View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Columbia University Academic Commons

INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY
COLUMBIA UNIVERSITY
WORKING PAPERS

ALTERNATIVE MODELS OF DYNAMICS IN BINARY TIME-SERIES-CROSS-SECTION MODELS: THE EXAMPLE OF STATE FAILURE

Nathaniel Beck
University of California, San Diego

David Epstein
Columbia University

Simon Jackman
Stanford University

Sharyn O'Halloran
Columbia University

August 2001

ISERP WORKING PAPER 02-01

ISERP
Institute for Social and Economic Research and Policy

Alternative Models of Dynamics in Binary Time-Series–Cross-Section Models: The Example of State Failure[1]

Nathaniel Beck[2] David Epstein[3] Simon Jackman[4]
Sharyn O'Halloran[5]

Prepared for delivery at the 2001 Annual Meeting of the Society for Political Methodology, Emory University. Draft of August 24, 2001

[1]Thanks to Gary King for some helpful discussions, as well as the unindicted co-conspirators Jonathan Katz and Richard Tucker. Epstein and O'Halloran acknowledge the support of the National Science Foundation under grant SES-99-05774.

[2]Department of Political Science; University of California, San Diego; La Jolla, CA 92093; beck@ucsd.edu

[3]Department of Political Science; Columbia University; NY, NY 10027; de11@columbia.edu

[4]Department of Political Science; Stanford University; Stanford, CA 94305; jackman@stanford.edu

[5]Department of Political Science; Columbia University; NY, NY, 10027; so33@columbia.edu

Abstract

This paper investigates a variety of dynamic probit models for time-series–cross-section data in the context of explaining state failure. It shows that ordinary probit, which ignores dynamics, is misleading. Alternatives that seem to produce sensible results are the transition model and a model which includes a lagged *latent* dependent variable. It is argued that the use of a lagged latent variable is often superior to the use of a lagged realized dependent variable. It is also shown that the latter is a special case of the transition model. The relationship between the transition model and event history methods is also considered: the transition model estimates an event history model for both values of the dependent variable, yielding estimates that are identical to those produced by the two event history models. Furthermore, one can incorporate the insights gleaned from the event history models into the transition analysis, so that researchers do not have to assume duration independence. The conclusion notes that investigations of the various models have been limited to data sets which contain long sequences of zeros; models may perform differently in data sets with shorter bursts of zeros and ones.

1 Introduction

Students of comparative politics and international relations have grown increasingly more methodologically aware when they model time-series–cross-section data with a binary dependent variable ("BTSCS") over the last half decade or so.[1] There are a number of possible avenues that researchers might use, and as methodologists we are in the infancy of understanding how these models work for typical political science applications.[2] The purpose of this paper is to see how a variety of different methods work in one particular application, the study of state failure.[3]

Obviously one cannot assess the statistical properties of any estimator, or compare the performance of a variety of estimators, by looking at one application. But it is important to see how the various methods, and their underlying statistical models, comport with real political science applications. So far the major political science test bed for comparing the various approaches has been the study of conflict in international relations, using the dyad-year design. It thus seems sensible to examine an application to comparative politics, though, as we shall see, some of the properties of the state failure data are similar to the IR conflict data.

The models used here have been discussed by us in previous papers (Beck, Katz and Tucker, 1998; Beck and Tucker, 1997; Jackman, 2000*b*) and so we do not go into detail on the etiology of the various models here. Since the paper is concerned with modelling dynamics, we focus on only one substantive model and do not consider model specification issues that are unrelated to dynamics.[4] In the next section we lay out the notation and the various

[1] This is not to say that all researchers were unaware of the methodological issues before that (for example, Londregan and Poole, 1990; Przeworski, Alvarez, Cheibub and Limongi, 2000). But these exceptional works typically did not lead other researchers to follow the paths they took.

[2] Scholars in other areas, particularly biometrics, have devoted much effort to binary panel data (see Diggle, Liang and Zeger, 1994). But if one looks at the paradigmatic applications of binary panel analysis in biometry, that is, whether someone in a panel is classified as ill over a repeated series of observations, it is clear that the biometric applications are different from the typical BTSCS applications that appear in IR and comparative politics. We can and have made much use of the work of biometricians, but we have to make sure that what is useful in biometrics is also useful in the various subfields of social science. Indeed, that is a principal endeavor of this paper.

[3] The State Failure Task Force has developed quite a complete data set on revolutionary wars, ethnic wars, regime changes, and genocides in order to study the causes of state breakdown in the postwar era. The data are described more fully in Section 3 below.

[4] Thus, in particular, we do not consider a number of issues in analyzing state failure discussed by King and Zeng (N.d.). We note that much of their discussion deals with issues arising from case-control designs, whereas here we use data from all nations. They, correctly, focus on the out-of-sample forecasting issue. For our purposes, and at the current moment, in-sample analyses suffice, though we agree that at the end of the day, it is out-

models, followed by a brief discussion of some features of these models that have received insufficient attention. Section 3 then discusses data and measurement issues. Section 4 presents the ordinary probit results and some non-specification based "fixes;" Section 5 discusses the various transitional models, with latent variable models described in Section 6. The concluding section generalizes the discussion beyond state failure.

2 Models and notation

We assume the data are generated as a binary dependent variable time-series–cross-section. Thus we assume that the number of units ("countries") , N, is fixed and all asymptotics are in the number of time periods, T, ("years"). While we make no specific assumptions about N or T, we assume that T is large enough (say more than 10) so that some time-series analysis is possible.[5]

Let the binary dependent variable be $y_{i,t}, i = 1, \ldots, N, t = 1, \ldots, T$. Since we do not deal with spatial issues, when we discuss the model for a generic observation, i, we suppress the first subscript and refer to the observations for that generic unit as y_t. For simplicity of notation, assume a rectangular data structure, where all countries are observed for the same time period. This simplifies notation and is not critical; the data set we analyze is in fact non-rectangular.[6] Let us assume we have some set of independent variables of interest, $\mathbf{x}_{i,t}$, which, when we can do so without confusion, we refer to as $\mathbf{x}_t$.

The "ordinary" probit[7] model assumes that all observations are indepen-

of-sample forecast performance that is most important. Finally, we note that King and Zeng implicitly use one of our preferred methods for treating dynamics in the state failure data and so there is no major disagreement between us on how one should model the dynamics of state failure.

[5] Almost all the binary longitudinal studies in biometrics are applications to panel data, which has asymptotics in N, not T (and typically, though not always, has a small T). The comparative politics applications (time-series–cross-section data) have fixed N and larger T than do the typical panel study; asymptotics are in T. Thus methods which are either good for, or made necessary by, binary panel data may not either work well or be necessary for BTSCS data, and vice versa (Beck, 2001).

[6] In dynamic models it is important that we somehow deal with interior missing data, since the models assume that the data analyzed are spaced at yearly intervals. The issue of missing data is orthogonal to the issue of modelling dynamics, and so in our own data analysis we are a bit cavalier about missing data. We fully agree with King and Zeng that correctly handling missing data issues is critical and agree that multiple imputations are the appropriate way to do this. Fortunately there is relatively little *interior* missing data for the variables we analyze. We return to this issue in Section 3.

[7] We focus on probit here because it works more easily with some later models. Almost everything we say here would also hold for logit.

dent, so we estimate

$$y_t^* = \mathbf{x}_t\beta + \epsilon_t \tag{1}$$

$$y_t = 1 \text{ if } y_t^* > 0 \tag{2}$$

$$y_t = 0 \text{ otherwise} \tag{3}$$

$$\epsilon_t \sim N(0,1) \tag{4}$$

This is the usual probit model which simply ignored dynamics. When we do not need to explicitly refer to the latent y_t^* we denote the probit model defined in Equations 1–4 by $y = \text{Probit}(\mathbf{x}\beta)$. Unless stated otherwise, we also assume independent identically distributed standard normal errors throughout.

Few analysts today would estimate the ordinary probit. Many, following common time series procedure, would simply add a lagged dependent variable to the model, yielding what we will call the "restricted transition" (the reason for this name will become obvious presently) probit:

$$y_t^* = \mathbf{x}_t\beta + \rho y_{t-1} + \epsilon_t \tag{5}$$

Note that this model simply shifts up the latent y by ρ when the lagged observed y is one (dropping the first observation for each country from the estimation). Because of the non-linear nature of the probit, this does not shift probabilities by a simple function of ρ. The restricted transition probit is often used simply because it looks like a standard time series method, but some analysts (for example, Londregan and Poole, 1990) have used this model (or models very much like it) because they theoretically believe that previous realized values of y are the determinants of current y.

The restricted transition model, however, is *NOT* the natural analog to the continuous dependent variable time series model with a lagged dependent variable. The right way to think about binary time series analogies of their continuous cousins is to write the times series model in terms of a continuous latent variable and then just take each period's realization of a zero or one as arising from a draw from the underlying normal distribution. We thus have the generalization of Equation 1 to

$$y_t^* = \mathbf{x}_t\beta + \rho y_{t-1}^* + \epsilon_t \tag{6}$$

(with Equations 2–4 remaining unchanged). In this model, the latent y^* follows a standard time series pattern. The difference between the two models is that in the restricted transition model it is the realized lagged values of y that affect current values, whereas in the latent lagged model, it is the underlying latent variable that shows persistence. The two models differ when a chance draw of the observed y is one even though the underlying latent y^* was small, or vice-versa, so that the chance of getting such an observed draw was low. The lagged latent model looks more like a standard time series model than does the restricted transition model; it is also much harder to estimate.[8]

Just as the ordinary probit is a special case of the restricted transition model with $\rho = 0$, so is the restricted transition model a special case of the full "transition" model employed in Jackman (2000*b*) (hence our choice of nomenclature[9]. This model is based on analyzing the transitions from a lagged y of zero or one to a current y of zero or one (based on simple first order Markov assumptions), allowing for different processes based on the lagged value of y. While in principle these two processes could be based on totally different independent variables, it is notationally most convenient (and also commonly, though perhaps incorrectly, assumed) that the same variables affect both transition processes, but with different parameters. With this simplifying assumption, the transition model has

$$P(y_t = 1 | y_{t-1} = 0) = \text{Probit}(\mathbf{x}_t \beta) \tag{7}$$
$$P(y_t = 1 | y_{t-1} = 1) = \text{Probit}(\mathbf{x}_t \alpha) \tag{8}$$

which can be writen more compactly as

$$P(y_t = 1) = \text{Probit}(\mathbf{x}_t \beta + y_{t-1} \mathbf{x}_t \gamma) \tag{9}$$

where

$$\gamma = \alpha - \beta. \tag{10}$$

Thus 5 is the case of 9 in which the constant coefficient in γ is ρ and the coefficients on all $\mathbf{x}$ terms in γ are 0. This model is well known in both biometrics and econometrics (for example, Amemiya, 1985; Ware, Lipsitz and Speizer, 1988) and has been used, for example, by Przeworski et al. (2000) in their work on transitions to democracy. In parlance, the restricted transition model is "observation-driven," while the lagged latent model is "parameter-driven."[10]

While the similarity of the transition model and the event history approach proposed in Beck, Katz and Tucker (1998) (BKT) has not always

[8] Estimation is via Markov Chain Monte Carlo (MCMC), as described in Appendix B. Estimation is difficult because we do not observe y^*, but only its sign.

[9] See Appendix A.

[10] Because we are in a binary dependent variable world, where we must assume that the variance of the underlying latent errors is one, it makes no difference whether we estimate the full transition model by probit (or logit), or whether we estimate models separately on the two subsets of data; either way the estimates of α, β and γ will be identical. This is different from the continuous dependent variable case, where the estimates of the variance of the errors differ depending on whether we do one big regression or two subset regressions.

been noticed, the two approaches have much in common. BKT propose to only model years where $y_{t-1} = 0$, that is, to only estimate Equation 7, dropping all observations where $y_{t-1} = 1$. They then note that estimating Equation 7 is equivalent to assuming no duration dependence in the isomorphic duration model. They thus propose to add a series of dummy variables, $t_{i,j,k}$ to Equation 7 where the dummies mark the time that has elapsed since the previous occurrence of an "event" ($y_{i,j-k} = 1$). We return to the similarity of the transition and event history approaches in the following subsection.

Discussion

The consequences of serial correlation in binary data are only partially understood. For instance, it is known that the probit estimates of β obtained under the assumption of independence (i.e., ignoring serial correlation of the error) remain consistent and asymptotically normal in the presence of serially correlated disturbances (Gourieroux, Monfort and Trognon, 1982; Poirier and Ruud, 1988), although the probit standard errors are no longer accurate. It should be noted that these results are for a single time series, not BTSCS data. The asymptotics here refer to the length of the time series (T), which in the BTSCS setting may not be large. Thus these asymptotic results provide little comfort, and, moreover, we are unaware of any characterizations of the finite sample properties of this estimator.

Gourieroux, Monfort and Trognon (1982) proposed tests of the null hypothesis of independence, against the alternative of ARMA disturbances. In particular, a score-based test of AR(1) errors is easily implemented, requiring only estimates of β obtained under the null of independence. The test has many parallels with well-known tests of serially correlated residuals from regression models for continuous dependent variables. In the case of a binary response model, they define generalized residuals as

$$\hat{\epsilon}_{it} = E(\epsilon_{it}|y_{it}, x_{it}, \beta)$$

with

$$E(\epsilon_{it}|y_{it} = 0, x_{it}, \beta) = \frac{-\phi_{it}}{1 - \Phi_{it}} \quad (11)$$

$$E(\epsilon_{it}|y_{it} = 1, x_{it}, \beta) = \frac{\phi_{it}}{\Phi_{it}} \quad (12)$$

$$(13)$$

where $\phi_{it} = \phi(x_{it}\beta)$ and $\Phi_{it} = \text{Probit}(x_{it}\beta)$. These residuals can be estimated given MLEs of β, $\hat{\beta}$. Their score test for AR(1) residuals is then

$$s = \sum_{i=1}^{n} \sum_{t=2}^{T_i} \hat{\epsilon}_{it} \hat{\epsilon}_{i,t-1} \quad (14)$$

which has variance

$$V(s) = \sum_{i=1}^{n} \sum_{t=2}^{T_i} \frac{\phi_{it}^2}{\Phi_{it}(1-\Phi_{it})} \frac{\phi_{i,t-1}^2}{\Phi_{i,t-1}(1-\Phi_{i,t-1})}. \tag{15}$$

Under the null of independence, $z = s/\sqrt{V(s)} \overset{\text{asy}}{\sim} N(0,1)$.

Assuming that this score test rejects the null of serially uncorrelated errors, we can either attempt to "fix" the errors or model the dynamics. We prefer the latter, and so only briefly discuss the former. Fixes which do not explicitly model the dynamics include Huber's (1967) robust standard errors (treating each country as a cluster) and Liang and Zeger's (1986) "generalized estimating equation (GEE)" Neither of these model the dynamics, in the sense that both use the ordinary probit for predicting y.[11] It is probably the case that either of these methods are an improvement on the ordinary probit (and almost certainly cannot hurt), but our interest is in attempting to model dynamics. We do, however, show the results from these two methods, and do find them more in line with estimates that we believe are superior to ordinary probit.[12]

While we present results on a model with serially correlated errors, we find this model as unappealing in the BTSCS context as in the standard time series context. In BTSCS terms the serially correlated error model (with AR1 errors) is

$$y_t^* = \mathbf{x}_t\beta + \epsilon_t \tag{16}$$

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t \tag{17}$$

where the ν are independently and identically distributed. We find this an odd model in the standard time series context, and equally odd here. The serially correlated errors model asserts that a one unit change in an *unmeasured* variable (in political science, errors are simply errors of the observer, that is,

[11]The Huber method simply fixes the standard errors of the ordinary probit, while leaving the estimated $\hat{\beta}$ intact; the GEE makes assumptions about the interrelationship of y_t and y_{t-1} and then uses those assumptions to perform "quasi-maximum likelihood." While this is a well known and often used method, it is a bit of a black box. While some political scientists, such as Zorn (2001) have found the GEE to be useful, it clearly is not an attempt to model the dynamics.

[12]We also do not discuss the use of fixed or random effects, which are another way to model the interrelationship of the observations without explicitly modelling the dynamics. Note that fixed effects would lead to our losing all observations on the approximately 80% of nations that never failed; all comparisons would be restricted to the timing of failure amongst those nations with at least one failure. Such a loss of information seems foolish for BTSCS data (Beck and Katz, 2001). The situation is very different for binary panels (that is, with small T), where random effects, or fixed effects using conditional logit, might be the best we can do to model interrelationships amongst the observations for any unit.

variables that we happen not to have measured) immediately increases y by one unit, with that effect declining exponentially at a rate of $1-\phi$ per year. But a one unit increase in a measured variable increases y by β units, with no dynamics whatsoever, that is, the effect of a change in some measured independent variable is only felt immediately. Since what variables are in the $\mathbf{x}$ and what are in the ϵ are determined by what we can and choose to measure, why should the two types of variables be treated differently? The lagged latent model we prefer does not do so. Thus we do not pursue the serially correlated errors model in any detail here.[13]

The restricted transition and lagged latent models differ only in whether they include lagged realized y or the lagged latent y^* in the specification. Obviously the model with lagged realized y is much easier to estimate (much, much easier!), but the model with lagged latent y is becoming easier to estimate and it is also much easier to interpret than is the restricted transition model. Thus the long run effect of a unit change in an independent variable is easy to calculate ($\frac{\beta}{1-\rho}$) whereas it is hard to calculate this for the restricted transition model. For the latter, one would have to use simulation, since long run impacts depend heavily on the probability of the latent y^* being converted to an observed 0 or 1.

But we should not choose on the basis on convenience. The two models are very different theoretically. Calling an occurrence of $y_t = 1$ an event, the question is whether past events make future events more unlikely, even if the prior event was itself unlikely. Thus Londregan and Poole (1990) argue that coups themselves cause coups, and so the lagged number of coups belongs in the specification. A more standard time series argument is that it takes time for a change in an independent variable to fully work its way through the system, and a simple general model for this is one of exponential decay. This would lead to the lagged latent variable model.[14] To see this formally, we can proceed as in the derivation of the Koyck (1954) exponentially distributed lag model. We can thus write

$$y_t^* = \mathbf{x}_t\beta + \mathbf{x}_{t-1}\beta\rho + \mathbf{x}_{t-2}\beta\rho^2 \cdots + \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} \tag{18}$$

and then transform as Koyck did, yielding Equation 6.[15] While there clearly will be situations where the restricted transition model is preferred on purely

[13] This is to say that the issue of whether the errors in either the restricted transition or lagged latent variable models are uncorrelated is not important; if they are, then the estimates of these models which assume serially uncorrelated errors will be inconsistent. There must be a Lagrange multiplier (score) test for this which should be easy to implement, but as of this moment we do not know the exact form of such a test. If BTSCS are anything like standard time series, we suspect that after including either a lag of y or the latent y^* there will be relatively little remaining serial correlation of the error, and usually not enough to do any statistical harm. But that is a conjecture.

[14] A partial adjustment story would also lead to the lagged latent variable model.

[15] This is a bit different than the original model of Koyck, which proposed that the

theoretical grounds, it seems likely that the lagged latent model, like its time series cousins, will often be the default choice.

The restricted transition is also an odd choice in that, as noted above, it is a special case of the transition model (assuming that one of the independent variables is a constant). While it may be preferred by the data to the full transition model, this is easy to test for and hence an odd place to start. It may be that the only parameter that differentiates transitions from 0 to 1 and 1 to 1 is the constant, but that should be a conclusion, not an assumption. The full transition model, on the other hand, seems quite sensible; should events following events be modelled the same as events following non-events? The answer seems obvious.

The message of the transition model is that we need to think about two separate theoretical processes, one of which tells us why events occur for the first time and the second of which tells us why they persist. The theories underlying these two processes may be similar or different. In BKT we argued that the process which leads to continuation of peace is different than the process which leads to continuation of war, and hence should be modelled differently. There may be some situations where the two transition processes are identical; again, this can be tested for, and should be a conclusion, not an assumption.

In many cases, we may have more interest in, say, the transition from a non-event to an event, or we may have a better understanding of the theory that drives such a transition. If this is the case, there is absolutely nothing lost by focusing only on those transitions, that is, on estimating models using only data until the first event is observed, dropping all the latter years of sequences of events. (In the epidemiology world, this distinction is between modelling incidence and prevalence.) If one believes that the observation-driven transition model is correct, *nothing* is lost by modelling only transitions to first events and using the appropriate subset of the data to estimate that model.[16] The only difference between BKT and transition

effect of the $\mathbf{x}$ died out exponentially but the errors are serially uncorrelated (and so have only immediate effect). This leads to a complicated model with a moving average error term (with a restricted MA coefficient). But if we make the reasonable assumption that the effect of the errors die out at the same rate as the effect of the measured independent variables, we end up with the simpler model with serially uncorrelated errors (Beck, 1991).

[16] We ignore the tricky problem of second spells of non-failure following a failure. In BKT we had several suggestions for modelling this. But the simplest assumption is that second events are independent of first events, so the first year of a non-event simply marks a new "spell" of non-events. In the empirical analysis below, there is no indication (based on trying methods detailed in BKT) that second spells of non-failure are different than first spells. But one would want to test for this, not assume it. Note the transition model, and most other models, assume that the probability of an event is only conditional on the prior year's observation rather than the entire event history that preceded that observation. Thus the probability of emerging from, say, the fifth failure in a nation is

models is that BKT does not attempt to model the transition from 1 to 1 or 1 to 0, and it attempts a more general model of the transition process from 0 to 1; but it is very clearly a model of that transition, and so at least the first half of the observation-driven transition model (Equation 7) is just a special case of what was proposed in BKT. It is easy to test whether this specialization is correct, that is, do the $t_{i,j,k}$'s belong in the specification?[17]

Before discussing how the various models work with the state failure dataset, it is appropriate to describe the data and to lay out the independent variables that will be used in all of our specifications.

3 Data

Our dependent variable for this study is "state failure," which captures severe political crisis exemplified by such recent events as Bosnia, Somalia and Afghanistan. In these instances, violent conflict or humanitarian crisis so weakened the institutions of governance that they could no longer exercise civil authority or maintain political order. While there exist many theories as to the conditions that generate failures—ranging from poverty to rising expectations to the presence of extractable natural resources—most observers are in agreement that the factors which cause failure are different from the factors which end it. Once the spark has been lit and failure sets in, the theory goes, the security dilemma that arises is so great that participants will not be able to reinstall a peaceful regime absent outside third-party guarantees of safety to all sides. Thus failures are a good candidate for studying via transition models, so that we do not a priori force the beginnings and ends of failure to have equal but opposite causes.

The data consist of annual observations on 147 countries between 1955 and 1997. As mentioned above these data are not rectangular, as some countries did not exist for the entire time period. Indeed, there were 50

assumed to be the same as the probability of emerging from the first failure.

[17]To see this, note that the standard formula for discrete time event history data is that the probability of observing a spell of length t, that is, non-failures in years $1,\ldots,t-1$ and then failure in year t is $P(y_t = 1, y_{t-1} = 0, y_{t-2} = 0, \ldots, y_1 = 0) = P(y_t = 1|y_{t-1} = 0)P(y_{t-1} = 0|y_{t-2} = 0 \ldots P(y_2 = 0|y_1 = 0)P(y_1 = 0)$. The discrete time event history approach estimates each of these terms as a probit (or other binary dependent variable model). Note that the transition model is a special case of this, since it assumes that $P(y_t = 1|y_{t-1} = 0) = P(y_{t-1} = 1|y_{t-2} = 0) = \cdots = P(y_2 = 1|y_1 = 0)$ whereas the event history approach allows these probabilities to differ as a function of time (the time dummies, the discrete time analogies of the baseline hazard in Cox's (1972) semi-parametric model). Note that the transition model drops observations for $t = 1$ since we cannot observe $t = 0$ data to condition on; the event history approach just uses the unconditional data at $t = 1$. While this difference can lead to an annoying difference in sample period for the two types of analysis, this is hardly a major issue if care is taken.

countries coded in 1955, as opposed to 138 today. One reason for this is that the recent political restructuring undertaken by many former Socialist Bloc countries has created a number of new states; for example, Azerbaijan only arrived in 1994. Moreover, some countries changed name during the time period; for instance, Czechoslovakia is in the data set until 1992 and then reenters as the Czech Republic in 1993. New states and states that changed identity are treated as new cases in our analysis.

While the complete collapse of state authority is rare—only 18 cases have occurred in the last 45 years—partial and sporadic failures are much more common, comprising 90 cases in the same time period. A set of coding rules was therefore needed to identify significant loss of government authority and the breakdown of the rule of law. Accordingly, states can fail in any of four ways: they can experience an ethnic war, a revolutionary war, an adverse regime change, or a genocide/politicide. These are defined as follows.[18]

- *Revolutionary wars* (50 episodes/359 case-years) are episodes of violent conflict between governments and politically organized groups (political challengers) that seek to overthrow the central government, to replace its leaders, or to seize power in one region. Conflicts must include substantial use of violence by one or both parties to qualify as "wars."

  "Politically organized groups" may include revolutionary and reform movements, political parties, student and labor organizations, elements of the armed forces, or the regime itself. If the challenging group represents a national, ethnic, or other communal minority, the conflict is analyzed as an Ethnic war, below. At a minimum, each party must mobilize 1000 or more people (armed agents, demonstrators, troops) and an average of 100 or more fatalities per year must occur during the episode.

- *Ethnic wars* (60 episodes/692 case-years) are episodes of violent conflict between governments and national, ethnic, religious, or other communal minorities (ethnic challengers) in which the challengers seek major changes in their status. Most ethnic wars since 1955 have been guerrilla or civil wars in which the challengers have sought independence or regional autonomy. A few, like the events in South Africa's black townships in 1976-77, involve large-scale demonstrations and riots aimed at sweeping political reform that were violently suppressed by police and military. Rioting and warfare between rival communal groups is not coded as ethnic warfare unless it involves conflict over political power or government policy.

[18]Full definitions are provided in Esty, Goldstone, Gurr, Surko and Unger (1995).

As with revolutionary wars, the minimum thresholds for including an ethnic conflict in the problem set are that each party must mobilize 1000 or more people (armed agents, demonstrators, troops) and an average of 100 or more fatalities per year must occur during the episode. The fatalities may result from armed conflict, terrorism, rioting, or government repression.

- *Adverse or disruptive regime transitions* (87 episodes/255 case-years) are defined as major, abrupt shifts in patterns of governance, including state collapse, periods of severe elite or regime instability, and shifts away from democratic toward authoritarian rule. Abrupt but nonviolent transitions from autocracy to democracy are not considered state failures and, thus, are not included. Two criteria were used to identify potential transitions: an abrupt shift of 3 points or more on the Polity scales of Democracy or Autocracy scores, or a transition period defined by the lack of stable political institutions.[19]

- *Genocide/Politicide* (36 episodes/265 case-years) is the promotion, execution, and/or implied consent of sustained policies by governing elites or their agents—or in the case of civil war, either of the contending authorities—that result in the deaths of a substantial portion of a communal group or politicized non-communal group. In genocides the victimized groups are defined primarily in terms of their communal (ethnolinguistic, religious) characteristics. In politicides, by contrast, groups are defined primarily in terms of their political opposition to the regime and dominant groups.

  Geno/politicide is distinguished from state repression and terror. In cases of state terror authorities arrest, persecute or execute a few members of a group in ways designed to terrorize the majority of the group into passivity or acquiescence. In the case of genocide/politicide authorities physically exterminate enough (not necessarily all) members of a target group so that it can no longer pose any conceivable threat to their rule or interests.

We code *FAILURE* as a binary variable equal to one if any one of the four modes of failure is present for a given country in a given year; otherwise, the variable takes on the value zero. Due to the difficulties in determining the start and end dates of failure episodes, any string of three or fewer non-failure years between failure episodes was also coded as a failure. Overall, our data set contains 4596 country-years from 1955 through 1997. Of these, 849 were failures, or 18.47 percent overall. Figure 1 shows the number of

[19] Polity scores are taken from Jaggers and Gurr (1995), updated through 1997.

new failures each year, and the total number of failures present at any given point in time.

As might be gathered, these failures did not necessarily occur in isolation. Of the country-years with failure, 566 had only one mode of failure, 189 had two modes, 65 had three, and 11 displayed all four modes simultaneously.[20] Furthermore, some failure modes were less likely to occur in isolation. In particular, our data shows no instances of geno/politicides occurring by themselves; they are all coincident with at least one other mode of failure. In all, 108 different cases of failure occurred, with an average duration of 7.86 years. Of the 147 countries in the study, 68, or 46 percent, had no failures at all, accounting for 42 percent of the country-years in the data. Of the remaining 79 countries, the average percent of years in failure was 36 percent, ranging from 2.33 percent (Greece and Mexico) to 100 percent (Angola, Azerbaijan, Moldova and India), as shown in Figure 2.

The independent variables used to explain failure include *OPEN*, trade openness (defined as exports plus imports over GDP as a decimal), *INFMORT*, logged infant mortality, *POPDENS*, logged population density, and *DEMOC*, democracy, coded as 1 if the country's polity score is above 0. Summary statistics are provided in Table 1.

Table 1: Summary Statistics

Variable	*Mean*	*Std.Dev.*	*Min*	*Max*
OPEN	.61	.40	.02	.44
DEMOC	.310	.46	0	1.00
INFMORT	3.97	.96	1.45	5.38
POPDENS	3.51	1.46	.07	8.59
FAILURE	.18	.39	0.00	1.00
N=4596				

Missing Data

The variables we used contain some missing data; almost all of this was either for very small countries (usually not even in the Correlates of War list of countries) in their entirety, or for a variety of nations either before some period (typically 1960, but sometimes later) or after some period (typically either 1996 or 1997). Since our methods are only difficult to use if there is

[20]These unfortunates were Zaire 1964-65, Philippines 1972, Iran 1981, Somalia 1989-90, and Angola 1992-96.

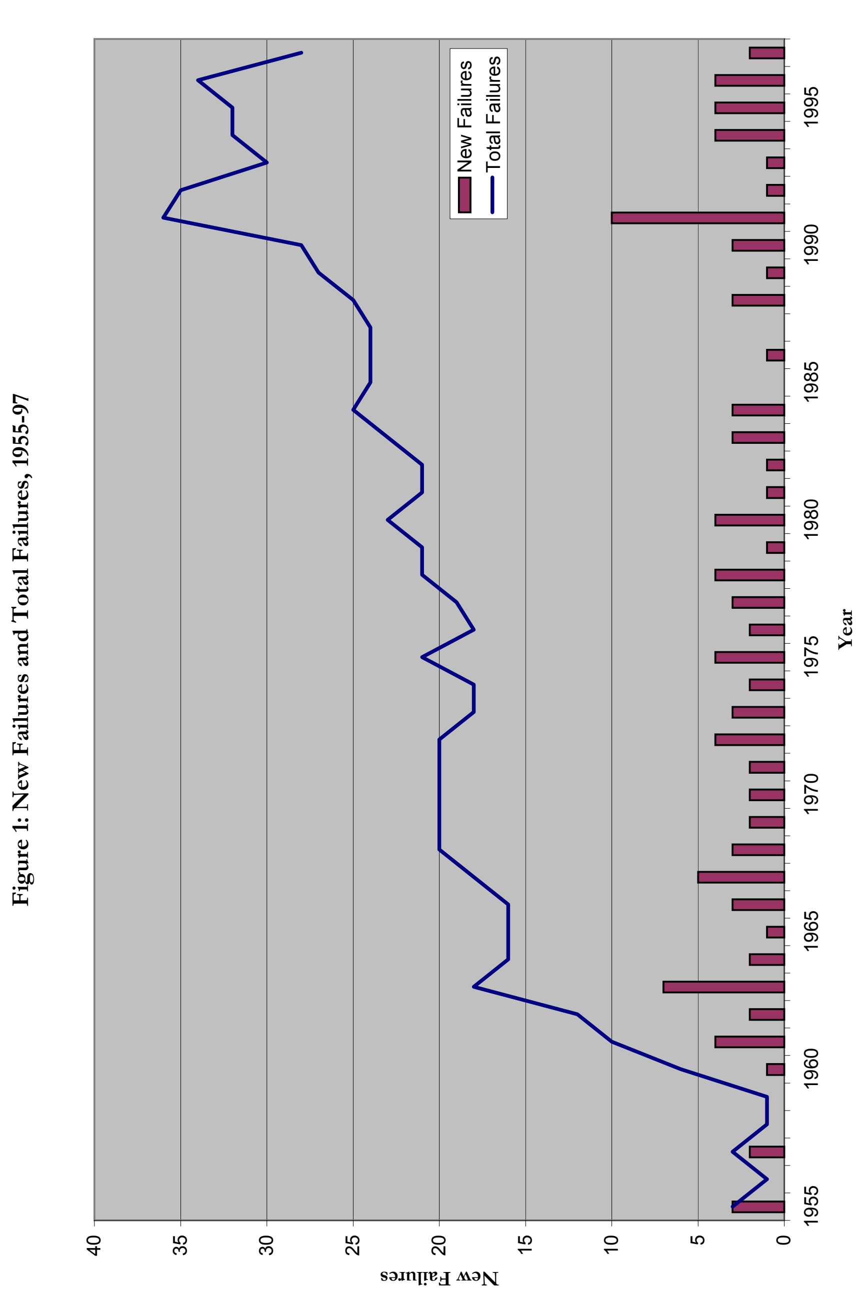

Figure 1: New Failures and Total Failures, 1955–97

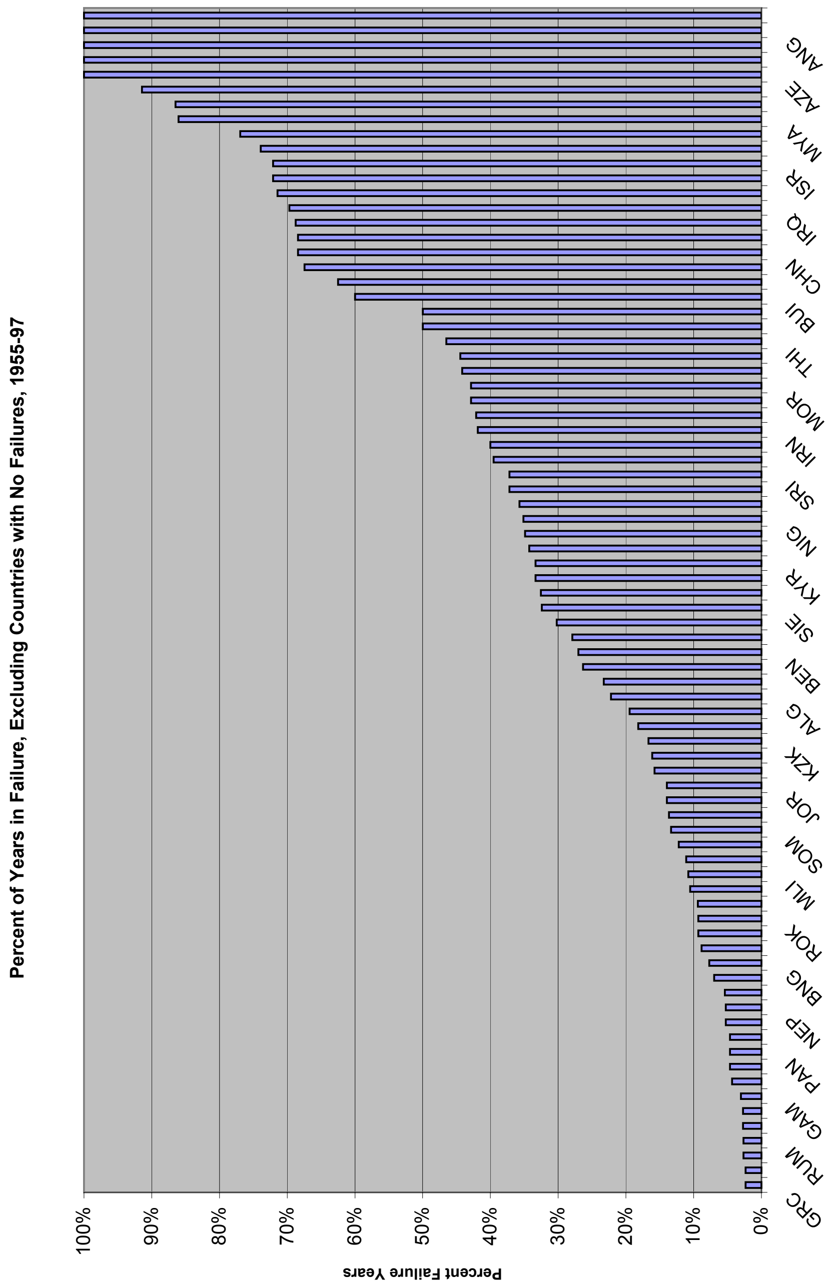

Figure 2: Percent of years in failure, excluding countries with no failures, 1955–97

missing data in the interior of a country's sample period, we simply dropped all observations with missing data that were either at the beginning or end of the overall observation period. In a few cases, such as Vietnam, entire countries were dropped from the analysis due to missing independent variables; for Vietnam the problem was the lack of reliable data for infant mortality or trade openness. There were a few missing interior observations on the *DEMOC* variable. Since *DEMOC* is a dummy variable indicating that a nation was democratic, and given what we know of the coding decisions in the Polity data set, we decided to treat those few remaining missing observations on *DEMOC* as zeros.[21]

4 Results: Naive models (and simple fixes)

We begin our presentation with the ordinary probit model. While we do not expect many political scientists would actually estimate this model, it does provide a baseline for comparison. Results are in Table 2; for completeness this table also shows the Huber robust standard errors (with clustering by country).

Table 2: Ordinary Probit Estimates of State Failure Model; All Failures

	Ordinary Probit			GEE	
Variable	$\hat{\beta}$	SE	Robust SE	$\hat{\beta}$	SE
OPEN	−.71	.08	.23	−.31	.12
DEMOC	−.40	.07	.26	−.36	.09
INFMORT	.25	.04	.13	.28	.09
POPDENS	.19	.02	.07	.19	.05
Constant	−2.12	.21	.68	−2.45	.50
$\hat{\rho}$				.86	
N=4596					

If we believed the ordinary probit, we would believe that all four independent variables have a highly statistically significant effect on the probability of state failure. Since we are always going to use the same specification, we

[21]There were approximately 10 such cases. While we did not choose our country list based on the Correlates of War list, had we done so we would have eliminated about half our missing data cases. Since our interest here is orthogonal to the missing data issue, we felt it made most sense to do what we did. If this were a more substantive paper we would clearly have to revisit the issue of missing data.

can generally just compare coefficients and standard errors for the different dynamic models. But to get a sense of what the various coefficients tell us about the probability of state failure, note that if we set all other variables at their mean, the probability of state failure for a democracy is .10 while it rises to .19 for a non-democracy; a confidence interval for this difference is (.06, .11). Looking at non-democracies, we if move *INFMORT* from its 75th percentile to it 25th percentile, we change the probability of state failure from .14 to .24; a confidence interval for this difference is (.08,.13). A similar move in openness from the 75th percentile to the 25th percentile increases the probability of state failure from .16 to .24; a confidence interval for this difference is (.06,.10). Finally, a similar move in population density increases the probability of state failure from .15 to .25; a confidence interval for this difference is (.08,.12).[22] Thus not only do the estimated impacts look very statistically significant (with z-scores ranging from 6 to 10), but they are also substantively large, since a change in the probability of state failure of even a few percent is substantively very meaningful.

For the ordinary probit model presented in Table 2, the score test overwhelmingly rejects the null of independent disturbances ($z = 50.94$); this is not surprising, given that the data comprise largely of uninterrupted sequences of non-failure and failure.[23] Figure 3 also makes this apparent, with the four clusters of residuals in each quadrant corresponding to the four possible pairs in our data: $(y_{t-1}, y_t) = (0, 0)$ in the bottom left and $(1, 1)$ in the top right. The few transitions in the data in the "off-diagonal" quadrants, where a negative generalized residual is followed by positive generalized residual, or vice-versa.

Clearly, then, assumptions of independence in these data are untenable, and we now consider models designed to directly tap the dynamics in the data or methods which attempt to "fix" the problems of ordinary probit. In this section we briefly consider the latter.

The Huber standard errors leave the underlying model intact but do correct for the statistical dependence of different yearly observations for the same country. In Beck and Katz (1997), we presented simulation results that showed the Huber standard errors (grouping on unit) to be much more accurate than the ordinary probit standard errors in the presence of serially correlated errors. The Huber standard errors show that the ordinary probit overstates confidence (z-scores by a factor of 3 or 4). The effect of democ-

[22]All computations other than the Markov Chain Monte Carlo computations were done using Stata Version 7 (with some graphs and more complicated statistics produced using Splus); the probabilities were calculated using Clarify. Individual probabilities based on the ordinary probit have a standard error of about one percent.

[23]To be precise, 97.44 percent of non-failures were followed by failures, while 89.84 percent of failures were followed by further failures.

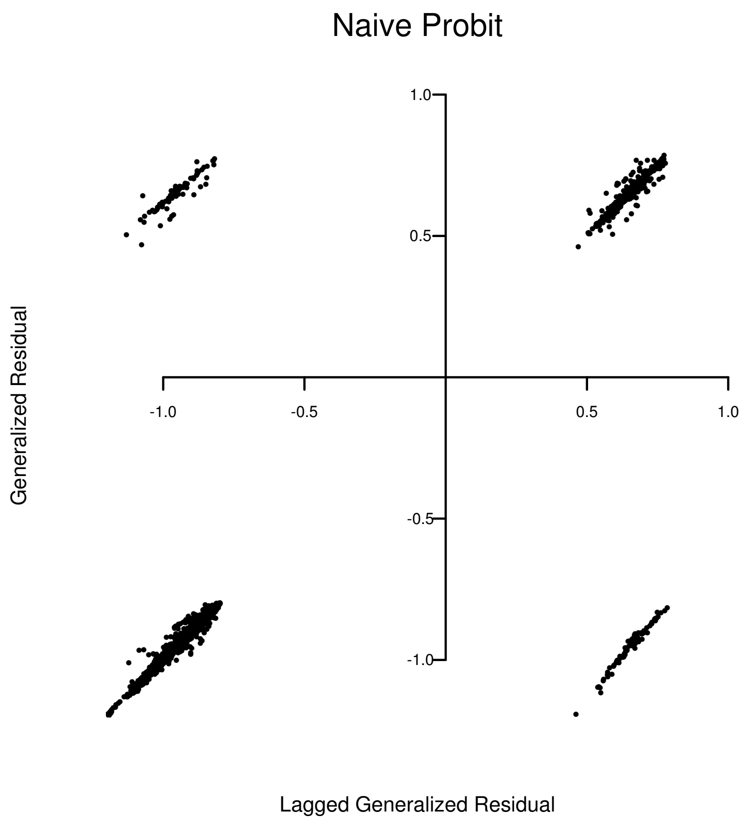

Figure 3: Residuals vs Lagged Residuals, Ordinary Probit

racy on state failure is no longer statistically significant and the three other z-scores now range from 2 to 3.

While the ordinary probit results with Huber standard errors yield more realistic z-scores than do the basic ordinary probit estimates, the Huber procedure does not attempt to model the dynamics, nor does it change any insights about either estimated coefficients or the way in which the independent variables affect state failure. Surely if one were limited to the ordinary probit model, one would at least use Huber standard errors. But we are not so limited.

For completeness we also present the GEE results, assuming that the dependent variable is correlated as a first order autoregressive process. Not surprisingly, the observations are highly correlated (.86). The z-scores for the GEE analysis are closer to the more realistic Huber z-scores than what was obtained from the ordinary probit. But only the *DEMOC* coefficient changes very much. As noted above, the GEE uses the same model to estimate the probability of failure as does the ordinary probit. While the GEE results here are probably superior to the ordinary probit, we can do better by trying to model the dynamic process, that is, allowing the dynamics to affect the probability of state failure. We turn to these more reasonable specifications now.

5 Results: Transition models

The standard transitional model (Equation 9) is both easy to estimate and interpret. The model results are Table 3.

Table 3: Transition Model; All Failures; Duration Independent

	$y_{t-1} = 0$		$y_{t-1} = 0$	
Variable	$\hat{\beta}$	SE	$\hat{\beta}$	SE
OPEN	$-.39$	.16	$-.45$	.22
DEMOC	$-.55$	.17	.55	.23
INFMORT	.17	.08	.10	.12
POPDENS	.08	.04	.07	.06
Constant	-2.12	.21	.75	.67
	N=3632		N=817	

We note again that one would obtain the exact results in Table 3 if one first did a probit of state failure using only observations following a non-failure (this would give the two left columns) and then the same probit using only observations following a failure (dropping the first observation for each country).

Since we have already seen how the various coefficients affect the probability of state failure in the ordinary probit analysis, we can limit ourselves to comparing coefficient estimates in comparing the ordinary probit and transition models. The first thing to note is that the constant term is almost three points greater in transitions from a state failure than in transitions to a state failure; thus knowing whether a nation is transitioning from a previous failure or a previous non-failure has about seven times the effect on current state failure as knowing whether a nation is a democracy.

The transition model does more than simply shift the intercept in the ordinary probit as a function of prior state failure; it also allows all coefficients to differ depending on whether a nation is transitioning from a non-failure or a failure. While the coefficients of three of our four independent variables do not change as a function of prior state failure, the coefficient on *DEMOC*, perhaps our most interesting political variable, changes dramatically.[24] Democracies are significantly less likely to fail if they had not failed last year, but they are significantly more likely to fail if they failed last year. That is, democracy keeps states from failing, but once they do fail, it actually makes them more likely to continue to fail. While the difference is probably not statistically significant, the ordinary probit, if anything, *understates* the role of democracy in preventing state failure. Thus, for example, a democracy (with all other independent variables set at their medians) is about 10 points less likely to fail than a non-democracy if it was not failing last year, but is about 10 points more likely to fail if it was failing last year.

For completeness, we note that the z-scores of the substantive independent variables are much smaller than the corresponding scores for the ordinary probit, and only slightly larger than those obtained with either Huber standard errors or the GEE model. We also note that the Huber standard errors for the transition model (not shown) are within 10% of the standard errors in Table 3, indicating that the transition model does a good job of taking account of the clustering of observations within a country.

To further evaluate the performance of these models, we examine their ROC (Receiver Operating Characteristic) curves.[25] This curve plots a model's

[24]A test of the hypothesis that the other three variables do not change as a function of whether they are transitioning from a prior failure or not yields a $\chi^2_3 = .32, P < .96$. Estimates of the three restricted coefficients are similar to those in Table 3 with z-scores close to those in that table.

[25]The military origins of this measure account for its strange locution. Early signal theory was concerned with the ability of an operator sitting at a radar screen to perceive an enemy ship, say, and distinguish it from a friendly one. This insight was taken up by biomedical researchers interested in the ability of tests to generate true positives and

performance as one continuously changes the cutoff criterion for counting a given observation as a positive finding, with the percent of true positives (the test's sensitivity) on the y-axis and the false positives (1-specificity, in parlance) on the x-axis. A higher cutoff threshold will cut down on the false positive, but reduce the true positives as well. A more aggressive lower cutoff will catch more true positives but likely introduce more false positives.

Figure 4 shows ROC curves for the ordinary probit and transition models. Since our data is binary, flipping a coin could get the classification right 50 percent of the time. The 45 degree line therefore indicates the minimal possible model performance. As the ROC curve diverges from the 45 degree line, overall predictions improve in that fewer false positives and more true positives will be classified. Thus the area under the ROC curve, also known as the C statistic, gives a general measure of in-sample performance. As indicated in the figure, the ordinary probit estimation produces a C statistic of 0.72, while the transition model's C statistic is 0.96. Since the latter curve lies consistently to the northwest of former, by this measure the transition model dominates the ordinary probit.

Lest we become too sanguine about the relative performances of the models, though, Figure 5 shows the predicted transition probabilities for observations in each of the four possible classes: non-failures that remained non-failures in the next period, non-failures that transitioned to failures, and so on. As the figure indicates, the ordinary ("naive") probit model nearly always predicts that a country will be in non-failure the following year, even if it is currently in failure; its maximum probability of failure never exceeds about 60%. This should be no surprise, given that the ordinary probit estimator cannot distinguish which state the system is currently in, so its predictions are dragged down by the prevalence of 0's in the data set. By comparison, the transition model correctly discovers that a country in failure is likely to remain there, and likewise with a country currently in non-failure.

This superior performance in predicting 0-0 and 1-1 observations is the basis for the transition model's dominance in the ROC curve analysis; after all, these account for 96.2% of the data. On the other hand, it is also true that the transition model actually does *less* well in predicting state transitions; that is, the 0-1 and 1-0 pairs. Table 4 shows the four models' average predicted probabilities for each possible transition class. Just as a stopped clock is right twice a day, the ordinary probit's insistence that a country is always about 20 percent likely to be a non-failure makes it the best predictor for both types of transition. We do not pursue issues of out-of-sample prediction further here, and one would certainly not want to overemphasize the naive probit's ability to predict transitions. But it is clear that if one is

minimize false positives, and it has now found its way into the social science literature as well.

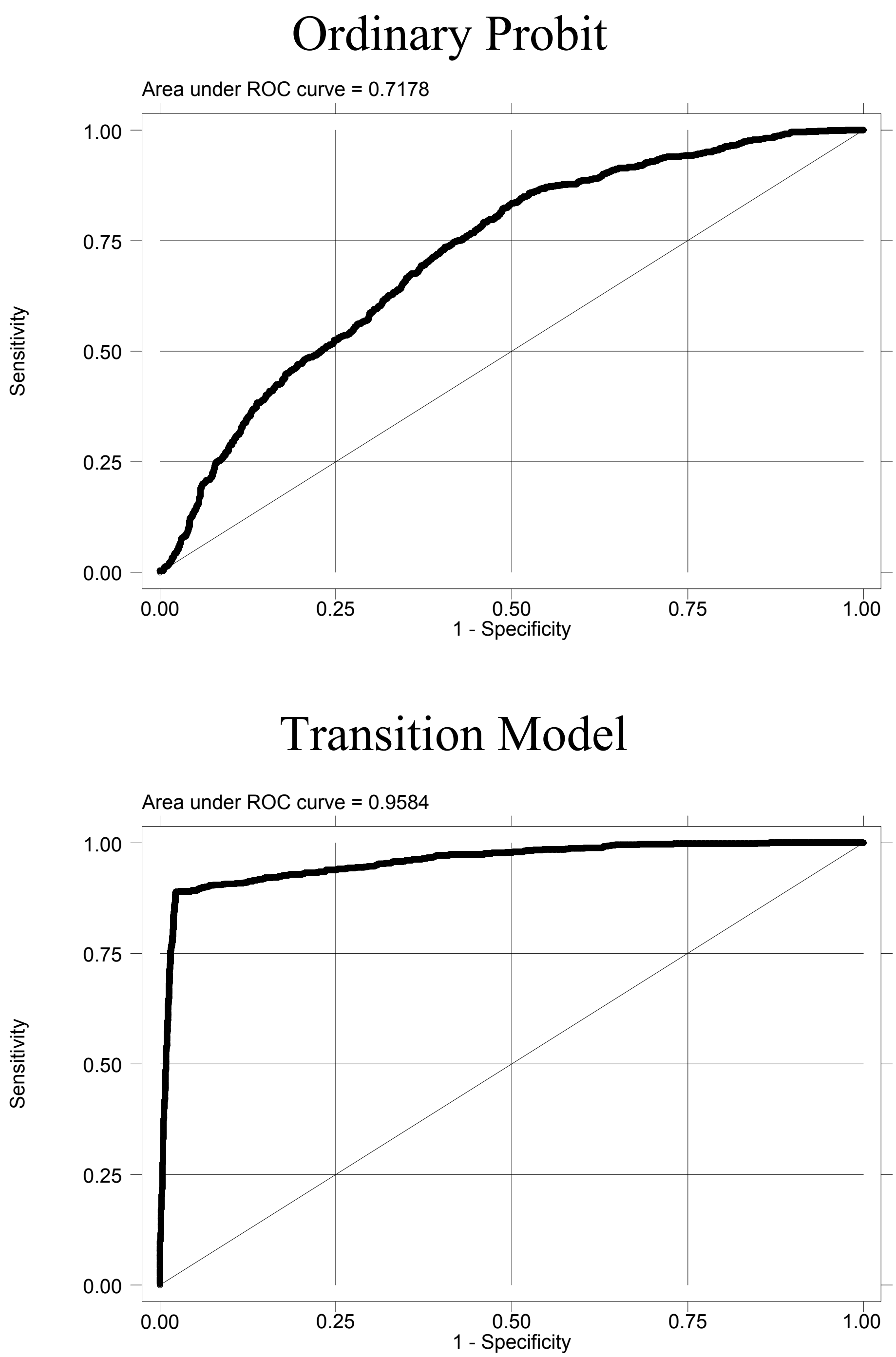

Figure 4: ROC Curves for Ordinary Probit and Transition Models

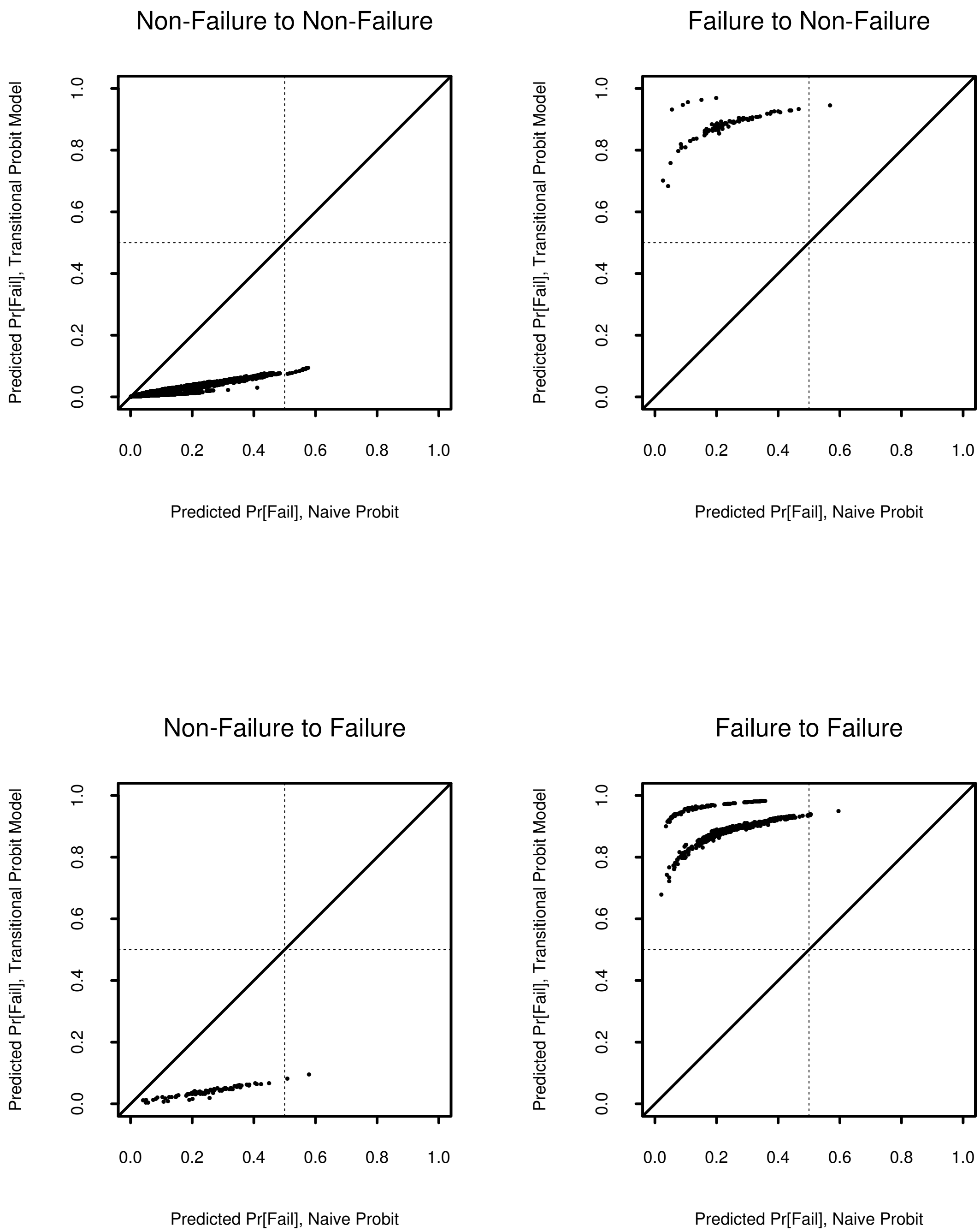

Figure 5: Comparison of Model Performance: Ordinary Probit vs. Transition

interested in forecasting, at the end of the day one should assess a model's performance in terms of (appropriately weighted) predictive power.[26]

Table 4: In-Sample Prediction Summary

Model	$0 \rightarrow 0$	$0 \rightarrow 1$	$1 \rightarrow 0$	$1 \rightarrow 1$
Ordinary Probit	0.17	0.25	0.23	0.25
Full Transition	0.03	0.04	0.88	0.90
Restricted Transition	0.03	0.04	0.89	0.90
Lagged Latent	0.07	0.10	0.65	0.70

Since the transition model is identical to estimating two separate probits depending on the prior state of *FAILURE*, the analysis is *identical* to what the BKT analyst would find in the absence of any duration dependence (except for the dropping of the first observation of any spell). Note that BKT assumed that one type of spell was either of more interest to the analyst than the other, or that the theory being tested applied to only one type of spell. Thus, in the IR dispute data, we analyzed spells of peace which were terminated by a dispute (which showed strong evidence of duration dependence), but did not analyze spells of disputes terminated by peace. This was partly because the data did not allow for the latter type of analysis; it contained only very short dispute spells. But it was also the case that the theory being tested related to the duration of spells of peace, not spells of disputes. Here we have enough data to examine both spells of non-failure and of failure, and the issue of the causes of transition from failure to non-failure is of interest.[27]

Let us begin with the transition model for spells of non-failure, that is, conditioning on $y_{t-1} = 0$. The estimates in the left columns of Table 3 assume duration independence, that is the probability of an exit from non-failure to failure (which given the conditioning in the data, is just P(*FAILURE*)) is assumed to not vary with t. As argued in BKT this is both a strong and testable assumption; if it is incorrect, then the results in the left columns of Table 3 will be wrong. To test the null hypothesis of duration independence for spells of non-failure, we added functions of time since last failure to the probit specification (both using dummy variables and splines). In all cases,

[26] King and Zeng (N.d.) make this point clearly in their assessment of the Task Force's previous claims regarding the predictive success of their model.

[27] Note that there is nothing in the transition model that forces us to use the same independent variables to model the transitions from $y_{t-1} = 0$ and 1; we could easily put together two entirely different models. But putting the two models in the one transition model is illusory, since, as we have seen, the transition model is really two independent models, based on the prior state of the binary dependent variable.

tests very clearly failed to reject the null hypothesis of duration independence. Thus the results for spells of non-failure, ignoring duration dependence, are not problematic.

The situation is different for spells of failure, that is, conditioning on $y_{t-1} = 1$. If we enter years since last non-failure (*FAILURE YEARS*) in the specification, we see in Table 5 that it is significant and it depresses the anomalous positive effect of democracy on length of spells of failure (to where it is no longer statistically significant).[28] Substantively, a nation with a spell of ten years of failure is about as unlikely to emerge from failure as is a democracy when compared to a non-democracy. At this point the results in Table 5 tell us there is some duration dependence in the data, but also more clearly tell us that we do not have a good model of when nations emerge from spells of failure. While this has something to do with our theories of why nations fail being better than our theories of why they emerge from failure, it also has something to do with the limited amount of data we have when conditioning on prior failure.

Table 5: Transition Model; Spells of Failure; Duration Dependence

Variable	$\hat{\beta}$	SE
OPEN	$-.35$	.22
DEMOC	.38	.24
INFMORT	.12	.12
POPDENS	.07	.06
FAILURE YEARS	.03	.01
Constant	.41	.68
	N=817	

6 Results: Lagged latent models

Before turning to the estimation of the lagged latent model, we present the results of estimating the apparently similar restricted transition model. While the difference between using lagged realized values of y and the lagged latent is not trivial, the models do appear superficially to be similar. Since

[28]We only show the linear analysis here. There is some indication that the effect of *FAILURE YEARS* is strongest early in a spell, and after about 10 years disappears. But with only 817 observations on years of failure, it is hard to be sure this effect is real; there is not enough data for the semi-parametric analysis we prefer, and so we limit ourselves to the model with the simple linear term.

this is a special case of the transition model, and we have already seen that the effect of democracy on state failure changes dramatically depending on whether we are modelling entry into failure or exit from failure, we know the transition model is preferred to the restricted transition model for this data set and specification. But even so, it is interesting to compare the restricted transition results, shown in Table 6, with our prior results.

Table 6: Restricted Transition Model; All Failures

Variable	$\hat{\beta}$	SE
OPEN	$-.37$	.12
DEMOC	$-.17$	.11
INFMORT	.20	.06
POPDENS	.10	.03
FAILURE lagged	3.12	.08
Constant	-2.87	.34
N=4449		

These results are clearly closer to those from the transition model than those from the ordinary probit model. We note that the coefficient on the *DEMOC* variable is negative but statistically not significant in the restricted transition model; this is because it is averaging the two opposite signed effects in the full transition model.

When comparing the coefficients of the restricted transition (or transition model) with those of the ordinary probit model, we must remember that the former models are analogies of distributed lag time series models whereas the latter is the analogue of a static model. Thus, for example, in the ordinary probit model, a move from non-democracy to democracy has an immediate effect on the probability of state failure of (negative) 10 points. This effect takes place all in one period. In the restricted transition or transition model, the effect takes place over time. But unlike the standard time series models, we cannot simply estimate the long run effect of a change in an independent variable as a simple function of its coefficient and the coefficient on the restricted transition. This is because the coefficient on the restricted transition only is relevant if a change in the independent variable is large enough to change the dependent variable from 0 to 1 or vice versa. One could compute long run effects (and associated standard errors) by simulation, but this is not quite as simple as dividing two coefficients.

The restricted transition (and transition) models both condition on prior state failure (or non-failure). Thus they make the substantive claim that failure itself makes future failure more likely (and non-failure itself makes future non-failure more likely). This is the identical claim made by Londregan and Poole (1990) about coups. From a policy perspective, this means that if we want to avoid state failure, we should attempt to keep states from failing by whatever means, which, if done, will prevent future state failure even with no changes in any of the variables which affect state failure. Failure breeds failure and non-failure breeds non-failure. This is what Heckman (1981) calls true state dependence.

Alternatively, there may simply be persistence in the underlying latent variable, that is, a change in an independent variable may only affect *FAILURE* over time, with the full impact being phased in exponentially. This seems like a generally plausible story. Note that if one accepts this story, then it does not matter whether a state failed or not last year, all that matters is the value of its latent propensity to fail last year. Thus a lucky state, which does not fail in spite of a high propensity to fail, is no less likely to fail this year because it was a non-failure last year. The policy implication of this model is that we must manipulate the relevant independent variables, not simply the outcome. We show the results of estimating such a lagged latent variable model (Equation 6) in Table 7.

Table 7: Lagged Latent Model; All Failures

	Short Run		Long Run	
Variable	Posterior Mean	SD	Posterior Mean	SD
OPEN	$-.179$	.046	$-.92$	.24
DEMOC	$-.087$	.038	$-.45$	.20
INFMORT	.065	.019	.33	.10
POPDENS	.041	.010	.21	.05
y^*_{t-1}	.805	.018		
Constant	$-.472$	.115		
N=4449				

If we compare the short run estimates in Table 7, that is the estimates of β in Equation 6, to the ordinary probit estimates in Table 2, they look quite different. But we must remember that the β in the lagged latent model are only short term effects, so the right comparison to the ordinary probit are the long-run effects, also presented in Table 7. The estimated long-run effects, as in continuous dependent variable time series analysis, are just $\frac{\hat{\beta}}{1-\hat{\rho}}$. It should be noted that the MCMC methodology makes it easy to compute the standard error of this long run effect. While the long run effects estimated by the lagged latent model are similar to the ordinary probit $\hat{\beta}$'s, the standard errors for the long run effects in the latent lagged model are much smaller than the standard errors from the ordinary probit, and comparable to those we have obtained in the various other dynamic models we have shown here. The ordinary probit assumes that a change in an independent variable is felt instantaneously; the estimate of ρ in the lagged latent model indicates it takes many years for a change in an independent variable to have its full impact. But the long-run impacts implied by the latent lagged model are considerably greater than those implied by the transition model (which, like the ordinary probit estimates, assume that all effects occur instantaneously).

The ROC curves for both the restricted transition and lagged latent models are nearly identical to the curve shown in the bottom half of Figure 4, with similar C statistics of 0.95 as well. This indicates that the ability to take history into account goes most of the way towards improving the models' performance. Exploring the lagged latent model a bit further, Figure 6 compares its predictions to those of the naive probit, similar to Figure 5 above. As shown, the lagged latent model is less extreme in its predictions than was the transition model, especially in predicting transitions away from failures. This difference is also apparent in the last two rows of Table 4: the restricted transition model produces transition probabilities quite similar to those of the full transition model, but the lagged latent has higher probabilities of both entering and leaving failure than either transition model; that is, it looks like a hybrid of the naive probit and the transition models. This last finding is due mainly to the fact that the predicted values of the y^* terms are not as extreme as the 0-1 lagged y terms, and so they will not shift the curves up or down by as great a factor.

For completeness, we also show the results of estimating a model with serially correlated errors. These results, in Table 8 show roughly the same dynamics as shown by the lagged latent model. This is similar to what we typically find in standard time series analysis. It is the case, however, that none of the substantive independent variables have a statistically significant impact, and all have substantive impacts similar to the short run estimates for the lagged latent model (but of course, the short and long run effects of the substantive independent variables are the same, which is why we do not like the serially correlated errors model). There is nothing in Table 8 to cause us to rethink our preference for the lagged latent model to the model with serially correlated errors.

7 Discussion and conclusion

Obviously one can draw no firm conclusions about general properties of models from one data set. We also warn that the data set we use has long se-

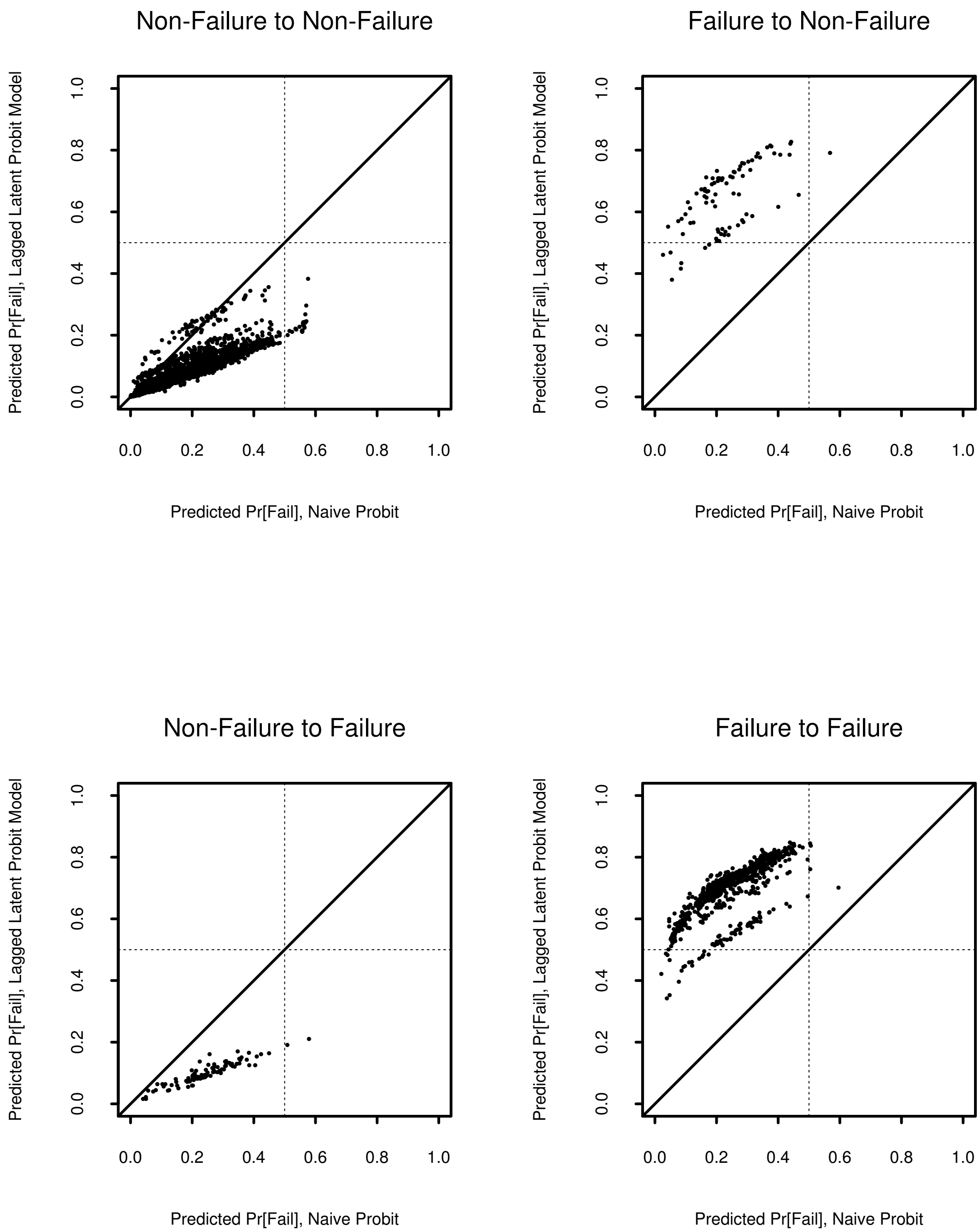

Figure 6: Comparison of Model Performance: Ordinary Probit vs. Lagged Latent

Table 8: Serially Correlated Errors Model; All Failures

Variable	Posterior Mean	SD
OPEN	$-.11$	.10
DEMOC	$-.10$	.09
INFMORT	.09	.07
POPDENS	.05	.05
$\hat{\rho}$	.86	.03
Constant	$-.72$	.40
N=4449		

quences of non-failure followed by various length sequences of failure; in this it looks a lot like the conflict data that we have analyzed elsewhere. Thus we have not tried the various models on data sets which consist of short sequences of 0's and 1's. While we intend to do this, as of now we have not found such a data set (at least that is not panel, all discussion here is for the BTSCS case).

There is no doubt that the ordinary probit should *not* be used if there is evidence of serious temporal correlation of the observations (within a unit). Evidence for this can either be the score test of Gourieroux, Monfort and Trognon we discussed, or perhaps a simple intuitive appeal that long sequences of 0's are unlikely to coexist with temporal independence.

While there are a variety of "fixes" for temporally dependent data, we prefer model-based approaches. Two appealing alternatives are the transition model (with the event history duration dependence extensions of BKT if necessary) or the lagged latent variable model. The latter is often thought too hard to estimate, but recent breakthroughs make it only very hard to estimate. While we should not underestimate the costs of estimating this model, it is very attractive in that it is the natural analogue of what we typically do in continuous dependent variable time series analysis. It also can be extended in some theoretically appealing ways. At this point we would suggest that BTSCS researchers faced with data like the state failure data use both of these approaches.

A Nomenclature

Since different disciplines have well accepted but differing terminologies for the models we discuss, and since some of our nomenclature is non-standard, Table 9 keys the disciplinal names to the various equations in the text.

Name	Equation	Dynamics	Interpretation	Estimation
Ordinary Probit	1–4	none	static	ML
Restricted Transition	5	ρy_{t-1}	state dependence	ML
Transition	9	$y_{t-1} x_t \gamma$	state dependence	ML
Lagged Latent	6	y^*_{t-1}	"latent Markov", propensity dependence, habit formation	MCMC
AR(1) errors	16–17	$\rho \epsilon_{t-1}$	latent serial dependence	MCMC/EM

Table 9: Nomenclature

B Identification, Estimation and Inference for the "Lagged Latent" Model

For clarity, we begin by briefly reproducing the derivation of the ordinary probit model, temporarily ignoring the i subscript indexing countries. Let $\theta_t = \Pr[y_t = 1]$ be the probability of a failure at time t. For ordinary probit, the probability of failure depends on covariates via a *latent regression function*

$$h(\theta_t) \equiv y_t^* = x_t\beta + \epsilon_t, \tag{19}$$

where x_t is a row vector of observations on k independent variables at time t, β is a vector of parameters to be estimated, $y_t^* \in \mathbf{R}$ is a latent dependent variable, observed only in terms of its sign, i.e.,

$$y_t = \begin{cases} 0, & \text{if} \quad y_t^* \leq 0 \\ 1, & \text{if} \quad y_t^* > 0 \end{cases} \tag{20}$$

and ϵ_t is a zero mean stochastic disturbance, identically and independently distributed for all t. For probit, we assume $f(\epsilon_t) = N(0, 1) \equiv \phi(), \forall\, t$; recall that the regression parameters β are identified only up to the scale factor σ, and so setting $\sigma = 1$ is a convenient normalization with no substantive implications. Note that with this latent regression approach we can express the joint probability for the observed data y_t in terms of the latent data, y_t^*: i.e.,

$$\Pr(y_1 = 1, \ldots, y_T = 1) = \Pr(y_1^* > 0, \ldots, y_T^* > 0)$$

Independence is a key assumption in the derivation of an ordinary probit model. In the present case, temporal independence means that

$$\Pr(y_1 = 1, \ldots, y_T = 1) = \Pr(y_1 = 1) \times \ldots \times \Pr(y_T = 1) = \prod_{t=1}^{T} \theta_t \tag{21}$$

or, in words, the joint probability for the data equals the product of the marginal probabilities, and so the log-likelihood can be simply written as sum of the observation-specific log-probabilities:

$$\ln \mathcal{L} = \sum_{t=1}^{T} \left[y_t \ln \theta_t + (1 - y_t) \ln(1 - \theta_t) \right]. \tag{22}$$

This log-likelihood is easily maximized to yield estimates of β with well-known asymptotic properties (consistency, and normality).

The "Lagged Latent" Model

Recall that the latent regression function for the "lagged latent" model is

$$y_t^* = x_t\beta + \rho y_{t-1} + \epsilon_t, \qquad |\rho| < 1$$

where the censoring rule in (20) links the latent and observed dependent variables. Given this model, the likelihood for the data given can now no longer be written as the product of the θ_t. Since y_t^* is a function of y_{t-1}^*, $\Pr(y_t = 1)$ is no longer independent of $\Pr(y_{t-r} = 1), \forall\ r \neq 0$; in turn, the joint probability of the data is no longer the product of the time-specific probabilities, Instead, the joint probability of the sequence of outcomes observed for dyad i:

$$\begin{aligned} \mathcal{L} &= \Pr[y_1, y_2, \ldots, y_T] \\ &= \int_{a_{i1}}^{b_{i1}} \int_{a_{i2}}^{b_{i2}} \ldots \int_{a_{iT_i}}^{b_{iT_i}} f_T(y^*|x\beta, \Sigma)\, dy_T^* \ldots dy_2^*\, dy_{i1}^*, \end{aligned} \tag{23}$$

where

$$(a_t, b_t) = \begin{cases} (-\infty, 0) & \text{if} \quad y_t = 0 \\ (0, \infty) & \text{if} \quad y_t = 1 \end{cases} \tag{24}$$

and $f_T(y^*|X\beta, \Sigma)$ is the T-dimensional probability density for the vector of latent variables $y^* = (y_1^*, \ldots, y_T^*)'$ (Poirier and Ruud, 1988, equation 2.8). For probit, this density is the multivariate normal PDF

$$(2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{\epsilon' \Sigma^{-1} \epsilon}{2}\right],$$

with $\epsilon = (\epsilon_1, \ldots, \epsilon_T)'$, where $\epsilon_t = y_t^* - \rho y_{t-1}^* - x_t\beta, \forall\ t = 2, \ldots, T$ and $\epsilon_1 = y_1^* - x_1\beta$.

The likelihood function in (23) poses a ferocious maximization problem, bearing a close resemblance to the intractabilities presented by the multinomial probit (MNP) model for qualitative choice. In MNP, the likelihood function becomes increasingly complex as the number of choices increases; each choice adds another dimension to the integral in the likelihood. Here we have a time-series probit model with the likelihood for each country $i = 1, \ldots, n$ involving integration of a T_i-dimensional Normal density. Statistical software packages such as GAUSS or S-Plus will evaluate integrals of bivariate Normal densities, but in any interesting time series setting T_i will be much larger than 2 or 3!

Estimation by Bayesian Simulation

The recent advent of simulation-based inference makes time-series probit models tractable. In particular, Markov chain Monte Carlo (MCMC) is attractive for this particular problem.[29] Rather than attempt to evaluate the high dimensional integral in the likelihood function in (23), the latent y^*_{it} themselves can be recovered by successively sampling from the sequence of conditional densities $f(y^*_t|y^*_{t-1}), t = 2, \ldots, T$. This sampling algorithm is an example of Gibbs sampling, the workhorse of MCMC. A review of Gibbs sampling need not detain us here,[30] but the key idea is that conditional on the latent y^*_{it}, the parameter-driven transitional model is simply a regression model with a lagged dependent variable. The MCMC algorithm proceeds by (1) generating *imputations* or, more formally, samples from the conditional distribution for the latent y^*_{it}, to yield a complete set of data; (2) using that complete set of data to estimate β and ρ, and then sampling from their implied conditional distributions.

In this case we seek the posterior distribution for the unknown parameters and latent data,

$$\pi(\beta, \rho, y^*|X, y), \tag{25}$$

recalling that X and y are the observed data. The MCMC approach begins by decomposing the joint posterior distribution into the two conditional distributions

$$g_1(\beta, \rho, |y^*, X, y)$$

and

$$g_2(y^*|\beta, \rho, X, y).$$

The MCMC algorithm here consists of successively sampling from each of these distributions, replacing β, ρ and y^* when they appear as conditioning arguments with the most recently sampled value for each. At the end of iteration m over each of the conditional distributions, the vector of sampled quantities $(\beta^{(m)}, \rho^{(m)}, y^{*(m)})'$ comprises the state vector of a Markov chain that has the joint posterior in (25) as its invariant distribution. When the Markov chain Monte Carlo algorithm has been run for a sufficiently lengthy period, each realization of the state vector is a draw from the joint posterior. These draws from the posterior distribution are saved and summarized for inference.

[29] Geweke, Keane and Runkle (1997) report that other simulation methods for dealing with the high-dimensional integrals required in multi-period probit models perform poorly as serial dependency increases; for instance, the Geweke-Hajivassiliou-Keane (GHK) simulator needs to be run for increasingly longer simulation runs as the magnitude of ρ increases. Indeed, a Markov Chain Monte Carlo (MCMC) approach generally outperforms the GHK simulator in the experimental conditions considered by Geweke, Keane and Runkle (1997).

[30] Jackman (2000*a*) provides a review of MCMC geared towards political scientists; models for discrete outcomes are among the examples used.

Conditional Densities for the Gibbs Sampler

The conditional distribution of y^* is a multivariate normal distribution with mean vector $X\beta$ and variance-covariance matrix Σ, truncated to the region $(a_1, b_1) \times \ldots \times (a_T, b_T)$, as defined in equation (24). Sampling from this truncated multivariate distribution can be accomplished by sequentially sampling from the conditional distributions for each y_t^*, where the conditioning is not just on the observed data and the parameters β and ρ, but also on the sampled values for $y_{r<t}^*$. For a probit model, each of these conditional distributions is a truncated univariate normal distribution. Given the model for the latent dependent variable

$$y_t^* = \rho y_{t-1}^* + x_t\beta + \epsilon_t, \;\; t = 2, \ldots, T$$

and the stationarity assumption $|\rho| < 1$, then if the covariates X are considered non-stochastic,

$$\mathrm{var}(y_t) = \rho^2 \mathrm{var}(y_{t-1}) + \mathrm{var}(\epsilon_t)$$

$\forall\, t = 2, \ldots, T$. But given the identifying normalization $\mathrm{var}(y_t) = 1$, this implies that $\mathrm{var}(\epsilon_t) = 1 - \rho^2, \forall\, t = 2, \ldots, T$. For probit, the latent disturbances have normal distributions, and so $\epsilon_t \sim N(0, 1 - \rho^2), \forall\, t = 2, \ldots, T$. Thus

$$y_t^* | y_{t-1}^* \sim N\left(\rho y_{t-1}^* + x_t\beta, 1 - \rho^2\right) I(a_t, b_t) \tag{26}$$

for $t = 2, \ldots, T$, where the function $I(\cdot, \cdot)$ is a binary (0,1) indicator function for the truncation bounds. The first observation of each unit-specific time series is sampled from

$$y_1^* \sim N(x_1\beta, 1) I(a_1, b_1)$$

Having generated the latent y^* by sampling from these distributions, we can update the estimates of β and ρ by simply running a regression of the y_t^* on y_{t-1}^* and X, $\forall\, t = 2, \ldots, T$. This regression yields a vector of parameter estimates $(\hat{\rho}, \hat{\beta})$, and a variance-covariance matrix $\sigma_\epsilon^2 (Z'Z)^{-1}$, where Z is the matrix formed by concatenating y_{t-1}^* and the matrix of covariates X, dropping the $t = 1$ observation within each unit. Note that σ_ϵ^2 is fixed at $1 - \rho^2$. With a diffuse prior, the update for β and ρ is obtained by sampling from the multivariate Normal distribution with mean vector $(\hat{\rho}, \hat{\beta})$ and variance-covariance matrix $\sigma_\epsilon^2 (Z'Z)^{-1}$; to enforce the stationarity constraint we would reject sampled values for ρ greater than 1, or less than -1, although we do not encounter any instances of the sampler attempting to visit this region of the parameter space cases with our data.

This Gibbs sampling scheme converges extremely quickly from a range of plausible starting values; the results in the text are based on 10,000 iterations, thinned by a factor of 10, and discarding the initial 1,000 iterations as burn-in. A `C` program implementing this sampler is available upon request.

Probit with AR(1) Errors

For the probit model with AR(1) errors, we again use MCMC methods. The latent regression function is as for ordinary probit

$$y_t^* = x_t\beta + \epsilon_t \tag{27}$$

but with the following AR(1) process for the latent disturbances:

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t, \qquad |\rho| < 1 \tag{28}$$

To estimate this model we employ an MCMC procedure similar to the well-known Cochrane-Orcutt procedure for regression models with AR(1) disturbances.

1. With the current estimate of β, generate the generalized residuals

   $$\epsilon_t^* = E(\epsilon_t; y_t, x_t, \beta)$$

   Expressions for these quantities are defined above, in the body of the paper.

2. Sample from the conditional distribution for ρ:

   $$\rho \sim N(r, R^{-1})$$

   where

   $$\begin{aligned} r &= \frac{\sum_{t=2}^{T} \epsilon_t^* \epsilon_{t-1}^*}{R}, \\ R &= \sum_{t=2}^{T} \epsilon_{t-1}^{*2} \end{aligned}$$

3. Sample from the conditional distribution for $\beta \sim N(b, B)$, where

   $$\begin{aligned} b &= (X^{*'}X^*)^{-1}X^{*'}y^{**} \\ B &= \sigma_\nu^2(X^{*'}X^*)^{-1} \\ X^* &= (x_2^*, \ldots, x_T^*)' \\ y^{**} &= (y_2^{**}, \ldots, y_T^{**})' \\ x_t^* &= x_t - \rho x_{t-1}, \quad t = 2, \ldots, T \\ y_t^{**} &= y_t^* - \rho y_{t-1}^*, \quad t = 2, \ldots, T \\ y_t^* &\sim N(x_t\beta + \rho\epsilon_{t-1}^*, 1 - \rho^2)I(a_t, b_t) \\ a_t &= \begin{cases} -\infty & \iff y_t = 0 \\ 0 & \iff y_t = 1 \end{cases} \\ b_t &= \begin{cases} 0 & \iff y_t = 0 \\ \infty & \iff y_t = 1 \end{cases} \\ \sigma_\nu^2 &= 1 - \rho^2 \end{aligned}$$

We iterate this scheme 10,000 times

References

Amemiya, Takeshi. 1985. *Advanced Econometrics.* Cambridge: Harvard University Press.

Beck, N. 2001. "Time-Series–Cross-Section Data." *Statistica Neerlandica.* 55(2):110–32.

Beck, Nathaniel. 1991. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis.* 3:51–87.

Beck, Nathaniel and Jonathan N. Katz. 1997. "The Analysis of Binary Time-Series–Cross-Section Data and/or The Democratic Peace." Paper presented at the Annual Meeting of the Political Methodology Group, Columbus, OH.

Beck, Nathaniel and Jonathan N. Katz. 2001. "Throwing Out the Baby With the Bath Water: A Comment on Green, Kim and Yoon." *International Organizations.* 55:487–95.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science.* 42(4):1260–88.

Beck, Nathaniel and Richard Tucker. 1997. "Conflict in Time and Space." Center for International Affairs Working Paper 97-8, Harvard University, https://wwwc.cc.columbia.edu/sec/dlc/ciao/wps/tur01/.

Cox, David R. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B.* 34:187–220.

Diggle, Peter J., Kung-Yee Liang and Scott L. Zeger. 1994. *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Esty, Daniel C., Jack Goldstone, Ted Robert Gurr, Pamela Surko and Alan Unger. 1995. "Working Papers: State Failure Task Force Report." MacLean, VA: SAIC.

Geweke, John F., Michael P. Keane and David E. Runkle. 1997. "Statistical inference in the multinomial multiperiod probit model." *Journal of Econometrics.* 80:125–165.

Gourieroux, C., A. Monfort and A. Trognon. 1982. Estimation and Test in Probit Models with Serial Correlation. In *Alternative Approaches to Time Series Analysis*, ed. J. P. Florens, M. Mouchart, J. P. Raouit and L. Simar. Bruxelles: Publications des Facultès universitaires Saint-Louis pp. 169–209.

Heckman, James. 1981. Statistical Models for Discrete Panel Data. In *Structural Analysis of Discrete Data with Econometric Applications*, ed. Charles Manski and Daniel McFadden. Cambridge, MA: MIT Press pp. 114–78.

Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions. In *Proceedings of the Fifth Annual Berkeley Symposium on Mathematical Statistics and Probability*, ed. Lucien M. LeCam and Jerzy Neyman. Vol. I Berkeley, Ca.: University of California Press pp. 221–33.

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

Jackman, Simon. 2000*b*. "In and Out of War and Peace: Transitional Models of International Conflict." Department of Political Science, Stanford University.

Jaggers, Keith and Ted Robert Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research*. 31(4):469–82.

King, Gary and Langche Zeng. N.d. "Improving Forecasts of State Failure." *World Politics*. Forthcoming.

Koyck, L. 1954. *Distributed Lags and Investment Analysis*. Amsterdam: North-Holland.

Liang, Kung-Yee and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*. 73:13–22.

Londregan, John and Keith Poole. 1990. "Poverty, the Coup Trap, and the Seizure of Executive Power." *World Politics*. 42:151–83.

Poirier, Dale J. and Paul A. Ruud. 1988. "Probit with Dependent Observations." *Review of Economic Studies*. 55(4):593–614.

Przeworski, A., M. Alvarez, J. A. Cheibub and F. Limongi. 2000. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950-1990*. Cambridge: Cambridge University Press.

Ware, J. H., S. Lipsitz and F. E. Speizer. 1988. "Issues in the Analysis of Repeated Categorical Outcomes." *Statistics in Medicine*. 7:95–107.

Zorn, Christopher J. W. 2001. "Generalized Estimating Equation Models for Correlated Data: A Review With Applications." *American Journal of Political Science*. 45(2):479–90.

Recent ISERP Working Papers

02-01 “Alternative Models of Dynamics in Binary Time-Series-Cross-Section Models: The Example of State Failure,” Nathaniel Beck, Political Science, UC San Diego, David Epstein, Political Science, Columbia, Simon Jackman, Political Science, Stanford and Sharyn O’Halloran, Political Science, Columbia

02-02 “Substitutability Cross-Stream Between Oriented Markets: Conventions in the Wine Sector of France,” Harrison White, Sociology, Columbia University

02-03 “Link, Search, Interact: The Co-Evolution of NGOs and Interactive Technology,” Jonathan Bach, Center on Organizational Innovation, Columbia University and David Stark, Center on Organizational Innovation, Columbia University

01-01 “Pathways of Property Transformation: Enterprise Network Careers in Hungary, 1988-2000 Outline of an Analytic Strategy,” David Stark, Sociology, Columbia and Balázs Vedres, Sociology, Columbia

01-02 “Policy Space and Voting Coalitions in Congress: the Bearing of Policy on Politics, 1930-1954,” Ira Katznelson, John Lapinski, and Rose Razaghian, Political Science, Columbia

01-03 “Doing Fractions: An Analysis of Partisan ship in Post-Socialist Russia,” Andrew D. Buck, Sociology, Columbia

01-04 “Opposite-Sex Twins and Adolescent Same-Sex Attraction,” Peter Bearman, Sociology/ISERP and Hannah Brückner, Sociology, Yale

01-05 “On the Uneven Evolution of Human Know-How,” Richard R. Nelson, Business/SIPA, Columbia

01-06 “Self-Control for the Righteous: Toward a Theory of Luxury Pre-Commitment,” Ran Kivetz, Business, Columbia and Itamar Simonson, Business, Stanford

01-07 “Distributing Intelligence and Organizing Diversity in New Media Projects,” Monique Girard, ISERP, Columbia and David Stark, Sociology, Columbia

01-08 “Agricultural Biotechnology’s Complementary Intellectual Assets,” Gregory D. Graff, Agricultural and Resource Economics, Berkeley, Gordon C. Rausser, Agricultural Economics, Berkeley and Arthur A. Small, SIPA/Earth Institute, Columbia

For copies of ISERP Working Papers
visit http://www.iserp.columbia.edu/initiatives/working_papers/paper_program.html,
write to iserp@columbia.edu or call 212-854-3081.

EDITORIAL BOARD
Karen Barkey, Sociology
Peter Bearman, Sociology/ISERP
Alan Brinkley, History
Charles Cameron, Political Science
Alessandra Casella, Economics
Ester Fuchs, Political Science/SIPA
John Huber, Political Science
Ira Katznelson, Political Science/History
Herbert Klein, History
Mary Clare Lennon, Public Health
Mahmood Mamdani, Anthropology
Marianthi Markatou, Statistics
William McAllister, ISERP
Kathryn Neckerman, ISERP
Richard Nelson, Business/SIPA
Elliot Sclar, Architecture, Planning and Preservation/SIPA
Seymour Spilerman, Sociology
Charles Tilly, Sociology
Harrison White, Sociology

ADMINISTRATION
Peter Bearman, Director
Kathryn Neckerman, Associate Director
Leslie Wright, Assistant Director

Institute for Social and Economic Research and Policy
Columbia University
International Affairs Building
420 West 118 Street, 8th Floor
Mail Code 3355
New York, NY 10027
telephone: 212-854-3081
facsimile: 212-854-8925
e-mail: iserp@columbia.edu
URL: http://www.iserp.columbia.edu