Advance Access publication May 11, 2005 *Political Analysis* (2005) 13:280–291
doi:10.1093/pan/mpi021

Qualitative Leverage and the Epistemology of Expert Opinion

Sanford C. Gordon and Alastair Smith
Department of Politics, New York University,
726 Broadway, 7th Floor, New York, NY 10003
e-mail: sanford.gordon@nyu.edu
e-mail: alastair.smith@nyu.edu

We discuss the motivation for integrating qualitative information in statistical models of complex, partially observable causal mechanisms and suggest ways to minimize the dangers posed by Braumoeller and Kirpichevsky. We stress the importance of linking our statistical estimators to underlying theories of the data generating process, qualitative or otherwise.

1 Introduction

The purpose of this article is to summarize and sharpen the focus of our original (2004) argument about how to integrate qualitative information into the statistical analysis of complex, partially observable causal mechanisms. Often, different social science theories posit different causal paths to the same outcome. While we may observe, for a large number of cases, whether that outcome occurred, we often have a difficult time assessing which causal path may have been operative in specific cases. Still, we may be interested in determining the effect of some covariates X on the probability that a particular causal mechanism is operative.

Partial observability models (Poirier 1980; Braumoeller 2003) present a potentially attractive solution but suffer from two drawbacks: problems associated with identification and labeling, and convergence difficulties in the presence of skewed outcome data. In our original piece, we argued that in the presence of these challenges, it may be possible for a subset of cases to rely on expert judgments concerning whether, for a particular case, a specific causal mechanism was operative. In so doing, we can overcome the potential limitations of partial observability models.

As in that article, we rely here on two premises: First, we should avoid discarding potentially useful information when it is available. Second, to the greatest extent possible, our statistical estimators should reflect models of data generating processes suggested by theory. We wish to emphasize that our principal aim is to think carefully about how qualitative information such as that derived via informed expert opinion can sharpen

Authors' note: We gratefully acknowledge the helpful comments of Bear Braumoeller, Yevgeniy Kirpichevsky, and the Columbia University Political Analysis Seminar participants.

quantitative techniques. We do not advocate a single, universally applicable estimator as the solution to all substantive problems.

The great virtue of the comment in this issue by Braumoeller and Kirpichevsky (hereinafter BK) on our original article is to suggest thinking carefully about the "qualitative data generating process" (QDGP) by which expert opinion is encoded in qualitative judgments. In our original article, we proposed two possible estimators, nicknamed "trubit" and "throbit" (like the reader, we too cringe). These reflect two distinct QDGPs, but we have no a priori attachment to them in particular. We can think of numerous different models of how experts render such judgments, each reflecting a different set of epistemic premises. Thus, while we find the admonitions of BK unassailable, we would argue that if the model does not fit the data generating process suggested by the theory, we need a different model.

2 Qualitative Information and Partially Observable Causal Mechanisms

2.1 *A Brief Restatement of the Problem*

Our original article was not the first to suggest a technique for integrating statistical and qualitative inference. Rosenbaum and Silber (2001), for example, suggest the value of employing the "thick description" (Geertz 1973) of patient histories to improve a matching analysis of postoperative mortality. Tarrow (2004) suggests the desirability of supplementing quantitative analysis with qualitative (or vice versa) to "triangulate" different measures on the same research problem.

We suggest a third means of integration, or what we refer to as "leverage." Different social science theories often suggest different causal paths to the same outcome. In our original article's running example, following Vreeland (2003), a country's leader(s) might pursue a loan from the International Monetary Fund (IMF) to overcome domestic opposition to political reforms *or* because of the country's dire financial situation. Let us refer to these causal mechanisms as $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. For observation i we observe whether a loan is made ($Y_i = 1$) or not ($Y_i = 0$). Unfortunately, we do not generally observe whether $\mathcal{M}_1$ or $\mathcal{M}_2$ or both were operative.

Extant econometric approaches aggregate across the two mechanisms to estimate the probability of an event occurring. For most observations in our data, this is typically the best we can do. However, for some observations within our data, we may have qualitative information that the event resulted from $\mathcal{M}_1$ rather than $\mathcal{M}_2$, or vice versa. Rather than ignoring this information, we can improve the quality of our estimators by integrating it into the estimation procedure.

We proceed by developing the partial observability probit and its generalizations. Having discussed their pitfalls, we then examine how qualitative knowledge concerning a limited number of cases helps address these problems. We then discuss how variations in our premises regarding how qualitative information is encoded suggest different data generating processes and how these different premises can be productively incorporated into an econometric model.

2.2 *The Poirier Model and Braumoeller's Generalization*

The probit model is the econometric standard for approaching dichotomous data. The latent variable formulation provides a convenient launching point: $Y_i = 1$ if $Y_i^* > 0$ and $Y_i = 0$ if $Y_i^* \leq 0$, with $Y_i^* = X_i\beta + \varepsilon_i$. Here, $Y_i \in \{0, 1\}$ is the dichotomous dependent variable for observation i, $Y_i^* \in \mathbb{R}$ is a continuous latent variable, X_i represents a vector of covariates,

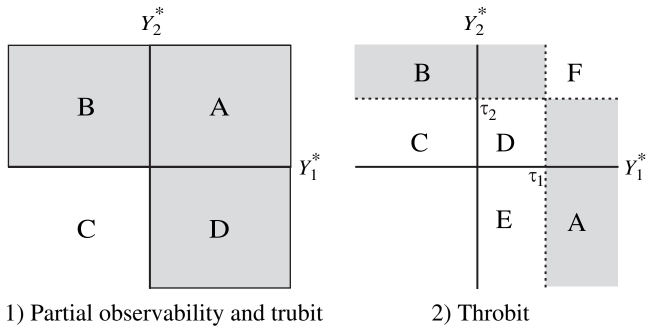

Fig. 1 Latent variables intuition: Trubit and throbit. If either or both latent variable exceeds zero, we observe the outcome $Y_i = 1$.

and ε_i is a $N(0, 1)$ stochastic error term. In this model the probability of observing a success, $Y_i = 1$, is $\Phi(X_i\beta)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Problems arise in the presence of multiple causal mechanisms when it is difficult to observe which is operative. Let Y_{ki} be an indicator variable reflecting whether mechanism $\mathcal{M}_k$ was operative. If we observed Y_{ki} directly, we could deploy the basic probit model to determine the effect of some covariates X on, for example, the probability that a leader pursued a loan for discretionary political reasons. The difficulty arises because all we generally observe is the presence of the outcome (a loan), and not the mechanism that caused it. Observing a loan, $Y_i = 1$, implies that *either* $Y_{1i} = 1$ *or* $Y_{2i} = 1$ (or both).

Poirier (1980) proceeds by positing the existence of multiple latent variables and derives the partially observable bivariate probit model as follows:

$$Y_i = 1 \text{ implies } Y_{1i}^* = X_{1i}\beta_1 + \varepsilon_{1i} > 0 \quad \textbf{or} \quad Y_{2i}^* = X_{2i}\beta_2 + \varepsilon_{2i} > 0;$$
$$Y_i = 0 \text{ implies } Y_{1i}^* = X_{1i}\beta_1 + \varepsilon_{1i} \leq 0 \quad \textbf{and} \quad Y_{2i}^* = X_{2i}\beta_2 + \varepsilon_{2i} \leq 0.$$

We show the derivation of the likelihood for this model in panel 1 of Fig. 1. This graphs the latent variables associated with mechanisms 1 and 2 on the horizontal and vertical axes. We have labeled the quadrants A though D. If we observe a success ($Y_i = 1$) then we know that either $Y_{1i}^* > 0$ or $Y_{2i}^* > 0$ (or both). This implies that the latent variables lie somewhere in quadrants B, A, or D. The contribution of observation i to the likelihood function is calculated by integrating the joint probability density of the latent variables over quadrants B, A, and D. Assuming the error terms associated with each latent equation are independent, i's contribution is $L_i = \Pr(Y_{1i}^* > 0) + \Pr(Y_{2i}^* > 0) - \Pr(Y_{1i}^* > 0) \times \Pr(Y_{2i}^* > 0) = 1 - \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$ if $Y_i = 1$. If observation i is a failure ($Y_i = 0$) then the latent variables lie in quadrant C; the likelihood for such an observation is $L_i = \Pr(Y_{1i}^* \leq 0) \Pr(Y_{2i}^* \leq 0) = \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$. The model can be estimated by standard maximum likelihood techniques (or alternatively by Bayesian MCMC).

As Braumoeller (2003) demonstrates, the Poirier model can be readily generalized such that arbitrarily complex logical combinations of partially observable causal mechanisms can be considered. He refers to this generalization as Boolean probit/logit. Unfortunately, identification conditions imply practical limits to the complexity of models that can be estimated. Loosely speaking, identification means that there is only one set of parameters (β_1 and β_2 in the current setting) that produce the same likelihood. To see the importance of identification, suppose that the same set of covariates that influence $\mathcal{M}_1$ also influence $\mathcal{M}_2$, but each covariate has a different effect on each mechanism's operation. In this setting, two sets of parameters generate an identical likelihood. Specifically, $L_i(Y_i = 1) = 1 - \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2) = 1 - \Phi(-X_{1i}\beta_2)\Phi(-X_{2i}\beta_1)$. Even if we could estimate this likelihood, we face a labeling problem in that we do not know whether the parameter estimates for the vector labeled β_1 should be associated with the first or second mechanism, and likewise for the vector labeled β_2. Obviously these ambiguities can make estimation and interpretation problematic.

Poirier derives minimal identification criteria for partial observability models. In particular there needs to be at least one covariate in X_1 that is not in X_2, or vice versa. Unfortunately in social science applications, exclusion restrictions—that is, finding a variable that strongly influences the process via one mechanism but not another—are often difficult to satisfy. As BK show in their response, when models are strongly identified and the number of 1s and 0s are balanced, then Boolean probit works well. However, when these criteria are not met, it can encounter difficulties.

2.3 *Improving Estimation Properties via Qualitative Leverage: A Class of Solutions*

Qualitative case studies can leverage quantitative analyses by providing anchors within the estimation to resolve identification and labeling issues. Returning to the IMF loan example, Vreeland (2003) argues that in 1975 Tanzania entered an IMF agreement because the government desperately needed a loan. In this case the government specifically sought to negotiate a loan that avoided the type of conditionality that would be consistent with a discretionary loan sought for domestic political reasons. As Vreeland (2003, p. 27) states, "The government did not want any conditions imposed, but it desperately needed foreign exchange." Such expert opinion allows us to dissect the probability space shown in Fig. 1 into finer regions. In this particular case, Vreeland uses his detailed knowledge of the case to make two claims. First, the loan was need-based. This implies that $Y^*_{2i} > 0$, i.e., the latent variables lie in either quadrants A or B of Fig. 1. Second, the government had no discretionary motive for the loan. This second claim further refines the probability space further: $Y^*_{1i} \leq 0$. This implies that the latent data lie within quadrant B.[1]

In 1983, Tanzania again sought an IMF loan. However, Vreeland states that "the government sought political support through the IMF conditionality to push through its preferred policies" (p. 29). This indicates that the discretionary mechanism was in effect, that is, $Y^*_{1i} > 0$. Importantly, however, Vreeland does not rule out that the government may have had need-based motives as well, as it was also desperately short of foreign currency reserves. Thus we remain ambivalent about whether Y^*_{2i} is positive or negative in this case.[2] In both these Tanzanian cases, Vreeland's detailed knowledge enables us to refine the likelihood function associated with each particular observation.

Under the Boolean probit model, the likelihood for each of the above observations is $L_i = 1 - \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$. In contrast, our qualitative information suggests that the likelihood function should be replaced by $L_i = \Phi(X_{2i}\beta_2)(1 - \Phi(X_{1i}\beta_1))$ in the first case and $L_i = \Phi(X_{1i}\beta_1)$ in the second case. We refer to a model that draws these distinctions as the truncated bivariate probit, or "trubit." Table 1 shows how expert opinion about causal mechanisms translates into the likelihood function for a variety of different scenarios. For example, the case labeled ABD corresponds to an absence of expert judgment as to the

[1]In BK's parlance, this corresponds to a case of trubit-c.
[2]In BK terms, this corresponds to an instance of trubit-u.

Table 1 The mapping between expert opinions and the likelihood function: Truncated bivariate probit

Case	Y_i	Y_{1i}^*	Y_{2i}^*	*Likelihood*	*Areas in Fig. 1*	*Description*
ABD	1	?	?	$1 - \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$	A+B+D	Vast majority of cases, mechanism is 1, 2, or both
D	1	+	−	$\Phi(X_{1i}\beta_1)(1 - \Phi(X_{2i}\beta_2))$	D	Cause is mechanism 1 and not mechanism 2
BA	1	?	+	$\Phi(X_{2i}\beta_2)$	A+B	Cause is mechanism 2
C	0	−	−	$\Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$	C	No success
B	1	−	+	$\Phi(X_{2i}\beta_2)(1 - \Phi(X_{1i}\beta_1))$	B	Cause is mechanism 2 and not mechanism 1
AD	1	+	?	$\Phi(X_{1i}\beta_1)$	A+D	Cause is mechanism 1
A	1	+	+	$\Phi(X_{1i}\beta_1)\Phi(X_{2i}\beta_2)$	A	Both mechanisms operative

cause of success. For observations that fall into this category, the likelihood assigned to each observation corresponds to that in the Boolean probit setting. However, when we have expert opinion as to the discernible cause then Table 1 tells us how to modify the likelihood function accordingly.

The ability to refine the likelihood function through qualitative evidence ameliorates the identification and labeling problems. Comparing case C to A, D, and DA is equivalent to running a standard probit model for cause 1. As such, there is no labeling or identification problem. Although qualitative case studies are often available only for a limited number of cases, they allow us to anchor the analysis. The labeling and identification obtained through these anchors prove to be particularly useful when we lack a justifiable exclusion restriction and when the dependent variable is unbalanced (many more 1s than 0s, or vice versa). Boolean probit requires us to treat all observations as if we are equally unknowledgeable about the causes of success. Note, however, that the trubit model collapses to Boolean probit when we have no informed opinion about any cases.

Our approach illustrates how information that is not available for every observation can still be employed in an estimation problem. The purpose of our original article was not to advocate any specific estimator, but rather to highlight this approach. We believe that one should never discard relevant information simply because it does not fit within the confines of a "canned" estimator. Rather, as far as possible, our estimator should be designed to account for the information available, given how our theory conceptualizes the problem at hand. The trubit estimator described here is just one illustration of how this might be done.

Threshold observability and the epistemology of expert opinion. The trubit model treats qualitative information with a broad brush, in that it provides no insight into the process by which experts form their opinions. If, however, the analyst has a theory of this process, it can be incorporated into the statistical model. The throbit, or threshold observability model, discussed in our original article represents a first attempt at modeling the formation of qualitative judgments concerning the operation of one or another causal mechanism. The epistemic premises on which throbit lies are crude but straightforward: Sometimes the operation of one causal mechanism is "glaringly obvious." If we conceive of the latent variable that underlies each causal mechanism as reflecting the incentives of a particular actor (as in the random utility formulation of the standard probit model), a discernible cause corresponds to an instance in which the incentive to take a particular action is overwhelmingly large *while the incentive corresponding to the alternative causal mechanism is not.* We conceive of this in terms of the latent variable associated with one cause surpassing a threshold value while that associated with another does not. Formally, an expert identifies mechanism k as a discernible cause iff $Y^*_{ki} > \tau_k$ **and** $Y^*_{\sim ki} \leq \tau_{\sim k}$ where $\tau_k > 0$ for all k.[3]

The throbit model further divides up the probability space of the latent variable, as shown in panel 2 of Fig. 1. The shaded area on the right, labeled A, corresponds to discernible causes by mechanism 1. The upper shaded area, labeled B, corresponds to discernible causes by mechanism 2. However, we assume the analyst is unable to discern between mechanism 1 and 2 in regions C, D, E, and F. In those cases, either the motive under each mechanism is not sufficiently strong (that is, $Y^*_{1i} \leq \tau_1$ and $Y^*_{21i} \leq \tau_2$ in regions C and D) or the incentive under both mechanisms is so strong that they cannot be

[3]BK suggest that modeling qualitative judgments in this fashion is a strict requirement, owing to the fact that a mechanism is discernible for a value of the latent variable just above the threshold and not just below. But this is true of all estimators that apply cutpoints, including standard probit and logit.

distinguished (that is, $Y^*_{1i} > \tau_1$ and $Y^*_{21i} > \tau_2$ in region E). Given expert judgments, the likelihood function is calculated by integrating over the appropriate areas of the probability space, as derived in our original article (see p. 240).

As we discuss below, throbit captures a particular qualitative data generating process. However, we are not wedded to it in particular. If the researcher believes qualitative assessments are generated by a different data generating process, the estimator should be derived with that process in mind rather than throbit.

3 The Objections of Braumoeller and Kirpichevsky

We now turn to the task of responding to and evaluating the specific objections of BK. A caveat is in order before proceeding: In our original piece, we argued that incorporating qualitative information could prove most helpful in situations in which the partial observability model encountered problems. Two such instances are when the outcome variable is highly skewed (a large proportion of successes or failures) and when the partial observability model is minimally identified.[4]

Most (though importantly, not all) simulations in BK's comment present ideal conditions for identification.[5] First, their simulations employ two, not one, exclusion restrictions. Finding a single variable believed *a priori* to condition one causal mechanism and not another is likely to be extremely difficult in social science applications; finding two is even more so. Second, their data consists of a 50/50 balance of successes and failures. Under these conditions, we can expect a well-behaved, globally log-concave likelihood function.

When competing on turf most favorable to the partial observability/Boolean model, we should not be surprised that given clear violations of the assumptions underlying the models we suggest, the estimator that discards potentially useful information rather than incorporating it incorrectly does indeed outperform the ones we offered. What, then, are we to make of these violations? As we will argue in each case, although BK should be applauded for pointing out these potential pitfalls, we believe they can be accommodated by fairly straightforward changes to the underlying estimator rather than by discarding the auxiliary qualitative information.

3.1 *Underreporting*

BK argue that our estimators implicitly assume that qualitative information has been gathered for each case, when in fact for the vast majority of cases, we would have no such information. They argue that ambiguity that results from the absence of qualitative information is fundamentally different from ambiguity that results from the failure of a considered expert judgment to produce discernment. This point is absolutely correct and is something about which we were insufficiently clear in our original article. As we note in Section 2.3, it is straightforward to accommodate underreporting in our estimators. When $Y_i = 1$ but there is no qualitative opinion (case ABD in Table 1), the contribution to the likelihood is $1 - \Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$. This applies to both the trubit and throbit estimators. In fact, *any* partial observability model that seeks to leverage qualitative information should collapse to the naive partial observability/Boolean likelihood in the absence of such auxiliary information.

[4]We should also note that given a sufficient number of discernible successes, the estimators we derived would be identified under conditions in which Boolean probit would not be—namely, when the same vector of covariates conditions both causal paths with no exclusion restrictions.

[5]They relax some of these in Section 5.

3.2 *Qualitative Experts Make Mistakes*

BK maintain that errors of expert judgment can create asymptotic bias in coefficient estimates. The question, we believe, is whether we can incorporate our concern about the possibility of misjudgment into our estimator. We believe this should be done via prior beliefs on the validity of expert opinion. It is easiest to demonstrate how in the context of trubit. Suppose, for example, we lack full confidence in our expert. The expert then identifies an observation as corresponding to case *D* in Table 1. Instead of representing that observation's contribution to the likelihood as the probability associated with quadrant D, we could instead weight the probabilities associated with quadrants A, B, and D to reflect our confidence in the expert, assigning the greatest weight to *D*. We could go one step further, manipulating the weights to determine the extent to which our estimates are robust to changes in our beliefs about the accuracy of qualitative judgments. We suggested a similar approach to encoding scholarly disagreement in the likelihood on page 248 of our original article.

In their simulation of expert misjudgment, BK rely on a coding scheme that we find implausible: The coding of discernible causes is *reversed* for a fraction of cases for which qualitative judgments have been offered. In practice, this means an expert may erroneously offer a conclusion that "causal mechanism 1 was certainly operative and causal mechanism 2 certainly was not," when precisely the opposite was true. The more likely scenario, in which an expert suggests that only one mechanism is operative when in fact both are, is ruled out, as is the scenario in which the expert equivocates when a unique causal path ought to have been discerned. In other words, the process of expert error they consider is one in which only *spectacular* errors of judgment are possible. This would be akin to a situation in which an expert asked to distinguish among black, white, and gray, when presented with black, responded "black" or "white" but never "gray."

3.3 *A Different QDGP May Have Generated the Qualitative Information*

BK suggest that if the cognitive process generating qualitative assessments differs from the one implied by the statistical model, that model will fail to recover consistent estimates of the underlying population parameters. This, of course, is correct and is especially relevant in their assessment of the trubit estimator. The simulations BK report employ data created by the throbit data generating process to test trubit. However, the throbit QDGP places much more structure on the knowledge of the cause than is assumed by the trubit structure. It is therefore not surprising that it biases coefficients. For an intuition as to the origin of this induced bias, consider a comparison of observations of failures ($Y_i = 0$) and observations assigned to mechanism 1 ($Y_i = 1$, $Y_{i1} > 1$). For this subset of observations, the trubit model effectively estimates $\mathcal{M}_1$ via a standard probit formulation. The likelihood of failure is given as $\Phi(-X_{1i}\beta_1)$ and the likelihood of success is estimated assuming that $\Phi(X_{1i}\beta_1)$. Yet given the assignment of causation under the throbit QDGP the likelihood of observing $Y_{1i} = 1$ is $\Phi(X_{1i}\beta_1 - \tau_1)$. As is to be expected, this leads to a bias in the parameter estimates of $\mathcal{M}_1$.

To examine the performance of trubit relative to Boolean probit under conditions consistent with its derivation, we performed a series of Monte Carlo studies. In each of 1000 iterations, we generated 1000 observations and modeled case studies as having been performed for around 10% of the successes ($Y_i = 1$). In particular, with a 10% chance we assigned cause to mechanism 1 if $Y^*_{1i} > 0$ (case AD in Table 1) and we assigned cause to mechanism 2 with 10% probability if $Y^*_{2i} > 0$ (case BA in Table 1). In the remaining successes we assumed no knowledge of cause (equivalent to case ABD in Table 1).

We then used BK's R version of Boolean probit and trubit-u to compare the efficiency of these estimators. Although both estimators performed well in recovering reliable estimates of the parameter values, trubit consistently produced smaller mean squared error. When the sample contained a relatively balanced sample of successes and failures, the efficiency of trubit relative to Boolean probit (measured in terms of the ratio of root mean squared errors) was between 1.16 and 2.10. Trubit's relative efficiency becomes even greater as the sample becomes skewed. For instance, when only 10% of the observations are successes, trubit's relative efficiency jumps to between 2.7 and 23.7. As discussed by BK, much of this improvement comes from avoiding the occasional "wild" Boolean probit estimates.

BK go on to suggest two alternative qualitative data generating processes to the ones we suggest in our original piece. For brevity, we focus on the first one. Let Q_{ki} be a dummy variable indicating qualitative evidence that mechanism k was operative. Then:

$$Q_{1i} = \begin{cases} 1 & \text{iff } \Phi(X_{1i}\beta_1) - \Phi(X_{2i}\beta_2) > \kappa_1 \text{ and } Y_i = 1 \\ 0 & \text{otherwise.} \end{cases}$$
$$Q_{1i} = \begin{cases} 1 & \text{iff } \Phi(X_{2i}\beta_2) - \Phi(X_{1i}\beta_1) > \kappa_2 \text{ and } Y_i = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

with κ_1, $\kappa_2 \sim U(0, 1)$. We find the spirit of this alternative to be quite appealing in its representation of *relative* discernibility. However, we believe Eq. (1) is somewhat problematic. Note that this formalization effectively requires that the qualitative expert has precisely the same cognitive model of the relationship between the X variables and the probability that each mechanism is operative that we, as quantitative researchers, would estimate. Also, conditional on κ_1 and κ_2, this setup is deterministic. However, one clear advantage of a qualitative researcher is his or her "deep" knowledge of the *unmodeled* (i.e., stochastic) determinants of each causal mechanism. These determinants are relegated to the residual in our statistical model, but the formalization in Eq. (1) presumes that the expert knows nothing about them. A better formalization of the spirit of this "relative discernability" model suggested by BK, which avoids these pitfalls, is as follows:

$$Q_{1i} = \begin{cases} 1 & \text{iff } Y^*_{1i} - Y^*_{2i} > \kappa_1 \text{ and } Y_i = 1 \\ 0 & \text{otherwise.} \end{cases}$$
$$Q_{1i} = \begin{cases} 1 & \text{iff } Y^*_{2i} - Y^*_{1i} > \kappa_2 \text{ and } Y_i = 1 \\ 0 & \text{otherwise,} \end{cases}$$

with $\kappa_1, \kappa_2 \in \mathbb{R}^+$. We assume that each threshold κ is a parameter to be estimated, although a more complicated model would treat each as a function of expert-specific characteristics (age, theoretical approach, etc.) and/or random effects. To form the likelihood, note that the probability of a success in which mechanism 1 is discernible can be expressed as

$$\begin{aligned} &\Pr(Y^*_{1i} - Y^*_{2i} > \kappa_1 \cap Y^*_{1i} > 0) \\ &\quad = \Pr(X_{1i}\beta_1 + \varepsilon_{1i} - X_{2i}\beta_2 - \varepsilon_{2i} > \kappa_1 \cap X_1\beta_1 + \varepsilon_1 > 0) \\ &\quad = \Pr(\varepsilon_{2i} - \varepsilon_{1i} < X_{1i}\beta_1 - X_{2i}\beta_2 - \kappa_1 \cap -\varepsilon_{1i} < X_{1i}\beta_1). \end{aligned} \tag{2}$$

Now consider two features of this expression: First, given error independence, $\varepsilon_2 - \varepsilon_1 \sim N(0, 2)$; and second, $\text{corr}(-\varepsilon_1, \varepsilon_2 - \varepsilon_1) = 1/\sqrt{2}$. Hence, the last line in Eq. (2) can be expressed in terms of the bivariate normal cdf with known $\rho = 1/\sqrt{2}$:

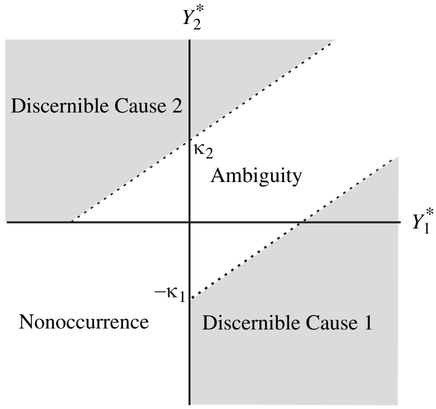

Fig. 2 Latent variables intuition for our modified version of Braumoeller and Kirpichevsky's relative discernibility model.

$$\Phi_2\left(\frac{X_{1i}\beta_1 - X_{2i}\beta_2 - \kappa_1}{\sqrt{2}}, X_{1i}\beta_1, \frac{1}{\sqrt{2}}\right).$$

A similar probability can be derived for a success via discernible mechanism 2. Given that the probability of a failure is still $\Phi(-X_{1i}\beta_1)\Phi(-X_{2i}\beta_2)$, the probability of an ambiguous success is simply the complement of the probability of failure or success associated with one or the other discernible mechanisms. Figure 2 displays the latent variable intuition of this particular estimator. In contrast with throbit, the cutting lines that distinguish ambiguous from discernible causality are parallel and at 45°.

Further, as noted in Section 3.1, in the presence of underreporting, we can simply employ the partial observability/Boolean probability of a success, $1 - \Phi(-X_1\beta_1)\Phi(-X_2\beta_2)$. And, as noted in Section 3.2, we can incorporate uncertainty about qualitative assessments via a probabilistic weighting of the Boolean success probability with the discernible success probabilities.

To demonstrate the efficacy of this model, which we label RD-throbit (*R*elative *D*iscernibility throbit), we ran a quick-and-dirty Monte Carlo simulation. We created conditions that would be maximally favorable to partial observability/Boolean probit and minimally favorable to this alternative estimator. To do this, we set the population parameters to assure 1) a 50/50 balance of successes and failures in the dependent variable; 2) a large sample size ($N = 2000$); 3) high values of κ_1 and κ_2 to make discernibility relatively rare (when all βs are set to one as in the original simulations, setting $\kappa_1 = \kappa_2 = 2.5$ implies that roughly 40% of successes would be discernible if subjected to qualitative analysis); 4) two exclusion restrictions (one per equation); and 5) a large degree of underreporting (90%, which implies that only about 3% of successes were, having been "examined," determined to be discernibly caused by mechanism 1 or 2).

Under these conditions, we should expect that 1) both Boolean probit and RD-throbit recover parameter estimates reliably, and 2) RD-throbit permit a small improvement in efficiency. Table 2 confirms these predictions: The mean parameter estimates over 1000 iterations are very close to their population values, but RD-throbit is between 5 and 15% more efficient (in terms of root mean squared error) than partial observability/Boolean probit. Note also that in this simulation, our estimation procedure successfully recovered unbiased estimates of RD-throbit's ancillary parameters.

Table 2 Monte Carlo comparison of partial observability/Boolean probit and "relative discernibility" throbit estimates

	β_1	β_2	β_3	β_4	β_5	β_6	$\ln(\kappa_1)$	$\ln(\kappa_2)$
Truth	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.92
Mean partial observability/Boolean	1.02	1.01	1.02	1.01	1.01	1.01	—	—
Mean RD-throbit	1.01	1.01	1.02	1.01	1.01	1.01	0.92	0.92
Ratio of RMSEs	1.07	1.15	1.14	1.05	1.13	1.11	—	—

We think that it would be premature to declare any sort of "victory" for this new model. Instead, we employ it as an example of how the analyst can move from a theory of the cognitive process by which experts encode their judgments into a probabilistic model. Note, however, that with several different models of that cognitive process, one can examine the extent to which parameter estimates are sensitive to different assumptions about how these judgments are formed. In applications, we might well conclude that the estimates derived via partial observability/Boolean probit are confirmed through the incorporation of qualitative assessments.

3.4 *Convergence*

Finally, BK argue that convergence difficulties of Boolean probit in our simulations were likely an artifact of our reliance on a quasi-Newton optimization algorithm (BFGS) in GAUSS. The R implementation of Boolean probit begins with traditional optimization techniques. It then switches to Sekhon and Mebane's (1998) genetic optimization algorithm (**genoud**) in difficult cases. We concede that the choice of optimizer may have affected convergence but feel this is somewhat beside the point. Given that the likelihood function of a fully identified version of the Poirier model or any of its descendents (Boolean probit, trubit, throbit, RD-throbit) ought to globally log-concave, it seems to us that relying on a particular statistical package or optimizer should not privilege any estimator a priori. Further, given the tenuous identification properties of partial observability models in practice, it may not be clear how to interpret the coefficients that a more sophisticated algorithm tells us reflect a global maximum once the simpler algorithm fails. This is likely to be even truer in real-world applications than in simulations.

4 Conclusion

In their comment, Braumoeller and Kirpichevsky perform two valuable services. First, they suggest a number of pitfalls of which to be cognizant when seeking to leverage qualitative judgments in the statistical analysis of complex causes. We believe that the problem of underreporting and the possibility of expert error can be incorporated into the likelihood functions we proposed in our original article without much difficulty.

Second, they have forced us to sharpen our understanding of the process by which qualitative judgments are formed. As political methodologists, each of us desires that our statistical estimators accurately capture the stochastic process that generated our data. Once we rely on qualitative judgments, this task becomes more complicated, as it requires that we expand our modeling endeavor to encompass the epistemology of expert assessments.

References

Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11:209–233.

Geertz, Clifford. 1973. "Thick Description: Toward an Interpretive Theory of Culture." *The Interpretation of Cultures*, chap. 1. New York: Basic Books.

Gordon, Sanford C., and Alastair Smith. 2004. "Quantitative Knowledge through Qualitative Leverage: Augmenting the Statistical Analysis of Complex Causes." *Political Analysis* 12:233–255.

Poirier, Dale J. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics* 12:209–217.

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

Sekhon, Jasjeet S., and Walter R. Mebane. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models." *Political Analysis* 7:189–213.

Tarrow, Sidney. 2004. "Bridging the Quantitative-Qualitative Divide." In "Rethinking Social Inquiry: Diverse Tools, Shared Standards," eds. Henry E. Brady and David Collier, Lanham, MD: Rowman and Littlefield, pp. 171–179.

Vreeland, James. 2003. *The IMF and Economic Development.* New York: Cambridge University Press.