13] That is, both Rosenbaum & Silber and Lieberman conceive of the analysis as iterative, but the former uses qualitative small-N research to inform the design of quantitative large-n research, while the latter uses quantitative large-N research to inform the design of qualitative small-N research.

rather than the toolkit itself. If public administration and public management scholars respond to the field's methodological shifts by abandoning qualitative work, or using qualitative work merely as an additional layer to quantitative analysis, the field may lose access to a great deal of important work and empirical settings.

This does not mean the *only* function of case studies produced using a common inferential logic as the quantitative analysis it complements need be addressing the limitations of econometric analysis. The design of mutually supportive mixed methods strategies ought to be informed by the nature of each method's weaknesses in a given context and the logic of common inferential strategies but need not focus exclusively on that challenge. A case study can, for example, confirm systematic results *and* explore mechanisms, leveraging the greater nuance that may not be available in a large-N study but a case analysis (e.g. process tracing) approach may provide. Nor does it mean the only function of an econometric analysis produced using a common inferential logic as the qualitative case study analysis it complements need be addressing the limitations of the case analysis. An econometric analysis can confirm that the dynamics in play in case studies hold more generally *and* explore systematic tendencies that may not be apparent in a small-N qualitative study, leveraging the larger set of observations to draw more nuanced systematic conclusions. While there may well be trade-offs in practice, in theory filling holes does not preclude complementary analyses from also adding layers, or exploring mechanisms, or performing any other analytic function.

Scholars who have different underlying models of what causation is (ontologies) and how we might learn about what causes what (epistemologies) are not likely to embrace each other's methods. The approach described in this paper is of most help to qualitative and quantitative scholars who agree on epistemic and ontological matters, not a means of resolving tensions between scholars who do not agree.

That said, ontological and epistemological divides exist within communities of primarily qualitative and primarily quantitative scholars, not merely between them. Many scholars recognize the benefit of multiple kinds of empirical strategies even if they believe the methods they use to be of primary usefulness. Mutually supportive mixed methods hold the promise of allowing communities of scholars to focus on what unites them rather than what divides them. While not the primary intent of this paper, it is possible that mutually supportive mixed methods may be a component of "track II diplomacy" that in practice can help bridge ossified methodological stalemates.

In some ways, this paper's main thrust is simple: when the purpose of one method of empirical enquiry – e.g. qualitative case study research – is "filling holes" in a large-N observational analysis in addition to "adding layers" to that analysis, what cases a researcher chooses and how cases are constructed and analyzed depend on the nature of the hole to be filled. This choice requires going beyond simply considering the differences between "most similar" and "most different" strategies, or typical vs. extreme cases, etc. On case selection, it implies a particular kind of "prior stratification" (Seawright and Gerring 2008): a stratification endogenous to the particular econometric problem one faces. On case analysis, the choice implies an analytic approach that is determined in concert with a quantitative analytic approach. The same holds in the reverse, where quantitative large-N analysis is conceived of as complementary to qualitative case studies.

This two-way effect necessitates co-creation of a single integrated empirical strategy involving multiple empirical strands. Mixed methods have normally been imagined to coexist like layers of wallpaper, one building on top of the other. This article argues they ought coexist more like blending paint: the mixing needs to occur ex-ante, before the paint is applied to the wall. The mixing is likely also to be iterated, with the desired hue achieved by careful examination and revision of research design before the first brush stroke – quantitative or qualitative - is made.

Mutually supportive mixed methods may be helpful to scholars examining a wide range of questions. One place where the value of mutually supportive mixed methods can be seen in the context of studying public agencies is in the study of time-invariant or slow-moving features of agencies or broader contexts. Mutually supportive mixed methods can be useful both in exploring mechanisms and in estimating quantities of interest while controlling for unobserved confounds via fixed effects. Carefully designed case studies can provide the temporality that econometric analysis cannot in these contexts provided the case studies are carefully constructed so as to be mutually supportive with econometric analysis. In mixed methods work case studies need not merely be a helpful addition, providing colorful illustration to econometric work. Qualitative and quantitative analyses can be co-designed in concert, with the whole greater than the sum of the parts.

Bibliography

Adcock, Robert, and David Collier. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* (2001): 95 (3): 529-546.

Aghion, Philippe, and J Tirole. "Formal and real authority in organizations." *Journal of Political Economy* (1997): 105 (1): 1–29.

Andrews, Rhys, Malcolm J. Beynon, and Aoife M. McDermott. "Organizational Capability in the Public Sector: A Configurational Approach." *Journal of Public Administration Research and Theory* (2015): 26:2, 239-258.

Bennett, Andrew and Jeffrey Checkel. *Process Tracing*. Cambridge, UK: Cambridge University press, 2015.

Boyne, George A., Kenneth J. Meier, Laurence J. O'Toole, Jr., and Richard M. Walker. *Public Service Performance: Perspectives on Measurement and Management.* Cambridge, UK: Cambridge University Press, eds. 2006.

Blatter, J.K. and T. Blume. "Co-variation and Casual Process Tracing Revisited: Clarifying New Directions for Causal Inference and Generalization in Case Study Methodology." *Qualitative Methods,* 2008: 6: 29-34.

Blattman, Christopher, Julian Jamison, Tricia Koroknay-Palicz, Katherine Rodrigues, and Margaret Sheridan. "Measuring the Measurement Error: A Method to Qualitatively Validate Survey Data." *Journal of Development Economics* (2016), 99-112.

Brady, Henry E. and D. Collier. *Rethinking Social Inquiry.* New York: Rowman & Littlefield, 2004.

Bullock, Justin B., Justin M. Stritch, and Hal G. Rainey. "International Comparison of Public and Private Employees' Work Motives, Attitudes, and Perceived Rewards**."** *Public Administration Review,* (2015): 75:3, 479-489.

Center for Systemic Peace. State Fragility Index, 2014.

Ennser-Jedenastik, Laurenz. "The Politicization of Regulatory Agencies: Between Partisan Influence and Formal Independence." *Journal of Public Administration Research and Theory* (2016): 26:3, 507-518.

Favero, Nathan, Kenneth J. Meier, Laurence J. O'Toole, Jr. "Goals, Trust, Participation, and Feedback: Linking Internal Management with Performance Outcomes." *Journal of Public Administration Research and Theory* (2016): 26:2, 327-343.

George, A. and Andrew Bennett. *Case Studies and Theory Development in the Social Sciences.* Cambridge, MA: MIT Press, 2005.

Gerring, John and Lee Cojocaru. “Case Selection: A Diversity of Methods and Criteria.” 2015.

Gerring, John. *Case Study Research: Principles and Practices.* Cambridge: Cambridge University Press, 2007.

Gilardi, Fabrizio. *Delegation in the Regulatory State: Independent Regulatory Agencies in Action*. Northampton, MA: Edward Elgar, 2008.

Hall, Peter A. “Tracing the Progress of Process Tracing.” *European Political Science* (2013): 12:1, 20-30.

Honig, Dan. “When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation.” *International Organization*, Forthcoming Winter 2019.

Honig, Dan. *Navigation by Judgment: Why and When Top-Down Control of Foreign Aid Doesn’t Work.* New York, NY: Oxford University Press, 2018.

Humphreys, Macartan and Alan Jacobs. “Mixing Methods: A Bayesian Approach.” *American Political Science Review* (2015).

Ingold, Karin and Philip Leifeld. “Structural and Institutional Determinants of Influence Reputation: A Comparison of Collaborative and Adversarial Policy Networks in Decision Making and Implementation.” *Journal of Public Administration Research and Theory* (2016): 26:1, 1-18.

Jilke, Sebastian, Gregg G. Van Ryzin, and Steven Van de Walle. “Responses to Decline in Marketized Public Services: An Experimental Evaluation of Choice Overload.” *Journal of Public Administration Research and Theory* (2016): 26:3, 421-432.

King, G., R. Keohane, and S. Verba. *Designing social inquiry.* Princeton University Press, 1994.

Levy, Jack. “Case Studies: Types, Designs, and Logics of Inference.” *Conflict Management and Peace Science* (2008): 25:1, 1-18.

Lipsky, Michael. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service.* New York: Russell Sage Foundation, 1980.

Lynn, Laurence E., Jr., Carolyn J. Heinrich, and Carolyn J. Hill. *Improving Governance: A New Logic for Empirical Research.* Washington, DC: Georgetown University Press, 2001.

Lieberman, Evan. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* (2005): 99, no. 3: 435-52.

Mahoney, James. "After KKV." *World Politics* (2010): 62: 1, 120-47.

Mahoney, James. "Qualitative Methodology and Comparative Politics." *Comparative Political Studies* (2007): 40:2, 122-144.

Marvel, John D. "Unconscious Bias in Citizens' Evaluations of Public Sector Performance." *Journal of Public Administration Research and Theory* (2016): 26:1, 143-158.

Meier, Kenneth, Simon Camar Andersen, Laurence J. O'Toole Jr., Nathan Favero, and Soren C. Winter. "Taking Managerial Context Seriously: Public Management and Performance in U.S. and Denmark Schools." *International Public Management Journal* (2015): 18:1, 130-150.

Mill, John Stuart 1843. *A System of Logic Ratiocinative and Inductive – Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation.* Toronto, CA: University of Toronto Press, 1974.

Nielsen, Richard A. "Case Selection via Matching." *Sociological Methods and Research* (2016), 45:3, 569-597.

OECD. *Aid Effectiveness 2011: Progress in Implementing the Paris Declaration.* OECD Publishing, 2012.

O'Toole Jr., Laurence and Kenneth J. Meier. "Public Management, Context, and Performance: In Quest of a More General Theory." *Journal of Public Administration Research and Methods* (2014): 25:237-256.

Pavone, Tomasso. "The Method of Inductive Case Selection: How to Select Cases for Comparative Sequential Analysis." Forthcoming chapter in *The Case for Case Studies,* Woolcock, Widner, & Ortega-Nieto, eds. New York, NY: Cambridge University Press, 2017.

Polanyi, Michael. *The Tacit Dimension*. University of Chicago Press, 1966.

Plumper, Thomas and Vera E. Troeger. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis* (2007): 15: 124-139.

Rosenbaum, Paul R., and Jeffrey H. Silber. "Matching and Thick Description in an Observational Study of Mortality After Surgery." *Biostatistics* (2001), 2:2, 217-232.

Seawright, Jason. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools.* Cambridge University Press, 2016.

Seawright, Jason and John Gerring. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* (2008): 61:2, 294-308.

Stein, JC. "Information production and capital allocation: Decentralized versus hierarchical firms." *The Journal of Finance* (2002): 57 (5): 1891-1921.

Walker, Richard M., George A. Boyne, and Gene A. Brewer. *Public Management and Performance: Research Directions*. Cambridge, UK: Cambridge University Press, eds. 2010.

Williamson, Oliver E. "The Economics of Organization: The Transaction Cost Approach." *American Journal of Sociology* (1981): 87:3, 548-577.

Woolcock, Michael. 'Reasons for Using Mixed Methods in the Evaluation of Complex Projects', in Michiru Nagatsu and Attilia Ruzzene (eds.) *Philosophy and Interdisciplinary Social Science: A Dialogue* London: Bloomsbury Academic (2018, forthcoming)

Woolcock, Michael. "Using case studies to explore the external validity of 'complex' development interventions." *Evaluation* (2013): 19 (3): 229-248.