Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets

Malcolm Fairbrother

School of Geographical Sciences
University of Bristol

University Road
Bristol BS8 1SS
United Kingdom

m.fairbrother@bristol.ac.uk
+44 117 928 8303

30 March 2013

Abstract:
Increasing numbers of comparative survey datasets span multiple waves. Moving beyond purely cross-sectional analyses, multilevel longitudinal analyses of such datasets should generate substantively important insights into the political, social, and economic correlates of many individual-level outcomes of interest (attitudes, behaviors, etc.). This paper describes two simple techniques for extracting such insights, which allow change over time in y to be a function of change over time in x, and/or of a time-invariant x. The paper presents results from simulation studies assessing the techniques in the presence of complications likely to arise with real-world data, and concludes with applications to the issues of generalized social trust and postmaterialist values, using data from World/European Values Surveys.

Keywords: multilevel, comparative, longitudinal, World Values Survey, trust, values

Acknowledgements
For useful comments and suggestions, thanks to: Rima Wilkes, Andy Bell, Dmitriy Poznyak, David Manley, and Kelvyn Jones; fellow members of the University of Bristol's School of Geographical Sciences Spatial Modelling Group; and audiences at ECPR 2011 in Reykjavik, ESRA 2011 in Lausanne, the 2011 ASA Spring Methodology Conference at Tilburg University, and a 2012 symposium on "The Quality of Measurement" at the Technische Universität Dresden. Thanks also to Gethin Williams for help running simulations remotely and in parallel on multiple cores of a Linux server.

Word count (incl. this page, footnotes, and references, excl. appendix): 10,308

1 Introduction

Comparative survey data consist of observations on survey respondents drawn from multiple countries or sub-national jurisdictions such as states, provinces, or metropolitan areas.[1] Comparing respondents from different jurisdictions can help to clarify how people differ depending on their social and political contexts, and to test theories about the consequences of macro-social conditions for individuals' attitudes, beliefs, circumstances, and behaviors.

Particularly in the last decade, multilevel models have become a standard method of conducting analyses of this kind. Multilevel models, also known as mixed, random effects, or hierarchical linear models, are useful for analyzing observations which are grouped or clustered (for overviews see e.g., Gelman and Hill 2007; Steenbergen and Jones 2002). Recent applications to comparative survey data have examined and sought to identify the covariates of outcomes like voting (Gelman et al. 2008), political attitudes (Jamal and Nooruddin 2010), union membership (Martin and Brady 2007), religiosity (Karakoç and Başkan 2012), health (Jen, Jones, and Johnston 2009), poverty (Brady, Fullerton, and Cross 2009), and gender gaps in household labor (Iversen and Rosenbluth 2006). This diversity of applications demonstrates the broad appeal of the method.

But having found many intriguing cross-sectional associations, very few of these studies address the issue of change over time, and how the kinds of individual-level outcomes cited above may depend on time-varying social, political, and economic conditions. This paper therefore addresses the question of how to apply multilevel models to survey data that are not just comparative, but also longitudinal:

[1] For convenience, this paper will refer to countries, but the same principles apply to sub-national jurisdictions.

repeated cross-sectional data, where each cross-section includes a new sample of respondents drawn from the same set of higher-level units. Researchers have access to increasing numbers of datasets of this kind, and a small number of studies have already made use of multiple waves of data from one or more cross-national surveys fielded multiple times. Yet such studies have generally not made use of the longitudinal character of the data, and—for reasons discussed below—in some cases the approaches that have been used could even be misleading.

This paper describes two techniques for analyzing such data and extracting insights about the correlates of change over time; both techniques are simple, incremental extensions of existing multilevel modeling methods. The first is a means of distinguishing between cross-sectional and longitudinal relationships when working with a time-varying, macro-level covariate. Change in y can then be linked to longitudinal variation in some x. The second is the application of "growth curves"—a multilevel modeling technique previously used predominantly with data on individuals—to the level of whole societies. Here, change in y is a function of some time-*in*variant x.

The paper then presents results from a simulation study assessing the performance of these two techniques when applied to data with complications that could occur in real-world applications. Among other things, the simulations show that cross-sectional correlations are biased where a national-level variable is correlated with national-level random intercepts/error terms; the coefficients on the two longitudinal terms proposed here, in contrast, are unbiased even in the presence of correlated national-level regressors and errors. Further simulations show how statistical power varies with the number of waves of data per country.

Finally, to illustrate the kinds of insights that can be derived using the two techniques, the paper applies them to data from the World/European Values Surveys, and investigates the relationships between inequality and generalized social trust, and between living standards and postmaterialist values. Previous research has found cross-sectional associations between these pairs of variables; this paper tests whether these associations hold longitudinally as well.

2 The Promise of Extending Multilevel Models Longitudinally

Most existing studies fitting multilevel models to comparative survey data rely exclusively on cross-sectional comparisons, sometimes with few units and thus degrees of freedom at the higher (country) level. In fact, although the headline conclusions of many such studies rest entirely on comparisons across the higher-level units, in many instances there are fewer than 30 of them, and occasionally fewer than 20 (e.g., Anderson and Tverdova 2003; Brady, Fullerton, and Cross 2009; Iversen and Rosenbluth 2006; Mattes and Bratton 2007; Weldon 2006). The results of such cross-sectional comparisons will clearly be sensitive to small changes in either the set of higher-level units or the right-hand-side variables included in the analysis (Beck 2007: 97; Wilkes et al. 2007).[2]

In the context of a theoretical argument about a causal relationship, if a given covariate is cross-sectionally associated with an outcome of interest, checking for the presence of the same relationship longitudinally is a logical next step in validating that the relationship is not spurious. And longitudinal research is a direct way of

[2] On the other hand, Stegmueller (2013) presents simulations indicating that reliable inferences can be derived from multilevel models with as few as 15 higher-level units, and even fewer than that, using Bayesian/MCMC estimation.

studying social change, in the most general sense—a major task of the social sciences, and a central focus of many specific fields of research.[3] Currently, however, studies seeking to understand social change often make inferences about longitudinal relationships based on the results of cross-sectional research. Ruiter and van Tubergen (2009), for example, derive conclusions about the long-term process of secularization from their cross-sectional study of religious attendance. And on the basis of their cross-sectional study of unionization, Martin and Brady (2007) argue that when a country signs an agreement with the IMF, there is a decline in the probability of a given resident of that country being a union member. Conclusions about social change based on cross-sectional research are not necessarily wrong, but they are clearly dependent on the ambitious assumption that longitudinal and cross-sectional relationships are the same. As Gelman (2005: 461) puts it, "it is a big leap to interpret differences between countries as a potential effect of a change within a country."

The assumption of cross-sectional and longitudinal equivalence will in many instances not be valid. Researchers working with national-level time-series cross-sectional (TSCS) data have found that such data typically fail a standard Hausman (1978) test of whether cross-sectional and longitudinal relationships are equivalent (Wilson and Butler 2007). By implication, they are different, and treating them singly will be misleading.

The non-equivalence of cross-sectional and longitudinal relationships has had consequences for the statistical models typically used in analyses of TSCS data. TSCS data can be analyzed using multilevel models, where country-observations are the level-1 units, countries are the level-2 units, and level-2 random intercepts are

[3] In the words of De Boef and Keele (2008: 184), "research in all subfields of political science is devoted in large measure to understanding the causes and consequences of events, opinions, behavior, and institutional change as they unfold over time."

estimated from the data. In practice, however, multilevel models have been used much less than either models estimated using OLS with panel-corrected standard errors, or models using fixed effects (FE), which use country dummy variables instead of random intercepts (see Wilson and Butler 2007 for a discussion). The dummy variables in FE models hold constant and thus control for differences among countries, though this control comes at the cost that time-invariant country-level variables cannot be included, because the country dummies use up all the degrees of freedom. The FE approach thus allows for the estimation of only "within" effects, not "between" effects (see e.g., Kittel and Winner 2005), and cannot test for the correlates of cross-sectional differences across countries.

Recent methodological studies and tests have shown that the few drawbacks of multilevel models for analyzing TSCS data can be relatively easily avoided, however, and such models have several advantages vis-à-vis the FE approach: "vastly increased flexibility in model specification, potential improvements in model fit, and better accounting of uncertainty at all levels of analysis" (Shor, Bafumi, Keele, and Park 2007: 166; see also Bell and Jones 2012; Western 1998). The possibility of correlation between a time-varying covariate and the country random intercepts is sometimes cited as a reason not to use a multilevel approach, since correlation of this kind will produce a biased estimate of the coefficient on that covariate. But this problem is easily addressed through the inclusion in the model of the group mean of the time-varying covariate (see Bafumi and Gelman 2006; Bartels 2008; Shor et al. 2007; Skrondal and Rabe-Hesketh 2008; and the simulations below).

3 Characterizing Comparative Longitudinal Survey Data

This section briefly highlights some key features of comparative longitudinal survey data, and of multilevel models fitted to such data.

First, comparative longitudinal survey data are a sub-category of repeated cross-sectional survey data. The latter are not necessarily comparative, where comparative here means that respondents' membership in larger macro-level units is one of their key characteristics, insofar as it allows for comparisons among those higher-level units. A study of repeated cross-sectional survey data where all respondents are drawn from a single higher-level unit would not be comparative.

Second, although societies are by no means reducible to their component members, and survey data are collected on individuals, such data are often sought specifically because they are, in aggregated form, an effective means of measuring differences among societies and for tracking changes in societies over time. But despite the frequency with which survey data are considered in aggregated form, statistical analyses of aggregated data are less informative than those of disaggregated, individual-level data. An analysis based on aggregated data can even be highly misleading, and it risks committing an ecological fallacy, where aggregate-level relationships are erroneously assumed to hold also for the individual units on which the pre-aggregated data were collected (see e.g., Subramanian et al. 2009). To take one notable recent example, Gelman et al. (2008) show that voters in American states with higher average incomes have recently tended to vote more for Democratic than Republican candidates. While an analysis using data aggregated to the state level would therefore suggest that Democratic voters tend to be wealthy elites, that is not the case: a multilevel analysis reveals that, within states, Democratic voters tend to be lower income-earners than Republican voters.

Third, the data drawn from multiple waves of a comparative survey, where different individual respondents are sampled in each wave, are not generally referred to as panel data; each respondent is observed only once, and panel data are typically defined as data based on repeated observations on the same units. But insofar as societies are units of analysis at least as important as the individuals on which observations are made, then in an important sense the data derived from such surveys are in fact panel data, since societies are each observed multiple times.

Fourth, macro-comparative social scientists working with quantitative country-level data tend to draw a firm distinction between panel data and time series cross-sectional data. There are often just 15 to 20 units observed on repeated occasions in TSCS datasets, whereas typical panel datasets include hundreds or even thousands of units—such as a nationally representative random sample of individuals. Comparative longitudinal survey data are thus non-repeated observations on a large random sample of micro-level units (like panel datasets), and repeated observations on a small non-random sample of macro-level units (like TSCS data).

4 Existing Models of Comparative Longitudinal Survey Data: De Facto Pooling

This section briefly reviews the approach taken by previous studies in applying multilevel models to comparative longitudinal survey data.

A typical two-level multilevel model, where observations are nested within groups, can be represented as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\zeta} + \boldsymbol{\varepsilon}, \tag{1}$$

where **y** is a vector of outcomes, **X** is a design matrix comprising a column of ones and columns of covariates, **β** is a vector of coefficients, **Z** is a design matrix relating random predictors **ζ** to the data, and **ε** is a vector of residuals. Unlike the unit dummies in a fixed effects model, the elements of **ζ** are not parameters, but random variables conforming to a variance-covariance matrix whose values are parameters, and are estimated from the data. A model of this kind, where individual survey respondents (indexed i) are nested within higher-level societal units (indexed j), and there is one individual- and one group-level covariate, can also be written as:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_j + u_j + e_{ij}\,, \qquad \text{for } i = 1 \ldots I_j, j = 1 \ldots J,$$
$$\text{where } u_j \sim \mathrm{N}(0, \sigma_u^2) \text{ and } e_{ij} \sim \mathrm{N}(0, \sigma_e^2)\,. \qquad (2)$$

Each j has random intercept (or group-level disturbance) u_j. The x variables can be either variable across individuals within countries (and are subscripted ij) or invariant within but variable across countries (j). Some representations of multilevel models prefer to use a series of two or more equations (one at each level), and this paper adopts such an approach below, where useful.

Aside from Normally distributed data, multilevel models have long been extended to generalized linear models with binary, count, proportion, etc. outcomes (including event history or survival analysis), using a variety of estimation techniques and link functions.

Among the few previous studies that have incorporated multiple waves of comparative survey data, the dominant approach has been to fit a three-level model, where respondents (i) are nested within country-years (tj), nested in turn within countries (j). Such a model can include a coefficient capturing the effect on y of an x variable that varies both across countries and within countries over time. This x

variable is considered to be a characteristic of country-years, and is indexed *tj*, because it is constant for individuals within a given country-year, and non-constant across both countries and the country-years nested within a given country:

$$y_{itj} = \beta_0 + \beta_1 x_{itj} + \beta_2 x_{tj} + u_j + u_{tj} + e_{itj} \,,$$
$$\text{where } e_{itj} \sim \mathrm{N}(0, \sigma_e{}^2) \,,\ u_{tj} \sim \mathrm{N}(0, \sigma_{u1}{}^2) \,,\ \text{and } u_j \sim \mathrm{N}(0, \sigma_{u2}{}^2) \,. \qquad (3)$$

Jen et al. (2009) adopt this approach, for example, in their test of Wilkinson and Pickett's (2006, 2009) hypothesis that societal-level income inequality tends to lead people to have poorer health. Contrary to multilevel analyses of individuals nested in U.S. states (e.g., Kennedy et al. 1998), they find that across countries inequality does not correlate with poorer health. Solt (2008) uses the same approach, in studying the effect of inequality on political engagement.[4]

This approach investigates social change only in the limited sense that, in making use of all the available survey data, those data happen to have a longitudinal dimension. Though the x variable of interest varies both over time and across countries, these two dimensions of its variability are treated singly—making it impossible to know whether just one dimension is driving any covariation found with y, or even potentially whether the two associations have different signs. This approach effectively assumes that the cross-sectional and longitudinal relationships between x and y are the same: a single coefficient β linked to x_{tj} captures both. That assumption of equivalence, as discussed earlier, will often be unjustified.

[4] Country-years could also be cross-classified in years, in research contexts where there may be variable characteristics of years or survey waves that influence ("exogenously shock") many countries simultaneously (see Shor et al. 2007). Where the number of years or waves is few, on the other hand, fixed effects will be more appropriate for controlling for such shocks.

5 Distinguishing Between Cross-Sectional and Longitudinal Relationships

I now turn to extensions of existing multilevel modeling techniques that address the limitations of existing approaches. The first extension I describe allows for the simultaneous but separate analysis of cross-sectional and longitudinal relationships. The technical requirement for distinguishing between cross-sectional and longitudinal relationships is simple: calculating a mean, and subtracting that mean from the time-varying variable of interest x_{tj}. The technique thus group mean-centers the covariate.[5]

To identify separate longitudinal and cross-sectional associations between x_{tj} and y, calculate the mean of x_{tj} across all relevant years for each country. The coefficient on the country mean $\bar{x_j}$ captures the effect on y of enduring cross-national differences in x_{tj}. To capture the effect on y of variation over time within each country, $\bar{x_j}$ can then subtracted from x_{tj}. The resulting longitudinal component x_{tjM} (a country-year level variable) is group mean-centered, and is orthogonal to $\bar{x_j}$, such that the two coefficients can be estimated separately.[6]

The resulting model is:

$$y_{itj} = \beta_0 + \beta_1 x_{itj} + \beta_2 x_{tjM} + \beta_3 \bar{x_j} + \beta_4 time_{tj} + u_j + u_{tj} + e_{itj} \ . \qquad (4)$$

The original variable x_{tj} is thus entered twice, having been decomposed into $\bar{x_j}$ and x_{tjM}. Binary variables can be treated the same as continuous ones; the group mean simply becomes a proportion (see Enders and Tofighi 2007). The equation also

[5] The inclusion of the group mean in a model is sufficient to distinguish the two relationships even without centering (see Mundlak 1978), but centering yields more directly interpretable results (see Bell and Jones 2012).

[6] That the variables are orthogonal is clear given that their dot product is equal to zero, signifying a right angle between the Euclidean vectors they define. The value of $\bar{x_j} \cdot x_{tjM}$ is the sum across $j = 1 \ldots J$, $t_j = 1 \ldots T_j$, of $\bar{x_j} x_{tjM}$. Since by construction $E(x_{tjM}) = 0$ for each macro-unit j, the sum of $\bar{x_j} x_{tjM}$ within and across all such units is also zero.

includes a variable for time—this could be a set of year dummies and/or a linear effect, and could include a quadratic term too. The need for a time term arises from the possibility of a simultaneous but unrelated time trends in both x_{tj}/x_{tjM} and y. If both are increasing over time, including a control for time reveals whether y tends to grow faster for units experiencing a faster increase in x.

This extension, which allows for both a "between" and "within" effect, represents an important improvement on existing techniques: it provides a direct investigation of social change without assuming that the longitudinal relationship is the same as the cross-sectional one. This extension investigates the possible covariation between a shift in x over time and a shift in y over time. This relationship can be investigated even as the cross-sectional association between $\bar{x_j}$ and y is also estimated. Note that time-varying national-level variables can be the national average of some characteristic measured on individuals (e.g., education, social capital).

There are precursors to this extension. Previous studies and commentaries have remarked on the benefits of group mean centering when fitting multilevel models (see Enders and Tofighi 2007; Raudenbush 1989; Wu and Wooldridge 2005). As mentioned earlier, Bafumi and Gelman (2006) note that group-mean centering provides a simple resolution of concerns about correlation between one or more regressors and the random effects, or equivalently the higher-level errors. Bartels (2008) has recently argued for the benefits of country-mean centering in analyses of national-level TSCS data. And Moller, Alderson, and Nielsen (2009) use group mean centering in analyzing the drivers of cross-sectional and longitudinal variations in U.S. counties' levels of inequality. The author has recently applied this technique in a study with U.S. states as the macro-level unit (Fairbrother and Martin 2013), but otherwise its usefulness for analyses of comparative longitudinal survey data has not previously been noted.

6 Societal Growth Curves

The technique introduced in the previous section investigates the association between a shift in x_{tj} over time and a shift in y over time. But it does not investigate whether the absolute level of some time-invariant (continuous or categorical) variable x_j leads to faster versus slower change in y. Yet one would ideally want to test for such patterns, and this section introduces a technique for doing so: the application of "growth curve modeling" to the higher level units in comparative longitudinal survey data. An x_j variable of interest in this regard could be the country mean $\bar{x_j}$ discussed above, though it could also denote some other time-invariant macro-level characteristic, such as the share of a jurisdiction's territory located in the tropics; the identity of a country's former colonial ruler; or the type of welfare state regime.

"Growth curves," which entail multilevel models of measurement occasions nested within individuals, have long been applied to individual-level panel data, such as in psychology or education research (see e.g., Willett, Singer, and Martin 1998). Canonically, growth curve analyses have been used in studies of child or adolescent development, and have answered questions about how some feature y, such as performance on a test of mental or physical ability, changes as subjects age. Such models allow for the possibility that change in y over time varies not with change in x, but according to the fixed level of x—for example gender or some characteristic of an individual's family or home environment in his/her early years.

Mathematically, just as distinguishing between cross-sectional and longitudinal effects involved no more than a simple decomposition, so too do growth curves require only an interaction between an x_j variable and time.[7] A growth curve can be represented, using equations at two levels, as

$$y_{tj} = \beta_{0j} + \beta_{1j} time_{tj} + e_{tj} , \tag{5}$$

where y_{tj} represents j's score at time t ($t = 1 \ldots T$), $time_{tj}$ is a time-varying (continuous or categorical) variable indexing time for individual j, and β_{0j} and β_{1j} represent a random intercept and time slope, respectively for j. This is the level-1 equation. The intercept and slope parameters are random effects, which can vary across units, as indicated by their j subscripts. Two level-2 equations describing these parameters are then

$$\begin{aligned} &\beta_{0j} = \gamma_{00} + \gamma_{01}x_{1j} + \gamma_{02}x_{2tj} + u_{0j} , \\ &\beta_{1j} = \gamma_{10} + \gamma_{11}x_{3j} + u_{1j} , \text{ and} \\ &e_{tj} \sim \mathrm{N}(0, \sigma_e^2) , u_{0j} \sim \mathrm{N}(0, \sigma_{uo}^2) , \text{ and } u_{1j} \sim \mathrm{N}(0, \sigma_{u1}^2) . \end{aligned} \tag{6}$$

In the level-2 equations, the intercept term β_{0j} combines an overall intercept γ_{00}, time-invariant and time-varying covariates x_{1j} and x_{2tj} (with coefficients γ_{01} and γ_{02} respectively), and a random intercept unique to each j, denoted as u_{0j}. Similarly, the time slope β_{1j} is the sum of an overall slope γ_{10}, a time-constant covariate x_{3j} with coefficient γ_{11}, and an j-specific random disturbance u_{1j}. The covariance of the random intercepts and slopes may be estimated. The special capability of the growth curve model is that it allows for the modeling of time (β_{1j}) as a function of one or

7 In a non-longitudinal context, the inclusion of macro-level x_j variables in cross-sectional multilevel models of comparative survey data is routine. The inclusion of such variables in models fitted to longitudinal data and without an interaction with time presents no particular difficulties or complications, assuming sufficient degrees of freedom at the higher level.

more time-invariant covariate(s) of interest; the rate of change over time depends on time-invariant predictors (Collins 2006). Growth "curves" are often indeed curvilinear (with quadratic functions of time, or orthogonal polynomials) but can also be linear or discrete. The challenge of this technique can be interpretation, since the statistical significance of interaction effects depends on how covariates are entered, and can be subject to weak statistical power (Mathieu et al. 2012).[8]

For analyses of comparative longitudinal survey data, I propose the use of "societal growth curves."[9] A societal growth curve is, as the name implies, a growth curve at the societal level rather than the individual level. Since societies are observed multiple times in repeated cross-sectional survey data, the pattern of change of each one through time can be described using a curved or straight line. A societal growth curve illustrates the pattern of change over time—the rate of change and the height of the original starting point. Such an approach has recently been used to good effect by Poznyak et al. (2011), who investigate the characteristics of regions in Belgium where support for the far right has been rising faster than elsewhere.

How can growth curve analysis be applied to whole societies, where observations on those societies are collected via observations on individuals? A societal growth curve model fitted to repeated cross-sectional individual-level survey data can be specified as:

$$y_{itj} = \beta_{0j} + \beta_{1tj} + \beta_{2j} time_{tj} + e_{itj}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} x_{1j} + \gamma_{02} x_{2itj} + u_{0j}$$
$$\beta_{1tj} = \gamma_{11} x_{3tj} + u_{1tj}$$

[8] Barr, Levy, Scheepers, and Tily (2013) argue for including random slopes in multilevel models generally, where justified by the research design. Estimation should check the sensitivity of results to the inclusion of random slopes by country for time.

[9] I use the term "growth curve" because it is well-established for the analysis of individual-level data, not because social changes must be considered "growth," with its connotations of maturation and/or development.

$$\beta_{2j} = \gamma_{20} + \gamma_{21}x_{4j} + u_{2j} \; . \tag{7}$$

Here γ_{00} refers to the overall intercept, γ_{01} denotes the fixed effect of country-level covariate x_{1j}, γ_{02} denotes the fixed effect of individual-level covariate x_{2itj}, and u_{0j} is a country j-specific random intercept. In the second of the stage-2 equations, γ_{11} refers to the fixed effect of a country-year-level variable x_{3tj} and u_{1tj} is a country-year-level random intercept. In the third of the stage-2 equations, γ_{20} denotes the fixed effect for time (the average time slope across all units), γ_{21} to the effect of some time-invariant country-level covariate x_{4j} on the time slope, and u_{2j} to the random disturbance for each country j. Thus the outcome y for individual i in country j observed at time t is the sum of an intercept, fixed effects at the country, country-year, and individual levels, random intercepts for countries and country-years, an individual-level disturbance, and the effect of time—the latter comprising an overall trend, a country-level fixed effect, and a country-level random disturbance.

Where there are concerns about the possibility that a series is non-stationary, time series in x can be tested for stationarity using techniques for panel unit root tests (e.g., Palm, Smeekes, and Urbain 2011). The possible non-stationarity of y can be checked using the same techniques but applied to the ij random intercepts from a null model. The simulations that follow also address concerns about autocorrelation.[10]

7 Simulation Study

This section reports the results of simulations assessing the performance of the above two techniques, including where data exhibit a variety of complications which could be present in real-world applications. For a series of eight data-generating

[10] Some software can allow for autocorrelation in multilevel models.

processes, I investigate the bias and efficiency of the parameter estimates, and also the optimism of the fixed effects' standard errors—optimism being defined by Shor et al. (2007) as the ratio of the true sampling variability to the estimate of the variability provided by the SEs. I generate the data in R and fit the models with the "lme4" package (Bates, Maechler, and Bolker 2012), using restricted maximum likelihood, or maximum likelihood with a Laplace approximation when estimating binomial models, as described further below.

The base data-generating process (DGP 1) is:

$$y_{itj} = \beta_0 + \beta_1 \bar{x_j} + \beta_2 x_{tjM} + \beta_3 x_{itj} + \beta_4 time_{tj} + \beta_5 time_{tj} * \bar{x_j} + u_j + u_{tj} + e_{itj}, \quad (8)$$
$$\text{with } u_j \sim \mathrm{N}(0, \sigma_{u1}^2),\ u_{tj} \sim \mathrm{N}(0, \sigma_{u2}^2), \text{ and } e_{itj} \sim \mathrm{N}(0, \sigma_e^2)\ .$$

The time-varying covariate x_{tj} has a random time trend, and is partitioned into $\bar{x_j}$ and x_{tjM}. The idea is to mimic a national-level variable, such as GDP/capita or the Gini index for income inequality. The x_{itj} variable is a binary individual-level covariate. The intercept and the coefficients on $\bar{x_j}$, x_{tjM}, and $time_{tj}*\bar{x_j}$ are set to 1, and those on x_{itj} and $time_{tj}$ to -1.[11] The random effects variances are set at 2 for countries, 1 for country-waves, and 9 for the residual term—yielding variance shares that are roughly typical of real-world applications.

Each simulated dataset has 12,500 "respondents" nested in 25 "countries." To investigate any impacts of having just a few waves of observations per country, I "observe" 2, 5, or 20 randomly selected years out of a 30-year range; each country-year includes 250, 100, or 25 people respectively, yielding a total of 500 observations

[11] This combination of values keeps the logit-scale linear predictor reasonably small, thereby avoiding modeled probabilities too close to 0 or 1.

per country in all cases.[12] The R code used to conduct these simulations is available as an online appendix, and can be easily modified to test many different variations: different sizes of *N* at different levels, combinations of complications, different effect sizes, random effects variances, etc. For each DGP and combination of Ns, I simulated and fit a model to 1000 datasets.

Seven subsequent data-generating processes each modify the first (base) process, as follows:

2. The country random intercepts u_j are correlated with $\bar{x_j}$. Such correlation could arise, for example, from endogeneity or omitted variables (see e.g., Ebbes, Böckenholt, and Wedel 2004). The results in this case should be of particular interest, given the prevalence of cross-sectional multilevel analyses where the greatest interest is in the coefficient on some x_j. The covariance is set at 0.7 (for u_j with variance 2, and $\bar{x_j}$ with variance 1, such that the correlation is 0.49).

3. The country random intercepts u_j are correlated with x_{itj}. Since x_{itj} is an individual-level binary variable, the correlation is created between u_j and the country mean of x_{itj} (i.e., the probability of each "respondent" having $x_{itj} = 1$). The covariance is set at 0.7 between u_j and the mean national x_{itj} on the logit scale.

4. The country-wave intercepts u_{tj} are temporally autocorrelated. The autocorrelated u_{tj} are generated before country-waves are selected at random for "observation", such that the autocorrelation across successive *observed* years within a country will vary. The autocorrelation is set at 0.25 between units one lag apart.

5. The country random intercepts u_j are non-Normally distributed, whereas typically multilevel models, as here, are fitted on the assumption that the random intercepts are

[12] These numbers were selected to approximate those in the applications described in the following section, albeit—for the sake of saving computation time—with far fewer respondents per country-wave.

Normally distributed. In this DGP, they are distributed skew-Normally (using the "sn" package by Azzalini 2011).

6. The effects of x_{tjM} are lagged by 5 years, and the fitted model assumes no lag. Such a model misspecification is a real concern, since theory may provide very little guidance as to the timescale on which effects will be felt.

7. The probability of a country-wave's selection (from the range of possible country-waves) is positively related to the value of x_{tj}. Probability weights are proportional to the antilogit of x_{tj}, and a fixed number of country-waves per country are drawn.

8. The dependent variable is binary rather than Normal. The model is therefore fit as a generalized linear mixed model, with an antilogit link. To speed up convergence, "respondents" in the same country-waves with the same covariates (at all levels) are binned, yielding a shortened dataset with many Bernoulli trials per row rather than just one. A model fitted this way yields the same fixed and random effects parameter estimates as one where responses are not binned (see Subramanian et al. 2001).

Figures 1 through 3 present the results of the simulation study. In all three, within each of the eight data-generating processes, the results are presented from left to right for simulations with two, five, and twenty country-waves per country (represented by Os, dots, and Xs, respectively).

Figures 1 and 2 plot the mean and 95% coverage intervals of the estimated fixed effects coefficients and random effects variances, with the true value indicated by a horizontal line. (Fixed effects whose true values are -1 have had their sign reversed for presentation.)

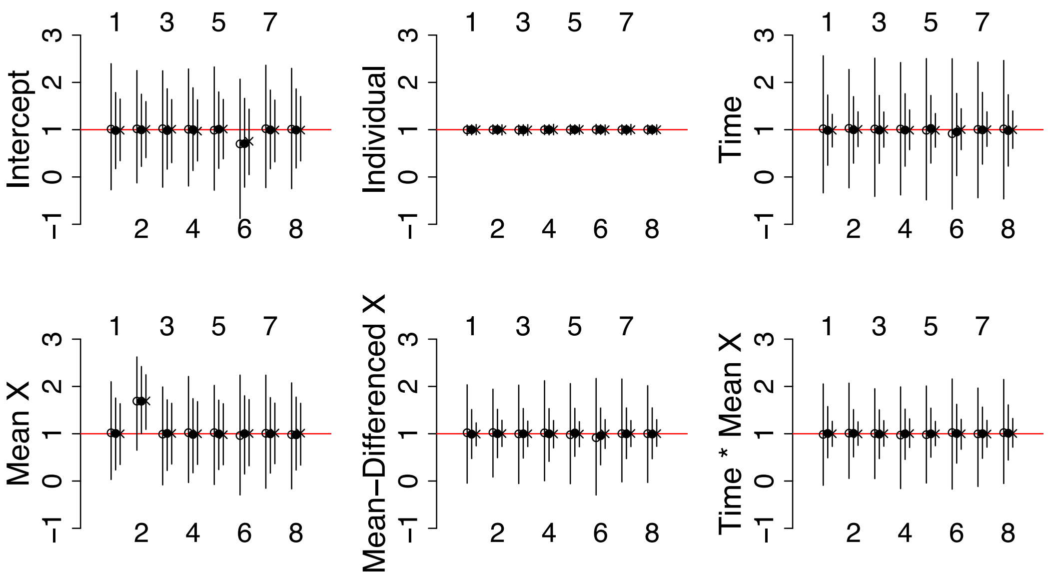

Fig. 1 Means and 95% coverage intervals for the beta coefficients from the simulations described in the text, for each of eight data-generating processes. Os, dots, and Xs are for simulations with, respectively, two and 250, five and 100, and 20 and 25 waves per country and respondents per country-wave. True values are all 1.

Figure 1 shows that the fixed effects coefficients are unbiased, except for the coefficient on $\bar{x}_j$ where $\bar{x}_j$ is correlated with u_j (DGP 2), and the intercept where the effects of x_{tjM} are lagged and the fitted model assumes no lag (DGP 6). In the latter instance, further simulations (not reported) indicate that the magnitude of the bias on the intercept varies positively with the magnitude of the effect of x_{tjM}. If data are generated with both the correlation in DGP 2 and the lag of DGP 6, the biases are combined but do not multiply each other. The variances of the coefficient estimates differ substantially across coefficients, given the different variances of the covariates, but not across DGPs. Observing more waves per country, for a fixed number of observations per country, yields tighter estimates of the coefficients on all variables, except x_{itj}. Observing even just five rather than two country-waves per country has substantial benefits for precision.

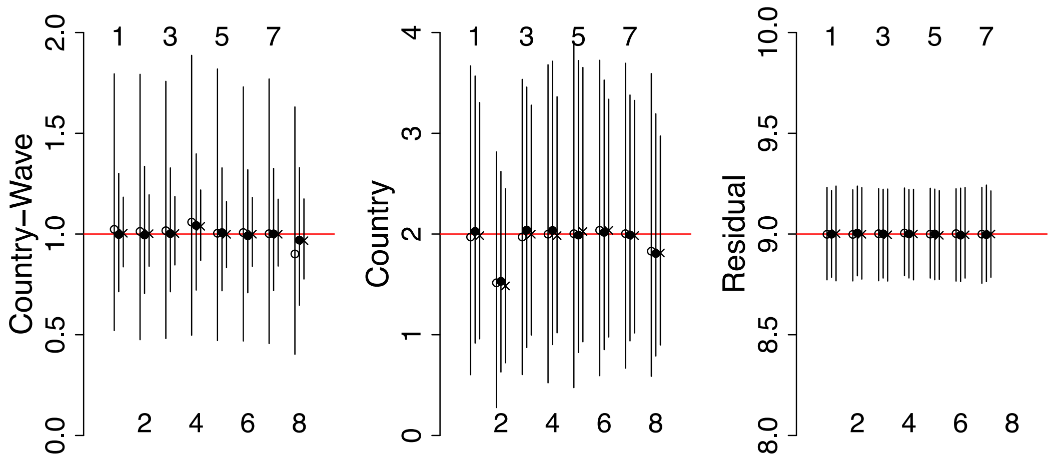

Fig. 2 Means and 95% coverage intervals of the estimated random effects variances from the simulation study. Os, dots, and Xs are for simulations with, respectively, two and 250, five and 100, and 20 and 25 waves per country and respondents per country-wave. True values are shown with a horizontal line.

In Figure 2, the random effects variances are generally unbiased, with the exception that the country-level variance is downward-biased for DGP 2, and to a lesser extent for DGP 8 (a binomial outcome). The dispersal of a fixed number of observations per country across a larger number of country-waves yields tighter estimates for the variances of the country-wave and country random intercepts.

Figure 3 plots the "optimism" of the standard errors returned for the fixed effects coefficients. The estimated SEs for all six fixed effects coefficients in most scenarios reflect the true variability of the estimates—the optimism is close to 1. In some cases, the SEs are slightly too large (over-conservative) or too small (anti-conservative). The scenario where the SEs are most misleading is DGP 4, where country-wave intercepts u_{tj} are temporally autocorrelated, though there is only a problem where the number of country-waves is large. The SEs here are substantially too small for the coefficients on Time, the time-varying x_{tjM}, and the interaction between Time and $\bar{x_j}$. There is thus some risk of committing a Type 1 error where countries are observed at many points in time and there may be autocorrelation.

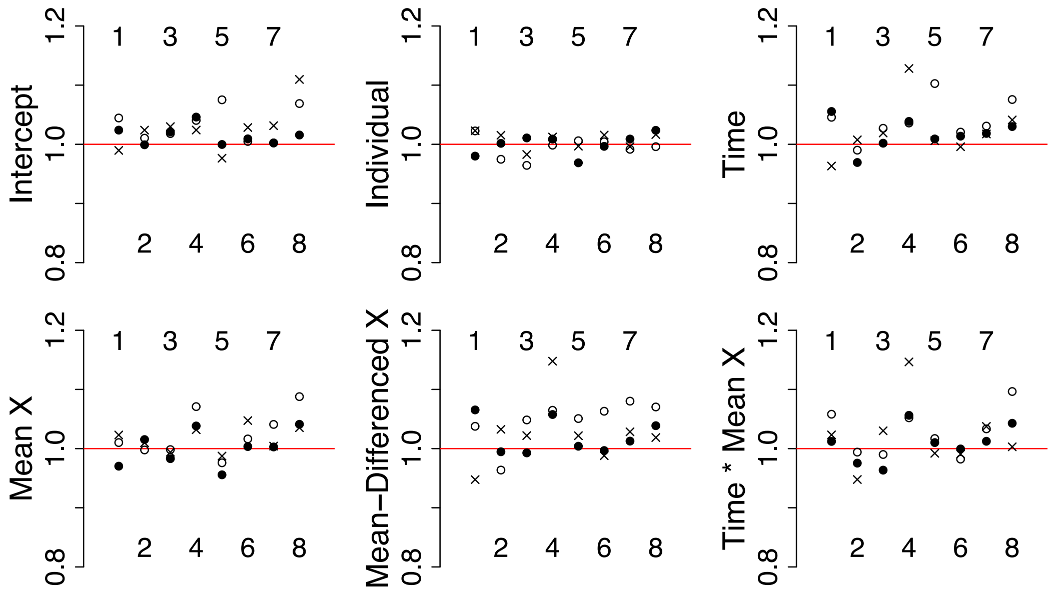

Fig. 3 Optimism of the standard errors returned in the simulation study (calculated as in Shor et al. 2007). Values above one indicate anti-conservative SEs, while values below derive from SEs that are over-conservative. Os, dots, and Xs are for simulations with, respectively, two and 250, five and 100, and 20 and 25 waves per country and respondents per country-wave.

Overall, then, with a little caution in interpreting the significance tests, these techniques appear relatively robust to real-world complications. Concerns about potentially inaccurate SEs could be further addressed using MCMC estimation, or in a maximum likelihood context with a bootstrap or other resampling technique (see e.g., Skrondal and Rabe-Hesketh 2009).

One final concern is statistical power, and the risk of committing a Type 2 error where the number of country-waves observed per country is small.[13] Figure 4 reports the results of additional simulations, similar to DGP 1 described above, but estimated using MCMC and setting the true coefficient on x_{tjM} to 0, 0.25, 0.5, 0.75, or 1.[14] Where it is 1, the effects of $\bar{x_j}$ and x_{tjM} are the same. Figure 4 therefore contrasts

[13] Thanks to an anonymous reviewer for suggesting a discussion of this issue.
[14] Estimation is done using uninformative priors and the "MCMCglmm" package in R (Hadfield 2010). MCMC chains are run for 13,000 iterations, with the first 3,000 iterations discarded for burn-in, and the remainder thinned by a factor of 10.

the results obtained using group mean-centering (and estimating both $\bar{x}_j$ and x_{tjM}) with those obtained by estimating x_{tj} alone.

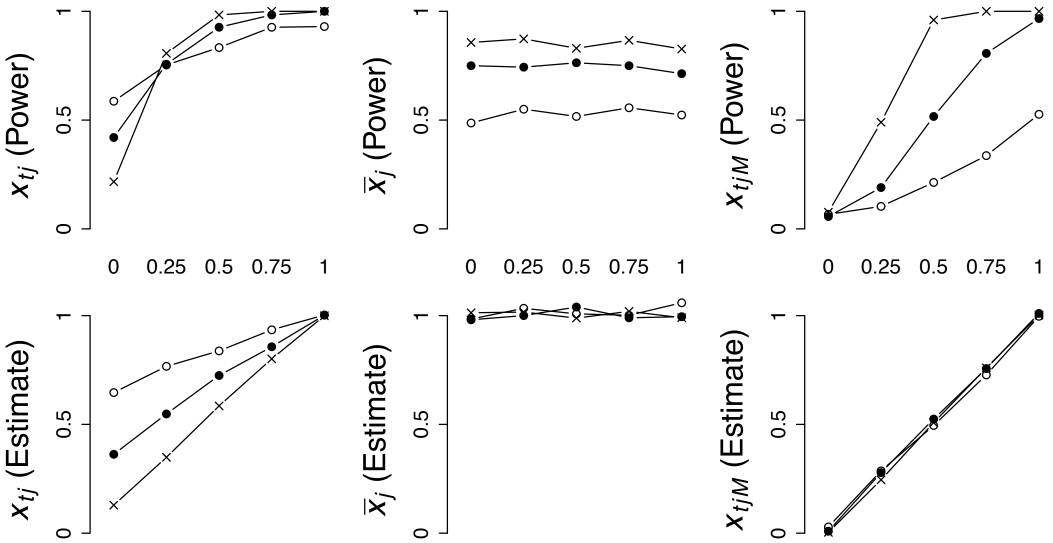

Fig. 4 Proportion of estimated 95% highest posterior density intervals which exclude 0 (top row), and mean estimates of the fixed effects coefficients (bottom row), for simulations with true x_{tjM} ranging from 0 to 1. De facto pooling yields estimates for x_{tj} (left column), and mean-centering for $\bar{x}_j$ and x_{tjM} (center and right columns). Os, dots, and Xs are for simulations with, respectively, two and 250, five and 100, and 20 and 25 waves per country and respondents per country-wave.

Figure 4 shows that the number of country-waves observed per country strongly influences the risk of a Type 2 error, above all for the coefficient on x_{tjM} but also to a lesser extent for the coefficient on $\bar{x}_j$. In some cases, the pooled effect of x_{tj} is more likely to be statistically significant than the partitioned effects of x_{tjM} and $\bar{x}_j$. But the bottom row of Figure 4 also shows how the mean estimate of the coefficient on x_{tj} varies with the number of country-waves per country, and represents a hard-to-interpret combination of x_{tj}'s two component parts (Raudenbush and Bryk 2002: 138). Where the true effect of x_{tjM} is 0, the coefficient on x_{tj} will often be statistically significant, without clarifying that only the cross-sectional component is driving the overall effect. In this sense, there is no simple solution to the problem of weak power where the number of country-waves is small; applied researchers will need to be careful not to over-interpret a statistically non-significant coefficient on x_{tjM}.

8 Applications

This section applies the techniques described above in analyzing World/European Values Survey data on two topics. Across five waves (which spanned from 1981 to 2008, depending on the country), these surveys included 445,319 individuals, in 97 countries, and 245 country-waves (countries have been observed in one to five waves, with developing and non-democratic countries noticeably less well represented).

Some samples were not drawn in ways that would make them nationally representative. Consequently, I follow Stevenson and Wolfers (2008) in excluding 17 country-waves from the first four waves, and using the same criteria I also exclude respondents surveyed in one country (Chile) in the fifth wave. These exclusions leave fewer than 97 countries and 245 country-waves, as specified below.[15]

8.1 Trust

The first topic is generalized social trust—that is, trust in a "generalized other" rather than a known relative, friend, or acquaintance. The literature regards trusting citizens as desirable: they tend to be more socially and politically engaged, generous in their charitable giving, tolerant, and happy with their lives, while countries with

[15] For neither application do I include an individual level covariate, though doing so would be simple; I do not do so here because missing data would reduce the N.

more trusting populations enjoy faster economic growth, better governance, and lower crime (Rothstein and Uslaner 2005: 41; Bjørnskov 2008: 271).

A number of previous studies have linked trust to equality, on the theoretical grounds that people tend to be more trusting of people similar to them, and inequality reduces the share of the population to whom a given person is similar (Uslaner 2002; Bjørnskov 2008). Empirically, in support of this theory, previous work has emphasized negative correlations between inequality and trust across nations (Bjørnskov 2008; Freitag and Bühlmann 2009; Knack and Keefer 1997; Rothstein and Uslaner 2005; Uslaner 2002, 2008). Largely on the strength of such cross-sectional correlations, existing literature presents inequality as "arguably the strongest determinant of the level of social trust" (Bjørnskov 2008: 271; see also Cozzolino 2011; Rothstein and Uslaner 2005; Uslaner and Brown 2005). Uslaner (2002: 186) argues boldly that "the level of economic inequality is the prime mover of generalized trust." On the other hand, Fairbrother and Martin (2013) find only a cross-sectional but not longitudinal relationship between inequality and trust across U.S. states, and no relationship of either kind for U.S. counties.

Here, at the international level, I further test the evidence for inequality as a major influence on people's trust in generalized others. For measuring trust, the relevant WVS/EVS question (A165) is: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" This question has been validated using cross-national differences in the prevalence of trusting behavior in experimental contexts (Johnson and Mislin 2012), and also by experimentally "losing" wallets in public places and counting the proportion returned to their "owners" (Knack and Keefer 1997). In both cases the proportion of self-reported trusters in a country correlates with observation-based measures of trusting behavior.

For data on the time-varying level of inequality in a country, I use Solt's (2009) Standardized World Income Inequality Database, version 3.0, which is based on data from the UN but improves on direct estimates of inequality using a custom missing-data algorithm to make observations more comparable with each other.

Figure 5 plots self-reported trust against income inequality, aggregating individual data for all possible countries and waves (except non-representative samples, as explained earlier). Visually, slopes of the logistic regression lines are suggestive of negative associations between inequality and trust both cross-sectionally and longitudinally, though the cross-sectional relationship (in the left panel) is clearer than the longitudinal relationship (center panel). Moreover, the slope of the line in the center panel could be a product of simultaneous but unrelated declines in both equality and trust over time. The right panel shows that trust has indeed been declining over time, and at roughly the same rate in nations whose mean levels of inequality, pooling across waves, are lower or higher than the median of countries' mean levels of inequality. The techniques proposed here test whether the inequality-trust link holds longitudinally when controlling for time and for the clustering of observations within countries and country-years.

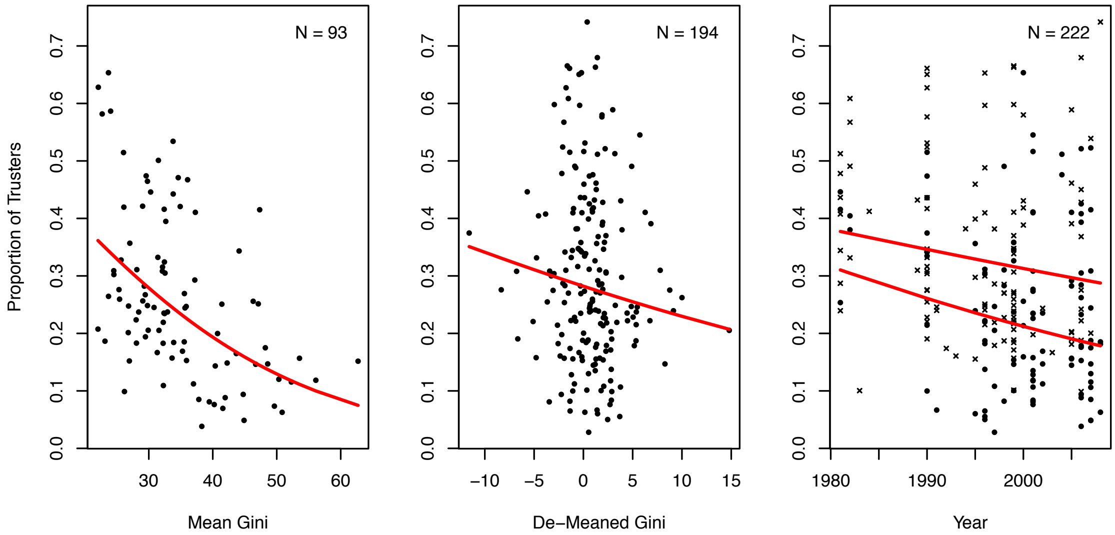

Fig. 5 Mean level of trust versus inequality (first two panels) or time (right panel), with a single-level logistic regression line (ignoring the clustering of country-waves within countries). In the right panel, crosses and dots denote observations from countries with less or more than median country-mean inequality, respectively, and the two lines are fit separately for these two groups of countries.

The model used here investigates the log-odds of a "Most people can be trusted" response as opposed to a "Can't be too careful" response. "Don't know" responses were dropped. A logit model was estimated via maximum likelihood and the Laplace approximation. Table 1 presents the model results. Model 0 is a null model with no covariates other than a constant, while Models 1 through 3 add covariates, and Model 4 includes controls for GDP/capita.

The main findings are that, across all specifications, the coefficient on Gini (mean) is negative and statistically significant at the 0.01 level, while the coefficient on Gini (differenced) is slightly less robust but still also generally negative and statistically significant. In contrast, the coefficient on the growth curve term is not statistically significant in any specification. The variance for the country random intercepts declines substantially with the inclusion of covariates at the country level, while the variance of the country-wave random intercepts shrinks slightly with the addition of the longitudinal (differenced) component of Gini.

Table 1 Logistic Multilevel Models of Trust

Model		0	1	2	3	4
Random Effects Variances	Country-Wave	0.083	0.081	0.079	0.078	0.071
	Country	0.553	0.540	0.405	0.408	0.381
Fixed Effects Coefficients	Time		-0.007* (0.004)	-0.003 (0.004)	0.011 (0.017)	-0.006 (0.008)
	Gini (Mean)			-0.046** (0.009)	-0.038** (0.014)	-0.033** (0.010)
	Gini (Diff.)			-0.018* (0.009)	-0.019* (0.009)	-0.016˚ (0.009)
	Gini (Mean)*Time				-0.000 (0.001)	
	GDP/cap (Mean)					0.019* (0.008)
	GDP/cap (Diff.)					0.005 (0.013)
	Intercept	-1.055** (0.087)	-0.932** (0.105)	0.579˚ (0.327)	0.324 (0.443)	-0.080 (0.444)
logLik		-515.5	-513.4	-501.1	-500.8	-480.3
N		351,971; 194; 79				345,709; 189, 76

** < 0.01, * p < 0.05, ˚ p < 0.1
Standard errors are in parentheses. N's are for respondents, country-waves, and countries.

The coefficient on the longitudinal component of Gini is only statistically significant at the 0.1 level if a random slope for time is included (with or without estimating a covariance between the random slope and the random country intercepts). A likelihood ratio test, however, suggests the random slopes are not warranted. In Model 4, if GDP/capita is entered in logged form, the coefficient on the mean component is no longer statistically significant, though other results are substantively similar, and the model fit is inferior.

The results here are therefore consistent with claims of a link between inequality and trust.[16] In countries where inequality has increased over time in recent decades, people have become systematically less trusting.

8.2 Postmaterialist Values

As the living standards of a country's residents rise to the point where most people's basic material needs are met, do their priorities shift to non-economic, higher-order concerns—about equality, self-determination, freedom, and quality of life? In articulating a theory that they do, Inglehart coined the concept of "postmaterialist" values, arguing that growing material affluence and security lead to more tolerant, egalitarian, participatory, and nurturing societies (see e.g., Inglehart and Abramson 1999; Inglehart and Baker 2000).

To measure people's possession of postmaterialist versus materialist values, the World/European Values Surveys includes an index (Y002) reflecting respondents' answers to a two-part question (E003 and E004). The question states: "If you had to choose, which one of the things on this card would you say is most important? And which would be the next most important?" The options are: "Maintaining order in the nation", "Give people more say", "Fighting rising prices", and "Protecting freedom of speech". Respondents who named "Give people more say" and "Protecting freedom of speech" were coded as "postmaterialists", while those who chose the other two options "materialists". Respondents who selected one of each pair (in either order)

---

[16] The coefficient on the longitudinal relationship in Model 4 is not significant at the conventional 0.05 level, in a two-tailed test. But given the nature of the arguments in the literature, a strong case could be made for a one-tailed test, according to which the coefficient is statistically significant.

were coded as “mixed”. The postmaterialist values index therefore has three (ordinal) response categories.

The hypothesis has come under criticism, such as from Davis and Davenport (1999) who say that postmaterialist values are “virtually unexplainable as a dependent variable.” Clarke and Dutt (1991) and Clarke, Dutt, and Rapkin (1997) contend that the postmaterialist-materialist index is too sensitive to current economic conditions and not to the more slow-changing conditions hypothesized by the concept’s defenders to be more important. There has also been much debate about the appropriate level of analysis in studies of postmaterialist versus materialist values (see e.g., Haller 2002), with critics of the index’s usefulness generally preferring individual-level research and advocates aggregate-level.

Here I link responses to the postmaterialism index to real GDP per capita (at purchasing power parity), as provided by the World Bank’s World Development Indicators. Figure 6 plots postmaterialist values against GDP per capita in the same two ways as did Figure 5 for trust and inequality. Again, the cross-sectional relationship is clear, with a strongly upward-sloping line showing the relationship between average living standards and postmaterialist values. In the center panel, the line is also upward-sloping, albeit less steep, and in the right panel postmaterialist values appear to be growing in prevalence in wealthier but not poorer nations.

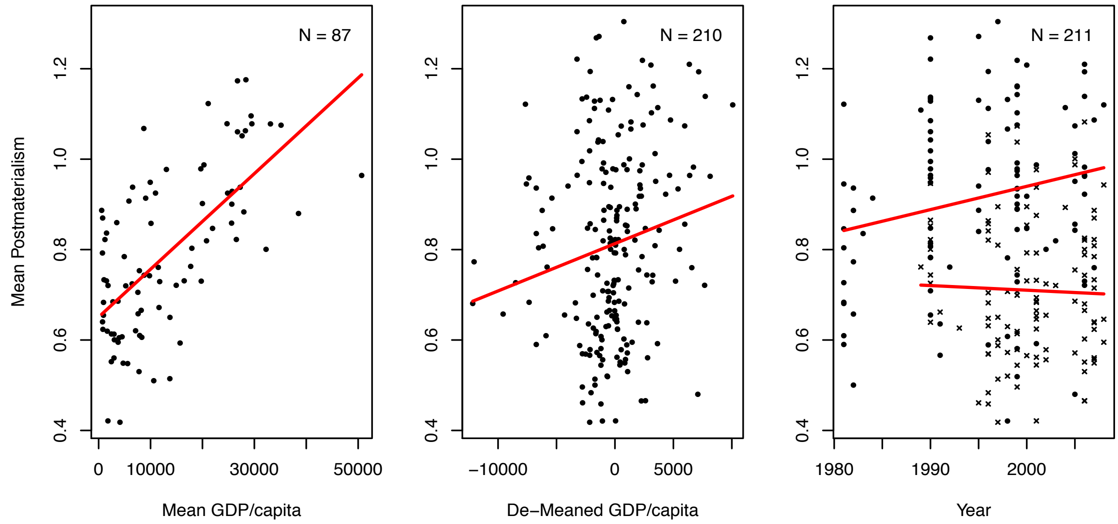

Fig. 6 Mean level of postmaterialist values versus GDP/capita (first two panels) or time (right panel), with a linear regression line (ignoring the clustering of country-waves within countries). In the right panel, crosses and dots denote observations from countries with less or more than median country-mean GDP/capita, respectively, and the two lines are fit separately for these two groups of countries.

A cumulative link/ordered logit model was estimated via maximum likelihood and the Laplace approximation, using the R "ordinal" package (Christensen 2012).[17]

Table 2 presents the results. Model 2 suggests there are statistically significant cross-sectional *and* longitudinal associations between national standards of living (GDP/capita) and the probability of postmaterialist versus materialist values. But robustness checks revealed that the inclusion of a dummy variable for being a (post-) socialist nation makes this relationship much weaker, and no longer statistically significant at even the 0.1 level (Model 3).[18] However, Model 4 clarifies with an interaction term that the longitudinal relationship is quite strong for (post-)socialist nations, while not significant for others. Moreover, as indicated by the other interaction term, the underlying growth curve for the two groups of countries differs

[17] The ordinal package's clmm function does not allow for the inclusion of random slopes, but the main results reported here for postmaterialist values and political discussion are not sensitive to the inclusion of random slopes (using MCMC estimation).
[18] Hungary was the only socialist nation included (in 1982) before the fall of Communism in 1989. Many countries in transition were surveyed in 1990 (and Poland in 1989).

substantially: ceteris paribus, postmaterialist values are declining for (post-)socialist countries, but increasing in others. In other models (not reported) the growth curve coefficient for all countries taken together is not significant.

Table 2 Cumulative (Logit) Link Multilevel Models of Postmaterialist Values

Model		0	1	2	3	4
Random Effects Variances	Country-Wave	0.117	0.114	0.112	0.113	0.104
	Country	0.303	0.190	0.179	0.120	0.124
Fixed Effects Coefficients	Time		0.009* (0.004)	-0.002 (0.005)	0.004 (0.006)	0.021** (0.007)
	GDP/cap (Mean)		0.323** (0.050)	0.312** (0.049)	0.308** (0.043)	0.311** (0.043)
	GDP/cap (Diff.)			0.456** (0.175)	0.173 (0.185)	-0.423 (0.265)
	(Post-)socialist				-0.566** (0.107)	0.124 (0.218)
	Time * (Post-)socialist					-0.039** (0.011)
	GDP/cap (Diff.) * (Post-)socialist					1.042** (0.360)
Threshold Coefficients	Materialist | Mixed	-0.81282 (0.065)	2.248 (0.473)	1.977 (0.467)	1.878 (0.429)	2.171 (0.436)
	Mixed | Postmaterialist	2.159 (0.065)	5.220 (0.473)	4.949 (0.467)	4.850 (0.429)	5.143 (0.436)
logLik		-334,218	-334,201	-334,198	-334,186	-334,180
N		372,312 respondents; 211 country-waves; 87 countries				

** < 0.01, * p < 0.05, ˚ p < 0.1
Standard errors are in parentheses. GDP/capita is entered in logged form, though main results also hold if it is not logged.

These results therefore provide mixed support for the theory of postmaterialist values. Specifically in the context of the rapid social change that was the transition from state socialism to capitalism, differences in processes of economic development appear to have had important consequences for people’s values. But these results also suggest that in many contexts there is little relationship between development and postmaterialist values, even if the cross-sectional relationship is strong and consistent.

9 Conclusions

The applications presented here, using statistical models embodying a multilevel view of repeated observations on nations, have shown how comparative longitudinal survey data can be put to work in testing hypotheses about the consequences of both time-varying and time-invariant macro-social conditions. The evidence for longitudinal relationships between inequality and generalized social trust, and between economic growth and postmaterialist values, represent important extensions and tests for the literatures on these topics.

Across the social sciences, a typical goal of much research is to explain social change. Repeated cross-sectional survey data would appear to be a natural tool for this work, and to some extent such data are already being put to use. However, existing applications do not fully exploit the properties of such datasets, and risk overlooking, or even deriving misleading inferences about, key relationships. Comparative longitudinal survey datasets should be a valuable resource for research on political and social change, in the broadest sense; they should help to answer big questions about why and how societies change, and why some change faster than others. Because of the increasing availability of these kinds of datasets, the usefulness of the techniques described here should only increase over time.

References

Anderson, Christopher J., and Yuliya V. Tverdova. 2003. “Corruption, Political Allegiances, and Attitudes Toward Government in Contemporary Democracies.” *American Journal of Political Science* 47[1]: 91–109.

Azzalini, Adelchi. 2011. R package “sn”. Version 0.4-17.

Bafumi, Joseph, and Andrew Gelman. 2006. “Fitting Multilevel Models When Predictors and Group Effects Correlate.” Paper prepared for the 2006 Annual Meeting of the Midwest Political Science Association, Chicago.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. “Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal.” *Journal of Memory and Language* 68: 255–278.

Bartels, Brandon L. 2008. “Beyond ‘Fixed Versus Random Effects’: A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data.” Unpublished manuscript. Available: http://home.gwu.edu/~bartels/cluster.pdf.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2012. R package “lme4”. Version 0.999999-0.

Beck, Nathaniel. 2007. “From Statistical Nuisances to Serious Modeling: Changing How We Think About the Analysis of Time-Series–Cross-Section Data.” *Political Analysis* 15:97–100.

Bell, Andrew, and Kelvyn Jones. 2012. "Explaining Fixed Effects: Random Effects Modelling of Time-Series Cross-Sectional and Panel Data." Unpublished manuscript. Available: http://polmeth.wustl.edu/media/Paper/FixedversusRandom_1_2.pdf.

Bjørnskov, Christian. 2008. “Social Trust and Fractionalization: A Possible Reinterpretation.” *European Sociological Review* 24(3): 271-283.

Brady, David, Andrew S. Fullerton, and Jennifer Moren Cross. 2009. “Putting Poverty in Political Context: A Multi-Level Analysis of Adult Poverty across 18 Affluent Democracies.” *Social Forces* 88[1]: 271-299.

Christensen, Rune Haubo. 2012. “ordinal” package. Version 5.22.

Clarke, Harold D., and Nitish Dutt. 1991. “Measuring Value Change in Western Industrialized Societies: The Impact of Unemployment.” *American Political Science Review* 85: 905-20.

Clarke, Harold D., Nitish Dutt, and Jonathan Rapkin. 1997. “Conversations in Context: The (Mis)Measurement of Value Change in Advanced Industrial Societies.” *Political Behavior* 19[1]: 19-39.

Collins, Linda M. 2006. “Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model.” *Annual Review of Psychology* 57: 505–28.

Cozzolino, Philip J. 2011. “Trust, Cooperation, and Equality: A Psychological Analysis of the Formation of Social Capital.” *British Journal of Social Psychology* 50: 302-320.

Davis, Darren W., and Christian Davenport. 1999. “Assessing the Validity of the Postmaterialism Index.” *American Political Science Review* 93[3]: 649-664.

De Boef, Suzanna, and Luke Keele. 2008. “Taking Time Seriously.” *American Journal of Political Science* 52[1]: 184–200.

Ebbes, Peter, Ulf Böckenholt, and Michel Wedel. 2004. “Regressor and Random-Effects Dependencies in Multilevel Models.” *Statistica Neerlandica* 58[2]: 161–78.

Enders, Craig K., and Davood Tofighi. 2007. “Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue.” *Psychological Methods* 12[2]: 121-38.

Fairbrother, Malcolm, and Isaac W. Martin. 2013. “Does Inequality Erode Social Trust? Results from Multilevel Models of US States and Counties.” *Social Science Research* 42[2]: 347–360.

Freitag, Markus and Marc Bühlmann. 2009. “Crafting Trust: The Role of Political Institutions in Comparative Perspective.” *Comparative Political Studies* 42(12): 1537-66.

Gelman, 2005. “Two-Stage Regression and Multilevel Modeling: A Commentary.” *Political Analysis* 13: 459-461.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, Andrew, David Park, Boris Shor, Joseph Bafumi, and Jeronimo Cortina. 2008. *Red State, Blue State, Rich State, Poor State*. Princeton: Princeton University Press.

Hadfield, Jarrod D. 2010. “MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package.” *Journal of Statistical Software* 33[2]: 1-22. Version 2.16.

Haller, Max. 2002. “Theory and Method in the Comparative Study of Values: Critique and Alternative to Inglehart.” *European Sociological Review* 18[2]: 139-58.

Hausman, Jerry A. 1978. “Specification Tests in Econometrics.” *Econometrica* 46: 1251–71.

Inglehart, Ronald, and Paul R. Abramson. 1999. "Measuring Postmaterialism." *American Political Science Review* 93[3]: 665-677.

Inglehart, Ronald, and Wayne E. Baker. 2000. "Modernization, Cultural Change, and the Persistence of Traditional Values." *American Sociological Review* 65: 19-51.

Iversen, Torben, and Frances Rosenbluth. 2006. "The Political Economy of Gender: Explaining Cross-National Variation in the Gender Division of Labor and the Gender Voting Gap." *American Journal of Political Science* 50[1]: 1–19.

Jamal, Amaney, and Irfan Nooruddin. 2010. "The Democratic Utility of Trust: A Cross-National Analysis." *The Journal of Politics* 72[1]: 45–59.

Jen, Min Hua, Kelvyn Jones, and Ron Johnston. 2009. "Global variations in health: Evaluating Wilkinson's income inequality hypothesis using the World Values Survey." *Social Science & Medicine* 68: 643–653.

Johnson, Noel D., and Alexandra Mislin. 2012. "How Much Should We Trust the World Values Survey Trust Question?" *Economics Letters* 116: 210-212.

Karakoç, Ekrem, and Birol Başkan. 2012. "Religion in Politics : How Does Inequality Affect Public Secularization?" *Comparative Political Studies* 45[12]: 1510-1541.

Kennedy, Bruce P., Ichiro Kawachi, Roberta Glass, and Deborah Prothrow-Stith. 1998. "Income Distribution, Socioeconomic Status, and Self-Rated Health in the United States: Multilevel Analysis." *British Medical Journal* 317: 917-921.

Kittel, Bernhard, and Hannes Winner. 2005. "How Reliable Is Pooled Analysis in Political Economy? The Globalization-Welfare State Nexus Revisited." *European Journal of Political Research* 44: 269–293.

Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112(4): 1251-1288.

Martin, Nathan D., and David Brady. 2007. "Workers of the Less Developed World Unite? A Multilevel Analysis of Unionization in Less Developed Countries." *American Sociological Review* 72: 562–584.

Mathieu, John E., Herman Aguinis, Steven A. Culpepper, and Gilad Chen. 2012. "Understanding and Estimating the Power to Detect Cross-Level Interaction Effects in Multilevel Modeling." *Journal of Applied Psychology*. 97(5): 951-966.

Mattes, Robert, and Michael Bratton. 2007. "Learning about Democracy in Africa: Awareness, Performance, and Experience." *American Journal of Political Science* 51[1]: 192–217.

Moller, Stephanie, Arthur S. Alderson, and François Nielsen. 2009. "Changing Patterns of Income Inequality in U.S. Counties, 1970–2000." *American Journal of Sociology* 114: 1037–1101.

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

Palm, Franz C., Stephan Smeekes, and Jean-Pierre Urbain. 2011. "Cross-sectional Dependence Robust Block Bootstrap Panel Unit Root Tests." *Journal of Econometrics* 163: 85-104.

Poznyak, D., K. Abts, and M. Swyngedouw. 2011. "The dynamics of the extreme right support: A growth curve model of the populist vote in Flanders-Belgium in 1987–2007." *Electoral Studies* 30: 672–688.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd ed. London: Sage.

Raudenbush, Steve. 1989. "'Centering' Predictors in Multilevel Analysis: Choices and Consequences." *Multilevel Modelling Newsletter* 1: 10-12.

Rothstein, Bo, and Eric M. Uslaner. 2005. "All For All: Equality, Corruption, and Social Trust." *World Politics* 58: 41–72.

Ruiter, Stijn, and Frank van Tubergen. 2009. "Religious Attendance in Cross-National Perspective: A Multilevel Analysis of 60 Countries." *American Journal of Sociology* 115[3]: 863–95.

Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis* 15: 165–181.

Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Skrondal, Anders , and Sophia Rabe-Hesketh. 2008. "Multilevel and Related Models for Longitudinal Data." Pp. 275-299 in J. de Leeuw, E. Meijer (eds.), *Handbook of Multilevel Analysis*. Springer.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2009. "Prediction in Multilevel Generalized Linear Models." *Journal of the Royal Statistical Society Series A* 172: 659–87.

Solt, Frederick. 2008. "Economic Inequality and Democratic Political Engagement." *American Journal of Political Science* 52[1]: 48–60.

Solt, Frederick. 2009. "Standardizing the World Income Inequality Database." *Social Science Quarterly* 90[2]: 231-242.

Steenbergen, Marco R., and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46[1]: 218-237.

Stegmueller, Daniel. 2013. "How Many Countries For Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches." *American Journal of Political Science*, forthcoming.

Stevenson, Betsey, and Justin Wolfers. 2008. "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*: 1-87.

Subramanian, S V, Kelvyn Jones, Afamia Kaddour, and Nancy Krieger. 2009. "Revisiting Robinson: The Perils of Individualistic and Ecologic Fallacy." *International Journal of Epidemiology* 38: 342–360.

Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. New York: Cambridge University Press.

Uslaner, Eric M. 2008. "Where You Stand Depends on Where Your Grandparents Sat: The Inheritability of Generalized Trust." *Public Opinion Quarterly* 72(4): 725-40.

Uslaner, Eric M., and Mitchell Brown. 2005. "Inequality, Trust, and Civic Engagement." *American Politics Research* 33(6): 868-894.

Weldon, Steven A. 2006. "The Institutional Context of Tolerance for Ethnic Minorities: A Comparative, Multilevel Analysis of Western Europe." *American Journal of Political Science* 50[2]: 331–349.

Wilkes, Rima, Neil Guppy and Lily Farris. 2007. "Comment on Semyonov, Raijman, and Gorodzeisky, ASR, June 2006: Right-Wing Parties and Anti-Foreigner Sentiment in Europe." *American Sociological Review* 72[5]: 831-840.

Wilkinson, Richard G., and Kate E. Pickett. 2006. Income Inequality and Population Health: A Review and Explanation of the Evidence. *Social Science & Medicine* 62: 1768–1784.

Wilkinson, Richard, and Kate Pickett. 2009. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Penguin.

Willett, John B., Judith D. Singer, and Nina C. Martin. 1998. "The Design and Analysis of Longitudinal Studies of Development and Psychopathology in Context: Statistical Models and Methodological Recommendations." *Development and Psychopathology* 10: 395–426.

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.

Wu, Y.-W. B., and P. J. Wooldridge. 2005. "The Impact of Centering First-Level Predictors on Individual and Contextual Effects in Multilevel Data Analysis." *Nursing Research* 54: 212–216.

Online Appendix A

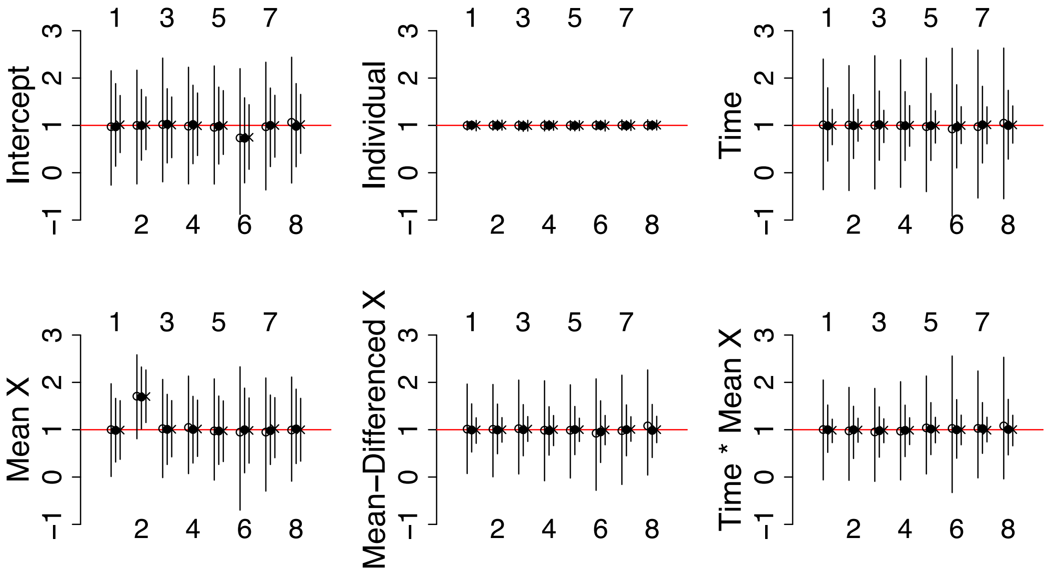

Fig. 7 Means and 95% coverage intervals for the beta coefficients from the eight types of simulations described in the text, but estimated using MCMC. Os, dots, and Xs are for simulations with, respectively, two and 250, five and 100, and 20 and 25 waves per country and respondents per country-wave. True values are all 1.

Online Appendix B

```
# R code to run Monte Carlo simulations for "Two
     Multilevel Modeling Techniques for Analyzing
     Comparative Longitudinal Survey Datasets"

# 30 March 2013

library(lme4.0)
library(multicore) # for mclapply
library(sn) # for rmsn

N <- 25 # set number of countries
n <- 500 # set number of people observed per country
T <- 30 # set number of years in study period
cwaves <- expand.grid(year=1:T, country=1:N) # create
     year and country indices
cwaves$cwave <- apply(cwaves, 1, function(x)
     100*x[2]+x[1]) # create cwave index

c.mer <- function(mod) { c(fixef(mod),
     vcov(mod)@factors$correlation@sd,
     c(unlist(lapply(VarCorr(mod), diag)),
     attr(VarCorr(mod), "sc")^2)) } # function to extract
     FEs, SEs, REs

# main model: yitj = B0 - B1*Xitj + B2*Xj - B3*timetj +
     B4*timetj*Xj + B5*(Xtj-Xj) + Uj + Utj + eitj
# alternative model for comparing de facto pooling with
     centering: logit(yitj) = B0 - B1*Xitj + B2*Xj -
     B3*timetj + B4*(Xtj-Xj) + Uj + Utj

dgp <- function(cwaves, N, T, n, s, ac, a, la, sam,
     waves, cof1) { # function to generate the data
     if(all(c(a!=0, sum(s-diag(diag(s)))!=0)))
     stop("Having both skew and non-zero covariance
     changes variance.")
     s <- s[rowSums(s)!=0,colSums(s)!=0] # to avoid an
     error when calling rmsn
     sh <- c(a, rep(0, ncol(s)-1)) # set shape parameters
     w <- s/(1-((2*(sh^2)/(1 + sh^2))/pi)) # determine
     scale parameter to get desired variances
     xi <- c(-sqrt(w)%*%sh/sqrt(1 + sh^2)*sqrt(2/pi)) #
     determine location parameter to get mean of 0
     dat <- rmsn(n=N, xi=xi, Omega=w, alpha=sh) #
     depending on s and a, generates correlated data, or
     Skew Normal country intercepts
     if (ncol(dat)<3) dat <- cbind(dat, 0)
     dat <- data.frame(cwaves, dat[cwaves$country,]) #
     expands to N*T rows
```

```
names(dat)[4:6] <- c("Uj", "Xj", "Xitj") # country-
level random intercept, country mean X, and
probability of Xitj being equal to one
dat$Utj <- suppressWarnings(unlist(by(dat,
dat$country, function(x)
as.numeric(arima.sim(list(ar=ac), n=T,
n.start=T))))) # set autocorrelation (if any)
dat$Xtj <- do.call("c", by(dat, dat$country,
function(x) (0:(T-1))*rnorm(1, 0.05, 0.05))) # time-
varying covariate with a random linear trend
dat$Xtj <- dat$Xj + dat$Xtj - as.numeric(by(dat,
dat$country, function(x) mean(x$Xtj)))[dat$country]
# country-wave-level covariate
dat$XtjM <- dat$Xtj-dat$Xj # mean-centred version of
Xtj
dat$XtjMlag <- unlist(by(dat, dat$country,
function(x) c(rep(NA,la),x$XtjM[1:(T-la)]))) #
creates a version of XtjM with lag la
if (sam) dat <- do.call("rbind", by(dat,
dat$country, function(x)
x[rep(sample(x$year[(la+1):T], waves,
prob=plogis(dat$Xtj)[(la+1):T]),n/waves),])) else
     dat <- do.call("rbind", by(dat, dat$country,
function(x) x[rep(sample(x$year[(la+1):T],
waves),n/waves),])) # waves*N*n rows
dat$Xitj <- rbinom(nrow(dat), 1, plogis(dat$Xitj)) #
individual-level (binary) covariate
dat$year <- dat$year/20 # simply to keep the
variation proportional
dat$y <- 1 - dat$Xitj + dat$Xj - dat$year +
dat$year*dat$Xj + dat$XtjMlag + dat$Uj + dat$Utj +
rnorm(nrow(dat), sd=3) # calculate Normally
distributed outcome
dat$ybi <- rbinom(nrow(dat), 1, plogis(1 - dat$Xitj
+ dat$Xj - dat$year + dat$year*dat$Xj + dat$XtjMlag
+ dat$Uj + dat$Utj)) # calculate (binary) outcome
if (cof1!=-1) dat$ybi <- rbinom(nrow(dat), 1,
plogis(1 - dat$Xitj + dat$Xj - dat$year +
cof1*dat$XtjMlag + dat$Uj + dat$Utj)) # alternative
model
dat[order(dat$cwave, dat$Xitj),]
}

bin <- function(dat) { # function to aggregate binary
     responses by combination of fixed and random effects
     dat2 <- by(dat, list(dat$Xitj,dat$cwave),
     function(x) c(sum(x$ybi==1),sum(x$ybi==0)))
     dat2 <- data.frame(do.call("rbind", dat2))
     names(dat2) <- c("successes", "failures")
     data.frame(dat2,
     dat[!duplicated(dat$cwave*100+dat$Xitj),])
     }
```

```
sims <- 1000 # set number of simulations per combination
     of conditions, for lme4

sim <- function(cwaves, N, T, n, s, ac, a, la, sam,
     logit, waves, cof1=-1) { # function to generate
     data, estimate a model with (g)lmer, and return the
     results
     dat <- dgp(cwaves=cwaves, N=N, T=T, n=n, s=s, ac=ac,
     a=a, la=la, sam=sam, waves=waves, cof1=cof1) #
     simulate data using dgp
     if (logit) { # fit logit model
          mod1 <- glmer(cbind(successes, failures) ~ Xitj
     + year*Xj + XtjM + (1 | cwave) + (1 | country),
     bin(dat), family=binomial) # with centering
          mod2 <- glmer(cbind(successes, failures) ~ Xitj
     + year + Xtj + (1 | cwave) + (1 | country),
     bin(dat), family=binomial) # no centering
          }
     if (!logit) { # fit linear model
          mod1 <- lmer(y ~ Xitj + year*Xj + XtjM + (1 |
     cwave) + (1 | country), dat) # with centering
          mod2 <- lmer(y ~ Xitj + year + Xtj + (1 |
     cwave) + (1 | country), dat) # no centering
          }
     c(c.mer(mod1), c.mer(mod2)) # return 6 FEs, 6 SEs,
     and 3 REs, for each model
     }

# DGP conditions:
s <- list(diag(c(2,1,0)),
     matrix(c(2,0.7,0,0.7,1,0,0,0,0), ncol=3),
     matrix(c(2,0,0.7,0,1,0,0.7,0,0.5), ncol=3))
     # correlation
ac <- c(0, 0.25) # autocorrelation in the country-wave
     random intercepts
a <- c(0, 10) # skewness of the random intercepts
la <- c(0, 5) # lag
sam <- c(FALSE, TRUE) # weighted selection
logit <- c(FALSE, TRUE) # binary outcome
waves <- c(2, 5, 20)
com <- expand.grid(s=s, ac=ac, a=a, la=la, sam=sam,
     logit=logit, waves=waves)
com <- com[com$a==0 | sapply(com$s, function(ss)
     sum(ss[upper.tri(ss)]))==0,] # to avoid having both
     skew and non-zero covariance
rownames(com) <- 1:nrow(com)

res <- apply(com, 1, function(x) mclapply(1:sims,
     function(y) sim(cwaves=cwaves, N=N, T=T, n=n, s=x$s,
     ac=x$ac, a=x$a, la=x$la, sam=x$sam, logit=x$logit,
     waves=x$waves)))
```

```
res <- lapply(res, function(x) do.call(rbind, x))

save(res, file="res.RData")


######## MCMC

library(MCMCglmm)

c.MC <- function(mod) {
     c(matrix(summary(mod)$solutions[,1:3]),
     apply(mod$Sol, 2, sd),
     summary(mod)$Gcovariances[,1],
     summary(mod)$Rcovariances[,1]) }
# function to extract FEs (post.mean, l-95% CI, u-95%
     CI), SEs, REs

priors1 <- list(R = list(V = 1, nu = 0.002), G = list(G1
     = list(V =1, nu = 0.002), G2 = list(V =1, nu =
     0.002)))

simMC <- function(cwaves, N, T, n, s, ac, a, la, sam,
     waves, logit, cof1=-1) { # function to generate
     data, estimate a model with MCMCglmm, and return the
     results
     dat <- dgp(cwaves=cwaves, N=N, T=T, n=n, s=s, ac=ac,
     a=a, la=la, sam=sam, waves=waves, cof1=cof1) #
     simulate data using dgp
     if (logit) { # fit logit model
          mod1 <- MCMCglmm(cbind(successes,failures) ~
     Xitj + year*Xj + XtjM, random = ~ cwave + country,
     data = bin(dat), family="multinomial2", verbose=F,
     prior=priors1)
          mod2 <- MCMCglmm(cbind(successes,failures) ~
     Xitj + year + Xtj, random = ~ cwave + country, data
     = bin(dat), family="multinomial2", verbose=F,
     prior=priors1)
          }
     if (!logit) { # fit linear model
          mod1 <- MCMCglmm(y ~ Xitj + year*Xj + XtjM,
     random = ~ cwave + country, data = dat, verbose=F,
     prior=priors1)
          mod2 <- MCMCglmm(y ~ Xitj + year + Xtj, random
     = ~ cwave + country, data = dat, verbose=F,
     prior=priors1)
          }
     c(c.MC(mod1), c.MC(mod2)) # return 6 FEs, 6 SEs, and
     3 REs, for each model
     }

simsMC <- 300 # set number of simulations per combination
     of conditions, for MCMCglmm
```

```
# DGP conditions for MCMC
# replicating eight main DGPs above, plus 5 additional
     (each for 2, 5, and 20 country-waves per country)

com <- com[c(c(1:4, 7, 9, 17, 33), c(1:4, 7, 9, 17,
     33)+64, c(1:4, 7, 9, 17, 33)+128),]
com$cof1 <- -1
cof1 <- seq(0, 1, by=0.25)
comMC <- expand.grid(s=s, ac=ac[1], a=a[1], la=la[1],
     sam=sam[1], logit=logit[2], waves=waves,
     cof1=cof1)[c(3*1:15-2),]
comMC <- rbind(com, comMC)
rownames(comMC) <- 1:nrow(comMC)

# run the simulations (parallelizing across as many cores
     as are available)

resMC <- apply(comMC, 1, function(x) mclapply(1:simsMC,
     function(y) simMC(cwaves=cwaves, N=N, T=T, n=n,
     s=x$s, ac=x$ac, a=x$a, la=x$la, sam=x$sam,
     waves=x$waves, logit=x$logit, cof1=x$cof1)))
resMC <- lapply(resMC, function(x) do.call(rbind, x))
     # convert from a list of lists, to a list of
     matrices

save(resMC, file="resMC.RData")

q("no")
```