Political Analysis (2004) 12:233–255
doi:10.1093/pan/mph017

Quantitative Leverage Through Qualitative Knowledge: Augmenting the Statistical Analysis of Complex Causes

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

Social scientific theories frequently posit that multiple causal mechanisms may produce the same outcome. Unfortunately, it is not always possible to observe which mechanism was responsible. For example, IMF scholars conjecture that nations enter IMF agreements both out of economic need and for discretionary domestic political reasons. Typically, though, all we observe is the fact of agreement, not its cause. Partial observability probit models (Poirier 1980, *Journal of Econometrics* 12:209–217; Braumoeller 2003, *Political Analysis* 11:209–233) provide one method for the statistical analysis of such phenomena. Unfortunately, they are often plagued by identification and labeling difficulties. Sometimes, however, qualitative studies of particular cases enlighten us about causes when quantitative studies cannot. We propose exploiting this information to lend additional structure to the partial observability approach. Monte Carlo simulation reveals that by anchoring "discernible" causes for a handful of cases about which we possess qualitative information, we obtain greater efficiency. More important, our method proves reliable at recovering unbiased parameter estimates when the partial observability model fails. The paper concludes with an analysis of the determinants of IMF agreements.

> A member shall be entitled to purchase the currencies of other members from the Fund . . .[provided] the member represents that it has a need to make the purchase because of its balance of payments or its reserve position or developments in its reserves.
>
> —International Monetary Fund Articles of Agreement

> [IMF] negotiations sometimes enable government leaders to do what they privately wish to do, but are powerless to do domestically.
>
> —Robert Putnam (1988, p. 457)

1 Introduction

The charter of the International Monetary Fund suggests that nations will enter IMF loan agreements out of need (Reichmann and Stillson 1978; Connors 1979; Gylfason 1987; Bird 1996). At the same time, scholars have recently argued that leaders often enter IMF programs for domestic political reasons. For example, leaders can use the strict conditions

Authors' note: We gratefully acknowledge the helpful comments of Bear Braumoeller, William Roberts Clark, Michael Gilligan, Gary King, Adam Przeworski, and James Vreeland.

attached to loans to overcome domestic opposition to reforms (Spaventa 1983; Remmer 1986; Vaubel 1986; Putnam 1988; Edwards and Santaella 1993; Bjork 1995; Dixit 1996; Vreeland 2000, 2003; Drazen 2001; Ramcharan 2003). In a broad sense, we can say that nations enter IMF loans *either* for discretionary (i.e., domestic political) reasons *or* out of economic need. The outcome (a loan) may be the same, but different causal paths lead there in different cases.

Statements such as "Either A_1 or A_2 cause B" reflect *causal complexity*, of which Braumoeller (2003) provides numerous examples from political science. When operational measures of A_1 and A_2 are easily observed, such propositions may be readily tested in the context of a "classic" linear-in-variables regression model, perhaps augmented with the judicious use of interaction effects. In many cases, however, sufficient conditions for an event's occurrence are difficult to observe. It may be difficult to tell which causal path produced the IMF loan, because leaders do not typically announce their intention to trade domestic sovereignty for a commitment to favored policies. Likewise, international relations scholars posit numerous sufficient conditions for war; in many cases, though, we cannot be certain whether the pursuit of territory (Huth 1996), miscalculation (Jervis 1976), diversion (Levy 1989; Downs and Rocke 1994, 1995; Smith 1996), or any other condition produced the result. To cite an example from the study of regulation, when a safety inspector visits a plant and finds no violations, we cannot tell whether compliance or nondetection produced this outcome (Feinstein 1989, 1990).

While underlying conditions may be difficult to observe, there may be independent variables that we can observe and that we believe increase or decrease the probability that one or the other underlying conditions is operative. As Braumoeller notes, however, simply lumping together all the independent variables associated with both mechanisms into a single linear-in-variables regression model does not properly capture the data-generating process implied by the theory, and it does not allow critical distinctions to be drawn among competing causes. Instead of this reduced-form approach, he advocates a family of models called Boolean probit, a generalization of Poirier's (1980) partially observable bivariate probit model. Unfortunately, neither author provides evidence of the estimators' performance in finite samples (via simulation, for example), which are plagued by weak identification and labeling problems. Both sets of problems stem from ambiguity in assigning the marginal impact of a particular independent variable to one or another cause—a difficulty that becomes particularly worrisome when we anticipate competing unobservable causes that each depend on the same exogenous variable. In such a setting, the estimates are identified only through functional form.

To improve our ability to estimate models of causal complexity, we propose using qualitative information to impose additional structure on the estimation process. Consider a model in which two difficult-to-observe causal paths could lead to an outcome of interest. Detailed qualitative studies can sometimes determine whether one or the other was responsible for a particular event. While for the majority of the sample the information as to the cause of an event is censored, in those cases in which the cause is known it can be used to anchor the statistical estimation and eliminate the ambiguity that plagues the partial observability approach.

In this paper, we first introduce a notation and terminology for causal complexity sufficiently general to encompass the various archetypes discussed by Braumoeller and others, and consider how we can translate a logical statement concerning the operation of necessary and sufficient antecedents into a statistical model. We then turn to situations where which of the relevant competing causal mechanisms is responsible for an event is generally censored. After reviewing the partial observability framework, we introduce two approaches for anchoring our statistical analysis with qualitative information. The first, a maximum likelihood approach called *threshold observability bivariate probit* (*throbit*), considers a situation in which the data-generating process occasionally produces information on discernible causes when one of two latent variables takes on a particularly large value. In other words, when a case is a particularly notable example of one mechanism, our ability to discern it provides exploitable information. A series of Monte Carlo simulations reveals several advantages of throbit over Boolean probit. The second approach, which we implement using Bayesian methods, is called *truncated bivariate probit* (*trubit*). It imposes less structure than throbit, although it exploits the same kind of case information. We conclude with an examination of a competing cause model of IMF agreements, comparing the Boolean probit, throbit, and trubit estimates.

2 Causal Complexity

We begin with some simple terminology and notation to span different causal statements of arbitrary complexity. We do so for two reasons: First, we wish to provide a framework with which to analyze such statements, redressing a deficiency in discussions of causality in political science; second, because even though our exposition focuses on one particular complex causal statement, the methods we discuss generalize easily to more difficult settings.[1]

Definition 1: $\mathcal{M}_j$ is the jth *causal mechanism* of B (for $j = 1, 2, \dots J$) if it consists of one or more conditions A_{jn_j}, jointly sufficient for B, where $n_j = 1, 2, \dots N_j$ indexes the nth condition necessary for mechanism j. Formally, $\mathcal{M}_j \equiv \cap_{n_j=1}^{N_j} A_{jn_j} \Rightarrow B$.

If J enumerates an exhaustive list of causal mechanisms, then $\cup_{j=1}^{J} \mathcal{M}_j \Leftrightarrow B$. We then say, following Mackie (1965), that A "causes" B if the former is an insufficient but necessary part of a condition that is itself unnecessary but sufficient for the result (an INUS condition).

"Either A_1 or A_2 is necessary to cause B" ($A_1 \cup A_2 \Leftrightarrow B$) suggests a causal statement with $J = 2$ (two causal mechanisms) and $N_j = 1$ for $j = 1, 2$. The statement "A_1 and A_2 are both necessary and jointly sufficient for B" ($A_1 \cap A_2 \Leftrightarrow B$) suggests a single causal mechanism for the result B. However, note that we could express this case as a pair of mechanisms, each of which causes B's complement: $\sim A_1 \cup \sim A_2 \Leftrightarrow \sim B$. In other words, if two conditions are necessary and jointly sufficient for an outcome, then the complement of either is by itself a sufficient condition for the outcome's nonoccurrence.[2]

To move from the formal notation to a statistical model requires (1) introducing uncertainty and (2) properly operationalizing causal mechanisms in terms of exogenous independent variables. With respect to the first requirement, as practitioners we are ordinarily content to make probabilistic rather than deterministic causal statements. With respect to the second, we define our independent variables X as a mapping from the set of relevant antecedent events and their complements to a subset of K-dimensional Euclidean space $\mathbb{R}^K$, where K is the number of independent variables necessary to properly operationalize the posited causal mechanisms.

---

[1]For a comprehensive treatment of causal inference, see Pearl (2000).

[2]The proof is trivial: By *Modus Tollens*, we have $\sim (A_1 \cap A_2) \Leftrightarrow \sim B$; by DeMorgan's Theorem, the left side of this statement may be expressed as $\sim A_1 \cup \sim A_2$.

Often we are concerned with dichotomous outcomes. Let Y be a dichotomous random variable, with $Y_i = 1$ in the event of a success and $Y_i = 0$ in the event of a failure. Given the above two requirements, a statistical model will take the form

$$\Pr(Y_i = 1) = F(X_i), \tag{1}$$

where the function $F(\cdot)$ takes on values constrained between zero and one.

2.1 *(Nearly) Full Observability*

A common technique used to analyze dichotomous-event data is the linear-in-variables probit model.[3] Assume the existence of a latent, unobserved continuous variable Y_i^*, whose value is a function of a vector of independent variables X_i, a parameter vector β, and a $N(0, 1)$ error term ε_i: $Y_i^* = X_i\beta + \varepsilon_i$. The event $Y_i^* > 0$ is necessary and sufficient to produce an outcome $Y_i = 1$. Hence, conditional on X_i the probability that $Y_i = 1$ is $\Pr(X_i\beta + \varepsilon_i > 0) = \Phi(X_i\beta)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. The probability that $Y_i = 0$ is the complementary probability, $1 - \Phi(X_i\beta)$. Estimation of β proceeds via maximum likelihood (King 1989).

Due to the continuity of ε, there are an infinite number of combinations of Xs and εs each jointly sufficient to make Y_i^* exceed zero, in turn resulting in the outcome $Y_i = 1$. This is a trivial point, but it illuminates the fact that it is often both possible and desirable to model complex causal statements in the context of this simple model. To wit, given the statement, "Either disastrous economic conditions or shared partisan identification with a challenger increases the probability that a citizen will vote against an incumbent president," there is little reason (assuming our measures are adequate) to abandon the old workhorse.

Likewise, causal mechanisms involving a confluence of factors can often easily be accommodated in the linear-in-variables context through the introduction of multiplicative interaction terms (Clark and Gilligan 2004). For example, consider the statement, "A_1 makes B more likely, but only in the presence of A_2." We would reject the null hypothesis corresponding to this prediction if the coefficient on the interaction between our operational measures corresponding to A_1 and A_2 was statistically significant and the coefficient on the independent effect of our measure for A_1 was not.[4]

2.2 *Partial Observability*

While the above models are simple to estimate and interpret, they rely on the requirement that the presence or absence of various antecedent conditions can in principle be operationalized by a vector of exogenous variables. Unfortunately, many theoretical models in political science reflect causal mechanisms in which conditions are unobservable. For instance, in the IMF example, we have posited that either domestic political considerations or genuine economic need induces leaders to pursue IMF loans. However, we cannot under normal circumstances observe whether one, the other, or both were operative. More generally, consider the following model:

$$\Pr(Y_i = 1) = \Pr(Y_{1i} = 1 \text{ or } Y_{2i} = 1), \tag{2}$$

[3]Little or nothing is gained from additional discussion of logit, so we forgo it.
[4]This assumes a suitable coding of independent variables.

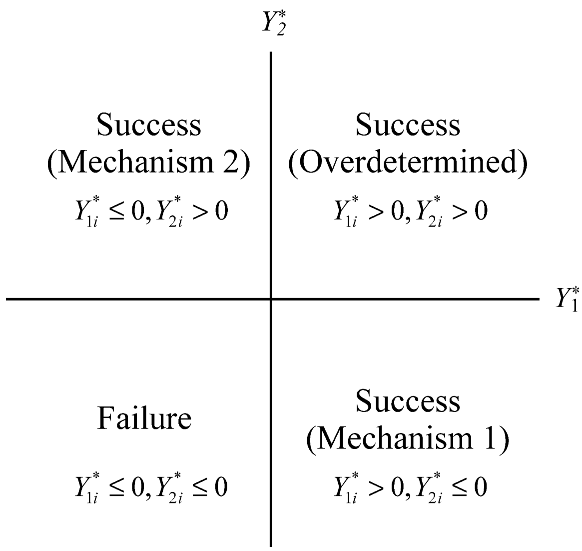

Fig. 1 Latent variables intuition for the partial observability bivariate probit. If either or both latent variables exceed zero, we observe the outcome $Y_i = 1$.

where neither Y_{1i} nor Y_{2i} is directly observed. Often Y_{1i} and Y_{2i} reflect choices made by strategic actors, so it is improper to consider them exogenous. Our theory may suggest, though, that the probability that each assumes a value of one is a function of some antecedents that can themselves be operationalized by vectors of exogenous variables.

To proceed, we posit the existence of two latent variables: Y_{1i}^* and Y_{2i}^*, with

$$\begin{aligned} Y_{1i}^* &= X_{1i}\beta_1 + \varepsilon_{1i} \\ Y_{2i}^* &= X_{2i}\beta_2 + \varepsilon_{2i}. \end{aligned} \tag{3}$$

We assume that ε_1 and ε_2 are drawn from a standard bivariate normal distribution with error correlation ρ. We observe $Y_i = 1$ if $\max\{Y_{1i}^*, Y_{2i}^*\} > 0$. In other words, a success occurs if and only if either latent variable takes on a positive value. Using our earlier notation, $\mathcal{M}_1$ is the mechanism corresponding to $Y_{1i}^* > 0$, and $\mathcal{M}_2$ is the mechanism corresponding to $Y_{2i}^* > 0$. Figure 1 depicts this model graphically. Successes are generated exclusively via the first mechanism in the lower right quadrant and exclusively via the second in the upper left. In the upper right quadrant, success is overdetermined. Only in the lower left quadrant, when neither latent variable exceeds zero, do we observe failure.

A method to estimate β_1 and β_2 is the *bivariate probit with partial observability* (Poirier 1980; Abowd and Farber 1982; Przeworksi and Vreeland 2002), which is a special case of a more general set of models referred to by Braumoeller (2003) as *Boolean probit*.[5] The Boolean probit methodology proceeds via maximum likelihood. Note that

$$\begin{aligned} \Pr(Y_i = 0) &= \Pr(\varepsilon_{1i} < -X_{1i}\beta_1, \varepsilon_{2i} < -X_{i2}\beta_2) = \Phi^2(-X_{1i}\beta_1, -X_{2i}\beta_2, \rho) \\ \Pr(Y_i = 1) &= 1 - \Phi^2(-X_{1i}\beta_1, -X_{2i}\beta_2, \rho), \end{aligned} \tag{4}$$

[5]Given two causal paths, Boolean probit and partial observability bivariate probit are synonymous. In the current context, we restrict our attention to such instances and employ the names interchangeably.

where $\Phi^2(\cdot)$ is the cumulative bivariate normal distribution function. Given these probabilities, we may fully specify a likelihood function, which may in turn be maximized using standard approaches.[6] This model is faithful to the basic framework of Eq. (1); F is just more complicated than previously.

Braumoeller (2003) has demonstrated the flexibility of Boolean probit for a wide variety of applications. There are two difficulties with this approach, however. The first is methodological. As he points out, the data requirements necessary for model identification can be quite severe. Poirier (1980) discusses this point at length: Different configurations of data can lead to ridges in the likelihood function or likelihoods with several modes. A minimal requirement for identification is that X_1 contain a regressor not in X_2 (or vice versa). Still, to our knowledge none of the authors cited above have examined the properties of these estimators via simulation (a task to which we turn below), limiting our ability to make statements about the utility of this class of estimators in practice. (Meng and Schmidt [1985] discuss the asymptotic efficiency of several variants of the partial observability model.)

The second problem is one of labeling and interpretation. It will nearly always be the case in social science examples that at least one independent variable will be a component of more than one causal mechanism. Consider the following application from the empirical analysis of regulatory outcomes: The Nuclear Regulatory Commission oversees the safety of civilian nuclear reactors in the United States. In order for us to observe a safety violation, it must be the case that (1) a violation has occurred (assuming no false positives), and (2) conditional on the violation's occurrence, it was detected. There are therefore two causal paths to an outcome of no observed violation: The violation did not occur, or it did occur and was not detected. Feinstein (1989, 1990) suggests a version of the partially observable bivariate probit, called *detection controlled estimation,* to estimate a facility's underlying violation rate given imperfect inspection.[7]

Now consider the analyst's decision of what independent variables to include in the model. Suppose we believed that one independent variable influenced both the probability of a violation's occurrence and the conditional probability of detection—for example, a change in the number of inspectors on hand. (Feinstein does not consider this variable directly, instead estimating fixed effects for individual inspectors present.) What could we conclude if the sign of the coefficient on the inspectors variable was positive in the detection equation and negative in the violation equation? A reasonable interpretation is that all else being equal, an increase in the number of inspectors improves the quality of inspections while deterring would-be violators.

It is possible, however, that we have mislabeled the equations: What we are calling the detection equation is actually the violation equation, and vice versa. Regulatory relationships have an inherently strategic quality, and this quality can produce all sorts of counterintuitive outcomes. Consider, for example, what might happen if inspectors who work in teams free ride on each others' efforts. In that case, an exogenous increase in the number of inspectors can simultaneously lead to an *increase* in the number of violations (if firms, anticipating the free riding, disinvest from compliance) and a *decrease* in the probability of detection (the direct effect of the free riding). This example may seem

[6]Our explication focuses on the case of two "substitutable" causes (A_1 *or* A_2), but to return to a point made earlier, the problem is isomorphic to one involving the conjunction of two causes (A_1 *and* A_2). In the current context, switching between the two models simply requires removing the negative signs preceding the arguments in the cumulative normal and recoding the dependent variable as its complement. For graphical intuition, reverse success and failure in Fig. 1 and rotate the page 180°.

[7]See also Brehm and Hamilton (1996) for a modified version of the detection control model.

far-fetched, but it is merely intended to show how the labeling problem can compound a problem of insufficient theorizing.

A third interpretation is that the statistical model is misspecified and that neither equation exclusively represents a single unobservable cause. In that case, we can still learn something about the influence of a change in the inspectors variable on *detected* violations. However, the model is simply providing the functional form for a complicated set of reduced-form nonlinear interaction effects.[8] More generally, while the partial observability/Boolean model may allow us to make more precise statements about the conditional impact of a particular independent variable on an observed outcome, employing the technique to make predictions about unobserved underlying conditions (e.g., the violation rate) may require a fair amount of faith in the validity of our model specification.[9]

3 Leveraging Qualitative Information for Quantitative Inference

In his recent book, Vreeland (2003) provides a detailed examination of why some nations choose to enter IMF loan agreements when suffering declining reserves and why other nations also enter IMF programs despite large robust reserves. In particular, he focuses on the cases of Tanzania, Nigeria, and Uruguay to explain how domestic and economic conditions drove the decision to turn to the IMF on different occasions. For instance, Vreeland demonstrates that while in 1975 Uruguay went under IMF programs due to economic need, its 1990 loan agreement was discretionary. In the context of the discussion above, for a limited number of cases Vreeland (2003) is able to differentiate between events where $Y_{1i} = 1$ and those where $Y_{2i} = 1$. For the majority of cases this information is censored.

Case studies and other qualitative data assist us in ascertaining the cause of an event.[10] Unfortunately, such qualitative evidence is rarely available for all cases in a sample. If it were, then partial observability would be a moot point. But even if such examples in which causes are revealed are few and far between, can we exploit that information in our large-N analysis?

We propose two methods to overcome the problems of identification and labeling inherent in the partial observability context. Both stake out an intermediate position between the partial observability model and the fully observable multivariate probit (one in which we observe dichotomous variables Y_{1i} and Y_{2i} directly; see Greene 2000, pp. 849–857). The idea behind both approaches is straightforward. Although for this class of problems information on which causal mechanism is responsible for an event is generally censored, we can occasionally discern that a cause was operative. Even if these instances are few and far between, we can use them as "anchors" to greatly reduce the identification and labeling problems.

[8]Compare the following traditional and Boolean probit models: (1) $\Pr(Y = 1) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$ and (2) $\Pr(Y = 1) = \Phi(\beta_{01} + \beta_1 X_1)\Phi(\beta_{02} + \beta_2 X_2)$. Nonlinearity implies that the marginal effect of X_1 will depend on the value of X_2 in both specifications, but in a more elaborate way for the second: For model 1, $\partial \Pr(Y = 1)/\partial X_1 = \beta_1\phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$ (i.e., the effect is maximized when $\Pr(Y_i = 1)$ is 0.5). For model 2, the effect is $\beta_1\phi(\beta_{01} + \beta_1 X_1)\Phi(\beta_{02} + \beta_2 X_2)$.

[9]Gordon (2003) avoids the labeling problem by formally modeling the strategic interaction between the regulator and regulated party directly, and testing the model's comparative statics on detected violations in the context of a structural estimation. While the technique prevents him from making statements about undetected violations, it does allow one to test the effect of independent variables on equilibrium regulatory behavior.

[10]For a discussion of the benefits of relating case studies and statistical analysis in the context of multivariate matching, see Rosenbaum and Silber (2001).

3.1 *A Maximum Likelihood Approach: Threshold Observability*

3.1.1 Intuition and Derivation

There are two latent variables, Y_1^* and Y_2^*, and our observed outcome variable is $Y_i = 1$ if $\max\{Y_{1i}^*, Y_{2i}^*\} > 0$. In the partial observability setting, there are three ways by which Y_i takes on values of one, each corresponding to the three success quadrants in Fig. 1. Now consider another case: One of the latent variables takes on an unusually large value, while the other does not. We define *unusually large* as exceeding some threshold $\tau > 0$. Such a situation might correspond to an observation for which we have detailed qualitative information that has not been collected in the vast majority of cases; in qualitative analyses such cases are often selected precisely because they are particularly illustrative of a specific causal mechanism. Since this information is generally censored we cannot systematically include it in the data matrix. For example, a case study may reveal that though the theory suggests that "$\mathcal{M}_1 : (Y_1^* > 0)$ or $\mathcal{M}_2 : (Y_2^* > 0)$ is a causal path to $Y_i = 1$," in this instance our descriptive knowledge from the case study informs us that it was really $\mathcal{M}_1$. In such an instance, we would say the mechanism provided a *discernible* cause of the outcome to occur.

Definition 2: $\mathcal{M}_j$ is a *discernible cause* of the event $Y_i = 1$ if $Y_{ji}^* > \tau_j$ and $Y_{\sim ji}^* \leq \tau_{\sim j}$.

In other cases we may observe $Y_i = 1$, but there may be a dispute among scholars as to the causal mechanism. Some may be vehement that A_1 was responsible, while others may be equally vehement that A_2 was at work. In such a case, ambiguity persists. Later we will discuss how such scholarly disagreement can be incorporated within a Bayesian context by specifying prior beliefs over competing causal mechanisms.

Definition 3: $\mathcal{M}_j$ is an *ambiguous cause* of the event $Y_i = 1$ if $\max\{Y_{1i}^*, Y_{2i}^*\} > 0$ and either (a) $0 \leq Y_{ji}^* \leq \tau_j$ and $Y_{\sim ji} < \tau_{\sim j}$ or (b) $Y_{ji}^* > \tau_j$ and $Y_{\sim ji}^* > \tau_{\sim j}$.

As it is currently constructed, the model suggests that moving from ambiguity to discernibility requires asymmetry between the latent variables. When the two latent variables are both positive and relatively small or both positive and relatively large, one cannot parse out a discernible causal mechanism from studies of individual cases. Discernibility emerges when one causal mechanism *stands out* relative to the other. Note that discernibility of one cause does not imply that the other cause is inoperative. Cases may emerge in which both latent variables are positive (i.e., the outcome is overdetermined) but for which only one causal mechanism is discernible.

Let $\delta_{ji} = 1$ if $\mathcal{M}_j$ is a discernible cause, and zero otherwise. From the above definitions, we can construct a likelihood function from the underlying probabilities. In the current setting,

$$\begin{aligned}
\Pr(Y_i = 0) &= \Phi^2(-X_{i1}\beta_1, -X_{i2}\beta_2, \rho) \\
\Pr(Y_i = 1, \delta_{1i} = 1) &= \Phi^2(X_{1i}\beta_1 - \tau_1, \tau_2 - X_{2i}\beta_2, -\rho) \\
\Pr(Y_i = 1, \delta_{2i} = 1) &= \Phi^2(\tau_1 - X_{1i}\beta_1, X_{2i}\beta_2 - \tau_2, -\rho).
\end{aligned} \tag{5}$$

The probability of observing a success with ambiguous causality, $\Pr(Y_i = 1,\ \delta_{1i} = 0,\ \delta_{2i} = 0)$, is simply the complement of the sum of the above probabilities. We refer to this model as the *threshold observability bivariate probit,* abbreviated as *throbit.*[11] Figure 2 provides graphical intuition.

[11]Readers may note that the threshold observability likelihood is equivalent to that of a censored bivariate three-category (per dimension) ordered probit model in which the censoring may be described as follows: If one (and only one) dependent variable takes on a value of two, we cannot tell whether the other dependent variable takes on a zero or one, and if both dependent variables take on values of two, we cannot distinguish that case from one in which either (a) the first dependent variable takes on a value of one and the second a value of zero or (b) vice versa. We thank an anonymous reviewer for pointing this out.

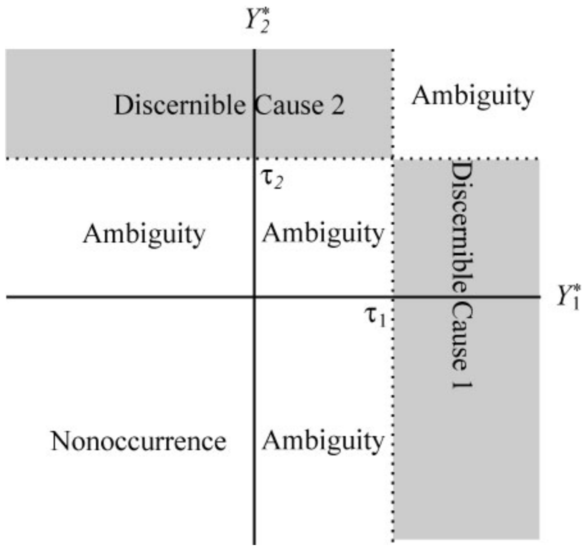

Fig. 2 Latent variables intuition for the threshold observability bivariate probit model. Throbit departs from Boolean probit by dividing $\{(Y_1^*, Y_2^*) \mid -\infty < Y_1^* < \infty, -\infty < Y_2^* < \infty\}$ into four regions instead of two, facilitating estimation. When Y_{1i} is unusually large relative to Y_{2i}, we say that the former discernibly caused $Y_i = 1$ (and vice versa).

Comparing Figs. 1 and 2, we immediately see the advantage the threshold observability model can provide over its partial observability (Boolean probit) cousin. In the less structured setting, one could relabel the axes and estimate the same model with a drastically different interpretation. The only things permitting us to identify the Boolean model are the nonlinear functional form and the inclusion of different regressor variables in X_1 and X_2. If X_1 and X_2 contain the same independent variable, our interpretation of its effect on one or the other unobserved causes hinges entirely on its interaction with the other variables in the model.

In the threshold observability model, case information about discernible causes serves the role of anchors, "locking-in" the dimensions by occasionally providing information about the correct dimensional orientation. In effect, the threshold observability model (and our alternative estimator, discussed below) replaces the identification problem with a more tractable censoring problem, making this *competing cause* model similar in spirit to *competing risks* duration models (David and Moeschberger 1978; Gordon 2002).

3.1.2 Monte Carlo Results: Comparing the Boolean/Partial Observability and Threshold Observability Models

As stated above, previous research on the partial observability probit models report the results of no simulations, thus limiting our ability to learn how the model actually functions in practice. Here we compare the performance of that model and the threshold observability specification in a series of Monte Carlo simulations.[12] In general, recovering estimates for ρ, the correlation between the error terms ε_1 and ε_2, requires a very large

[12]All simulations were conducted using *Gauss* 5.0 and the *Maxlik* module.

Table 1 Monte Carlo comparison of Boolean probit and throbit

	β_{10}	β_{1c}	β_{13}	β_{20}	β_{2c}	β_{23}	$ln(\tau_1)$	$ln(\tau_2)$
$N = 100$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.69
Mean Boolean probit	9.52	9.26	14.91	11.16	10.11	14.04	—	—
Mean throbit	1.40	1.60	1.67	1.25	1.47	1.49	0.76	0.80
Relative efficiency	18.41	10.36	13.18	30.77	27.58	35.28	—	—
$N = 500$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.69
Mean Boolean probit	1.05	1.05	1.07	1.05	1.04	1.06	—	—
Mean throbit	1.02	1.03	1.03	1.03	1.03	1.03	0.71	0.71
Relative efficiency	1.28	1.62	1.62	1.32	1.52	1.59	—	—
$N = 1000$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.69
Mean Boolean probit	1.02	1.02	1.03	1.02	1.02	1.03	—	—
Mean throbit	1.01	1.01	1.01	1.01	1.02	1.02	0.70	0.70
Relative efficiency	1.21	1.48	1.44	1.22	1.44	1.50	—	—
$N = 5000$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.69
Mean Boolean probit	1.00	1.00	1.00	1.00	1.01	1.01	—	—
Mean throbit	1.00	1.00	1.00	1.00	1.00	1.00	0.69	0.69
Relative efficiency	1.23	1.46	1.48	1.24	1.49	1.46	—	—

Note. Approximate proportion ambiguous success: 0.42. Approximate proportion discernible success for first reason: 0.05. Approximate proportion discernible success for second reason: 0.05.

sample size, so we constrain $\rho = 0$ for the experiments. Partially observable data were generated from the following parameterization:

$$\begin{aligned} Y^*_{1i} &= 1 + 1X_{ci} + 1X_{1i} + \varepsilon_{1i} \\ Y^*_{2i} &= 1 + 1X_{ci} + 1X_{2i} + \varepsilon_{2i} \\ Y_i &= \begin{cases} 1 & \text{if } \max\{Y^*_{1i}, Y^*_{2i}\} > 0 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{6}$$

with ε_{1i}, ε_{2i} fresh $N(0, 1)$ draws for each iteration.[13] Note that the variable X_c is common to both equations. Setting τ_1 and τ_2 permits us to distinguish discernible causes from ambiguous ones. For example, a large τ_2 and small τ_1 imply, all else equal, the greater discernibility of the first mechanism.

We report summary statistics from three sets of 1000 simulations with different sample sizes. In each, Xs are normal draws held fixed in repeated samples, but we manipulate their means and the value of τ_1 and τ_2 to control the proportion of failures, successes with ambiguous causes, and successes with discernible causes. Table 1 displays the results of the first set of simulations. Xs and τs are configured such that roughly half the sample

[13]We also ran versions of the simulation with different coefficient values. The relative performance of the two estimators was very similar to that reported here.

Table 2 Monte Carlo comparison of Boolean probit and throbit, skewed outcomes

	β_{10}	β_{1c}	β_{12}	β_{20}	β_{2c}	β_{22}	$ln(\tau_1)$	$ln(\tau_2)$
$N = 500$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	−0.63	−0.63
Mean Boolean probit	1.35	1.09	1.18	1.26	1.08	1.14	—	—
Mean throbit	1.15	1.05	1.08	1.16	1.05	1.08	−0.76	−0.67
Relative efficiency	1.69	1.50	1.89	1.46	1.52	1.76	—	—
Proportion successful convergence (Boolean probit): 0.871								
$N = 1000$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	−0.63	−0.63
Mean Boolean probit	1.13	1.04	1.07	1.12	1.02	1.05	—	—
Mean throbit	1.09	1.03	1.04	1.09	1.02	1.04	−0.65	−0.64
Relative efficiency	1.28	1.32	1.40	1.22	1.29	1.27	—	—
Proportion successful convergence (Boolean probit): 0.762								
$N = 5000$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	−0.63	−0.63
Mean Boolean probit	1.00	1.00	1.00	1.01	1.01	1.01	—	—
Mean throbit	1.00	1.00	1.00	1.01	1.00	1.00	−0.63	−0.64
Relative efficiency	1.21	1.22	1.25	1.21	1.24	1.23	—	—
Proportion successful convergence (Boolean probit): 0.997								

consists of successes and half failures. In approximately 10% of the successes the first causal mechanism is discernible, and in 10% the second is.

We report the mean Boolean probit and throbit estimates, as well as a measure of relative efficiency (Beck and Katz 1995): the ratio of Boolean probit's root mean squared error (RMSE) to that of throbit. A number greater than one, therefore, implies throbit's greater efficiency. The first thing to note is that for small sample sizes ($N = 100$), neither estimator performs particularly well. Comparatively speaking, however, throbit is considerably more efficient. We note also that in 3% of the Boolean probit estimations, the maximum likelihood procedure could not invert the Hessian at the extremum.

Compare these results with those when the sample size equals, respectively, 500, 1000, and 5000. Here both models perform considerably better. Throbit is considerably more efficient in terms of RMSE, but this is not surprising given that Boolean probit does not exploit information about discernible causes. Also, the throbit estimator successfully approximates the true value of $\ln(\tau_1)$ and $\ln(\tau_2)$. (We estimate the natural log of the τs to insure $\tau > 0$ at each step in the maximization algorithm.)

Efficiency is well and good but seems like an inadequate reason to prefer throbit to Boolean probit. After all, improvements in efficiency may come at considerable cost: the need to go out and uncover information about discernible causes, e.g., from case studies. Note, however, that the simulations whose results are posted in Table 1 represent a relatively easy case for both models: Roughly half of the observations have the dependent variable taking on a value of one. In many cases in political science, successes are comparatively rarer, leading to a skewed distribution of the dependent variable. To examine the case of skewed data, we turn to Table 2. In this set of experiments, each of 1000 simulations, the data and τs were set to produce a sample in which the dependent variable took on values of one only approximately 10% of the time. Of those cases,

Table 3 Monte Carlo simulation of throbit, "hard" cases for skewed outcomes

	β_{10}	β_{1c}	β_{12}	β_{20}	β_{2c}	β_{22}	$ln(\tau_1)$	$ln(\tau_2)$
$N = 500$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	−0.63	−0.63
Mean throbit	0.76	1.09	0.98	0.77	1.16	1.00	−0.82	−0.66
RMSE	0.07	0.02	0.02	0.05	0.03	0.02	0.12	0.03
Number of hard cases: 129								
$N = 1000$								
Truth	1.00	1.00	1.00	1.00	1.00	1.00	−0.63	−0.63
Mean throbit	0.92	1.06	1.01	0.85	1.04	0.98	−0.68	−0.66
RMSE	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.02
Number of hard cases: 238								

roughly a quarter were instances of success discernibly via the first mechanism, and a quarter discernibly by the second mechanism.

Because of poor performance by both models, we exclude consideration of cases in which $N = 100$. Turning our attention to results for other sample sizes, it turns out that the Boolean probit likelihood sometimes lacks a unique mode: Roughly 13% to 24% of the time for the cases in which $N = 500$ and $N = 1000$, it does not converge. To make the results for Boolean probit and throbit comparable, we include in Table 2 only summary statistics for trials in which the former successfully converged.[14] Restricting our attention to these experiments, we see similar results to those in Table 1. In general, throbit does at least as well if not better than Boolean probit in terms of bias, and considerably better in terms of relative efficiency.

How did the threshold observability model perform when Boolean probit could not? To answer this question, we turn to Table 3. Here we examine the performance of throbit in the 129 cases in which $N = 500$ and Boolean probit failed, and the 238 cases in which $N = 1000$ and the simpler model failed. In terms of bias and RMSE, throbit continues to do a good job recovering parameter estimates even for the "hard" cases in which Boolean probit fails outright.

Finally, we consider a third set of experiments representing the minimal identifying conditions for the partial observability model discussed in Poirier (1980). The latent regressions are

$$\begin{aligned} Y_{1i}^* &= 1 + 1X_{ci} + 1X_{1i} + \varepsilon_{1i} \\ Y_{2i}^* &= 1 + 1X_{ci} + \varepsilon_{2i}. \end{aligned} \tag{7}$$

In other words, there is a regressor variable common to both equations, and a second regressor appearing only in the first equation. According to Poirier, given sufficient variation in the data, the partial observability model (and thus Boolean probit) should be identified. However, as Meng and Schmidt (1985) show, the efficiency cost of the model compared with the more structured approaches they advocate is high when the model is just identified.

Unfortunately, almost half the time, Boolean probit fails to converge. Summary statistics for Boolean probit and throbit when the former did converge are reported in

[14]For these cases, throbit always converges.

Table 4 Monte Carlo comparison of Boolean probit and throbit, minimal identifying conditions

	β_{10}	β_{1c}	β_{12}	β_{20}	β_{2c}	$ln(\tau_1)$	$ln(\tau_2)$
$N = 500$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean Boolean probit	1.28	1.11	1.03	−0.69	0.63	—	—
Mean throbit	0.99	1.02	1.04	1.02	1.00	−0.00	−8.75
Relative efficiency	2.19	1.78	1.69	8.72	4.52	—	—
Proportion successful convergence (Boolean probit): 0.579							
$N = 1000$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean Boolean probit	1.12	1.05	1.01	0.32	0.85	—	—
Mean throbit	1.00	1.01	1.02	1.02	1.01	−0.01	−8.55
Relative efficiency	2.23	1.69	1.59	8.02	4.50	—	—
Proportion successful convergence (Boolean probit): 0.504							
$N = 5000$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean Boolean probit	1.04	1.02	1.01	0.86	0.96	—	—
Mean throbit	1.00	1.00	1.01	1.00	0.99	0.00	−8.31
Relative efficiency	2.07	1.55	1.71	6.07	3.41	—	—
Proportion successful convergence (Boolean probit): 0.523							

Table 4. For this subset of cases, throbit drastically outperforms its Boolean counterpart. Boolean probit coefficients are often biased whereas throbit coefficients are not, and throbit is anywhere from 55% to 872% more efficient. Table 5 presents throbit results for the data sets in which Boolean probit failed outright. Throbit continues to perform well, its estimates approximating the βs with low mean squared error.[15]

Finally, note that in the partial observability model, a more restrictive case in which both latent regressions contained the same independent variables would not be identified.[16] Because of the dimensional anchoring provided by discernible causal information, a similarly specified throbit model would be identified. Though we do not reproduce the results here, additional simulation revealed that this is in fact the case.

The intuition underlying the threshold model is that, while in general the cause of an event is censored, in some cases we have additional evidence (that cannot be systematically integrated into the statistical model) that points to one cause over another. As we saw in the simulations above, being able to anchor a small subset of the successes allows us to greatly improve the efficiency of our estimation and overcome the labeling and weak identification problems of Boolean probit. Unfortunately, throbit as we have presented it places a somewhat rigid structure on the data that in many cases might be inappropriate. For example, it does not allow scholars to assign cases in which both causal mechanisms are operative (the upper right quadrant in both figures). Moreover, we might be interested in making probabilistic claims about which causal mechanism is operative

[15]We note in passing that throbit does less well recovering estimates for $\ln(\tau_2)$.

[16]A similar problem crops up in Heckman-style selection models, necessitating exclusion restrictions. Sartori (2003) provides an alternative approach to this issue by assuming that the errors in both the selection and outcome equations are identical.

Table 5 Monte Carlo simulation of throbit, "hard" cases associated with minimal identifying conditions

	β_{10}	β_{1c}	β_{12}	β_{20}	β_{2c}	$ln(\tau_1)$	$ln(\tau_2)$
$N = 500$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean throbit	0.96	1.01	1.03	1.15	1.06	−0.02	−7.88
RMSE	0.01	0.01	0.01	0.02	0.01	0.01	0.21
Number of hard cases: 421							
$N = 1000$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean throbit	0.97	1.01	1.03	1.08	1.03	−0.01	−8.10
RMSE	0.01	0.00	0.00	0.01	0.01	0.00	0.18
Number of hard cases: 496							
$N = 5000$							
Truth	1.00	1.00	1.00	1.00	1.00	0.00	−6.91
Mean throbit	0.99	1.00	1.01	1.03	1.01	−0.00	−8.40
RMSE	0.00	0.00	0.00	0.01	0.00	0.00	0.17
Number of hard cases: 477							

based on our prior beliefs. Bayesian methods provide a natural framework for incorporating this type of information.

3.2 *A Bayesian Approach: Anchoring via Truncation*

3.2.1 Bayesian Markov Chain Monte Carlo (MCMC) Estimation and Data Augmentation

A full recounting of the methodology of Bayesian simulation is beyond the scope of the current paper.[17] Here we briefly review the basic concepts in the context of the univariate, linear-in-variables probit model (Albert and Chib 1993; Chib and Greenberg 1998). As above, we have $Y_i^* = X_i\beta + \varepsilon_i$, with $Y_i = 1$ if (and only if) $Y_i^* > 0$, and $\varepsilon_i \sim N(0,1)$.

The objective of the Bayesian approach is to update our beliefs about parameters having observed the data. In their seminal paper, Tanner and Wong (1987) show that the Bayesian estimation procedure is greatly simplified by treating realizations of the latent variable Y_i^* as parameters to be estimated, a method referred to as data augmentation. The MCMC procedure exploits the fact that if the latent data were known, the Bayesian update on β would be simple. In particular, assume a normally distributed prior for β with mean β_0 and variance V_0. Then, conditional on the latent data Y^*, the posterior distribution of $\beta \mid Y^*$ is normal with mean $\hat{\beta}$ and variance $\hat{V}$, where $\hat{\beta} = \hat{V}(V_0^{-1}\beta_0 + \sum_{i=1}^{N} X_i' Y_i^*)$ and $\hat{V} = (V_0^{-1} + X'X)^{-1}$.

The conditional distribution of the latent data, $Y^* | \beta, Y$ is similarly straightforward. Let $TN_{[p,q]}(\mu, \sigma^2)$ denote the truncated normal distribution with (untruncated) mean μ and variance σ^2, lower truncation point p, and upper truncation point q. Then if $Y_i = 1$, $Y_i^* \sim TN_{[0,+\infty]}(X_i\beta, 1)$; and if $Y_i = 0$, then $Y_i^* \sim TN_{[-\infty,0]}(X_i\beta, 1)$. Thus, given Y^* the distribution

[17] Jackman (2000) provides an excellent introduction to MCMC techniques and their application in political science.

of β is straightforward and given β the distribution of Y^* is straightforward. MCMC techniques exploit this to simulate the posterior density of β by iterating between random draws of β and the latent data. The algorithm is referred to as a Gibbs sampler:

1. Set $g = 1$ and choose starting values, $\beta^{(1)}$ and $(Y^*)^{(1)}$.
2. Randomly draw $\beta^{(g+1)}$ from distribution $\beta|Y^{*(g)} \sim N(\hat{\beta}, \hat{B})$.
3. Randomly draw $(Y^*)^{(t+1)}$ from distribution $Y^*|\beta^{(g+1)}$.
4. Let $g = g + 1$ and go to 2.

The initial draws from the sampler are strongly influenced by its starting values. However, after a suitable "burn-in" period, the draws from the sampler represent draws from the posterior distribution of β. Therefore the draws from the sampler can be used to approximate the desired distribution.[18]

3.2.2 The Bivariate Truncation Model

Return to the latent variables intuition of the partial observability/Boolean probit: $Y^*_{1i} = X_{1i}\beta_1 + \varepsilon_{1i}$ and $Y^*_{2i} = X_{2i}\beta_2 + \varepsilon_{2i}$, with $Y_i = 1$ if max $\{Y^*_{1i}, Y^*_{2i}\} > 0$. For ease of exposition, we will assume zero correlation between the εs, although modeling these errors as correlated is fairly simple (Smith 1999). The Bayesian implementation of the univariate probit model is readily extended to this setting. Since we explicitly model the latent data, it is straightforward to incorporate all our relevant knowledge about truncation, which can be done on a case-by-case basis. Suppose, for instance, that $Y_i = 0$. This implies that $Y^*_{1i} \leq 0$ and $Y^*_{2i} \leq 0$. In step 3 of the Gibbs sampler we would draw latent variables such that both were truncated below zero. The extension to successes ($Y_i = 1$) is more interesting.

Suppose $Y_i = 1$. Without any additional knowledge, this implies that either $Y^*_{1i} > 0$, $Y^*_{2i} > 0$, or both.[19] This is the Boolean probit setting. Additional knowledge of particular cases can be readily integrated into the model. Suppose, for example, that qualitative evidence suggests that cause 1 (and explicitly not cause 2) is responsible for causing success in a particular observation. The appropriate truncation for this observation is $Y^*_{1i} \sim TN_{[0,\infty]}(X_{1i}\beta_1, 1)$ and $Y^*_{2i} \sim TN_{[-\infty,0]}(X_{2i}\beta_2, 1)$. If, alternatively, mechanism 1 causes the success but mechanism 2 cannot be ruled out, then the appropriate truncation becomes $Y^*_{1i} \sim TN_{[0,\infty]}(X_{1i}\beta_1, 1)$ and $Y^*_{2i} \sim TN_{[-\infty,+\infty]}(X_{2i}\beta_2, 1)$. Other causal patterns map into

[18] In practice, there are numerous implementation issues associated with ensuring that the chain has converged. In the results reported we ran a chain of 200,000 iterations following a burn in period of 100,000 iterations. The chain exhibits a high level of autocorrelation. To reduce the impact of autocorrelation, we thinned the chain by examining only every 20th draw. To test for convergence we ran seven parallel chains of 150,000 iterations from different starting values and discarded the first 10,000 iterations of each chain. We then calculated the within-chain variance, $W = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}(\theta_{ij}-\bar{\theta}_j)^2$, and the between-chain variance, $B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\bar{\theta}})^2$, where m is the number of chains, n is the number of draws from each chain, θ_{ij} is the ith draw from chain j of parameter θ, $\bar{\theta}_j$ is the mean of parameter θ in chain j, and $\bar{\bar{\theta}}$ is the mean of parameter θ across all chains. Gelman and Rubin (1992) propose the statistic $\sqrt{\hat{R}} = \sqrt{\frac{(1-1/n)W+(1/n)B}{W}}$, which converges to 1 if the multiple chains converge to the same distribution. Gelman et al. (1995, p. 322) suggest that $\sqrt{\hat{R}}$ less that 1.1 or 1.2 are acceptable. In our analyses, for no parameter does $\sqrt{\hat{R}}$ exceed 1.0003.

[19] Although Y^*_{1i} and Y^*_{2i} can be readily simulated by repeatedly drawing untruncated normal variates until conditions on p and q are met (a technique called an "acceptance/rejection" algorithm), this is less efficient than the following approach: Calculate $A = \Phi(X_{1i}\beta_1)(1 - \Phi(X_{2i}\beta_2))$, $B = \Phi(X_{2i}\beta_2)(1 - \Phi(X_{1i}\beta_1))$, and $C = \Phi(X_{1i}\beta_1)\Phi(X_{2i}\beta_2)$. With probability $\frac{A}{A+B+C}$, $Y^*_{1i} \sim TN_{[0,\infty]}(X_{1i}\beta_1, 1)$ and $Y^*_{2i} \sim TN_{[-\infty,0]}(X_{2i}\beta_2, 1)$; with probability $\frac{B}{A+B+C}$, $Y^*_{1i} \sim TN_{[-\infty,0]}(X_{1i}\beta_1, 1)$ and $Y^*_{2i} \sim TN_{[0,+\infty]}(X_{2i}\beta_2, 1)$; and with probability $\frac{C}{A+B+C}$, $Y^*_{1i} \sim TN_{[0,+\infty]}(X_{1i}\beta_1, 1)$ and $Y^*_{2i} \sim TN_{[0,+\infty]}(X_{2i}\beta_2, 1)$. Noting that if $x \sim TN_{[p,q]}(\mu, \sigma^2)$ and u is a uniform random number, then $\tilde{x} = \mu + \sigma\Phi^{-1}(\Phi((p - \mu)/\sigma) + u(\Phi((q - \mu)/\sigma) - \Phi((p - \mu)/\sigma)))$ represents a random draw of x, the latent data can be drawn directly without the need for acceptance/rejection methods.

different truncation rules in an analogous manner. Once the truncations are established, the posterior distribution of β_1 and β_2 can be simulated using the Gibbs sampler, noting that in both steps 2 and 3 we need to make draws corresponding to both latent regression equations. Based on the sampling rules, we refer to this model as the *truncated bivariate probit* model, abbreviated *trubit.*

The Bayesian method is highly appropriate for integrating additional knowledge into the estimation procedure, such as that which would be obtained via qualitative study. Further, the setup described allows any prior additional information about *likely* causes to be used in assigning truncation of the latent data. We considered above the case in which qualitative evidence discernibly suggested one causal mechanism. However, the method is readily adapted when there is scholarly disagreement. For instance, analyses of case studies looking at a particular observation might suggest that mechanism 1 was 90% likely to have caused the success, mechanism 2 was 5% likely, and both mechanisms caused success with 5%. This is easily simulated in step 3 of the Gibbs sampler. With probability 0.9, Y^*_{1i} is positive and Y^*_{2i} is negative, with probability 0.05 draw Y^*_{2i} positive and Y^*_{1i} negative, and with probability 0.05 draw both latent variables as positive.

As in the throbit setting, qualitative evidence allows us to place additional structure on the estimation problem, this time via the truncation of the latent data. Once again the estimation procedure serves to anchor the underlying dimensions of a complex causal model. The chief difference between the throbit and trubit estimators stems from the intuition underlying when operative causes are observable in principle. The threshold observability model is well suited when a subset of the data consists of what might be called "glaringly obvious" cases. For instance, while isolating the cause of a war is in general difficult, in some cases a consensus might exist. As an example, many scholars point to the 1982 Falklands conflict as an example of a diversionary war, as discussed by Levy and Vakili (1992). Of course, even in this perhaps glaringly obvious case, there is dissent. Welch (1993) argues that the war resulted from a failure to achieve a normatively appropriate settlement through negotiations. The Bayesian context of the trubit estimator allows the researcher to include such scholarly disagreements by specifying probability distributions over the different possible truncation patterns.

Note that while we have discussed threshold observability in the maximum likelihood context and bivariate truncation in a Bayesian setting, we could have done the reverse, because both methods are amenable to either approach.[20] The advantage of the Bayesian methodology is its natural interpretation in terms of bringing additional information to bear on a problem while rigorously quantifying our uncertainty about that information; the advantage of the maximum likelihood approach, as usual, is superior understanding of its asymptotic and convergence properties.

4 Motives for IMF Loans

We have illustrated our arguments throughout with reference to the competing motives that leaders have to sign IMF agreements. We now turn to the determinants of each, estimating models using Boolean probit, throbit, and trubit. Since the focus of this paper is methodological, we will not delay with a long justification of our independent variables. For a substantive description of this problem see Smith and Vreeland (2003).

[20]As a separate exercise, we conducted a Monte Carlo simulation of a maximum likelihood version of the bivariate truncation model. Its performance, in terms of relative efficiency, fell somewhere between that of Boolean probit and the threshold observability model.

The unit of analysis is the nation-month. Our data consist of 11,816 observations from 1971 onward. For each, we measure whether a nation went under IMF agreement ($Y_i = 1$) or remained free of IMF encumbrances ($Y_i = 0$) (IMF *Annual Report* 1971–2002). We do not examine the decision to exit IMF programs, although the analysis could be readily expanded to address that question (Przeworski and Vreeland 2002).[21] Because entry into agreements is rare in the data set (only 225 entries, or just under 2%), the dependent variable is highly skewed, making conditions for successful estimation more demanding (as the Monte Carlo simulation revealed).

Several independent variables are included in the domestic/discretionary equation (i.e., $\mathcal{M}_1$). First, we consider the subject nation's domestic political institutions. We employ the winning coalition measure, *W*, of Bueno de Mesquita et al. (2002, 2003). This variable, scaled from zero to one, gauges the number of supporters to whom a leader is beholden in order to retain power. As *W* increases, the ease with which a leader can implement unpopular reform efforts decreases, raising the incentives to seek exogenous reforms through IMF programs. We also employ their *effective selectorate* measure. This captures the size of the pool of potential supporters from which a leader draws her winning coalition, multiplied by $1 - W$. This variable takes its greatest values in corrupt universal franchise systems, where potentially almost anyone can (with very low probability) be brought into a leader's small coalition of supporters. The effective selectorate measure is smallest when either the winning coalition is large or the pool of available supporters is small, as is often the case in juntas or monarchies in which effectively only the military or aristocracy have political rights to choose leaders. This measure describes the degree of a leader's policy autonomy, larger values of which obviate the need to sacrifice domestic sovereignty for policy purposes through the IMF.

Next, we consider the role of a leader's tenure in office. Previous research (Bueno de Mesquita et al. 2003) suggests that the probability that a leader exits from power at a particular point in his tenure given survival in office to that point (i.e., the hazard rate) is steeply declining over time for autocratic leaders (who are able gradually to consolidate their support coalitions) and much less steeply declining for democratic ones. With respect to discretionary IMF loans, democratic leaders should thus be more willing to enter agreements early in their tenure, when their electoral consequences are a far-off prospect. In contrast, autocrats, unencumbered by electoral fears, are more likely to enter agreements later in their tenure. To test this hypothesis, we include a measure of tenure and its interaction with *W*. (Leadership data are taken from Bueno de Mesquita et al. 2003.)

The first equation also included a measured of economic growth taken from the World Bank Development Indicators (2001). When economic growth stalls, leaders are mostly likely to desire reforms.

Finally, we consider secular change in global and domestic attitudes about the acceptability of submitting to IMF loans. For global change, we control for the age of the IMF program (calendar year minus 1975), hypothesizing a secular increase in propensity to apply. With respect to local change, the marginal impact on sovereignty is decreasing with each successive loan, so we control for the total number of the country's prior loans.

For the equation operationalizing $\mathcal{M}_2$ (economic need), we included the regime-type variables, as well as measures of annual levels of debt service (percent of GNP), largest

[21]In their 2002 article, Przeworski and Vreeland treat the IMF itself as a strategic actor, something we do not undertake here. Further, they consider agreement entry, persistence, and exit, whereas we restrict our attention to the agreement entry decision.

monthly change in exchange rate over the previous three months, and level of international reserves (log of total reserves in terms of number of months of imports, from IMF International Financial Statistics quarterly data).[22] Data on economic conditions are taken primarily from the World Bank Development Indicators (2001) and the IMF's International Financial Statistics (2003).

We assigned discernible cause to a number of cases. In his book on IMF politics, Vreeland (2003) distinguishes several cases in which IMF loans were surely sought for domestic political reasons (e.g., Uruguay 1990), and several others in which they were surely sought for economic need (e.g., Uruguay 1974). We use his case studies as the basis for assigning discernible causes. We have 33 instances of loan agreements corresponding to the first mechanism and 17 via the second (175 loans are ambiguous).

Table 6 shows estimates for each cause of IMF agreements. The first, second, and third columns of estimates correspond, respectively, to maximum likelihood estimates (MLEs) for Boolean probit and throbit, and the MCMC estimates for trubit. We experienced some difficulty getting the likelihood estimators to converge for Boolean probit.[23]

Of course, unlike the simulations above, in this case we do not know the true underlying model. This makes discernible model comparisons difficult. That said, Boolean probit yields very slightly weaker parameter estimates than throbit. In the discretionary equation, the significant variables are year, number of prior agreements, and economic growth. The alternative models also support this conclusion. These results are consistent with Vreeland's (2003, chapter 2) observation: Initially there is a stigma attached to the loss of sovereignty associated with an IMF loan, but this diminishes with time and experience. Contrary to expectations, neither political institutions nor either of the tenure variables is a statistically significant determinant of discretionary loans.

Switching to the need equation, the biggest determinants are debt service and level of reserves. In all models these factors are highly significant. The throbit model also suggests that currency devaluation increases the likelihood of a need-based loan.[24] In the Boolean probit estimates, this relationship was statistically insignificant.

The MCMC estimates for the bivariate truncation model differ slightly from the throbit MLE, although both yield similar substantive conclusions.[25]

We also estimated a version of the throbit model in which the error correlation is not assumed to be zero. We estimated the ρ to be -0.18, although it is not statistically significant. The estimates of the other parameters remain nearly unchanged in this specification (as one might expect), with the coefficients varying by about 1% between models.[26] We do not report these results here.

[22]We also examined change in reserves. While the substantitive results did not change markedly with the inclusion of this variable, we experienced considerable problems achieving convergence. Further, the posterior distribution obtained from the Gibbs sampler for this variable was highly skewed.

[23]As a robustness check, we estimated models in both *Gauss* 5.0 using the *Maxlik* routine and *Stata* 8.0 using its ML procedures. Using the Newton-Raphson algorithm, *Gauss* versions of throbit typically converge within about 11–12 iterations with uninformative starting values. Using *Stata*, we estimated identical parameters for throbit even with different starting values. Estimating coefficients for Boolean probit proved more difficult. With uninformative starting values, the *Gauss* version of the model did not converge. Using *Stata* and specifying the "difficult" option (which switches between steepest ascent and Newton-Raphson), coefficient estimates depended on starting values. Those reported in Table 6 use the threshold observability estimates as a vector of initial values.

[24]Exchange rates are measured as the amount of currency to buy one SDR (an IMF basket of currencies). Hence the positive coefficient indicates that IMF programs are more likely following devaluation.

[25]The MCMC estimates used normally distributed priors with mean zero and variance ten. MLE estimates of the Trubit model yield similar results.

[26]In the model with error correlation, we estimate $\text{atanh}(\rho)$, which is bounded between -1 and 1.

Table 6 Competing causes for IMF agreements: Boolean probit, threshold observability, and bivariate truncation estimates

	Boolean probit MLE	*Threshold observability MLE*	*Bivariate truncation MCMC*
Mechanism 1: Discretionary political reasons			
Constant	-2.455^{**}	-2.603^{**}	−2.469 (.120)
	(.172)	(.186)	−2.465 0.000
W	−.139	−.064	.154 (.147)
	(.199)	(.223)	.153 .853
Effective	−.269	.138	.118 (.154)
Selectorate	(.324)	(.230)	.118 .780
Tenure	−.015	−.012	.002 (.010)
	(.024)	(.017)	.002 .597
W^*Tenure	−.002	.002	−.007 (.018)
	(.041)	(.029)	−.007 .346
Year	$.020^{**}$	$.017^{**}$	.011 (.004)
	(.007)	(.006)	.011 .997
Cumulative	$.209^{**}$	$.193^{*}$	.182 (.047)
Agreements	(.075)	(.066)	.182 .999
GDP growth	$-.022^{**}$	−.015	−.025 (.006)
	(.008)	(.008)	−.025 0.000
Mechanism 2: Economic need			
Constant	−4.411	-2.285^{**}	−2.496 (.216)
	(3.100)	(.310)	−2.489 0.000
W	2.099	.244	.150 (.219)
	(3.061)	(.341)	.144 .752
Effective	2.464	.238	.183 (.245)
Selectorate	(3.111)	(.348)	.184 .773
Debt service	$.063^{**}$	$.062^{**}$	.049 (.010)
	(.014)	(.010)	.049 1.000
Reserves	$-.444^{**}$	$-.600^{**}$	−.393 (.091)
	(.104)	(.118)	−.393 0.000
ΔExchange rate	.059	$.110^{*}$	.081 (.070)
	(.069)	(.053)	.089 0.881
$\ln \tau_1$		$-.719^{**}$	
		(.147)	
$\ln \tau_2$		$-.405^{**}$	
		(.149)	
Observations	11816	11816	11816
Loglikelihood	−1049.1	−1187.4	—

Note. In the MLE estimates the first number corresponds to the parameter estimate and the second number is the estimated standard error. In the MCMC estimates, we report the mean, standard deviation, median, and proportion of the draws greater than zero (from the 100,000 draws of the sample). This latter statistic can be thought of as equivalent to the probability of significance (α, or $1 - \alpha$) in classical hypothesis tests. $^{*}p < .05$; $^{**}p < .01$ (two-tailed tests).

By anchoring the data through the assignment of "extreme" cases, throbit and the bivariate truncation model overcome the problems of weak identification and labeling. The use of evidence outside of the data matrix allows us to sharpen parameter estimates and allows the determinant of each cause of an event to be distinguished more readily.

5 Why It Matters

We have shown that we can improve the quality and interpretation of estimates in models in which underlying causal mechanisms are largely unobserved, but one might nevertheless question the value of this. A direct reply is that these methods allow us to estimate and distinguish between competing causes. While this is in itself an important goal, we believe it is the use to which these models can be put in determining the consequences of events that is even more important.

Selection models have gained increasing importance in political science, and with good reason. Many of the processes we study involve explicit selection. For instance, if we study how someone answers a survey question, we need to remember that those citizens who choose to answer are not a random draw from the population (Berinsky 1999). If we examine who wins wars, then again, we need to remember that wars between nations are not a random draw of all the possible disputes between nations (Reiter and Stam 2002). The example above shows that nations entering IMF are not a random sample either. If we want to examine the *effects* of IMF loans, whether in terms of leader survival or economic performance, we need to account for the decision to enter an IMF agreement in the first place.

The most common methods to control for selection are Heckman-type models (Heckman 1976, 1979; Achen 1986; Dubin and Rivers 1989) in which the decision to enter the sample (for instance, to sign an IMF agreement) is modeled as a probit simultaneously with the outcome variable of interest. Such models have allowed social scientists to make remarkable progress. Unfortunately, when there are multiple mechanisms through which an actor might enter the sample, treating selection into the sample as a single probit equation is inappropriate. In the context of IMF loans, theory suggests that the impact of IMF programs on leader survival and economic growth should be very different depending on whether entry into the IMF program was triggered by actual economic need or a leader's desire to exploit IMF conditionality to implement domestic reforms. When democratic leaders take discretionary loans, we expect them to enhance political survival and help promote economic growth. In contrast, when democratic leaders take loans through necessity, we expect it to be detrimental to their survival and to impede economic growth. Since the two mechanisms have competing consequences, it is unsurprising that the aggregate impact of IMF programs is ambiguous. (There is great controversy as to the impact of IMF programs: see Payer 1974; Reichmann and Stillson 1978; Connors 1979; Gylfason 1987; Pastor 1987; Khan 1990; Haggard and Kaufman 1992; Edwards and Santaella 1993; Conway 1994.) Simply controlling for being under an IMF program, without differentiating as to why a leader signed in the first place, does not control for the appropriate contingent circumstance. Using similar estimates to those above, Smith and Vreeland (2003) create measures of the propensity to enter IMF for different reasons. In particular, they estimate the probability that a loan was discretionary rather than need based. These propensity measures strongly influence the impact of IMF programs on leader survival. In particular, loans taken for discretionary purposes strongly enhance survival, while those taken in response to need damage a democrat's tenure.

6 Conclusion

We provide a method for improving interpretation of and identification for statistical models of complex causes in which the underlying causes are difficult to observe. In short, information on operative causes that is generally censored but that is available in a handful of cases can help anchor the estimation procedure and substantially reduce attendant difficulties in large-*N* studies. Note that the threshold observability and bivariate truncation models we discuss may be easily generalized to problems far more complex than the one (A_1 or A_2 lead to B) that we have explicated here. Braumoeller's generalization of Poirier's partial observability bivariate probit model to such complex settings is clearly a step toward greater understanding of these processes, but the ambiguous nature of causality implied by the partial observability/Boolean models creates the problems discussed above. Our solution(s), of course, should not be taken as a direct criticism of Braumoeller's recent research on Boolean processes, but as a suggestion for facilitating interpretation of a class of estimators potentially of great use to social scientists.

References

Abowd, John M., and Henry S. Farber. 1982. "Job Queues and the Union Status of Workers." *Industrial and Labor Relations Review* 35:354–367.

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments.* Berkeley: University of California Press.

Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88:669–679.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data in Comparative Politics." *American Political Science Review* 89:634–647.

Berinsky, Adam J. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science* 43:1209–1230.

Bird, Graham. 1996. "Borrowing from the IMF: The Policy Implications of Recent Empirical Research." *World Development* 24:1753–1760.

Bjork, James. 1995. "The Uses of Conditionality." *East European Quarterly* 29:89–124.

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

Brehm, John, and James T. Hamilton. 1996. "Noncompliance in Environmental Reporting: Are Violators Ignorant, or Evasive, of the Law?" *American Journal of Political Science* 40:444–477.

Bueno de Mesquita, Bruce, James Morrow, Randolph Siverson, and Alastair Smith. 2002. "Political Institutions, Policy Choice and the Survival of Leaders." *British Journal of Political Science* 32:559–590.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph Siverson, and James D. Morrow. 2003. *The Logic of Political Survival.* Cambridge, MA: MIT Press.

Chib, Siddhartha, and Edward Greenberg. 1998. "Analysis of Multivariate Probit Models." *Biometrika* 85:347–361.

Clark, William Roberts, and Michael J. Gilligan. 2004. "A Simple Multivariate Statistical Test for Necessary or Sufficient Conditions." Unpublished manuscript, New York University.

Connors, Thomas A. 1979. "The Apparent Effects of Recent IMF Stabilization Programs." *International Finance Discussion Paper*, No. 135. Washington, DC: Board of Governors of the Federal Reserve System.

Conway, Patrick. 1994. "IMF Lending Programs: Participation and Impact." *Journal of Development Economics* 45:365–391.

David, Herbert A., and Melvin L. Moeschberger. 1978. *The Theory of Competing Risks.* New York: MacMillan.

Dixit, Avinash K. 1996. *The Making of Economic Policy: A Transaction-Cost Politics Perspective.* Cambridge, MA: MIT Press.

Downs, George W., and David M. Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War." *American Journal of Political Science* 38:362–380.

Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations.* Princeton, NJ: Princeton University Press.

Drazen, Allan. 2001. "Conditionality and Ownership in IMF Lending: A Political Economy Approach." Mimeo.

Dubin, Jeffrey A., and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research* 18:360–390.

Edwards, Sebastian, and Julio A. Santaella. 1993. "Devaluation Controversies in the Developing Countries: Lessons from the Bretton Woods Era." In *A Retrospective on the Bretton Woods System,* eds. Michael D. Bordo and Barry Eichengreen. Chicago, IL: University of Chicago Press, pp. 405–455.

Feinstein, Jonathan S. 1989. "The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences." *Journal of Political Economy* 97:115–154.

Feinstein, Jonathan S. 1990. "Detection Controlled Estimation." *Journal of Law and Economics* 33:233–276.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis.* London: Chapman and Hall.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7:457–511.

Gordon, Sanford C. 2002. "Stochastic Dependence in Competing Risks." *American Journal of Political Science* 46:200–217.

Gordon, Sanford C. 2003. "On the Political Strength of Regulated Industries and the Conditional Nature of Regulatory Autonomy." Unpublished manuscript, New York University.

Greene, William H. 2000. *Econometric Analysis,* 4th ed. Upper Saddle River, NJ: Prentice Hall.

Gylfason, Thorvaldur. 1987. "Credit Policy and Economic Activity in Developing Countries with IMF Stabilization Programs." *Princeton Studies in International Finance*, No. 60.

Haggard, Stephan, and Robert R. Kaufman. 1992. *The Politics of Economic Adjustment: International Constraints, Distributive Conflicts and the State.* Princeton, NJ: Princeton University Press.

Heckman, James. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5:475–492.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

Huth, Paul K. 1996. *Standing Your Ground: Territorial Disputes and International Conflict.* Ann Arbor, MI: University of Michigan Press.

International Monetary Fund. 1971–2002. *Annual Report*, various issues. Washington, DC: International Monetary Fund.

International Monetary Fund. 2003. *International Financial Statistics on CD-ROM, February 2003.* Washington, DC: International Monetary Fund.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44:375–404.

Jervis, Robert. 1976. *Perception and Misperception in International Politics.* Princeton, NJ: Princeton University Press.

Khan, Mohsin S. 1990. "The Macroeconomic Effects of Fund-Supported Adjustment Programs." *IMF Staff Papers* 37:195–231.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* New York: Cambridge University Press.

Levy, Jack. 1989. "The Diversionary Theory of War: A Critique." In *Handbook of War Studies,* ed. Manus I. Midlarsky. Boston: Unwin Hyman, pp. 259–288.

Levy, Jack, and Lily I. Vakili. 1992. "Diversionary Action by Authoritarian Regimes: Argentina in the Falklands/Malvinas Case." In *The Internationalization of Communal Strife,* ed. Manus I. Midlarsky. New York: Routledge, pp. 118–146.

Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2:245–264.

Meng, Chun-Lo, and Peter Schmidt. 1985. "On the Cost of Partial Observability in the Bivariate Probit Model." *International Economic Review* 26:71–85.

Pastor, Manuel, Jr. 1987. *The International Monetary Fund and Latin America: Economic Stabilization and Class Conflict.* Boulder, CO: Westview.

Payer, Cheryl. 1974. *The Debt Trap: The IMF and the Third World.* New York: Monthly Review Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference.* New York: Cambridge University Press.

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

Przeworski, Adam, and James Raymond Vreeland. 2002. "A Statistical Model of Bilateral Cooperation." *Political Analysis* 10:101–112.

Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42:427–460.

Ramcharan, Rodney. 2003. "Reputation, Debt and Policy Conditionality." Working Paper, Research Department, International Monetary Fund.

Reichmann, Thomas M., and Richard T. Stillson. 1978. "Experience with Programs of Balance of Payments Adjustment: Stand-by Arrangements in the Highest Tranches, 1963–72." *IMF Staff Papers* 25:293–309.

Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War.* Princeton, NJ: Princeton University Press.

Remmer, Karen L. 1986. "The Politics of Economic Stabilization, IMF Standby Programs in Latin America, 1954–1984." *Comparative Politics* 19:1–24.

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

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions." *Political Analysis* 11:111–138.

Smith, Alastair. 1996. "Diversionary Foreign Policy in Democratic Systems." *International Studies Quarterly* 40:133–153.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43:1254–1283.

Smith, Alastair, and James Vreeland. 2003. "The Survival of Political Leaders and IMF Programs: Testing the Scapegoat Story." Presented at the Yale University Conference on the Impact of Globalization on the Nation-State from Above: The International Monetary Fund and the World Bank, April 25–27.

Spaventa, Luigi. 1983. "Two Letters of Intent: External Crises and Stabilization Policy, Italy, 1973–77." In *IMF Conditionality,* ed. John Williamson. Washington, DC: Institute for International Economics, pp. 441–473.

Tanner, Martin A., and Wing Hung Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82:528–540.

Vaubel, Roland. 1986. "A Public Choice Approach to International Organization." *Public Choice* 51:39–57.

Vreeland, James Raymond. 2000. "The Institutional Determinants of IMF Programs." Presented at the Leitner Work-in-Progress Seminar, Yale University, December 4.

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

Welch, David A. 1993. *Justice and the Genesis of War.* New York: Cambridge University Press.

World Bank. 2001. *World Bank Development Indicators on CD-ROM.* Washington, DC: World Bank.