

Temporal Dependence and the Sensitivity of Quantities of Interest: A Solution for a Common Problem

Online Appendix

Laron K. Williams
Department of Political Science
University of Missouri
williamslaro@missouri.edu

Overview

In this *Online Appendix* I describe the procedures used to produce the survey, meta-analysis, and a brief discussion on the linear probability model.

Survey Description

I used the Web of Science database to collect all studies that cite either Beck, Katz and Tucker (1998) or Carter and Signorino (2010) from the following six journals: *American Political Science Review* (February 2017, Issue 1), *American Journal of Political Science* (January 2017, Issue 1), *the Journal of Politics* (July 2017, Issue 3), *International Studies Quarterly* (March 2017, Issue 1), *International Organization* (Summer 2017, Issue 3), and *the Journal of Conflict Resolution* (November 2017, Issue 10). I excluded those pieces that were not addressing substantive questions (such as methodology papers).

For each piece, I code whether they actually implement the BKT or CS suggestions (instead of simply citing them), how they deal with temporal dependence (such as splines or cubic polynomials), how they interpret the TDVs (if necessary), whether they calculate QIs for their primary theoretical variable, whether a scenario used to calculate the QI is provided, and the value of the t in that scenario. I summarize the patterns from this survey in the manuscript.

Meta-Analysis

From the universe of citations that I survey above, I randomly sample 20 studies from each of the six journals listed above. If the study was not quantitative in nature, did not use the methods suggested by BKT/CS, did not generate QIs, or did not offer complete replication materials, it was excluded and another study was randomly selected.

Linear Probability Model

Angrist and Pischke (2009) offer an intriguing solution to the problem of compression outlined in the manuscript. The linear probability model (LPM)—estimated via ordinary least squares regression—provides a marginal effect of X on Y that is constant across all values of X , and any other X . Though there are some other issues associated with estimating an LPM on a non-continuous outcome (see Long 1997), Angrist and Pischke (2009) raise the possibility that the marginal effect from OLS is comparable to the average marginal effect from the nonlinear model. Indeed, they provide some preliminary evidence from an application of the effects of childbearing on mothers' labor supply that supports this notion.

In Table A.1 I estimate the LPM on the 20 models from the meta-analysis to see if these values are generally comparable across studies of international relations. In the *Nonlinear QI* column I depict the average of the in-sample partial effects from the nonlinear model (either logit or probit in this sample). The other two columns depict the range of in-sample partial effects, and the marginal effect from the LPM.

The clear pattern from Table A.1 is that the LPM marginal effect is quite close to the average in-sample partial effect estimated via logit or probit. There are a few notable case studies where they depart a great deal from each other (most notably the Salehyan 2008 and Fuhrmann and Kreps 2010 studies), but beyond that the LPM does an admirable job in retrieving the “average” effect in the sample. In fact, all the studies but one (Salehyan 2008) have LPM marginal effects that are within a standard deviation of the mean in-sample nonlinear partial effect.

If scholars are focused on quickly generating an easy-to-interpret measure of the average effect, then the LPM is an obvious alternative. The choice to estimate an LPM comes with clear drawbacks, however, as the final column reveals. Only a few studies produce predicted probabilities that fall almost entirely in the reasonable boundaries of 0 and 1. Some (Fordham and Asal 2007, Fuhrmann and Kreps 2010, and Weeks 2012) fail to produce even a majority of reasonable predictions from the LPM.

Table A.1: Comparison of Quantities of Interest from Nonlinear and Linear Probability Models

Citation	Variable	Nonlinear QI ^a	Probability ^b	LPM QI (%) ^c	± 1 SD ^d	% Nonsense ^e
Gelpi and Feaver (2002: 789)	Percent Veteran	-.029	.034	-.017 (49)	1	9.3
Meinke, et al. (2006: 188)	Ideological Distance	-.570	.693	-.605 (60)	1	29.6
Salehyan and Gleditsch (2006: 356-357)	Refugees	+.007	.040	+.016 (88)	1	25.5
Fordham and Asal (2007: 41)	Major Power Democracy	+.036	.006	+.029 (63)	1	56.5
Gleditsch, et al. (2008: 492)	Civil War	+.009	.024	+.012 (81)	1	37.5
Salehyan (2008: 797)	Refugees in Initiator	+.008	.013	+.042 (95)	0	38.0
Ostby, et al. (2009: 314-315)	Distance to Neighboring Conflict	-.034	.051	-.035 (32)	1	0.1
Fuhrmann and Kreps (2010: 851)	Violent Conflict	+.021	.001	+.049 (91)	1	59.7
Cunningham, et al. (2012: 80)	Violent Factions	+.087	.177	+.085 (59)	1	33.8
Kleinberg, et al. (2012: 536)	Trade Concentration	-.0002	.001	-.0003 (11)	1	41.6
Owsiak (2012: 64)	Settled Borders	-.015	.065	-.022 (22)	1	0.3
Weeks (2012: 342)	Machine	-.0006	.002	-.0004 (17)	1	59.3
Fuhrmann and Sechser (2014: 929)	Deployment	-.002	.004	-.001 (30)	1	23.0
Miller (2014: 926)	Dependence Score	-.012	.0002	-.012 (19)	1	1.4
Peterson (2015: 713, 718)	Distortion	+.0003	.002	+.0002 (79)	1	43.8
Wu (2015: 798)	Labor Endowment	+.008	.293	+.009 (74)	1	38.1
Bapat, et al. (2016: 262)	GDP Per Capita	+.084	.036	+.024 (10)	1	15.2
Bapat and Zeigler (2016: 347)	Terrorists	+.096	.009	+.047 (25)	1	31.8
Bayer and Urpelainen (2016: 612)	Democracy	+.031	.010	+.018 (56)	1	32.3
Casper (2017: 983)	Program Initiation	+.050	.052	+.031 (48)	1	31.4

^a Mean in-sample quantity of interest from logit or probit.

^b Mean in-sample baseline predicted probabilities.

^c Quantity of interest from linear probability model (with percentile based on in-sample nonlinear QI in parentheses).

^d Does QI from LPM fall within a standard deviation of mean in-sample nonlinear QI?

^e Percentage of baseline predicted probabilities from LPM falling outside reasonable bounds (< 0 or > 1).

In Table A.2 I provide those same comparisons for the three applications explored in the manuscript. The same general patterns discussed above in the meta-analysis apply equally to these three studies. Besides Way and Weeks (2014), the marginal effect from the LPM is quite close to the mean marginal effect from nonlinear model. The significant drawback is that the LPM generates nonsensical predictions for almost (or more than) a majority of observations.

Table A.2: Comparison of Quantities of Interest from Nonlinear and Linear Probability Models: Three Applications

Cite	Nonlinear QI ^a	Range ^b	LPM QI (%) ^c	± 1 SD ^d	% Nonsense ^e
Way and Weeks (2014)	+0.007	[0, .64]	+0.041 (97)	1	53.2
Cunningham (2013)	+0.203	[.023, .561]	+0.199 (57)	1	46.7
Flores-Macias and Kreps (2013)	+0.199	[.005, .619]	+0.209 (61)	1	45.8

^a Mean in-sample quantity of interest from logit or probit.

^b Range of in-sample quantity of interest from logit or probit.

^c Quantity of interest from linear probability model (with percentile based on in-sample nonlinear QI in parentheses).

^d Does QI from LPM fall within a standard deviation of mean in-sample nonlinear QI?

^e Percentage of baseline predicted probabilities from LPM falling outside reasonable bounds (< 0 or > 1).

Compression

In practice, this means that the size of the quantity of interest will depend on the baseline probabilities. To demonstrate how the range of baseline probabilities influence the size of these changes, I set up a series of simulations of 2000 observations drawn from a uniform distribution over various ranges of probabilities (e.g., $\text{Pr} = [0.3, 0.7]$). Thus, the sample size and standard deviations are nearly constant across probability ranges. I then calculate the changes in predicted probability for a 1-unit increase in X_k ($\beta_{X_k} = 1$) across different ranges of baseline probabilities. Another way to think of these baseline probabilities is representing observations with different values of the independent variables. I present box-whisker plots of the substantive effects for the various simulations in Figure A.1.

As Figure A.1 shows, compression heavily influences the quantities of interest in two ways. Compression produces considerable variation in quantities of interest across the sample. An identical shift in X produces 2000 different quantities of interest, depending on the observation's location along the CDF. Figure A.1 also shows that compression generates sensitivity, in that changes to the scenario produce changes in the quantities of interest. The quantities of interest are sensitive to the scenario (in this case, probabilities) and range from nearly 0 to 0.25. First, it is clear that scholars depicting substantive effects with simulation scenarios that generate probabilities near 0.5 (such as those in the $[0.3, 0.7]$ range) will demonstrate, on average, larger effects than those with consistently low or high ranges. Second, the sensitivity within these ranges increases as one shifts away from 0.5. Essentially, this identifies the situations where the substantive effects are most sensitive to changes in the scenario. The box-whisker plots in Figure A.1 show that the effects of a key theoretical variable will be much more sensitive to the values of the control variable for probability ranges near 0 ($[0, 0.4]$) or 1 ($[0.6, 1.0]$) compared to 0.5 ($[0.3, 0.7]$). While econometricians have long emphasized that quantities of interest in binary models depend on the scenarios chosen (e.g., Maddala 1983: 23-24), it has only recently triggered a debate in political science regarding the proper way to handle "hyper-sensitivity" (e.g., Nagler 1994), including whether interaction variables are actually necessary (Berry, DeMeritt and Esarey 2010; Rainey 2016).

For example, consider the often-studied outcome of international disputes. Using a relatively liberal definition of conflict including threats, demonstrations and uses of force on the universe of dyads from 1816-1991 results in conflicts in only 4.7% of the dyad-years (Russett and Oneal 2001). A low baseline probability such as this corresponds to values of $X\beta$ that are largely negative, and quite small. The patterns of duration dependence in international conflict are typically negative, indicating that the baseline probability of a dispute is highest immediately following a dispute, and then declines with the passage of time. Figure A.3 shows the cumulative density function for the logistic distribution (left panel) and an example of negative duration dependence (right panel) (Russett and Oneal 2001). With these two pieces of information in hand, it is clear that the quantities of interest are quite sensitive to the scenario chosen (in this case, t). One can deduce that the largest substantive effect for a variable will occur at extremely small values of the TDVs (since this configuration of values increases $X\beta$ toward 0) and decrease at higher values of TDVs (as the $X\beta$ becomes more negative). Furthermore, the figure on the right ignores the variability of partial

Figure A.1: Box-Whisker Plots of the Change in Predicted Probability for a 1-Unit Increase ($\beta = 1$) across Ranges of Probabilities

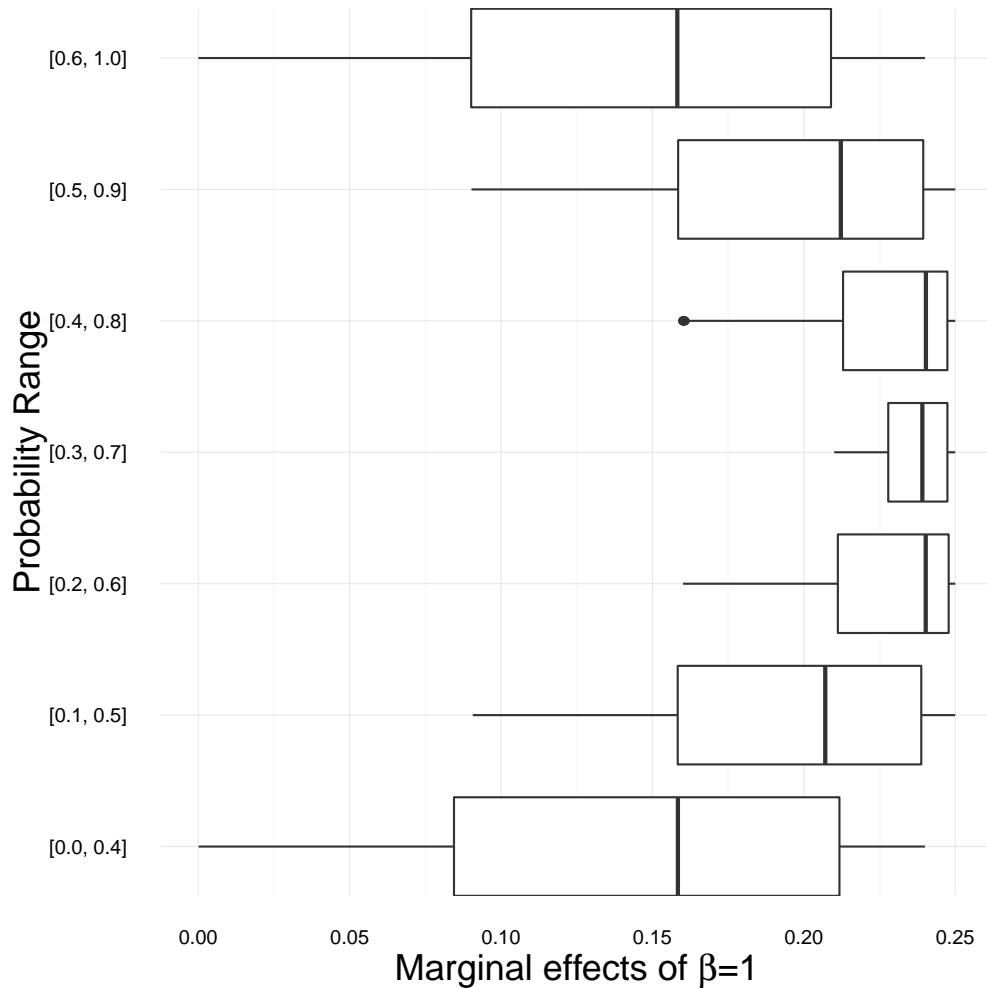
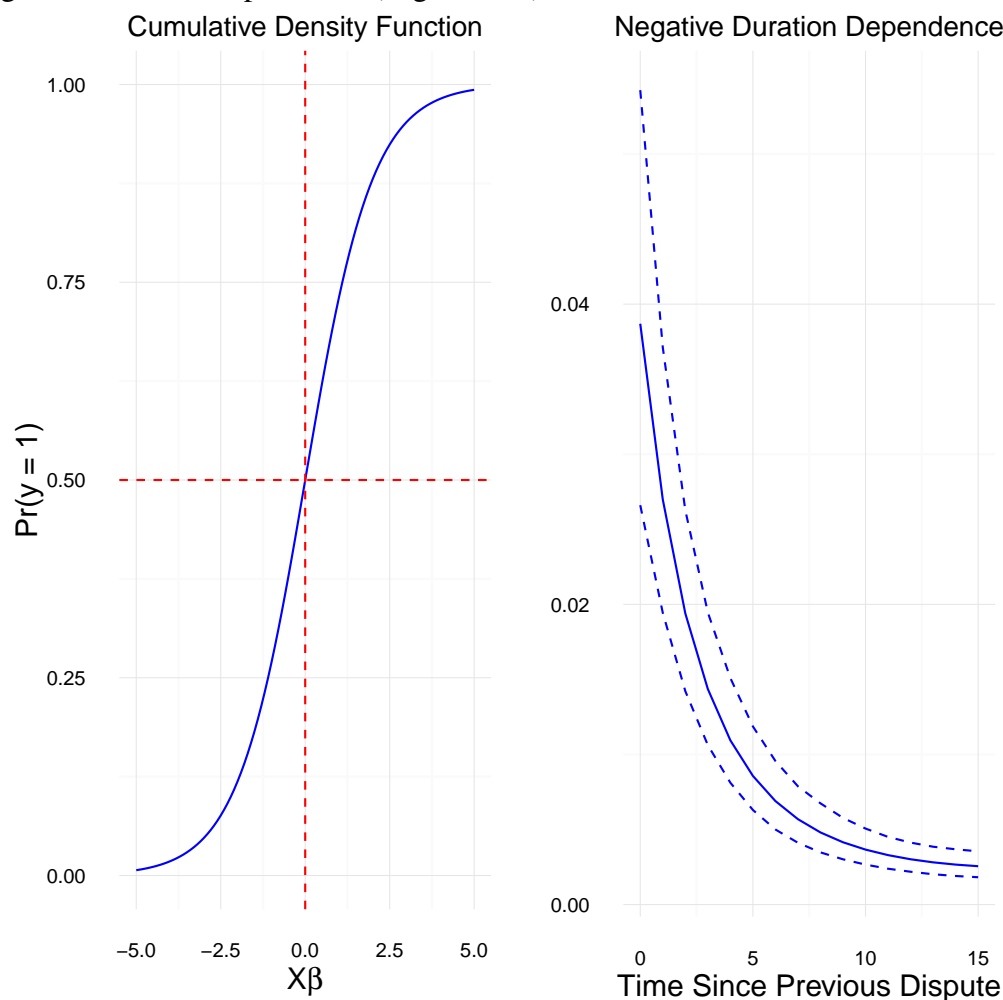


Figure A.2: *

Note: Each range of probabilities is drawn from a uniform distribution ($N = 2000$).

effects in the sample. While other simple scenarios—such as low probability outcomes coupled with positive duration dependence—may be easy to conceptualize, outcomes with probabilities closer to 0.5 and non-linear functional forms of duration dependence (such as splines) complicate this conceptualization considerably. In the absence of a more careful exploration of quantities of interest (see the recommendations below), without this information, the reader cannot assess how the substantive effects vary as a function of the underlying hazard rate.

Figure A.3: Cumulative Density Function for the Logistic Distribution (Left Panel) and an Example of Negative Duration Dependence (Right Panel)



Stata Program for Estimating Compression Effects

In this section I briefly describe a Stata program that automates the calculation of the measures of partial effects discussed in the manuscript.¹ The command `sdpe` calculates in-sample partial effects and provides the meaningful values of the distribution of partial effects, including the mean, standard deviation, minimum and maximum. Optionally, the user can specify a variable that she suspects is causing some conditionality (such as *time*) and `sdpe` will provide these values across each unique value of *time* (the option `comp(time)`).

Flexibility was a priority in the creation of the program. Toward this end, `sdpe` calculates par-

¹If this manuscript is accepted for publication, I will upload this program and its help file to Stata's Statistical Software Components (SSC) archive.

tial effects following a wide variety of the most common versions of logit and probit models (e.g., such as population-averaged, random effects and fixed effects). Moreover, the code for specifying the variables used to calculate the partial effects is based on King, Tomz and Wittenberg's (2000) `setx` command, so many users are already familiar with the basic syntax. One can specify multiple variables (such as lower-order and interaction terms) to calculate the marginal effects and users can set the variables to a wide variety of functions (such as minimum, maximum, mean, median, percentiles).

For example, consider Oneal and Russett's (1997: 286) model of dispute initiation. Assume that we were interested in the change in the probability of dispute initiation as the lower democracy (`dem_low`) in the dyad changes from a value of -10 (full autocracy) to +10 (full democracy). We first estimate the logit model, and then calculate the partial effects with `sdpe`. The command also returns the values depicting the distribution as local macros, so it is easy to use these values for other purposes (perhaps in a `foreach` loop or a `postfile`).

```
. quietly logit dispute dem_low dem_hi growth_low allies contig capratio /*
> */      depend_low d_depend_hi democ_chg autoc_chg _prefail t t2 t3

. sdpe dem_low -10 10
```

Partial Effect				
Percentiles		Smallest		
1%	-.1947876	-.2252312	Obs	19455
5%	-.0792117	-.2252281	Sum of Wgt.	19455
10%	-.0351589	-.2252119	Mean	-.0158471
25%	-.012184	-.2251843	Std. Dev.	.0333926
50%	-.0043752		Variance	.0011151
		Largest	Skewness	-3.985807
75%	-.0027483	-5.60e-21	Kurtosis	20.18231
90%	-.001052	-2.71e-21		
95%	-.0003925	-2.38e-21		
99%	-7.91e-07	-1.76e-21		

```
. return list
```

```
scalars:
      r(maxpe) = -1.76063656626e-21
      r(minpe) = -.2252312004566193
      r(sdpe)  = .0333926326349945
      r(ape)   = -.0158471045645255
```

Concerns about temporal conditionality warrant an examination of these values across t .

```
. sdpe dem_low -10 10, comp(t)
```

In addition to the partial effects that are presented as the default, including the `comp(t)` option produces the subgroup partial effects across values of t . Also notice that these values are returned

	t	mean_pe	sd_pe	min_pe	max_pe
4.	3	-.033351	.0338249	-.2251843	-5.60e-21
5.	4	-.0245958	.0257234	-.2110711	-2.71e-21
6.	5	-.0187978	.0201508	-.21299	-1.76e-21
8.	7	-.0121491	.0135267	-.1866422	-2.38e-21
10.	9	-.0083461	.0081977	-.0618138	-2.91e-11
11.	10	-.0070998	.0070606	-.0552553	-2.79e-11
12.	11	-.0062957	.0062309	-.0504371	-2.17e-11
13.	12	-.0056591	.0056165	-.0470269	-2.60e-11
14.	13	-.0051796	.0051725	-.0441802	-3.28e-11
17.	16	-.0044701	.0042622	-.0390172	-3.97e-09
18.	17	-.0043427	.0042111	-.0381188	-3.36e-09
19.	18	-.004234	.0037288	-.0234532	-1.11e-08
20.	19	-.0042892	.0037831	-.0228583	-9.83e-08
21.	20	-.0042654	.0036314	-.0252584	-9.32e-09
23.	22	-.0043461	.0033658	-.0201519	-.0001619
25.	24	-.0042921	.0034485	-.0232595	-.0001728
26.	25	-.0043438	.0034096	-.0240389	-.000049
27.	26	-.0047253	.0035974	-.0239113	-.0001649
28.	27	-.0048843	.0036623	-.0236125	-.0000243
29.	28	-.0048636	.0036301	-.022619	-.0002629
30.	29	-.0047258	.003554	-.0262076	-.0001317
32.	31	-.0046785	.0033526	-.0205246	-.0002908
35.	34	-.0043329	.0030916	-.0151019	-.0002701
36.	35	-.0041402	.0032442	-.02269	-.0002366

. return list

scalars:

r(maxpe) = -1.76063656626e-21
r(minpe) = -.2252312004566193
r(sdpe) = .0333926326349945
r(ape) = -.0158471045645255

matrices:

r(tempond) : 36 x 5

as a matrix called `r(tempcond)`, so it is easy to produce a simple figure depicting the temporal conditionality. The following commands use the values in the returned matrix to graphically depict the mean and standard deviation of the partial effects across values of t .

```
. preserve
.       clear
.       matrix tempcond = r(tempcond)
.       svmat tempcond, names(col)
number of observations will be reset to 36
Press any key to continue, or Break to abort
obs was 0, now 36

.       twoway (line mean_pe t) (line sd_pe t, lpattern(dash)), /*
>       */ ytitle("Partial Effect") legend(label(1 "Mean") label(2 "S.D.))

. restore
```

