Drawing Accurate Inferences About the Differences Between Countries in Time-Series Cross-Section Data

Jonathan Kropko

June 27, 2010

Abstract

Time-series cross-section (TSCS) data contain a cross-section of N panels, measured at each of T time points, and have two variances: a variance across the panels, and a variance across time. TSCS data increase the power of statistical analyses by increasing the sample of N by a factor of T, but it is easy for tests of substantive questions to confuse the variation of a dependent variable within the cross-section with the variation of the variable across time. In this paper, a method called the between effects estimation routine (BEER) is developed which considers only the variance across panels by "de-pooling" the data, running OLS on individual cross-sections, and averaging results across the cross-sections. The method is used on a model of regional authority in 21 countries from 1950 to 2006. Commonly used methods for TSCS data- pooled OLS, panel corrected standard errors, random effects, fixed effects, fixed effects with vector decomposition, and Mundlak's (1978) approach- are also used to estimate the regional authority model, and the results are compared. BEER provides consistent estimates of the cross-sectional effects at a particular point in time, and can compare time points to one another. No single method, however, provides a complete analysis of the relationships in any TSCS dataset, and a combination of methods must be used.

1 Introduction

Time-series cross-section (TSCS) data contain a cross-section of N panels, measured at each of T time points, and have two variances: a variance across the panels, and a variance across time. Studies that use TSCS data are becoming increasingly prevalent in political science. In American politics, many studies consider the 50 states over a time span. In comparative politics and international relations, these data take the form of countries and years. TSCS data increase the power of statistical analyses by increasing the sample of N by a factor of T, but it is easy for tests of substantive questions to confuse the variation of a dependent variable within the cross-section with the variation of the variable across time.

Traditional statistical approaches often fail to speak to our concerns as political researchers.[1] In economics, researchers utilize TSCS data to determine how economic indicators affect other economic indicators, but are rarely interested in the characteristics of the panels themselves. Driver and Muñoz-Bugarin (2009), for example, use a dataset that consists of seven European countries, Italy, the Netherlands, Belgium, France, the UK, Germany, and Sweden, from 1993-2006 to determine the effect of capital investment on the unemployment rates. By utilizing a fixed effects model the coefficient on capital investment is consistent for the dynamic effect, but variables which are fixed across time drop out of the model. Fixed effects in these models are "nuisance parameters," which account for cross-sectional variation but have no substantive interpretation. They contain all time-invariant variables at once but explain none of them. In fact, economists have seen this attribute of fixed effects models as one of its primary strengths; fixed effects correct the time-invariant omitted variable bias in a pooled model, while researchers must resort to instrumental variables in a cross-section (Cameron and Trivedi 2005, p.607).

As political scientists, we are often much more interested in time-invariant independent variables, and in the variance between countries in a given year. An EU scholar might point out that the post-communist legacy in Germany, traditional workers' rights in France, and ethnic division in Belgium may all explain part of the variation in unemployment rates. Leaving these important time-invariant variables to nuisance parameters is not an acceptable model. Fixed effects also change the interpretation of variables that vary across countries and years: changes in capital investment may effect unemployment, but the actual levels of capital investment from country to country may have an effect as well. Political scientists have recognized the weaknesses of fixed effects approaches for some political questions. Kittel (1999, 2006) calls for greater emphasis on comparative case studies since commonly used statistical models for TSCS data do not ameliorate the problem of having a small-N in the cross-section.

Many treatments of linear models for TSCS data work within the regression paradigm of explaining variance in the dependent variable. These treatments tend to ignore or brush aside the fact that a TSCS data structure has two variances on which an R^2 statistic can be computed. Bartels (2009) presents a discussion of linear estimators for TSCS data that rightly focuses on the difference between the two variances. Each predictor has two effects

[1] In addition, a great deal of confusion arises with TSCS methodology because different disciplines use inconsistent terminology. For example, in econometrics, a dummy variable for a panel is called a fixed effect and a slope estimate for an independent variable is called a coefficient. In psychometrics, a dummy variable for a panel is called a random effect and a slope estimate for an independent variable is called a fixed effect. Political scientists experience more grief from the terminology because the field is interdisciplinary, and researchers commonly receive methods training in both economics and psychology departments. Furthermore, the treatment of TSCS data is a subject which is of concern to researchers in many disciplines, but discussion across disciplines has been limited. This lack of collaboration is partly due to different focuses for the use of TSCS data, but it is partly because this kind of data structure has many different names: TSCS, pooled time-series, longitudinal, repeated measures, and wave study data. All of these data structures are special cases of panel or pooled data, which are special cases of multilevel or hierarchical data.

on the dependent variable: the effect on the variance over time is called the within effect, and the effect on the variance between panels is called the between effect. There is no reason why these two effects should necessarily be equal, and there are many situations in fact when we would expect the two effects to be in opposite directions. For example, consider the effect of ethnic conflict within a country on the level of descriptive ethnic representation in the legislature. Between countries, it may be the case that countries with greater levels of ethnic conflict have lower levels of descriptive ethnic representation, since ethnic minorities may use conflict to address institutional disenfranchisement. However, it may also be the case that within a typical country, increases in ethnic conflict are associated with increases in descriptive representation as conflicts are translated into election outcomes. In this case, we expect the between effect of ethnic conflict on descriptive representation to be negative, but we also expect the within effect to be positive. Models which drop the between effect may be incomplete for our theory, and models that average these two effects may be meaningless.

In this paper, I discuss a methodology for political research which focuses solely on the between effects of predictors. This method, called the between effects estimation routine (BEER), considers only the variance across panels by "de-pooling" the data, running OLS on individual cross-sections, and averaging results across the cross-sections. I demonstrate the use of this method on a model of regional authority in 21 countries from 1950 to 2006. The data are described in section 2.1. Commonly used methods for TSCS data are described in section 2.2, and each one is used to estimate the regional authority model. These results are compared, but none of them provide adequate estimates of between effects. The formulation of the method is presented in section 3. I estimate the regional authority model using this method, report the results in section 4, and discuss the different inferences that can be made with this method compared to the conclusions drawn using the other methods in section 5. In section 6 I conclude.

2 Background

2.1 Data: Regional Authority in 21 Countries, 1950-2006

In order to demonstrate the characteristics of each commonly used estimator for TSCS data, I will estimate the same linear model using each of these common methods. The example data focus on federalism in a comparative context. The dependent variable is the regional authority index (RAI) compiled by Gary Marks, Liesbet Hooghe and Arjan H. Schakel (2008), which measures the strength of regional governments within a country relative to the federal government. The index is provided for 21 countries, measured annually, beginning in 1950 and ending in 2006. No countries enter or drop out of the analysis during this time span.[2] Table 1 contains a list of the countries included in the analysis and the over-time average value of the RAI for each country.

Table 1: Countries in the Sample and Average Regional Authority Index

Country	RAI	Country	RAI	Country	RAI
Australia	18.3	Greece	3.1	New Zealand	5.6
Belgium	22.3	Iceland	0.0	Norway	7.4
Canada	22.6	Ireland	1.5	Sweden	11.3
Denmark	7.9	Italy	13.6	Switzerland	19.5
Finland	2.6	Japan	8.2	Turkey	4.2
France	10.8	Luxembourg	0.0	United Kingdom	9.1
Germany	29.1	Netherlands	13.7	United States	23.1

There are many theories of why some countries adopt greater levels of regional authority than others which deserve great deal of attention and evaluation. My purposes in this paper, however, are to demonstrate the use of a new method to test hypotheses about the differences between countries. I am devoting less space than is appropriate to the explication of theory, and with apologies, I list some theories here which depend only on differences between countries, and make no claim about how individual countries change over time.

Federalism is expensive to run successfully, so we can expect that countries with higher levels of RAI are also wealthier. Some countries may have higher levels of regional authority because strong regions are needed for administration: countries which are geographically large have more land to administer, and countries which have

[2]The number of countries is kept constant across time here so that changes in between effects over time cannot be due to panels entering or dropping out of the analysis. All of the methods discussed and presented in this paper, however, can be used on unbalanced panels as well as balanced panels.

large populations have more people to serve. Therefore countries with greater areas and larger populations should also have higher levels of regional authority. A country may also have greater regional authority because segments of the population have a desire for strong regional governance. Specifically, a country which has regions where a majority of people are part of an ethnic minority, such as Quebec in Canada, should have greater regional authority. In addition, regional-based political parties have an incentive to institute greater regional authority to increase their political power. Finally, authoritarian regimes will be less willing to cede power to regional governments, so we expect countries with lower Freedom House democracy scores to have correspondingly lower levels of RAI.[3]

Note that these hypotheses speak only about the differences between countries. Theories regarding changes within countries are different questions. We can hypothesize about the effects when a country becomes richer over time, or larger, or more populous, ethnically fragmented, politically regionalized, or democratic. However, it is not at all clear that these effects should be necessarily equal, similar, or even in the same direction as the between effects suggested above.

For the analyses presented here, RAI is regressed on six independent variables:

1. the country's GDP per capita in each year, measured in 2006 U.S dollars,
2. the country's population in each year, in thousands of people,
3. the country's area, in units of 1000 sq. km.,
4. a measure of the country's ethnic fragmentation which expresses the probability that two people chosen at random from the country belong to different ethnic groups,
5. the proportion of seats in the country's federal legislature in each year which are controlled by regional political parties, and
6. the country's Freedom House democracy score in each year.

These data were collected by Gary Marks and Liesbet Hooghe, and are used here with their permission. The measure of ethnic fragmentation was compiled by Fearon (2003), which considers one time point for each country. To account for the skewness of population, area, and GDP per capita across countries, the common (base 10) logarithm of these three variables is used. The interpretation of the effects of these variables now changes: a one-unit increase in a $\log_{10}$ transformed variable is equivalent to a 10-fold increase in the untransformed variable. Since data on ethnic fragmentation is not measured over time, this variable is treated as a time-invariant predictor in the analyses in the next section. Summary statistics for these six independent variables are listed in table 2.

Table 2: Summary Statistics for the Variables in the Regional Authority Analysis

Variable	Mean	S.D.	Min	Max
GDP per capita, $\log_{10}$	3.8	0.5	2.4	4.8
Population, $\log_{10}$	7.1	0.7	5.2	8.5
Area, $\log_{10}$	2.4	0.8	0.4	4.0
Ethnic fragmentation	0.24	0.18	0.02	0.6
Seats held by regional parties (%/100)	0.03	0.8	0	0.6
Freedom House democracy score	11.3	1.3	2	12

2.2 Commonly Used Estimators for Modeling TSCS Data

This section focuses on demonstrating how various estimators of models for TSCS data produce different results, and why. A summary of the of how these approaches have been used in the literature is not the focus, although Wilson and Butler (2007) have provided an extensive literature review of how TSCS methods have been used in political science.

All models for TSCS data use the following equation:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + u_i + e_{it}. \quad (1)$$

[3]Freedom House uses a coding scheme that awards lower scores to the most democratic countries, but this scale is reversed for this study.

This model is sometimes generally called the pooled model (Cameron and Trivedi 2005, p. 699). The subscript i refers to the panel (or country), and there are N total panels. The subscript t refers to the time point. There are T_i total time points for panel i, and $\sum_{i=1}^{N} T_i$ total observations. x_{it} represents the independent variables that vary across panels and over time, and z_i represents the independent variables which vary across panels but are fixed over time. The overall residual for the model is $\varepsilon_{it} = u_i + e_{it}$, where u_i is the unobserved variance of y_{it} that varies across panels but is fixed across time. These values are also called unit effects, or unobserved heterogeneity. e_{it} is the idiosyncratic error, which varies over time and across panels. In the discussion in this section, I assume that the residuals are independent of the regressors. This assumption would imply that $\text{Corr}(x_{it}, u_i) = 0$, which may be controversial, but only if endogeneity is a concern for a cross-sectional OLS regression.

There are many options for researchers who have TSCS data and want to choose an estimator for the model in equation 1. Among these options are:

1. Pooled OLS
2. Pooled OLS with panel corrected standard errors (Beck and Katz 1995)
3. Random effects
4. Fixed effects
5. Fixed effects with vector decomposition (Plumper and Troeger 2007)
6. Mundlak's (1978) approach of separating the between and within parts of each independent variable

In this discussion, I aim to demonstrate what each of these estimators does to and assumes about the data. Each of these models is run on the example data introduced in section 2.1, and the results are presented in table 4.

Before proceeding, it will be useful to define two terms. In section 1, a between effect and a within effect were defined. Any variable x_{it} that varies across panels and over time can be separated into two parts: a between variable and a within variable. The between variable is the the average value of the variable over time for each panel:

$$x_i^B = \frac{\sum_{t=1}^{T_i} x_{it}}{T_i}. \tag{2}$$

The within variable is the difference between the variable and the between variable:

$$x_{it}^W = x_{it} - x_i^B. \tag{3}$$

The within variable centers x_{it} around its mean for each country, so that average differences between countries are removed. The differences from year to year, however, are preserved. Notice that, by these definitions, any variable which varies over time and cross-sectionally is the sum of its between and within parts:

$$x_{it} = x_i^B + x_{it}^W. \tag{4}$$

Table 3 provides an example of how to calculate the between and within parts of a TSCS variable. Suppose we have data on the population of two towns from 2004 to 2008. Both towns are growing, but town 2 is always bigger than town 1. The between version of population is the average population of each town over the five year period. This variable captures the fact that town 2 is bigger than town 1. The within version of population is the difference between the populations and each town's average population. This variable preserves year to year differences within each town- the population of town 1 grew by 200 people between 2004 and 2005, and so did the within version of population- but eliminates meaningful differences between town 1 and town 2.

Each of the six commonly used methods listed above makes assumptions about the effects of these between and within variables. Recognizing these assumptions is vital to selecting a correct model to answer political questions with TSCS data.

Pooled OLS

Pooled OLS is simply an OLS regression on the TSCS data, with no adjustment for the time or repeated cross-sectional nature of the data. Like all standard OLS models, the variance of the dependent variable is assumed to be a scalar value, rather than a quantity with separate time and cross-sectional dimensions. The model does not distinguish between the unit effects and the overall residual, so the model which is estimated is

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + \varepsilon_{it}. \tag{5}$$

Table 3: Example of Between and Within Parts of a Variable in TSCS Data

Town ID	Year	Population	Pop. Between	Pop. Within
1	2004	2500	3200	-700
1	2005	2700	3200	-500
1	2006	3200	3200	0
1	2007	3400	3200	200
1	2008	3800	3200	600
2	2004	6850	7480	-630
2	2005	7100	7480	-380
2	2006	7350	7480	-130
2	2007	7800	7480	320
2	2008	8300	7480	820

If substitute for x_{it} using equation 4, the model becomes

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1(x_i^B + x_{it}^W) + \beta_2 z_i + \varepsilon_{it} \\ &= \beta_0 + \beta_1 x_i^B + \beta_1 x_{it}^W + \beta_2 z_i + \varepsilon_{it}. \end{aligned} \tag{6}$$

The common econometric objection to pooled OLS is that it returns biased standard errors by assuming that residuals within a panel are uncorrelated (Cameron and Trivedi 2005, p. 702). But there is another problem to consider. The effects of the between and within parts of x_{it} on y_{it} are constrained to be equal. As a result, β_1 is a weighted average of the between and within effects of x_{it}. The weights depend on the amount of cross-sectional variance in y_{it} compared to the time variance. If the between and within effects of x_{it} should be different for substantive reasons, as in the example in section 1 of ethnic conflict and descriptive ethnic representation, then it is unclear from the pooled OLS results alone which effect β_1 most resembles, and whether the average has any meaningful substantive interpretation.[4] Pooled OLS results for the regional authority data are listed in the first column of table 4.

Pooled OLS with panel corrected standard errors

One recommendation in political science to correct the estimates from pooled OLS is to use panel corrected standard errors, or PCSEs (Beck and Katz 1995). It is important to note that PCSEs only change the estimates of standard errors, and do not change the coefficient point estimates from pooled OLS. PCSEs correct the standard errors for heteroskedasticity and contemporaneous error correlation, which is correlation between the errors for two different panels at the same point in time. This approach does not address any bias that may be present in the pooled OLS coefficient point estimates. The results of the regional authority pooled OLS model with PCSEs is listed in the second column of table 4. Note that the coefficient estimates are the same as the point estimates for pooled OLS, but that standard errors are different. If we suspect that coefficients are biased for the pooled OLS model, then they must be biased for the PCSE model as well.

Random effects

Random effects treat the unit effects as a separate part of the residual, where:

$$u_i \sim N(0, \sigma_u^2). \tag{7}$$

Random effects estimators will compute the between variance, σ_u^2, the within variance, σ_e^2, and the ratio between the two. Like pooled OLS, effects are weighted averages of between and within effects, although the weights are calculated through a slightly different formula. Random effects are commonly used by researchers who want to account for the unit effects in the data while allowing the effects of time-invariant predictors to be estimated. However, in general, random effects place more weight on the within effects of each covariate than pooled OLS does. Results for the regional authority model estimated with random effects are listed in the third column of table 4. Note that the coefficients and standard errors are different from the results from the pooled OLS model.

[4]Furthermore, the effect of the time-invariant predictor z_i cannot be trusted when z_i is correlated with x_i^B, since the mediation effect is constrained.

Fixed effects

Unlike the previous three approaches, fixed effects estimators disaggregate the between and within variation. The interpretation of results from a fixed effects model are unambiguous, because the between variance is eliminated from a fixed effects model entirely: the unit effects are directly subtracted out of the model. This behavior of fixed effects is not seen as a disadvantage, but as its best characteristic. Econometricians are always concerned with minimizing the variance of the residuals in a linear model, and fixed effects remove a significant portion of this variance by design (Cameron and Trivedi 2005, p. 704).

The dependent variable in a fixed effects model is constrained to be the within version of y_{it}, and this constraint is mathematically equivalent to including a dummy variable for each panel. These dummy variables control for, but do not explain, the between variation. Since the between variation drops out of the model, all time-invariant predictors drop out as well. Fixed effects models can only answer questions about how changes in an independent variable over time associate with changes in a dependent variable over time, and because time-invariant predictors never change over time, they have no explanatory value by construction.

Proofs that fixed effects are consistent should not be taken as a directive to use fixed effects by political scientists whose research questions consider the differences between panels. Fixed effects are consistent only for within estimates, and say nothing about the between-panel effects. Various reformulations of the fixed effects model, such as error correction models which account for autocorrelation and time trends within the residuals (Keele and DeBouf 2008) are still fixed effects models, and therefore continue to have no bearing on questions regarding the differences between panels.

The results for the regional authority model estimated through fixed effects are listed in the fourth column of table 4. Note that unlike the previous three estimators, the coefficients are listed only on the "within" row for each covariate. Also note that ethnic fragmentation, which is time-invariant, drops out of the model.

Fixed effects with vector decomposition

Fixed effects with vector decomposition (FEVD) is a method designed by Plumper and Troeger (2007) which estimates a model in three steps. First, a fixed effects model is run, and the unit effect estimates produced by the model are saved. Variables which are fixed across time or have little variance over time are excluded from this first step. Second, these unit effect estimates are regressed on the independent variables which are time-invariant or rarely changing over time. The residual, which is called a decomposed unit effect (denoted in table 4 as η), is saved from this regression. Finally, a pooled OLS model is run which includes all of the time-variant, time-invariant, and rarely changing predictors, as well as η.

The idea behind FEVD is that the controversial assumption of random effects, that the unit effects are uncorrelated with all of the predictors, can be circumvented by separating the time-fixed variables from the unit effects. The unit effect estimates produced by the fixed effects model represent the unmodeled cross-sectional variance in the data. By regressing the unit effects on the time-invariant predictors, the residual η by definition is uncorrelated with the time-invariant predictors. In the final model, η theoretically contains all cross-sectional variance outside of the time-invariant predictors.

It is important to note that the FEVD estimates of time-variant predictors are within estimates only. Ideally, they should return the same values as the estimates from the fixed effects model. These estimates, however, are not exact since η is derived from a regression on estimated unit effects. During each of the three steps, estimation error can throw final estimates slightly off.

The fifth and sixth columns of table 4 contain results for FEVD estimates of the regional authority model. Both models specify that ethnic fragmentation is time-invariant. The model in the sixth column also specifies that area is a rarely changing variable. Note that the effects of time-variant predictors are very similar, though not exactly equal to, the fixed effects estimates in the fourth column. The standard errors for these estimates are also very low compared to the fixed effects estimates, and the R^2 returned by the model is much higher. The estimates for the time-invariant and rarely changing predictors are between effects, but they might suffer from omitted variable bias because the between effects for the other variables remain excluded.

Mundlak's approach of separating the between and within parts of each independent variable

Mundlak (1978) recommended running a random effects model in which each predictor is broken into its between and within parts. Mundlak criticizes the pooled OLS and random effects estimators for averaging these effects, stating that "any matrix combination of the 'within' and 'between' estimates is generally biased." The significance

Table 4: Results from Pooled OLS, Panel Corrected Standard Errors, Random Effects, Fixed Effects, Vector Decomposition, and Mundlak Estimators.

Variable		POLS	PCSE	RE	FE	FEVD (1)	FEVD (2)	Mundlak (RE)	Mundlak (POLS)
GDP p.c., $\log_{10}$	B	2.08(.32)***	2.08(.18)***	2.78(.20)***	-	-	-	17.8(13.3)	17.8(1.9)***
	W				3.62(.24)***	3.62(.13)***	3.62(.13)***	3.62(.24)***	3.62(.56)***
Pop., $\log_{10}$	B	6.02(.26)***	6.02(.09)***	-1.74(1.2)	-	-	-	6.15(1.71)***	6.15(.25)***
	W				-7.57(1.6)***	-7.57(.19)***	-7.57(.19)***	-7.57(1.6)***	-7.57(3.7)**
Area, $\log_{10}$	B	1.09(.23)***	1.09(.08)***	2.88(1.4)**	-	-	7.65(.12)***	1.58(1.6)	1.58(.22)***
	W				-5.67(3.2)*	-5.67(.12)***	-	-5.67(3.2)*	-5.67(7.4)
Ethnic frag.	B	11.8(.90)***	11.8(.51)***	16.4(6.5)**	-	26.3(.40)***	15.2(.36)***	6.85(6.6)	6.85(.94)***
Reg. party seats	B	30.3(2.1)***	30.3(1.5)***	18.6(1.3)***	-	-	-	45.5(19.4)**	45.5(2.8)***
	W				17.5(1.3)***	17.5(.84)***	17.5(.84)***	17.5(1.3)***	17.5(2.9)***
Freedom House	B	1.79(.12)***	1.79(.12)***	0.23(.07)***	-	-	-	0.63(2.0)	0.63(.29)**
	W				0.16(.07)**	0.16(.05)***	0.16(.05)***	0.16(.07)**	0.16(.17)
Constant		-65.8(2.3)***	-65.8(1.8)***	-1.08(7.7)	62.0(12.4)***	55.8(1.8)***	26.5(1.5)***	-113.7(33.1)***	-113.7(4.8)***
Groups		-	21	21	21	21	21	21	21
$N \times T$		1197	1197	1197	1197	1197	1197	1197	1197
σ_u		-	-	5.21	15.7	-	-	-	4.36
σ_e		-	-	2.05	2.05	-	-	-	2.05
ρ		-	-	.866	.983	-	-	-	.818
η		-	-	-	-	1(.01)***	1(.01)***	-	-
R^2 between		-	-	.290	.220	-	-	.738	-
R^2 within		-	-	.409	.420	-	-	.420	-
R^2 overall		.645	.645	.301	.164	.946	.946	.708	.708

Note: FEVD model 1 treats ethnic fragmentation alone as a time-invariant independent variable. FEVD model 2 also treats area, which is rarely changing, as invariant. Mundlak (POLS) separates the independent variables into between and within parts and runs a pooled OLS model which includes each of these variables. Mundlak (RE) also includes a country-level random effect, as with the random effects estimator.

* $p < .1$, ** $p < .05$, *** $p < .01$

of the between effects provided a test of the existence of unit effects, but the between variables themselves were not supposed to be interpreted. Two political scientists however, Zorn (2001) and Bartels (2009), have used Mundlak's approach and have explained the substantive interpretation of the between variables. The effect of the between variables can be interpreted as the effect of a one-unit change in x between panels at a typical time point in the data.

Mundlak recommended that all of the independent variables be broken up into their between and within parts so that separate effects can be estimated for each part. He also suggested that random effects be used to control for the remaining, unmodeled unit effects. In table 4, the seventh and eighth columns contain the results from two specifications of Mundlak models. In both models, the between and within effects are separated. The model in the seventh column is estimated by random effects and the model in the eighth column is estimated by pooled OLS. In both models, note that the coefficient point estimates and standard errors for the within variables are nearly identical to the results from the fixed effects model. The standard errors for the between effects are much higher, however, for the random effects model than they are for the pooled OLS model.

Table 4, as a whole, provides an illustration of the behavior of each estimator. Pooled OLS and random effects are different weighted averages, typically falling between the two effects estimated in the Mundlak models. Fixed effects is consistent, but only for the within estimates. FEVD estimates within effects with some estimation errors, and between effects with omitted variable bias. Standard errors in the Mundlak models are sensitive to whether unit effects are within the residual or not. Furthermore, these between effects are average effects over time, and obscure conclusions about how these effects might change at different points in time.

None of these approaches seems to be particularly satisfying for a researcher who is concerned with the differences between panels at recent, or at particular points in time. In the next section, I develop a methodology which is designed to analyze between effects directly. Unlike the Mundlak approach, which averages between effects over time, this methodology considers the between effects at specific points in time.

3 Methodology

The easiest way to obtain between effects is to perform a simple OLS regression on a cross-section. If we take one year out of a TSCS dataset, then time is guaranteed to be controlled in the model, and therefore there is no risk of confounding the between effects with any time effect. These regressions will often suffer from the inefficiencies of a small sample size. Returning to the regional authority data, we can remove one year, say 1960, and conduct an OLS regression on those data alone. These results are presented in table 5.

Table 5: OLS Results for Regional Authority in 1960

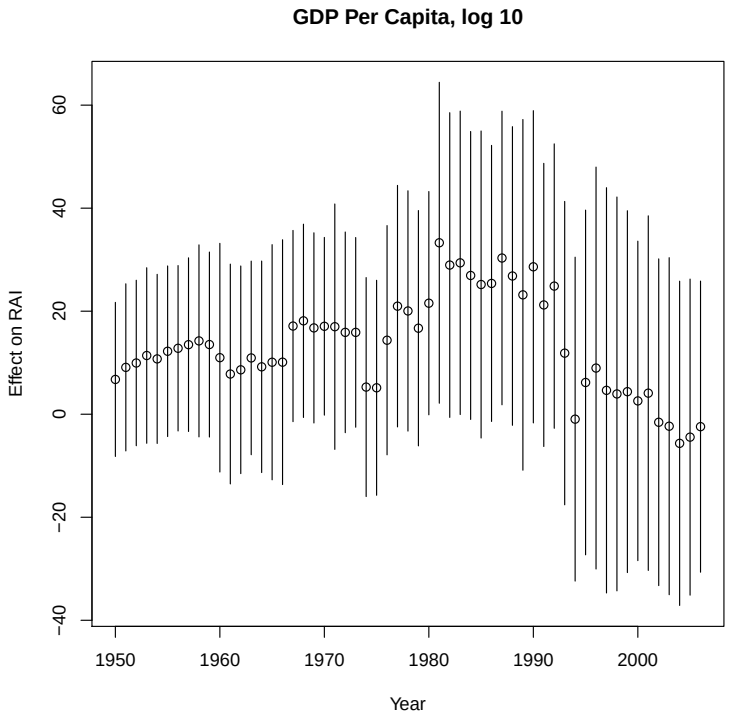
Variable	Coefficient(S.E)
GDP per capita, $\log_{10}$	11.0(11.3)
Population, $\log_{10}$	6.26(2.3)*
Area, $\log_{10}$	1.42(2.0)
Ethnic Fragmentation	13.42(9.0)
Regional Party Seats	53.6(50.4)
Freedom House	1.26(1.2)
Constant	-91.3(33.4)*
N	21
R^2	.664
Adj. R^2	.520

* $.01 < p < .05$

The sample size here is only 21, which is quite small by the standards of the law of large numbers, but rather big compared to the one-year sample size of many TSCS datasets in political science. The effect of population is in the hypothesized direction and achieves marginal significance. The effects of GDP per capita, area, ethnic fragmentation, the seats controlled by regional parties, and the Freedom House democracy score are in the right direction according to theory, but are not significant. These results are not satisfying: not because variables of interest fail to achieve significance, but because we have so much more information than is contained in these 21 points, but is unused.

If we repeat this analysis for each year in the data, we may see similar results, but we may see different results. What is guaranteed is that, no matter what the results are, the standard errors will be higher than they should be given the total amount of data we have. Even if we find a significant effect, we should be able to be even more conclusive about the magnitude of that effect. Figure 1 graphs the coefficient point estimate and 95% confidence interval for GDP per capita in every year in the data.

Figure 1: OLS Coefficient Point Estimates for GDP Per Capita, with 95% Confidence Intervals, From 1950 to 2006



The intuition behind the method presented here is that there is valuable information in looking at the entire timespan as we do in figure 1, and that there is a way to leverage this information to improve our inferences about the between effects. Until about 1995, figure 1 illustrates a relationship between wealth and federalism which is clearly very strong, even though few individual cross-sectional results are significant. The figure also shows a drastic change around the millennium which researchers will want to examine closely, although these effects individually are not significantly different from the stronger ones in the 1990s.

The between effects estimation routine takes advantage of the following two observations about between effects. First, if we see a similar effect being iterated in several proximate time points, then we can be more certain about the effect than we can by looking within any one time point alone. Second, between effects may change over time, therefore between effects need to be calculated with respect to a particular time point of interest. In short, a researcher who is interested in between effects needs to specify a particular time point in which to estimate these effects. Results from different time points are weighted with respect to this time point, and are averaged. The weight for a particular time point should depend on three considerations:

1. Similarity: time points which are empirically similar to the time point of interest should be weighted more heavily than time points which are less similar.

2. Proximity: time points which are proximate to the time point under consideration should also be weighted more heavily than time points which are further away.

3. Sample size: time points with larger cross-sectional sample sizes should be weighted more heavily than time points with smaller sample sizes.

Weights take the form $w_t \times n_t$, where n_t is the cross-sectional sample size in a time point t, and $w_t \in [0, 1]$ is a discount factor which depends on the similarity and proximity of time point t to the time point under consideration in the analysis. Using these weights, we can calculate an "updated" sample size. If $w_t = 1$ for every time point t, then every time point is weighted only according to its corresponding n_t, and the updated sample size is the total number of observations in the data. If $0 < w_t < 1$ for at least some time points t, then the updated sample size is bigger than the sample size within the time point under analysis, but is less than the total number of observations in the data.

To calculate similarity, a t-test is used which tests the null hypothesis that a point estimate in time point j, $\hat{\beta}_j$, with estimated variance $\hat{\sigma}_j^2$ and sample size n_j is equal to the point estimate from the time point k, $\hat{\beta}_k$, with variance $\hat{\sigma}_k^2$ and sample size n_k. The test statistic is

$$t_{j,k} = \frac{\hat{\beta}_j - \hat{\beta}_k}{\sqrt{\frac{(n_j-1)\hat{\sigma}_j^2+(n_k-1)\hat{\sigma}_k^2}{n_j+n_k-2}}\sqrt{\frac{1}{n_j}+\frac{1}{n_k}}}, \tag{8}$$

which is distributed by the student's T distribution with $n_j + n_k - 2$ degrees of freedom. To make the weights less sensitive to small differences, the t statistic is divided by 10, and the p value is calculated. This calculation is performed for each covariate in the analysis, and the mean of all of the p values is taken as the measure of the similarity between two time points.

Proximity is now included in the calculation by taking the absolute difference between each time point and the time point under consideration, in terms of the unit of time being used. For example, the proximity score for the comparison of years 2000 and 2010 is 10 years. This proximity score is set as the exponent of the similarity score, so that the weights are lower as years become less proximate.

The total updated sample size can be easily calculated after computing similarity and proximity. For example, in figure 2, data from 1950 through 1954 is used, and 1952 is the year under consideration. The years 1950-1954 each include a sample size of 10, and have similarity scores with 1952 of 0.8, 0.6, 1, 0.85, and 0.5 respectively. To

Figure 2: An Example of How to Calculate the "Updated" N for OLS Results in 1952.

1950 → 1951 → **1952** ← 1953 ← 1954

N	10	10	10	10	10
Similarity to 1952	0.8	0.6	1	0.85	0.5
Years away from 1952	2	1	0	1	2
Weight	$(0.8)^2$ =0.64	$(0.6)^1$ =0.6	$(1)^0$ =1	$(0.85)^1$ =0.85	$(0.5)^2$ =0.25

Total "updated" N = 0.64(10) + 0.6(10) + 1(10) + 0.85(10) + 0.25(10) = 33.4

calculate an updated sample size, each year's similarity score is raised to its proximity to 1952. The contribution of 1950 to the updated sample size for 1952 is $10 \times (0.8)^2 = 6.4$. 1951 contributes $10 \times (0.6) = 6$, 1952 contributes its entire sample size of 10, 1953 contributes $10 \times (0.85) = 8.5$, and 1954 contributes $10 \times (0.5)^2 = 2.5$, for a total updated sample size of 33.4.

Coefficient point estimates at time point t are calculated by taking a weighted average of the OLS coefficient point estimates from every time point in the set $\{1, 2, \ldots, T\}$:

$$\beta_t = \frac{\sum_{j=1}^{T} w_{t,j} \cdot n_j \cdot \hat{\beta}_j}{\sum_{j=1}^{T} w_{t,j} \cdot n_j}, \tag{9}$$

where n_j is the sample size in time point j, $\hat{\beta}_j$ is the vector of OLS coefficient point estimates from time point j, and

$$w_{t,j} = {s_{t,j}}^{p_{t,j}}, \tag{10}$$

where $s_{t,j}$ and $p_{t,j}$ are respectively the similarity and proximity scores between time points t and j. A more detailed discussion of this formula and the calculation of standard errors is presented in the appendix.

By averaging, both the coefficient standard errors and the point estimates will be different from the OLS results within a single time point. If many time points yield similar results to the time point of interest, then the average has the effect of increasing the model sample size, thereby lowering the standard errors. In other words, we will become more certain about the size of an effect if we see it duplicated again and again. If proximate time points yield different results than the time point under consideration, then the point estimates are adjusted to better reflect a current trend. In very noisy situations, the standard errors may increase. If one time point is an aberration, then we want to be cautious about accepting evidence from that time point as the truth if things return to "normal" later. If we observe wildly different effects over time, then we should become less certain about the size of the averaged effect.

These averages can be conceptualized as a Bayesian learning algorithm which uses conjugate priors. Conjugacy make it possible for these averages to be computed quickly and exactly, without relying on computationally intensive simulation techniques. The prior and posterior distributions used to compute the averaged point estimates and standard errors in a time point are listed in the appendix.

The between effects estimation routine can also provide evidence to help researchers find the time points in which the between effects change. Generally, the question of when the effect of an independent variable on a dependent variable changes over time is the focus of a large statistical literature on Bayesian change-point models (Chib 1997). Whether an effect changes at a particular time point can be tested in a time series by interacting the independent variable with an indicator for observations after the time point, as in the following regression equation:

$$y_t = \alpha + \beta_1 x_t + \beta_2 x_t \times I(t > 1980) + \varepsilon_t. \tag{11}$$

In this equation, the function $I(t > 1980)$ is 1 if the year under consideration is later than 1980, and 0 otherwise. The effect of x on y is β_1 for years before 1980, and $\beta_1 + \beta_2$ for years after 1980. When β_2 is significant, then 1980 can be called a change point. Bayesian change-point models try to determine the probability that any of the time points in the data are change points, and change point models for TSCS data in political science have been discussed by Western and Kleykamp (2004), Spirling (2007) and Park (2009). It is important to note, however, that all of these models, including the applications to TSCS data, posit change points on within effects only. These models are placing interactions on the coefficients of fixed effects models, and are treating the between effects like noise.

When we derive change points for between effects, we have an informational advantage over change point models for within effects. Models that rely on fixed effects consider exogenous time-variant predictors, and exogenous variables are known and have no standard error. The similarity score introduced in equation 8 takes the differences in point estimates into account, but also considers the standard errors of each estimate. By computing a similarity score for every pair of distinct time points, we can build a $T \times T$ matrix which contains these similarity scores. If we subtract each score from 1, then this matrix is a dissimilarity matrix for the data, which can be analyzed using many different statistical measurement techniques. Latent dimensions of the variance among time points can be estimated, or latent groups of time points can be identified. For example, suppose we have a model in which election outcomes are regressed on defense spending in western democracies. We calculate similarity between all pairs of years between 1960 to 2000, and we find that the years cluster into two discernable groups: before and after the year 1992. That could be evidence that the fall of the Soviet Union caused the importance of defense spending as an explanation for election outcomes to change.

As with any method, researchers need to be aware of some problems that arise when using the between effects estimation routine. First, since the method depends on a series of cross-sectional OLS models, potentially with very small sample sizes, the number of covariates must be less than the minimum number of observations in any one year. This requirement may present severe difficulties especially when some data are missing and are treated with listwise deletion. Additionally, due to collinearity with the constant, any variable that is fixed across panels but varies over time will drop out. Variables which are constant across panels in even one time point, for example an indicator variable for EU membership before the EU exists, will not be estimated at years in which there is no cross-sectional variation. Sharp changes in point estimates may be due to the addition or loss of panels rather than a more significant change in the underlying data generating process. Finally, although alternative OLS specifications, such as robust or cluster-robust standard errors, are allowed, the method can be, but has not yet, been formulated for limited dependent variables.

In section 4, the between effects are derived for the regional authority data. The effects are estimated for several different years including the most recent year in the data, and the results are examined for potential change points.

4 Results

RAI is regressed on the common logarithms of GDP per capita, population, and area, as well as the measures of ethnic fragmentation, seats controlled by regional parties, and the Freedom House democracy score within each of the 56 years in the data. These OLS estimates are then used to compute between effects in the most recent year, 2006, and also in 1960, 1970, 1980, 1990, and 2000. These between estimates are listed in table 6, along with their updated sample sizes.

Table 6: Between Effect Estimates at Particular Years

Variable	1960	1970	1980	1990	2000	2006
GDP per capita, $\log_{10}$	11.9(1.5)**	15.08(1.8)**	21.3(1.4)**	22.5(2.0)**	2.97(2.0)	0.00(2.8)
Population, $\log_{10}$	6.44(.25)**	5.90(.30)**	5.86(.23)**	5.82(.32)**	6.76(.33)**	7.00(.47)**
Area, $\log_{10}$	1.32(.22)**	1.57(.26)**	1.53(.20)**	1.52(.28)**	0.38(.29)	0.20(.41)
Ethnic Fragmentation	11.7(.90)**	10.3(1.1)**	9.60(.82)**	9.56(1.2)**	5.05(1.2)**	4.28(1.7)*
Regional Party Seats	52.1(3.6)**	45.5(4.3)**	32.2(3.3)**	36.4(4.6)**	49.7(4.7)**	50.6(6.6)**
Freedom House	2.07(.23)**	1.13(.27)**	0.19(.21)	-0.44(.30)	2.02(.31)**	2.74(.43)**
Constant	-105.0(4.6)**	-105.2(5.4)**	-123.2(4.3)**	-124.0(6.0)**	-74.8(6.0)**	-71.5(8.5)**
N	21	21	21	21	21	21
"Updated" N	314.38	266.90	342.61	245.71	239.48	169.04

* $p < .05$, ** $p < .01$

Through 1990, GDP per capita has an effect in which countries with higher levels of GDP per capita are positively and significantly associated with higher levels of RAI, but this effect has completely disappeared in recent years. In 1960, a ten-fold increase in GDP per capita is associated with a 11.9 increase in RAI, on average. By 1990 this effect has grown to 22.5, however by 2006 this effect has dropped to 0. A ten-fold increase in population consistently is associated with about a 6 or 7 point increase in RAI, but a ten-fold increase in area in 1960 is associated with less than a 2 point increase in RAI, and this effect disappears by 2006. Similarly, moving from the minimum amount to the maximum amount of ethnic fragmentation in 1960 is associated with a 11.7 point increase in RAI on average, but this effect shrinks, and by 2006 this increase is only associated with a 4.28 point increase in RAI. A legislature in which all the seats are held by regional parties in 1960 has on average a 52.1 point higher level of RAI than a legislature with no regional party seats. In contrast to ethnic fragmentation, this effect has endured: although the effect slips to 32.2 in 1980, it recovers to 50.6 by 2006. In 1960, as the Freedom House index becomes one-unit more democratic, RAI increases by 2.1 points on average. In 1980 and 1990 this effect disappears, but by 2000 and 2006 the effect once again is above 2.

For the most part, the between effects confirm the hypotheses laid out in section 2.1. In 1960 and 1970, all of the variables have effects in the hypothesized direction and achieve significance. It appears that population is a powerful logistic reason for regional authority, while area does not exhibit as strong of an effect when controlling for population. It also appears as though ethnic fragmentation is becoming a less important determinant of regional authority, especially when compared to the consistently strong effect of regional parties. This may indicate that ethnic conflicts are increasingly becoming institutionalized through political parties. Finally, the effect of the Freedom House democracy score is the most difficult one to understand, but the disappearance of the effect in 1980 and 1990 may be due to a lack of variation in the quality of democracy across the 21 cases during this time.

The results in table 6 list the between effect estimates in 6 representative years in the data, but it is illustrative to consider the movement of these estimates over the entire time span. The dots in the graphs in figure 3 represent the OLS estimates in every year, and the vertical lines represent the 95% confidence intervals for each point estimate. The between effect estimates with updated sample sizes are overlaid as a line on the OLS results, and the 95% confidence regions for the estimates are shaded. These graphs show how the between effects use more information than the OLS estimates, and therefore have lower standard errors. But they also show how the between effects change over time. It is immediately clear by looking at the graphs that the effects of GDP per capita, area, and ethnic fragmentation diminish over time. It is also clear that the between effects of regional party seats and the Freedom House score are weakest in the mid to late 1980s, but recover by the 2000s.

One important observation about the trends illustrated in figure 3 is that the effects for each independent variable do not move independently of one another; rather they tend to move together. The effects are basically stable for the first 15 years or so, and thereafter changes in the effects of different independent variables seem to occur around the same time. It is possible that these changes are a result of a changing underlying process which generates the observed distribution in regional authority. It seems as if there are eras in the data, and a measurement model can be used to gain some insight into the locations of these eras. As discussed in section 3, we can compute similarity scores for each pair of years in the data and build a similarity (or dissimilarity) matrix for the years. If the years can be aligned on a latent structure, then a method such as factor analysis or multidimensional scaling can provide an estimate of the dimensionality of the latent space and of how the years are aligned on this space. Cluster analysis can provide an estimate of how the years are grouped on this space.

It is important to note that these similarity measures depend only on the similarity of OLS results between two years. The proximity of one year to another is not considered in this calculation. If two consecutive years are very similar, it is strictly because the data in one year resemble the data in the next year.

For the regional authority data, similarity scores are calculated for every pair of years between 1950 and 2006. These scores are converted to dissimilarity scores and are assembled in a matrix. To illustrate the capabilities of this approach, I perform an average linkage cluster analysis on this matrix. The dendrogram displayed in figure 4 lists the individual years on the left-hand border, and shows the number of unique groups for any level of dissimilarity. A unique group is denoted by a single line. For example, if we take any dissimilarity greater than 0 to indicate a unique group, then there are 56 groups for the 56 years: therefore there are 56 unique lines on the left side of the dendrogram. The first two years to be grouped together are 1965 and 1966.

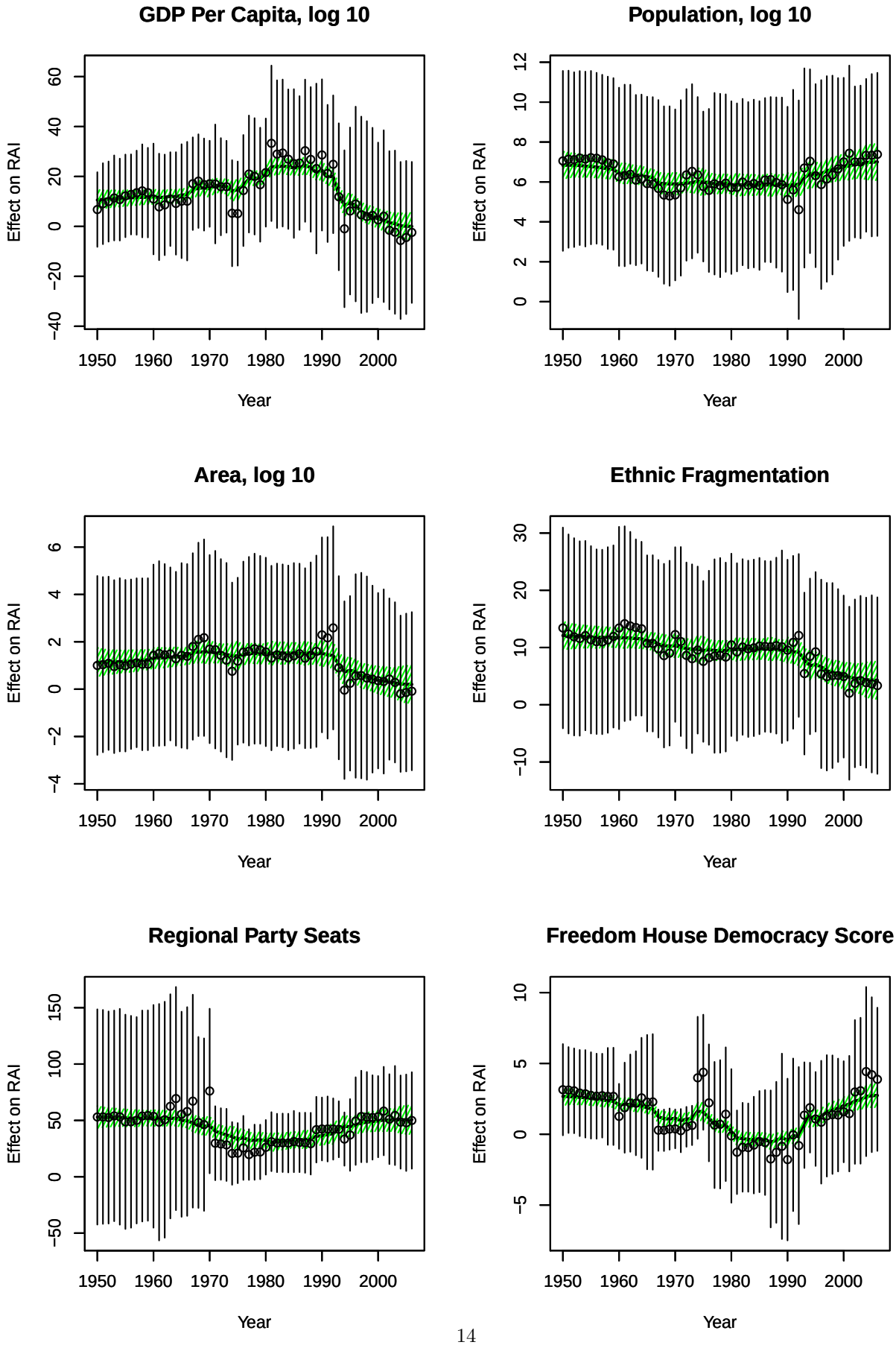
If we take a dissimilarity score greater than .1 to indicate a unique group, then there are 6 unique groups among the years in the data. These groups are listed in table 7. Note that although proximity is not specified as a criterion

Table 7: Groups from Average Linkage Cluster Analysis on Year Dissimilarity, Six Group Solution

Group	Years
1	1950-1966
2	1967-1973, 1976-1980
3	1974, 1975
4	1981-1992
5	1994, 1995
6	1993, 1996-2006

for similarity, the years are grouped for the most part in chronological order. Intuitively, these results make sense, and they accord with the trends we see in the graphs in figure 3. These groups should not be taken as a strict definition of the eras in the data. They can, however, inform additional theory by pointing to particular time points at which the underlying data generating process for the cross-section changes. The next step in the analysis of these data can be an investigation of special circumstances in 1966-1967, for instance, which lead to the changes we observe in the between effects.

Figure 3: Between Effect Coefficient Estimates, From 1950 to 2006



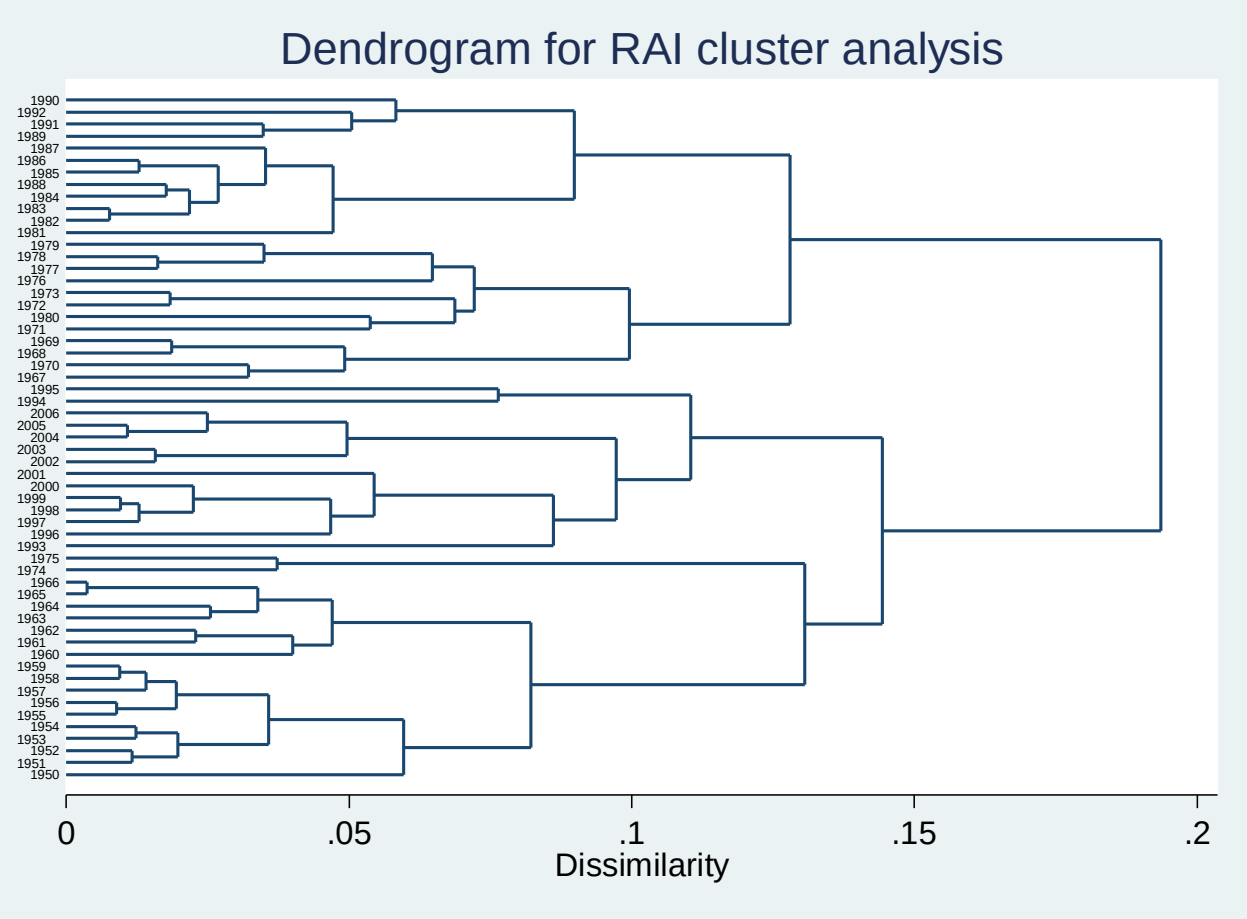


Figure 4: Dendrogram for the Cluster Analysis of Between Effects on RAI Over Time.

5 Discussion: What Does the Between Effects Estimation Routine Shows Us That Other Methods Do Not

As discussed in section 1, between effects in TSCS data consider the association between the levels of an independent variable on the levels of the dependent variable and ignore how these variables change over time. Within effects consider how changes in the independent variable affect changes in the dependent variable, but ignore the levels of these variables. Since there is no reason why a between effect should be equal, or even at all similar, to a within effect, political theories need to be refined to formulate hypotheses about the cross-sectional variation and the time variation in the data separately.

As shown in section 2.2, each of the current methods commonly used to estimate TSCS models take a different approach to estimating the between and within effects. Pooled OLS and panel corrected standard errors average the between and within effects. Random effects also averages these effects, but converges more quickly to the within effects as the time variation in the data increases relative to the cross-sectional variation. The family of fixed effects estimators considers only the within variation, as does fixed effects with vector decomposition, although FEVD estimates the between effect for time-invariant and rarely changing independent variables with some omitted variable bias. Mundlak's approach estimates the between and within effects separately, but the between effects are averaged over the time span.

Unlike any of the methods listed above, BEER provides estimates of the between effects only, and ignores the within effects. Furthermore, this method illustrates how between effects change over time, rather than assuming causal homogeneity over the time span.

These differences are important, and can have profound implications for the conclusions a researcher will draw from the statistical analysis.

In table 4 in section 2.2, the results from several competing methods are reported for the regional authority

data. The results from the between effects estimation routine are reported in section 4 in table 6 and in figure 3. What would we have concluded differently had we run one of the competing models in table 4 instead of BEER?

Had we run pooled OLS with or without panel corrected standard errors, we would have derived positive and significant effects for all six independent variables. In the earlier years in the data, these results more or less match the results from BEER, although we would derive stronger effects for GDP per capita and regional party seats using POLS than using the between effects routine. We would miss the fact that in later years the effects of GDP per capita and area disappear, and we would conclude that GDP and area are strong determinants of regional authority even though it is no longer true. However, there would be a more serious problem with our analysis: we would be unable to determine the extent to which the positive results we see from POLS apply to the within variation as opposed to the between variation. If we haven't refined our theory to consider the differences between levels and changes, then we are in danger of interpreting these positive effects as within effects. For population and area, these positive and significant results directly contradict the negative and significant within results estimated by fixed effects.

Had we run random effects, we would have encountered the same problems that POLS has in averaging the between and within effects. Again, it is not clear to what extent these effects apply to levels and to what extent these effects apply to changes. The effect of population is negative but is not significant, and is incorrect both from a between and a within perspective: it is consistently positive and significant using the between effects estimation routine, and is negative and significant using fixed effects.

Had we used any version of fixed effects, ethnic fragmentation would have dropped out of the model and the other variables would have measured changes rather than levels. For population and area, fixed effects provides estimates which are negative and significant while BEER provides estimates which are positive and significant.

When ethnic fragmentation is specified to be the only between variable, FEVD still contradicts BEER in regards to population and area, and the effect of ethnic fragmentation is estimated to be much stronger. When area is treated as a between variable as well, both area and ethnic fragmentation are too strong, although they are in the same direction as the estimates from BEER.

Mundlak's models are more in accord with BEER than are any of the other methods. However, both models are still incomplete in comparison to BEER. If we run Mundlak's model with POLS, we find positive and significant results for every covariate, and obscure the lack of significance for GDP per capita and area in the later part of the time series and for the Freedom House score in the middle of the time series. If we run Mundlak's model with random effects, we find positive but non-signficant effects for GDP per capita, area, ethnic fragmentation and the Freedom House score, obscuring the fact that all of these variables are significant during at least some part of the time series.

In short, each of the competing methods discussed in section 2.2 provides inferences which are, at least in part, different from the inferences drawn from BEER. Researchers, therefore, must take the choice between statistical models seriously because the differences between models will lead researchers to draw different conclusions.

6 Conclusion

Although the between effects estimation routine and each of the competing methods presented here produce different conclusions about the relationships in the data, it must be emphasized that none of these methods present a complete picture. BEER should not be used in place of the other methods, but rather it should be used *with* the other methods. An effective statistical investigation of TSCS data should present theory about both the between and within effects, test within hypotheses with some type of fixed effects model, test the between effects with a method like BEER, test causal homogeniety of the between effects across time using Mundlak's method, and test causal homogeniety across levels and changes using random effects or pooled OLS. One concern might be that running several analyses can give a researcher leeway to report only the most favorable results, and obscure the real relationships. However, if a researcher adopts a design which requires reporting the results from all of these methods, and understands the focus of each method, then the presentation of results will be more nuanced and complete than can be achieved using any single method alone. It is not the case with TSCS data that a relationship simply exists or does not exist; relationships look different over time than they do in a cross-section, and need to be examined from all perspectives.

7 Appendix: Formulation of the Model

The goal is to estimate the parameters for a linear regression data at one time point, t, by using additional information from other time points. The sufficient statistics for the multivariate normal distribution assumed for

the errors in an OLS regression are the coefficients , β, and the residual variance, s^2. The variance-covariance matrix of the coefficients, $V(\hat{\beta}) = \hat{\Sigma}$, is determined by the coefficient estimates and the residual variance. However, in this application, I will treat β_t and Σ_t as unknown. The logic is that information from other time points may change both the coefficient point estimates, and also their standard errors and correlations. The advantage of a Bayesian setup is that information from other time points can be summarized in the prior, which can then be used to update the estimates of β_t and Σ_t. First, in section 7.1, the OLS parameters are expressed in a way that conjugate priors exist and can easily be applied. Second, in section 7.2, a strategy is discussed to summarize information from different time points within the prior distributions.

7.1 Bayesian Representation of the Regression Parameters

We can express the joint probability of β_t and Σ_t in terms of the conditional probability of β_t given Σ_t and the marginal probability of Σ_t:

$$P(\beta_t, \Sigma_t) = P(\beta_t|\Sigma_t) \cdot P(\Sigma_t). \tag{12}$$

Given data X_t and y_t that contain n_t observations and k covariates, the conditional probability of β_t is the likelihood function for β, which follows the multivariate normal distribution,

$$P(\beta_t|\Sigma_t) = L(\beta_t|\Sigma_t, X_t, y_t) = N_k\left(\hat{\beta}_t, \frac{\Sigma_t}{n_t}\right), \tag{13}$$

where

$$\hat{\beta}_t = (X_t'X_t)^{-1}X_t'y_t. \tag{14}$$

Σ_t follows the inverse-Wishart distribution (Gill 2008, p.570-571):

$$P(\Sigma_t) = L(\Sigma_t|X_t, y_t) = W^{-1}(n_t, \hat{\Sigma}_t), \tag{15}$$

where W^{-1} denotes the inverse-Wishart PDF, and

$$\hat{\Sigma}_t = \left[\frac{n_t}{n_t - k}(y_t - X_t\hat{\beta}_t)'(y_t - X_t\hat{\beta}_t)\right](X_t'X_t)^{-1}. \tag{16}$$

When expressed with the likelihood functions listed in equations 13 and 15, the maximum likelihood estimates for β_t and Σ_t are the same as the OLS estimates for these parameters. This setup for the distributions of β_t and Σ_t may seem overly complicated compared to the standard OLS equations, but this presentation makes it easier to derive the posterior distribution when information from other time points are included as a prior distribution.

Let $L(\beta_t, \Sigma_t|X_t, y_t)$ represent the probability distribution for β_t and Σ_t that depends only on data from time point t. Information from different time points allow us to develop prior beliefs about the parameters, which are contained in the prior distribution $P(\beta_t, \Sigma_t)$. Bayes' theorem allows us to calculate the posterior distribution:

$$P(\beta_t, \Sigma_t|X, y) \propto L(\beta_t, \Sigma_t|X_t, y_t) \cdot P(\beta_t, \Sigma_t)$$

$$= L(\beta_t|\Sigma_t, X_t, y_t) \cdot L(\Sigma_t|X_t, y_t) \cdot P(\beta|\Sigma) \cdot P(\Sigma). \tag{17}$$

Conjugate priors are available for this likelihood function. Following Gill (2008, p.81-84), Gelman et al (2004, p. 87-88), and Robert (2001, p. 189-190), when the following prior distributions are used,

$$\beta_t|\Sigma_t \sim N_k\left(\bar{\beta}, \frac{\Sigma_t}{n_o}\right), \quad \text{and} \quad \Sigma_t \sim W^{-1}(a, B), \tag{18}$$

the posterior conditional distribution for β_t is

$$\beta_t|(\Sigma_t, X_t, y_t) \sim N_k\left(\frac{n_o\bar{\beta} + n_t\hat{\beta}}{n_o + n}, \frac{\Sigma_t}{n_o + n_t}\right), \tag{19}$$

and the posterior marginal distribution for Σ_t is

$$\Sigma_t|(X_t, y_y) \sim W^{-1}\left(a + n_t, \left[B^{-1} + \hat{\Sigma} + \frac{n_o n_t}{n_o + n_t}(\hat{\beta} - \bar{\beta})(\hat{\beta} - \bar{\beta})'\right]^{-1}\right). \tag{20}$$

Table 8: Formulas to Calculate Posterior Arguments From the Prior and Data Arguments.

Argument	Prior	Data	Posterior
Sample Size	n_o	n_t	$n_o + n_t$
Coefficient means	$\bar{\beta}$	$\hat{\beta}$	$\frac{n_o\bar{\beta} + n_t\hat{\beta}}{n_o + n_t}$
Inverse Wishart Degrees of Freedom	a	n_t	$a + n_t$
Inverse Wishart Scale Matrix	B	$\hat{\Sigma}$	$\left[B^{-1} + \hat{\Sigma} + \frac{n_o n_t}{n_o + n_t}(\hat{\beta} - \bar{\beta})(\hat{\beta} - \bar{\beta})'\right]^{-1}$

In order to update the parameters from a regression on data from time point t, we only need to calculate the four posterior arguments using the values of these arguments from the prior distribution and from the regression. The formulas to calculate these posterior arguments are listed in table 8.

Let n^P, β^P, a^P, and B^P be respectively the sample size, coefficient mean vector, inverse Wishart degrees of freedom, and inverse Wishart scale matrix for the posterior distribution. The mean of the inverse Wishart distribution can be used for the variance-covariance matrix of the coefficients (Robert 2001, p. 190):

$$E(\Sigma^P) = \frac{(B^P)^{-1}}{a^P - k - 1}. \tag{21}$$

The distribution of the updated coefficients is

$$N_k\left(\beta^P, \frac{\Sigma^p}{n^P}\right), \tag{22}$$

therefore β^P contains the coefficient point estimates, and the coefficient standard errors are contained in the square root of the diagonal of $E(\Sigma^P)/n^P$.

7.2 Accumulating Information Over Time Within the Prior Distributions

Information is accumulated through time by using the posterior results for one time point as the prior distribution for the next time point. As described in section 3, two chains are computed which converge on the time point of interest. One chain starts at the first time point and moves forwards through the time point which precedes the one under consideration. The second chain begins at the last time point and updates backwards through the time point under consideration. The posterior results from the first chain are then used as a prior distribution on the posterior results from the second chain.

Suppose that there are T time points in the data which can be labeled times 1 through T. The posterior results for time point $t \in \{1, 2, \ldots, T\}$ can be calculated by computing the forward-moving and backward-moving chains. Unless informed priors are desired, uninformed priors are used for time 1, yielding $\bar{\beta} = \hat{\beta}_1, n = n_1, a = n_1$, and $B = \hat{\Sigma}_1$. These posterior estimates for time 1 are used as the priors for time 2, and β^P, n^P, a^P and B^P are recalculated using the formulas in table 8. Then these estimates are used as the priors for time 3, and so on, through time $t-1$. Independently, uniformed priors are used for time T, which yields posteriors which are used to compute parameters for time $T-1$, and so on, through time t. Finally, the posterior results for $t-1$ are used as a prior on the results for t.

As noted by Gill, the prior sample size parameter n_o should not necessarily represent the empirically derived prior sample size, rather it is "intended to be a reflection of prior precision relative to the sample size that is tunable by the researcher to reflect prior confidence" (2008, p. 83). When combining information from all of the time points in the data, I reduce the sample size in time point j by multiplying it by a weight parameter $w_{t,j} \in [0, 1]$. This weight, defined in section 3, depends on both the similarity and the proximity of time point j to the time point of interest in the analysis.

The scale matrix of the inverse Wishart matrix, B, must be computed using the algorithm described above. The other three arguments, however, can be updated more efficiently. It can be shown that by combining the forwards-moving and the backwards-moving chains, the posterior or "total updated" sample size is the weighted average of the scaled sample sizes from every time point in the data:

$$n_P = \sum_{j=1}^{T} w_{t,j} \cdot n_j. \tag{23}$$

The posterior estimate for the inverse Wishart degrees of freedom is equivalent to posterior sample size. The posterior coefficient means at time point t are also a weighted average of coefficient point estimates:

$$\beta_t^P = \frac{\sum_{j=1}^{T} w_{t,j} \cdot n_j \cdot \hat{\beta}_j}{n_P}. \tag{24}$$

References

Bartels, Brandon L. 2009. "Beyond 'Fixed Versus Random Effects': A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data." Under review, *APSR*.

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

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Chib, Suddhartha. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics*. 86:22141.

DeBouf, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science*. 52(1):184-200.

Driver, Ciaran and Jair Muñoz-Bugarin. 2009. "Capital Investment and Unemployment in Europe: Neutrality or Not?" *Journal of Macroeconomics*. In Press.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*. 8(2):195-222.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Second Ed. Texts in Statistical Science. New York: Chapman and Hall.

Gill, Jeff. 2008. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Statistics in the Social and Behavioral Sciences Series. New York: Chapman and Hall.

Hooghe, Liesbet, Gary Marks and Arjan Schakel. 2008. "Operationalizing Regional Authority: A Coding Scheme for 42 Countries, 1950-2006." *Regional and Federal Studies*. 18(2,3):123-142.

Kittle, Bernhard. 1999. "Sense and Sensitivity in Pooled Analysis of Political Data." *European Journal of Political Research*. 35: 225-253.

—. 2006. "A Crazy Methodology? On the Limits of Macro-Quantitative Social Science Research." *International Sociology*. 21(5):647-677.

Marks, Gary, Liesbet Hooghe and Arjan Schakel. 2008. "Measuring Regional Authority." *Regional and Federal Studies*. 18(2,3):111-121.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*. 46(1): 69-85.

Park, Jong Hee. 2009. "Joint Modeling of Dynamic and Cross-Sectional Heterogeniety: Introducing Hidden Markov Panel Models." Working paper.

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

Robert, Christian P. 2001. *The Bayesian Choice*. Second Ed. New York: Springer.

Spirling, Arthur. 2007. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis*. 15(4).

Western, Bruce and Meredith Kleykamp. 2004. "A Bayesian Change Point Model for Historical Time Series Analysis." *Political Analysis*. 12(4):354-374

Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More To Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications". *Political Analysis*. 15:101-123.

Zorn, Christopher. 2001. "Estimating Between- and Within-Cluster Covariate Effects, With an Application to Models of international Disputes." *International Interactions*. 27:433-445.