International Journal of Conflict and Violence

Vol. 8 (2) 2014

urn:nbn:de:0070-ijcv-20142109

ISSN: 1864-1385

International Journal of Conflict and Violence – IJCV

The International Journal of Conflict and Violence (IJCV) is a peer-reviewed periodical for scientific exchange and public dissemination of the latest academic research on conflict and violence. It was included in the Social Sciences Citation Index (SSCI) in March 2011. The subjects on which the IJCV concentrates have always been the subject of interest in many different areas of academic life. Consequently, the journal encompasses contributions from a wide range of disciplines including sociology, political science, education, social psychology, criminology, ethnology, history, political philosophy, urban studies, economics, and the study of religions. The IJCV is open-access: All text of the IJCV is subject to the terms of the Creative Commons Attribution-NoDerivatives License. The IJCV is published twice a year, in spring and in fall. Each issue will focus on one specific topic while also including articles on other issues.

Managing Editor

Julia Marth
University of Bielefeld
Institute for Interdisciplinary Research on Conflict and Violence
P.O. Box 100131
33501 Bielefeld
Germany
editorial.office@ijcv.org
www.ijcv.org

Copy-Editing

Meredith Dale, Berlin
medale@t-online.de

Typesetting

Florian Rudt, Bielefeld
frudt@rudtpublishing.de

Design by

meier stracke gbr, büro für gestaltung
Ernst-Rein-Str. 40
33613 Bielefeld
Germany
post@meier-stracke.de
www.meier-stracke.de

Editorial

Letter from the Editors

Dear Reader,

We are pleased to present issue 8 (2). The focus section is edited by Helmut Thome (University of Halle-Wittenberg, Germany) and Steven F. Messner (University at Albany, United States), and focuses on methodological constraints, options, and solutions in longitudinal research on criminal violence. We are honored that Helmut Thome, Steven Messner, and the contributors offer their scientific knowledge and evidence in the interests of gaining a deeper understanding of one of the most promising methodological approaches in conflict and violence research, i.e. longitudinal data analyses.

Sincere thanks to the editors and authors for opening up the potential of this specific kind of research for advances to better assess and understand criminal violence in a long term perspective. We hope this issue will be discussed and quoted in the interdisciplinary community of researchers.

The open section again fulfills its aim to foster contributions on diverse topics in the field of conflict and violence, starting with an analysis of Southeast Asian media reception of the Israeli/Palestine conflict. The second paper shows how social cohesion activities have the potential to change disparaging attitudes in Cyprus. The third and final contribution examines teen dating violence in Switzerland from various angles.

Enjoy reading and critical thinking

June 2015

Andreas Zick　　Steven F. Messner　　Gary LaFree　　Ekaterina Stepanova

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014284
IJCV: Vol. 8 (2) 2014

Guest Editorial: Methodological Issues in Longitudinal Analyses of Criminal Violence

Helmut Thome, Institute of Sociology, University of Halle-Wittenberg, Germany
Steven F. Messner, Department of Sociology, University at Albany, State University of New York, United States

Vol. 8 (2) 2014

ISSN: 1864–1385

Guest Editorial: Methodological Issues in Longitudinal Analyses of Criminal Violence

Helmut Thome, Institute of Sociology, University of Halle-Wittenberg, Germany
Steven F. Messner, Department of Sociology, University at Albany, State University of New York, United States

This guest editorial introduces the Focus Section on Methodological Issues in Longitudinal Analyses of Criminal Violence. Longitudinal designs offer distinctive advantages for purposes of making causal inferences with observational data, but significant challenges must be confronted as well. This editorial highlights some of the more important methodological issues that arise, describes in general terms selected approaches for dealing with them, and indicates how the papers included in this focus section skilfully apply methodological techniques for longitudinal analyses to address substantively important issues pertaining to criminal violence.

1. Background

Much of the quantitative research conducted by criminologists is still based on data collected from regional units (such as districts or nation-states) at a single point in time. Assumed structural relationships between dependent and independent (predictor) variables are regularly specified in single-equation regression models, the parameters of which are estimated with Ordinary-Least-Squares (OLS) techniques. The impact of each of the predictor variables is indicated by its slope coefficient and its standard error. The quality of these estimates depends (particularly but not exclusively) on certain characteristics of the "errors", of the differences between the observed and the "expected" values predicted for each case on the basis of the values they have on the independent variables (impact factors). The most emphasized assumptions are that the errors should be normally distributed (or nearly so) around an expected value of zero with constant (homoscedastic) variance, and that they should not be correlated with each other or with any of the predictor variables. The last assumption is violated if impact factors not included in the model not only correlate with the dependent variable but also with one or more of the predictor variables actually included in the model (the problem of "omitted-variable bias").

Another assumption underlying causal inferences drawn from such models is often overlooked: At the time of measurement, the data should be in a state of equilibrium, in other words any more or less recent change in the predictor variable *X* should have completely unfolded its impact on the dependent variable *Y* at the time of measurement. If this is not the case and the speed of the dis- and re-equilibration process covaries with the level of *X* (for example, cases with higher *X*-values may adjust more quickly than cases with lower *X*-values), the estimated slope coefficient will be distorted (if there is no covariation with *X* only the intercept will be distorted).

Time-series data (in principle) allow the researcher to uncover the dynamics of the causal processes and discriminate between short-term change effects and long-term level effects. However, one might also be interested in the impact of variables which are quite stable over time (within the period of observation) but may vary considerably between individual or regional units of analysis. Furthermore, "process" effects may differ from "structural" effects attributed to the (seemingly) same variable. For example, in Germany during the last decades of the nineteenth century rapid social change (in terms of increasing urbanization and industrialization) was accompanied by, and apparently spawned, a strong increase in the rate of aggravated assault and battery across the roughly one thousand rural and urban districts. This can be explained in terms of Durkheimian assumptions concerning the anomic consequences inherent in rapid social change. This change, however, led to new social-structural arrangements and cultural features (the erosion of "collectivism"), which induced a *decrease* in violent crime. Cross-sectional comparisons between rural and urban districts in the German Empire reveal an interesting difference: already during this period of rapid change the average assault rate in large cities (though also rising) remained considerably below the rate in rural areas. In other words, the effects of "urbanization" differed from those of "urbanity" (cf. Thome 2010).[1]

In a similar vein Phillips (2006) differentiates (transitory) "flow" and (lasting) "stock" effects; the first are triggered by fluctuating changes noticeable over time, the second are more readily observed via time-stable variations across the units of analysis.[2] So, for example, a rise in unemployment may have an immediately negative effect on certain crime rates (like that of burglary), because more people stay at home and thereby decrease the opportunity for that type of crime. On the other hand, if people stay unemployed for longer periods of time they might become more motivated to commit criminal acts themselves.

Pooling cross-sectional and time-series data generally expands the possibilities for examining broader ranges of theoretically interesting impact factors and causal dynamics. This benefit, however, comes at the cost of increased data heterogeneity, making it more difficult to determine unbiased and efficient parameter estimates. It may become a rather challenging task to develop a model design which adequately balances the claims of substantive theory and the requirements of sound statistical analysis. While numerous models are to be found in the literature, none of them counts as "the best" under all circumstances, and quite often not even within the specific circumstances encountered in a well-defined research project. In a guest editorial we cannot present a detailed overview on such models; instead we will outline only some of the main alternatives that are considered or applied in the articles included in our focus section. Here and in the extant literature, much discussion is devoted to the respective merits and deficiencies of "fixed-effects" versus "random-effects" modelling strategies as they depend on characteristics of the data and on the kind of substantive hypotheses to be addressed. In the following paragraphs we describe some of the major alternatives and characteristic features that are involved in these modelling strategies.

Whereas in purely cross-sectional analyses we have a data set in which all the *N* cases (individuals, organizations, regional units etc.) are ordered row by row with their variable values given column by column, a Time-Series Cross-Sectional (TSCS)[3] data set consists of *NxT* cases, where each unit i ($i = 1, \dots, N$) displays its T ($t = 1, \dots, T$) time-specific values of all the variables measured successively row by row. The error assumptions underlying OLS regression analysis with purely cross-sectionally distributed data (T=1) are regularly violated by the pooled-data set to an extent that exceeds the limits set by the "robustness" assumption often applied in justifying OLS estimation techniques even in the case of "minor" departures from the regular error assumptions.

Nevertheless, for heuristic purposes one may start with a "completely pooled" model (as in the paper by Raffalovich and Chung included in this focus section) in which all the *NxT* cases are combined into one homogenous data matrix without making any structural distinctions with respect to

1 This observation was confirmed by extended regression analyses including additional indicators of the relative weight of collectivism vs. individualism. Instead of being overcome after a while, "anomie" might become "chronic" (Durkheim) or "institutionalized" (Messner and Rosenfeld 2013), in the sense of turning into a structural (besides a temporal) property of a social system. In the second half of the twentieth century the structural properties of individualism may also have been evolving towards strengthening its "disintegrative" (and therefore criminogenic) components over its "cooperative" components (Messner et al. 2008; Thome and Stahlschmidt 2013).

2 One may, however, encounter time-specific changes in the level of certain impact factors which affect all cross-sectional units in the same way, such as changes in prevention and incapacitation policies introduced by legislation in a centralized state and invariantly implemented across its regional units.

3 The TSCS label is often used to refer to pooled data for which $T > N$ or N not much larger than T. The label "Cross-Sectional Time-Series" analysis accordingly refers to a data set with N considerably larger than T (also referred to as "Panel Analysis"). But this terminology is not uniformly applied in this way.

cross-section or time dependencies. Thus, the measurements assigned to the *i*-th unit at time *t* count as measurements of one specific case drawn independently from the measurements of any *other* case constituted by the same unit at time $t \pm j$ or, equally, by another unit $i \pm n$ and time point *t,* and so forth. Under this assumption of time and cross-sectional independence one may also assume that the errors are not auto-correlated over time or space and that they have equal variance over all cases (homoscedasticity). On this basis an OLS regression model could be estimated in the same way as an OLS regression with purely cross-sectionally varied data. These assumptions, however, are empirically unrealistic. Even though there might be no "spatial" correlation across units (at the same or over different time points), the over-time measurements of any given unit will regularly be auto-correlated. It is also more realistic to assume that the error-variances and co-variances are not the same across all units (heteroscedasticity).

More realistic assumptions are introduced by the "Kmenta" pooling model (Kmenta 1986), which allows for unit-specific error-variances, errors correlated over time (auto-correlation), and "contemporaneous correlation" between errors of different units at the same time. The coefficients of such a model are to be determined by Estimated (Feasible) Generalized Least Squares (EGLS, FGLS) procedures. A major restriction of this model is the assumption that the vector of parameters to be estimated should be constant for all units at all points of time, including the intercept. The last component in this restriction (referring to the intercept), in particular, is often quite unrealistic. In many (probably most) cases, criminologists have to deal with (regional) units which across all time-specific measurements exhibit sizable and persistent level differences in the dependent variable (like assault or homicide rates), which cannot be explained by the predictor variables, because they are produced by "omitted" (unknown or unavailable) impact factors not included in the model. These level differences might be (and often are) correlated with the included predictor variables. In such cases, the base level (common intercept) and, more importantly, the slope coefficients estimated by EGLS according to the Kmenta or similar models would be largely distorted (Hsiao 1986; Stimson 1985, 919–21).

One approach to deal with this problem is the so-called "fixed-effects" modelling design. Here the time-invariant level differences not accounted for by the predictor variables are represented in the regression model by unit-specific intercepts.[4] They can be calculated as the slope coefficients of *N* dummy variables[5] D_{jt} additionally introduced into the regression equation (Least Squares Dummy Variable [LSDV] models). Each is coded with the value of "1" for each point of time for a specific unit $j = i$ and the value of "0" for all the other units $j \neq i$, where *i* runs from 1 (the first unit) to *N* (the last unit).[6] If *N* is large, it is recommended not to use dummy variables but to transform the dependent and all the independent variables by subtracting the observed values from their respective means calculated separately for each unit over all the time-specific measures available (for a detailed description see Alecke 1995, 11–15; for an application see the contribution by Entorf and Sieger in this focus section). With this transformation it becomes even more obvious that in *fixed-effects* modelling the estimation of the slope coefficients is based exclusively on the *within*-variation given for each unit over time. The *between*-variation across the units gets neutralized, levelled off, not used in the estimation of the slope coefficients (usually assumed not to vary over time and units).[7] This has the advantage of eliminating or

4 The model might be extended by the inclusion of time-specific intercepts (equal for all units) representing, for example, seasonal or business-cycle effects not captured by the predictor variables. See the contribution by Raffalovich and Chung, who elaborate such a model extension, including tests to check its appropriateness.

5 Or with *N*-1 dummies if one wants to have a reference unit (a "common" intercept) with zero values on all the dummy-variables included in the equation.

6 The equation can also be expanded by including lagged dependent variables on the right-hand side (thus presumably reducing serial correlation in the errors), and it can be modified by using first differences (for example, $\Delta X = X_t - X_{t-1}$) in order to deal with non-stationarity (cf. Beck and Katz 1996, 2011). However, adding lagged dependent variables may induce endogeneity bias and is particularly problematic with small T (cf. Nickell 1981).

7 LSDV models can not only be extended to include time-specific effects, but also transformed, for example, into Seemingly Unrelated Regression Equations (SURE models) which take into account slope coefficients that vary over units and allow for "contemporaneous correlation" between time-specific errors across individual units (cf. Alecke 1995, 24–28). Model designs and testing procedures that help to take into account such additional variants of structural effects are presented by Raffalovich and Chung in this focus section.

reducing the heterogeneity bias rooted in omitted variables that vary across units.[8] But this gain comes with a loss of estimation efficiency due to the reduction of variance in the explanatory variables included in the model. In addition, the impact of factors that do not vary over time cannot be estimated at all. To overcome these deficiencies Plümper and Troeger (2007) have proposed a three-stage "fixed effects vector decomposition" (FEVD) model which allows for retention of some of the between-unit variation in order to permit the estimation of effects attributable to time-invariant variables and a more efficient estimation of the effects attributed to "almost" time-invariant variables. This appears to be a rather attractive modelling strategy preserving the bias-reducing features of *fixed-effects* modelling while reducing the loss of efficiency by recovering some of the between-unit variance. However, the FEVD modelling strategy has received some rather critical comments as well (see Bell and Jones 2015; Breusch et al. 2011; and the replies in Plümper and Tröger 2011), and there are other versions of decomposition models which are either interpreted within the framework of *fixed-effects* or of so-called *random-effects* (RE) modelling (cf. Bell and Jones 2015); one of them is applied by Thames and McCall in their contribution in this focus-section. With regard to RE modelling (also referred to as Error Components [EC] modelling), we will not get into details here but point out at least some of its basic characteristics.

In the RE approach, unit- and/or time-specific effects which stem from sources outside the predictor variables actually included in the regression model are not "fixed" into unit and time-specific intercepts but treated as components of the error structure; that is, they are treated as random variables with mean zero and constant variance. The total error in the RE model thus has three components: "error systematic to space (cross-section), error systematic to time, and error systematic to both" (Sayrs 1989, 33). These three components have to be disentangled (under various assumptions) so that their systematic (but not fixed) effects can be combined into a single vector of slope coefficients. The partitioning of the error covariance matrix rests on the assumption that unit effects are captured as serial correlations that are constant at all lags over time. This in turn requires the restriction that the covariates *X* and the unit effects are uncorrelated and that there is neither spatial nor time-serial autocorrelation that would confound the constant serial correlation indicative of the unit effects (Stimson 1985, 924–25). There are several strategies to check for violations of assumptions and, if need be, to modify or expand the model in such a way as to allow for spatial and serial autocorrelation. So, for example, ARMA variations of the GLS model have been proposed to allow for serial autocorrelation (Stimson 1985, 925–29, 938–45; Sayrs 1989, 36–39). Whatever the specific characteristics of the applied models are, the components of the overall error matrix have to be identified in several steps, and the data (the values of the dependent and the independent variables) have to be transformed accordingly.

In a final step the required EGLS estimates are provided by an OLS regression performed on the transformed data. A weighting factor used in this final transformation reflects the relative size of the within- and between-error variances disentangled and estimated in previous steps. The final OLS estimates are thus a weighted average of the previously calculated within- and between-estimates. The larger the T, the more weight is given to the within estimates. In the limiting case of $T \rightarrow \infty$ the estimated regression coefficients of the fixed-effects LSDV model coincide with those of the EC model. Generally, the larger the T and the smaller the N, the more the advantages of FE modelling (minimizing bias) come to bear (Beck and Katz 1996, 4, fn. 7). On the other hand, with larger N and smaller T the efficiency gains achieved by EC *(random coefficient)* estimation become more paramount. But one should always keep in mind that the EC estimates (unlike the LSDV esti-

8 We know from ordinary cross-sectional regression analysis that the effect estimates are biased if omitted impact factors are correlated with included predictor variables. So it seems obvious that "to the extent that there are omitted characteristics that vary over time [in addition to those that vary only across units], the within-unit estimators will also be biased" (Phillips 2006, 952). Phillips and Greenberg (2008, 54, fn. 3) also note that if there are a small number of waves, "the fixed effects estimates are not necessarily unbiased no matter how many cases the researcher has. Random effects estimates, on the other hand, are consistent as the number of cases increases without limit, regardless oh how many observation times there are in the panel."

mates) are biased if the unit effects correlate with the predictor variables. The null-hypothesis of no correlation can be checked, for example, by the Hausman Test (Greene 1993, 479–80). The result of this test might confront the researcher with a difficult choice: either to maximize efficiency or to minimize bias. Much more discussion would be needed here, and we can only briefly draw attention to core issues. Bell and Jones, for example, note that "the Hausman test is not a test of FE versus RE; it is a test of the similarity of within and between effects" (Bell and Jones 2015, 144). They also "see the FE model as a constrained form of the RE model, meaning that the latter can encompass the former but not vice versa" (143). Beck and Katz (2007) strongly recommend considering the possibility of unit-to-unit variation in the model parameters, in other words the application of *random-coefficient* models (RCM) whenever a TSCS pooling format is given. And they present evidence from Monte-Carlo simulation studies demonstrating that in such cases Maximum-Likelihood estimation methods perform better than FGLS techniques.

2. Applications in the Focus Section

The initial paper by Thome ("Cointegration and Error Correction Modelling in Time-Series Analysis: A Brief Introduction") provides an introduction to cointegration and error-correction modelling in time-series analyses. The overarching substantive issues under investigation are how to distinguish between deterministic and stochastic trend-components, and how to avoid the associated dangers of spurious regression or spurious non-causality. The paper outlines some of the basic features and practical steps of cointegration modelling as a strategy for dealing with these issues, and illustrates this strategy with data on U.S. homicide rates and divorce rates, and with German data on sentencing and imprisonment.

In "Models for Pooled Time-Series Cross-Section Data," Raffalovich and Chung explain how modelling strategies for pooled data sets can also be conceptualized within the framework of *Multilevel/Hierarchical Linear Modelling* approaches. The authors use this analytic framework to develop a step-by-step testing strategy for identifying theoretically interpretable heterogeneities inherent in their pooled data set comprising $N = 40$ nations and $T = 56$ yearly measurements of homicide rates (dependent variable), divorce rates, and per-capita income (independent variables) between 1950 and 2005. They start with a "completely pooled" model implying that all countries over all time-points are identical in all unmeasured respects (perfect homogeneity given in the matrix of $NT = 2{,}240$ cases). They then test successively for country-specific, time-specific and time/country-specific effects, and finally for the possibility that the slope coefficients to be estimated for each of the predictor variables may vary across time and/or across countries. They apply log-likelihood ratio tests and FGLS estimation methods. Since the time-series are non-stationary they use first differences (yearly changes) for each variable. They also include the lagged dependent variable on the right-hand side of the regression equations. Raffalovich and Chung use this variable to control for time-dependencies but abstain from theoretical interpretations concerning the sign and magnitude of the respective coefficients.[9] They conclude with observations about how the models under consideration may help mitigate threats to validity that commonly arise in pooled time-series cross-section data analysis.

The general topic addressed by Thome – testing and modelling the over-time dynamics of structural relationships in a TSCS setting – is also the focal concern in Christoph Birkel's paper, "The Analysis of Non-Stationary Pooled Time Series Cross-Section Data".[10] If the time-series data for two or more variables exhibit trend components (non-stationarity) these variables will correlate even if they are causally unrelated. A common device to avoid such "spurious causality" (or "spurious regression") is to transform these time-series into their first (or higher-order) differences. This may however produce another problem: "spurious non-causality", where two trending series may be structurally related in the long run, but not in their

9 For a detailed discussion of the use of lagged dependent variables to model effect dynamics see Beck and Katz (2011).

10 Readers not familiar with the concepts of non-stationary, unit-root processes, cointegration and error-correction models are referred to the introductory paper by Helmut Thome, which has been included here to facilitate access to Birkel's contribution short-term movements extracted by differencing. On the other hand, two variables may be structurally related in their short-term movements, but not with regard to their long-term level relationship (as exemplified by the series analysed by Raffalovich and Chung, and also by Thome). There are several testing and modelling strategies that help the researcher not to fall victim, one way or the other, to the spuriousness trap. There are various forms of unit-root tests to check for the presence of *stochastic* (instead of deterministic) trends (*integrated* processes) in a set of time-series data, and also to check for so-called "cointegration", in the sense of corresponding (causally related) stochastic trend movements across two or more time-series. If the hypothesis of cointegration has been confirmed we can estimate not only the long-term level relationship between a predictor and the dependent variable, but also the parameters identifying the time-path of the "re-equilibration" process leading to the final level change ("error-correction models"). Birkel gives a detailed overview on various testing and modelling strategies, whose applicability and adequacy in each case depend on the substantive questions to be pursued and on given characteristics of the pooled data set. These characteristics include: the size of the sample (the number of units and time-points), cross-section dependencies, level-shifts and structural breaks caused by external events, and the degree of homogeneity assumed for residual variances and covariances and for short- and long-run parameters (e.g., the short-run dynamics may differ across units, but the long-run effects might still be homogeneous). How this array of pertinent or less pertinent modelling and estimation strategies can be evaluated and put to use in practical research, and the often uncertain and risky decisions that have to be made in this context, are exemplified in Birkel's analysis of a pooled set of time-series data (year by year from 1971 to 2004) for the eleven West German federal states. Trending robbery rates are the dependent variable; the predictor variables include per-capita income, per-capita consumption, and clearance rates (as well as demographic control variables). Birkel concludes that the available methodological procedures perform reasonably well with sufficient sample size, but notes that this qualification can create difficulties in practical situations, and points to areas where future development is needed.

As we have already mentioned, conventional RE (error-composition) models derive common slope coefficients from weighted averages of within- and between-variance components. But the framework of RE modelling has also been used to construct "decomposition" models (in the literature also referred to as "hybrid" models) which disentangle within- and between-effect estimates, thus providing two sets of slope coefficients (Phillips 2006, Bell and Jones 2015). Such models help the researcher to gather empirical evidence that may support or refute substantive hypotheses regarding different modes of causal dynamics, such as those briefly indicated at the beginning of our editorial: temporary process effects versus lasting structural effects, flow versus stock effects. Such distinctions may also be conceptualized within the framework of multi-level analysis (Bell and Jones 2015): as context effects (possibly attributed to the regional units) versus individual effects resulting from the over-time variations within these contexts – or the other way round. Which of the two, cross-sectional units or time-points, should be assigned to the "higher" or "lower" level depends upon the specific hypotheses to be examined and the relative size of *N* and *T.* In "A Longitudinal Examination of the Effects of Social Support on Homicide Across European Regions", Thames and McCall apply such decomposition models to examine the impact that "social support" and other predictor variables (relative deprivation and unemployment plus demographic control variables) exert upon homicide rates. Their study examines these structural relationships across 197 Western European and (separately) 50 Eastern European regions at three time points: 2000, 2005, and 2009. The results of their analyses offer reasonably robust evidence in support of social support, thereby complementing and extending prior work based on cross-sectional data.

In criminological research predictor variables (like GNP per capita or unemployment rates) are usually treated as *exogenous* variables that impact some dependent variable (like assault or homicide rates). But various types of *endogeneity* may also be involved in such overall causal structures. In "Does the Magnitude of the Link between Unemployment and Crime Depend on the Crime Level? A Quantile Regression Approach", Entorf and Sieger consider, for example, the possibility that the effect of unemployment on various types of crime depends on the level of crime given in a regional environment. They refer to opportunity theory, which suggests that "those who become unemployed in a low-crime area have higher incentives to commit a crime than those in high-crime regions, because they would face less effective prevention of potential victims and lower competition from other criminals than those in high-crime areas". On the other hand they note that the "stigma-based hypothesis … predicts low marginal effects … in low-crime areas, because here any potential detection bears a higher risk of stigma than in regions where criminal behaviour is more common". They examine these opposing hypotheses by applying a "quantile regression" approach, rarely used so far in criminological research. This modelling strategy allows estimation of different sets of regression coefficients depending on pre-defined quantile (percentile) levels of the dependent variable. The authors base their study on a pooled data set with yearly measurements from 2005 to 2009 gathered from 301 rural districts and 111 urban municipalities in Germany. They apply the conventional mean regression approach and compare its results with the findings from quantile regressions specified for the 5-, 25-, 50-, 75-, and 95-percent quantiles of their dependent variables (burglary, car theft, assault rates). Their main focus is on the effect of unemployment rates, but they also include other variables (like household income and clearance rate) among their predictors. The results obtained by these different approaches confirm that "conventional mean regressions might produce misleading results".

3. Outlook

The papers in this focus section underscore the promise of longitudinal analyses in research on criminal violence. Incorporating time into the design of studies can provide unique forms of leverage to facilitate inferences about causal processes. Moreover, the methodological foundations for longitudinal research have developed dramatically over recent decades, as reflected in the increasingly sophisticated approaches to statistical modelling. At the same time, debates are ongoing about the relative benefits and costs of various strategies, and there are often no easy solutions to some of the more difficult challenges. We hope that this focus section will stimulate further interest in longitudinal analyses of criminal violence and in the development of methodologies to advance such analyses.

References

Alecke, Björn. 1995. *Regressionsanalyse mit Panel-Daten*, Volkswirtschaftliche Diskussionsbeiträge 221. Münster: Institut für Wirtschafts- und Sozialgeschichte.

Beck, Nathanael, and Jonathan N. Katz. 1996. Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models. *Political Analysis* 6:1–36.

Beck, Nathanael, and Jonathan N. Katz. 2007. Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments. *Political Analysis* 15:182–95.

Beck, Nathanael, and Jonathan N. Katz. 2011. Modeling Dynamics in Time-Series-Cross-Section Political Economy Data. *Annual Review of Political Science* 14:331–52.

Bell, Andrew, and Kelvyn Jones. 2015. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods* 3:133–53.

Birkel, Christoph. 2015. The Analysis of Non-Stationary Pooled Time Series Cross-Section Data. *International Journal of Conflict and Violence* 8 (2): 222–242.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas. 2011. On the Fixed-Effects Vector Decomposition. *Political Analysis* 19:123–34.

Entorf, Horst, and Philip Sieger. 2014. Does the Magnitude of the Link between Unemployment and Crime Depend on the Crime Level? A Quantile Regression Approach. *International Journal of Conflict and Violence* 8 (2): 262–283.

Greene, William H. 1993. *Econometric Analysis*, 2d ed. New York et al.: Macmillan.

Hsiao, Cheng. 1986. *Analysis of Panel Data.* New York: Cambridge University Press.

Kmenta, Jan. 1986. *Elements of Econometrics*, 2d ed. New York: Macmillan.

Messner, Steven F., and Richard Rosenfeld. 2013. *Crime and the American Dream*, 5th ed. Belmont, CA: Wadsworth. 1st ed. 1994.

Messner, Steven F., Helmut Thome, and Richard Rosenfeld. 2008. Institutions, Anomie, and Violent Crime: Clarifying and Elaborating Institutional-Anomie Theory. *International Journal of Conflict and Violence* 2 (2): 163–81.

Nickell, Stephen J. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica* 49:1417–26.

Phillips, Julie A. 2006. Explaining Discrepant Findings in Cross-Sectional and Longitudinal Analyses: An Application to U.S. Homicide Rates. *Social Science Research* 35:948–74.

Phillips, Julie A., and David F. Greenberg. 2008. A Comparison of Methods for Analyzing Criminological Panal Data. *Journal of Quantitative Criminology* 24:51–72.

Plümper, Thomas, and Vera E. Troeger. 2007. Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis* 15:124–39.

Plümper, Thomas, and Vera E. Troeger. 2011. Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments. *Political Analysis* 19:147–64.

Raffalovich, Lawrence E., and Rakkoo Chung. 2014. Models for Pooled Time-Series Cross-Section Data. *International Journal of Conflict and Violence* 8 (2): 209–221.

Sayrs, Lois W. 1989. *Pooled Time Series Analysis*. Beverly Hills: Sage.

Stimson, James A. 1985. Regression in Space and Time: A Statistical Essay. *American Journal of Political Science* 29:914–47.

Thames, Kelly M., and Patricia L. McCall. 2014. A Longitudinal Examination of the Effects of Social Support on Homicide Across European Regions. *International Journal of Conflict and Violence* 8 (2): 243–261.

Thome, Helmut. 2010. Violent Crime (and Suicide) in Imperial Germany, 1883–1901: Quantitative Analyses and a Durkheimian Interpretation. *International Criminal Justice Review* 20 (1): 5–34.

Thome, Helmut. 2014. Cointegration and Error Correction Modelling in Time-Series Analysis: A Brief Introduction. *International Journal of Conflict and Violence* 8 (2): 199–208.

Thome, Helmut, and Stephan Stahlschmidt. 2013. Ost und West, Nord und Süd: Zur räumlichen Verteilung und theoretischen Erklärung der Gewaltkriminalität in Deutschland. *Berliner Journal für Soziologie* 23:441–70.

Helmut Thome
helmut.thome@soziologie.uni-halle.de

Steven F. Messner
s.messner@albany.edu

International Journal
of Conflict and Violence

urn:nbn:de:0070-ijcv-2014230
IJCV: Vol. 8 (2) 2014

Cointegration and Error Correction Modelling in Time-Series Analysis: A Brief Introduction

Helmut Thome, Institute of Sociology, University of Halle-Wittenberg, Germany

Vol. 8 (2) 2014

ISSN: 1864–1385

Cointegration and Error Correction Modelling in Time-Series Analysis: A Brief Introduction

Helmut Thome, Institute of Sociology, University of Halle-Wittenberg, Germany

Criminological research is often based on time-series data showing some type of trend movement. Trending time-series may correlate strongly even in cases where no causal relationship exists (spurious causality). To avoid this problem researchers often apply some technique of detrending their data, such as by differencing the series. This approach, however, may bring up another problem: that of spurious non-causality. Both problems can, in principle, be avoided if the series under investigation are "difference-stationary" (if the trend movements are stochastic) and "cointegrated" (if the stochastically changing trend-movements in different variables correspond to each other). The article gives a brief introduction to key instruments and interpretative tools applied in cointegration modelling.

Criminologists often use time-series data to describe long-term developments of crime. Such data can also be used to identify and model assumed structural relationships between crime rates (treated as dependent variables) and factors like unemployment or divorce rates (treated as independent, explanatory variables). The adequacy of the specific analytical techniques and statistical models applied in such analyses has to be judged with regard to certain features – problems and possibilities – inherent in the given data. One of those features that need careful consideration is the absence or presence of trend components. Two or more time-series, each of them exhibiting a persisting upward or downward trend, will always correlate with each other (positively or negatively) even in cases where no causal relationship between them exists. On the other hand, if we eliminate the trend components the remaining series will likely be uncorrelated even in cases where their levels are structurally related to each other. Usually, however, there are more alternatives available than choosing between spurious causality and spurious non-causality. Often, level changes may proceed in a temporarily changing pattern, switching from upward to downward movements, speeding up or slowing down in this or that direction, in other words they might be "stochastic" (rather than "deterministic"). If two (or more) series that show such unsteady, stochastic trend movements still correlate with each other, then we can be quite confident that there is indeed a structural (causal) relationship between them; otherwise their unsteady trend movements would not be corresponsive across the series under inspection. This paper gives a brief introduction into certain statistical strategies and techniques that can be used (or should not be used) in testing and modelling structural relationships between time series exhibiting some type of trend development.

1. Deterministic versus Stochastic Trend Components

A trend component is usually represented in one of the following two ways: either "deterministically" as a linear or non-linear function of time or "stochastically" as a so-called *unit-root process*. A simple example of the first variant would be:

I wish to thank Christoph Birkel for his helpful comments on an earlier draft of this paper, which also draws upon Thome (1996, 2005). Birkel's own article (pp. ###–### in this issue) presupposes a basic understanding of the concepts and procedures outlined in the present paper.

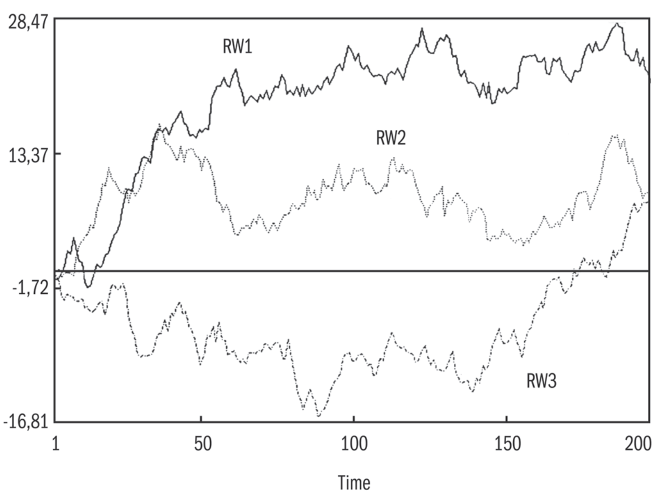
(1) $z_t = \alpha + \gamma t + \varepsilon_t, \quad t = 0,1,2,3 \ldots,$

where *t* is a time-index, α the initial level of the time series *Z(t)*, and ε symbolizes a random ("error") input with constant variance and an expected value $E(\varepsilon_t) = 0$. When the trend coefficient γ is known (estimated), the series can be detrended by calculating $z_t - \gamma t = \alpha + \varepsilon_t$. However, if the trend is in fact not deterministic but has been generated by a random process, this procedure of modelling and detrending the series would be inappropriate.[1] The simplest model of such a random process producing a trend is given by the following equation:

(2) $z_t = z_{t-1} + \varepsilon_t \leftrightarrow z_t = z_0 + \Sigma\varepsilon_t$

If the errors ε(t) are distributed normally with constant variance[2] around the expected value $E(\varepsilon_t) = 0$, and if they are also uncorrelated with each other and with Z_{t-1}, (if they are "white noise"), this process is called a *simple random walk* (RW). Figure 1 represents three realizations of this type of processes exhibiting temporary upward and downward movements along the time axis.

Figure 1: Three realizations of a random walk



All these RW realizations start with the value $z_0 = 0$ and then successively add the accumulated random shocks $\Sigma\varepsilon_t$ (t = 1,2,3 … 200) according to equation (2). Note that even a *simple* random walk (without "drift," see below) may, within a limited period of time, appear to produce an overall trend component and/or a cyclical movement.

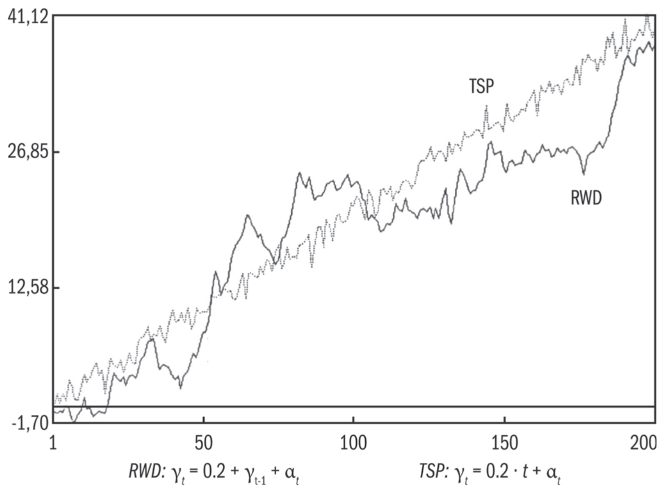
Stochastic trend components of this kind can be eliminated by calculating the first differences: $\Delta Z_t = Z_t - Z_{t-1}$. Consequently such a process is called a *difference-stationary process* (DSP) and contrasted with the *trend-stationary process* (TSP) given in equation (1). In some cases, the first-order differencing may not be sufficient to produce a stationary process, which however might still be achieved by differencing the series of first differences, and possibly the second differences as well and so forth, thereby leading to *second* or *higher order differences*, $\Delta^p Z_t$. Difference-stationary processes are also referred to as *Integrated Processes* of order *p:* I(p)-processes. The more technical term "unit-root process" is derived from the mathematics of difference equations, which cannot be introduced here (a "unit-root" of 1 is the formal requirement of difference-stationarity).

Equation (2) can be extended by adding further components, in particular a constant term μ, a so-called *drift* parameter. This is a deterministic linear trend component (with μ as the slope coefficient), thus the time series moves more and more away from its original level, but since the trend component is embedded within a random walk process the fluctuations around this long-term trend line increase with time. Figure 2 represents such a *random walk with drift* (RWD), contrasted with a TSP according to equation (1).

1 See Nelson and Kang (1981, 1984), Banerjee et al. (1993), and Raffalovich (1994).

2 But note that the variance of the time series Z(t) is $t \times \sigma^2$, so it increases with time.

Figure 2: Realization of a random walk with drift (RWD) and of a trend-stationary process (TSP)



Among the problems that arise when fitting a deterministic trend (according to equation 1) to a random walk process, we find the following (see Nelson and Kang 1981, 1984; Banerjee et al. 1983):

(1) If the series has been generated by a simple random walk (without *drift*), an OLS regression on the time index *t* produces a spurious coefficient of determination which does not decrease by increasing sample size. Standard tests of significance of the slope coefficient (based on Student's *t* statistic) tend to be largely biased in an upward direction. The correct null hypothesis (stating that the slope coefficient of the time index should be zero) will be rejected in a large majority of cases. (2) If the random walk contains a *drift* component, a coefficient of determination larger than zero (produced by an OLS regression on *t*) makes some sense, but it also tends to be overestimated. (3) The autocorrelations of the residuals resulting from such "spurious" regressions tend to exhibit an artificially cyclical pattern, whose period length and standard deviation depend (positively) on sample size (the length of the series observed).

As the name suggests, calculating the first- or higher-order differences is the appropriate way of detrending a difference stationary process. However, two problems have to be considered before doing this. (1) The elimination of the respective trend components forestalls any possibility to identify and test the level relationship which might connect two or more series ("co-integration", see below). (2) Since differencing eliminates or reduces the weight of low-frequency components in general, it not only eliminates the trend, but also any cycles present in the series, no matter if they are deterministic or emanate from some stationary second- or higher-order autoregressive process.

With regard to causality, one has to be aware of additional problems. Imagine that we have two simple random walks, Y_t and X_t, according to equation (2), which have been produced independently from each other (by way of simulation experiments, for example). If we regress the Y-series on the X-series

(3) $y_t = \alpha + \beta x_t + \varepsilon_t$

the theoretically expected slope coefficient is, of course, $\beta = 0$. But we are very likely to obtain a slope coefficient which departs significantly from zero, and this likelihood will increase with the length of the series (Banerjee et al. 1993, 74 ff.). This is another instance of "spurious regression" (Granger and Newbold 1974). And again this problem cannot be solved by detrending the series with a polynomial function of time before running the regression or by including the time index t in the set of regressors.

These observations taken together suggest the following approach: If one wants to identify or test structural (causal) relationships between seemingly trending time series, one should not start by detrending the data at all. Instead, one should first test the assumption that the series to be analyzed are *difference-stationary*, that the trend in each series is stochastic (this can be done by *unit-root* testing, as explained in section 3 below). A necessary (but not sufficient) condition for a structural relationship between such series is that they are integrated processes of the same order. If this turns out to be the case, an assumed structural relationship between the series can be identified and tested with the help of *cointegration models* (see section 2). When such a hypothesis has been confirmed, the temporal dynamics in which the level of one series is adjusted to the changing level of the other series can be identified with the help of *error correction models.* Before illustrating the application of this strategy in section 3, the key concept of "cointegration" is briefly outlined in the next section.

2. The Concept of Cointegration

Two or more time series are said to be co-integrated if two conditions prevail: first, each of the series must be integrated to the same order; second, there must exist at least one linear combination among the series which is stationary. If there is only one such linear combination, it can easily be obtained by regressing one series upon the other(s):

$$(4) \quad y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

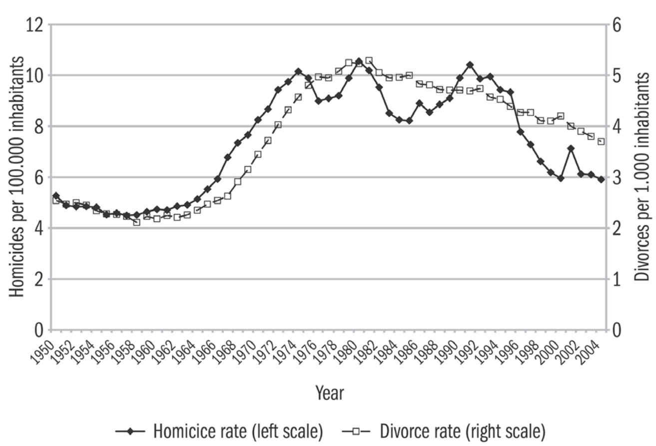
The residuals of this (static) regression are a linear combination of the Y- and X-series. Generally, linear combinations of two or more first-order integrated series are again integrated to the first order. But under specific conditions – if the stochastic trend components in each series evolve correspondingly (in "co-integration" with each other) – they will be stationary. Consequently, if the residuals of the estimated equation (4) prove to be stationary, we have a strong indicator for a causal relationship between the variables involved. Without engaging in formal derivations the following line of reasoning can be developed:

Imagine two time series, each dominated by stochastic trend components. If they are integrated to *different* orders they cannot be structurally related to each other in their long-term development.[3] If they are integrated to the same order, their stochastically evolving trend components might be (causally) related to each other – or not. If we find a close correspondence between the trending up and down movements in different series, either positively or negatively, we can be quite confident that there is indeed a structural, causal relationship between these series, precisely because of their stochastic nature. Without being causally connected, stochastic up and down movements could not be expected to move on in close correspondence across different series. And if we observe such a correspondence in stochastic movements, we can be quite confident that a causal relationship exists. This correspondence – or the lack of it – is revealed in the residuals of the (static) cointegration regression. If they are stationary, a *moving equilibrium* relationship must exist between the series under investigation. Externally induced departures from the equilibrium spawn more or less rapid readjustments. Differencing the series to level stationarity before performing the regression analysis would eliminate this long term co-movement, preventing it from being detected. One would thus fall victim not to spurious causality (or regression) but to spurious non-causality.

3. Modelling Cointegration: Two Examples

In our first example we look at the homicide and divorce rates of the United States from 1950 to 2005 (see Figure 3).[4]

Figure 3: U.S. homicide and divorce rates



3 However, a variable Y (such as confidence in the future) may remain level-stationary as long as there is a persistent upward or downward trend in another variable X, such as a steadily growing gross national product. In such a case, the structural relationship might be tested by regressing Y on the first differences of X or the growth rates derived from X. Though a "co-integration" model can be formally specified only for two or more series integrated to the same order, a specific series might be differenced in advance, based on theoretical argument, before being included in the cointegration equation.

4 I am grateful to Steve Messner, who made these data available to me

In criminological research, divorce rates have repeatedly been treated as indicators of certain aspects of social disorganisation or institutional anomie considered to be conducive to criminal violence (Beaulieu and Messner 2010; Land et al. 1990; Pratt and Cullen 2005). Neither the homicide nor the divorce rates seem to follow a deterministic trend. So, we first check if they can be modelled as integrated processes of the same order. The unit-root tests we apply here is the *augmented* Dickey-Fuller Test (ADF Test), which takes into account the auto-correlation of residuals (Dickey and Fuller 1979, 1981).[5] In its simplest form the equation to be estimated for testing-purposes is

$$(5) \quad y_t = \rho_a y_{t-1} + \varepsilon_t$$

The null hypothesis (that the Y-series is a first order integrated [difference-stationary] process) implies a coefficient $\rho = 1$, the alternative hypothesis (level stationarity) implies $\rho < 1$. The alternative of a stationary *autoregressive* process would usually need to be modelled with the inclusion of a constant term. So, in order to be "fair" against the alternative, in praxis equation (5) is usually expanded into equation (6):

$$(6) \quad y_t = \alpha_b + \rho_b y_{t-1} + \varepsilon_t$$

To be on the safe side, it is often recommended to also include the time index among the regressor variables, particularly if the series shows (or seems to show) a prevailing upward or downward movement within the observation period:

$$(7) \quad y_t = \alpha_c + \gamma t + \rho_c y_{t-1} + \varepsilon_t$$

In this way one can check if there is sufficient evidence for the presence of a stochastic trend component even though a deterministic trend component is given a chance to be identified as well. If one wants to rule out the possibility that there is also a deterministic trend component (besides the stochastic trend component), one would have to test the combined hypothesis $\rho=1$ *and* $\gamma=0$.

Under the null hypotheses, the OLS estimates of the ρ parameters are not normally distributed. Dickey and Fuller (1981) have applied Monte Carlo simulations in order to establish the distribution of these estimates under various conditions: does the equation tested include a constant or a time index, does the true process include or not include a drift parameter, a deterministic trend component etc.? The ratio of the difference between observed and expected ρ-coefficient divided by the standard deviation is usually symbolized by the letter τ (to differentiate it *from Student's t*). The critical values of this test statistic and the significance level α associated with them, each specified for the varied conditions just mentioned, are available in the literature (see, for example, Hamilton 1994) and computer programs like *STATA*.

When we apply this testing strategy to our series of homicide and divorce rates, the assumption that both series are difference-stationary (integrated) processes of order 1 is confirmed.[6] When estimating the ρ-coefficient on the basis of equation (7) we get $\hat{\rho}_c = 0.95$ for homicide rates and $\hat{\rho}_c = 0.98$ for divorce rates. These observed coefficients depart only $\tau = 1.21$ and $\tau = 1.19$ standard deviations from their expected value of $\rho = 1.0$; consequently, the error risk for rejecting the null hypothesis would be $\alpha \geq 10$ percent.[7] In addition, we tested the combined hypothesis $\rho=1$ *and* $\gamma=0$; it was confirmed for both series.

Since the two series are apparently integrated processes of the same order, they are also candidates for co-integration. We thus regress the homicide series on the divorce series according to equation (3). The estimated slope coefficient is highly significant, the coefficient of determination is

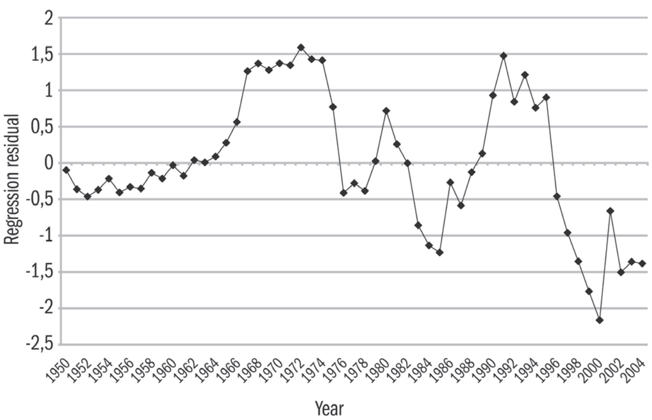
5 There is quite some discussion in the literature about the appropriateness of unit-root testing in general and of specific drawbacks or advantages of alternative testing strategies (for example DeJong et al. 1992; Hamilton 1994). Generally, one can say that these tests get more problematic the shorter the length of the inspected time series is (some authors even argue that they are useless with less than 100 observations).

6 I am grateful to Christoph Birkel, who carried out these tests with the help of the *STATA* computer program.

7 The critical value chosen to accept or reject the null hypothesis depends on whether one would like to reject the null hypothesis (which is made more convincing the lower the α-value) or rather not reject it (which is made more convincing the higher the α-value; commonly recommended in this case are values of $\alpha \geq 10$ percent).

$R^2 = 0.80$. However, the estimated residuals do not represent a stationary process. As shown in Figure 4, there is a clear upward trend between 1950 and the mid-1970s, and a slight and fuzzy decline afterwards. The results of the ADF Tests applied to the residual series confirm this impression. The conclusion thus is that the changing level of homicide rates is not causally related to the changing level of divorce rates. Apparently, what we have here is an example of *spurious regression.*

Figure 4: Residuals of cointegration model with homicide and divorce rates

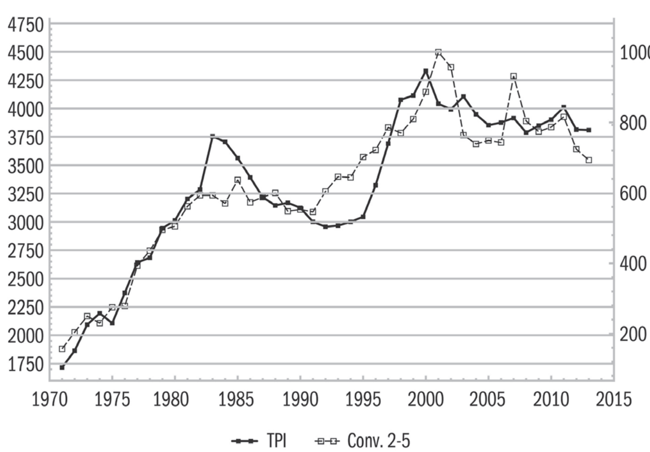


A word of caution however is in order at this point. If there are strong theoretical arguments in support of a structural relationship between the two variables (against the detected spuriousness), one might consider alternative testing procedures before giving up the hypothesis. For example, one might apply models that combine stochastic and deterministic trend conceptions by assuming that deterministic trends change their functional form stochastically over time (see Perron 1989; Perron and Vogelsang 1992). Alternative testing procedures have also been used by Christoph Birkel in his article published in our present issue. A very helpful overview on various testing and modelling strategies concerning cointegration is presented by Enders (2010).

Also, the bivariate cointegration model might be underspecified in our example: there might be additional input factors (apart from divorce rates) that should be included in the model (see Messner et al. 2011). In such cases, not all of the input factors included in a cointegration model need to be integrated processes of the same order. Greenberg (2001) even finds a cointegrating relationship between US homicide and divorce rates using yearly data for the period between 1946 and 1997. He notes, however, that "the parallel movement may have weakened some in recent years" (ibid., 302). This weakening apparently continued in the following years until the end of our observation period (2005), thus tilting the results toward "spurious regression" (for the bivariate relationship).

We now take a look at two other series: (a) the total number of prison inmates (TPI) who, in their great majority, serve short-term sentences, and (b) the (smaller) number of perpetrators sentenced to relatively long periods of imprisonment between two and five years (CONV2-5) in the German state of Hesse between 1971 and 2013 (see Figure 5).[8]

Figure 5: Total number of prisoners (left scale) and number serving two to five years (right scale)



8 I am grateful to Rainer Metz (*GESIS-Leibniz Institute,* Cologne) who made these data available to me and also carried out the statistical analyses (selectively) reported below. For an extended analysis and theoretical discussion of law enforcement and incarceration practices see Metz (2013)

Law-making as well as law enforcement and sentencing practices are influenced by public opinion and political opportunities following short- and long-term fluctuations. Public discussions often focus on more severe and cruel criminal acts, but such discussions may also increase the readiness and determination to prosecute and incarcerate people for minor crimes as well. On the other hand, they may lead to a redirection of policing and prosecution resources to concentrate on more serious crimes. In Germany we have no detailed and comprehensive statistics on length of imprisonment actually served compared to original sentence. So one might ask (among other things) if the (relatively small) number of convictions covering some limited range of severity (like, as in our example, two to five years of incarceration) will in the long run closely correspond (or not correspond) to the development of the total number of prison inmates. How indicative or representative are these convictions for the long-term development of the total number of imprisoned persons? As it turns out in our particular example, eliminating the seemingly linear trend component in both series before doing any kind of regression analysis does not lead to zero-correlations between the two residual series. Nonetheless, it is advisable not to eliminate the trend or drift component right at the beginning of the analysis, but to check if the two series are in fact cointegrated, whether they have a long term equilibrium relationship with regard to their (stochastic trend) levels.

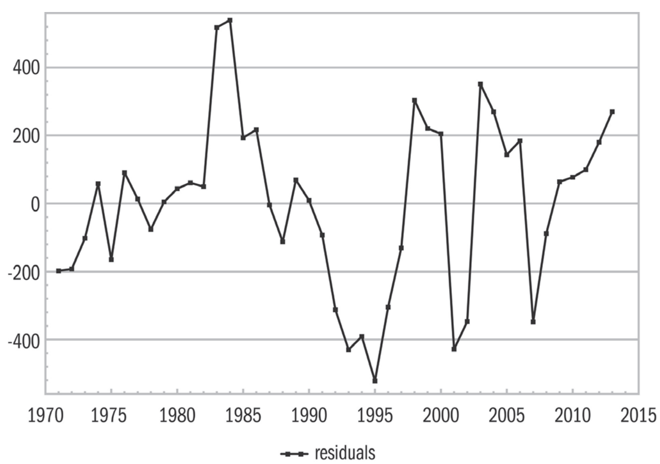
Thus, we first apply ADF Tests to both series according to equation (7). Their results support the assumption that they incorporate first-order integrated (difference-stationary) processes.

In the next step we run the co-integration regression (see equation 3 above) with the following result:

(8) $TPI_t = 1437.7 + 3.033(CONV2to5)_t + e_t$

The slope coefficient of 3.033 implies that an increase in the number of people sentenced to (relatively) long-term imprisonment will uplift the total number of prisoners by a factor of 3.03 (with a coefficient of determination $R^2 = 0.86$). For example, an addition of ten more persons convicted in this category will be followed by an increase of thirty in the total number of prisoners (whatever the mechanism in this process might be). But before we accept this hypothesis, we must check whether or not the residuals of this regression are stationary (see Figure 6)

Figure 6: Residuals of cointegration model with total numbers of prisoners and those serving two to five years



The results show that they are; the unit-root test confirms the impression received from visual inspection of the residuals: the null hypothesis of a unit-root can be rejected with a risk of $\alpha < 0.001$.

So far, we have produced evidence that these two series are cointegrated, but the static regression says nothing about the dynamics of the re-equilibration processes. Engle and Granger (1987) have suggested that this process can be specified in a so-called error correction model, which we present here in the form adapted to our example[9] and with parameter estimates obtained by OLS regression:[10]

9 In other cases, lagged terms of the dependent and/or the independent variable might need to be included as well.

10 We skip here any discussion about diagnostic statistical tests concerning the adequacy of the model specification and the estimation procedure (for example Greene 1993, 216 ff.). The long-term effect parameters of equation (8) and the short-term effect parameters of equation (9) can also be estimated simultaneously in one equation (see Wolters 1995, p. 153; Wagner and Hlouskova 2007).

(9) $\Delta TPI_t = 0.876 \times \Delta(CONV2to5)_t - 0.314 \times Z_{t-1} + e_t$

As already mentioned, the symbol **Δ** indicates differencing, $\Delta TPI_t = TPI_t - TPI_{t-1}$. Consequently, the dependent variable and the regressor variables in this model are stationary. The variable Z_{t-1} is given by the estimated Residuals $TPI_t - 1437.7 - 3.0333(CONV2to5)_t$ derived from equation (8). These residuals represent departures from the equilibrium relationship. The negative coefficient of -0.314 thus gives the rate of readjustment, year by year, towards the new equilibrium level *(re-equilibration)*, no matter whether the disequilibrium has been induced by a change in the independent variable or by a change in the dependent variable caused by some (non-specified) intervention impacting directly the dependent variable. The system is in a state of equilibrium if $Z_t = 0$, i.e., $TPI_t - 3.03(CONV2to5)_t = 1437.7$. Let us suppose that at some year *t* the number of people sentenced to between two and five years will increase from 200 to 220. Then, according to equation (8), the equilibrium level in TPI would change from 3.03 × 200 + 1437.7 = 2044 to 3.03 × 220 + 1437.7 = 2104, an increase of 60. The level change in the number of convicted persons (in this category) is immediately (in the same year) answered (according to equation 9) by an expected increase of 0.876 × 20 = 17.5 in the total number of prisoners (TPI),[11] thus reducing the dis-equilibrium from 60 to (60 – 17.5) = 42.5. This remaining disequilibrium will (according to equation 9) in the next year be reduced to (42.5 – 0.314 × 42.5) = 29.15, in the subsequent year to (29.15 – 0.314 × 29.15) = 20 and so forth.

4. Concluding Remarks

Within criminological research, time series data are often used to depict the long-term development of crime and to test hypotheses seeking to explain such developments. In order to identify and test structural relationships that may exist among two or more theoretically relevant time-series, one has to take into account the specific components and features inherent in such data. The point of departure in this paper has been the distinction between deterministic and stochastic trend-components and the danger (related to these components) of falling victim to either spurious regression or spurious non-causality. Subsequently, some of the basic features and practical steps in cointegration modelling have been outlined as a strategy which helps to identify and test structural relationships between trending time series without getting entrapped into spurious causality or non-causality. The practical application of this strategy has been illustrated by analysing American homicide and divorce rates given for the years 1950 to 2005, and German data on the number of sentenced and imprisoned people in the years 1971 to 2013. The purpose of this paper has been to outline the basic ideas behind the concepts of unit-root testing and cointegration modelling, which are useful instruments in analysing time-series data relevant to criminological research. The analyses presented here could have been extended both with regard to substantive as well as methodological and technical issues. This however would have gone beyond the (didactically defined) scope of the article.

11 Note that the interpretation of the regression coefficient (here: 0.876) does not change when the two variables have been transformed by the same filter, here by taking first differences.

References

Banerjee, Anindya, Juan J. Dolado, John W. Galbraith, and David F. Hendry. 1993. *Co-Integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Oxford: Oxford University Press.

Beaulieu, Mark, and Steven F. Messner. 2010. Assessing Changes in the Effect of Divorce Rates on Homicide Rates across Large U. S. Cities, 1960–2000: Revisiting the Chicago School. *Homicide Studies* 14:24–51.

DeJong, David N., John C. Nakervis, Nathan E. Savins, and Charles H. Whiteman. 1992. The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors. *Journal of Econometrics* 53:31–50.

Dickey, David A., and Wayne A. Fuller. 1979. Distribution of Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74:427–31.

Dickey, David A., and Wayne A. Fuller. 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* 49:1057–72.

Enders, Walter. 2010. *Applied Econometric Time Series* (3d ed.). New York: Wiley.

Engle, Robert F., and Clive W. J. Granger. 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55:251–76.

Granger, Clive W.J., and Paul Newbold. 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2:111–20.

Greenberg, David F. 2001. Time Series Analyses of Crime Rates. *Journal of Quantitative Criminology* 17:291–327.

Greene, William H. 1993. *Econometric Analysis*, 2d. ed. New York et al.: Macmillan.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton: New University Press.

Land, Kenneth C., Patricia L. McCall, and Lawrence E. Cohen. 1990. Structural Covariates of Homicide Rates: Are There Any Invariances across Time and Social Space? *American Journal of Sociology* 95:922–63.

Messner, Steven F., Benjamin Pearson-Nelson, Lawrence E. Raffalovich, and Zachary Miner. 2011. Cross-National Homicide Trends in the Latter Decades of the Twentieth Century: Losses and Gains in Institutional Control? In *Control of Violence: Historical and International Perspectives on Violence in Modern Societies*, ed. W. Heitmeyer, H.-G. Haupt, S. Malthaner, and A. Kirschner, 65–89. New York: Springer.

Metz, Rainer. 2013. Zeitreihen und Strafvollzugsprognosen. In *Rechtspsychologie, Kriminologie und Praxis. Festschrift für Rudolf Egg zum 65. Geburtstag*, ed. A. Dessecker and W. Sohn, 399–434. Wiesbaden: Eigenverlag Kriminologische Zentralstelle.

Nelson, Charles R., and Heejoon Kang. 1981. Spurious Periodicity in Inappropriately Detrended Time Series. *Econometrica* 49:741–51.

Nelson, Charles R., and Heejoon Kang. 1984. Pitfalls in the Use of Time as an Explanatory Variable in Regression. *Journal of Business & Economic Statistics* 2:73–82.

Perron, Pierre. 1989. The Great Crash, The Oil Prices Shock, and the Unit Root Hypothesis. *Econometrica* 57:1361–1401.

Perron, Pierre, and Timothy J. Vogelsang. 1992. Testing for a Unit Root in a Time Series with a Changing Mean: Corrections and Extensions. *Journal of Business and Economic Statistics* 10:467–70.

Pratt, Travis C., and Francis T. Cullen. 2005. Assessing Macro-level Predictors of Crime: A Meta-analysis. In *Crime and Justice: A Review of Research*, ed. M. Tonry, 373–450. Chicago: Chicago University Press.

Raffalovich, Lawrence E. 1994. Detrending Time Series: A Cautionary Note. *Sociological Methods and Research* 22:492–519.

Thome, Helmut. 1996. Trends, Cycles, and Co-integration: Some Issues in Modelling Long-term Development in Time Series Analysis. *Historical Social Research/Historische Sozialforschung* 21:3–23.

Thome, Helmut. 2005. *Zeitreihenanalyse: Eine Einführung für Sozialwissenschaftler und Historiker*. Munich and Vienna: Oldenbourg.

Wagner, Martin, and Jaroslava Hlouskova. 2009. The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study. *Economic Reviews* 29:182–223.

Wolters, Jürgen. 1995. Kointegration und Zinsentwicklung im ESW: Eine Einführung in die Kointegrationsmethodologie und deren Anwendung. *Allgemeines Statistisches Archiv* 79:146–69.

Helmut Thome
helmut.thome@soziologie.uni-halle.de

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014255
IJCV: Vol. 8 (2) 2014

Models for Pooled Time-Series Cross-Section Data

Lawrence E. Raffalovich, Department of Sociology, University at Albany, State University of New York
Rakkoo Chung, Department of Sociology, University at Albany, State University of New York

Vol. 8 (2) 2014

ISSN: 1864–1385

Models for Pooled Time-Series Cross-Section Data

Lawrence E. Raffalovich, Department of Sociology, University at Albany, State University of New York
Rakkoo Chung, Department of Sociology, University at Albany, State University of New York

Several models are available for the analysis of pooled time-series cross-section (TSCS) data, defined as "repeated observations on fixed units" (Beck and Katz 1995). In this paper, we run the following models: (1) a completely pooled model, (2) fixed effects models, and (3) multi-level/hierarchical linear models. To illustrate these models, we use a Generalized Least Squares (GLS) estimator with cross-section weights and panel-corrected standard errors (with EViews 8) on the cross-national homicide trends data of forty countries from 1950 to 2005, which we source from published research (Messner et al. 2011). We describe and discuss the similarities and differences between the models, and what information each can contribute to help answer substantive research questions. We conclude with a discussion of how the models we present may help to mitigate validity threats inherent in pooled time-series cross-section data analysis.

The analysis of pooled time-series cross-section (TSCS) data has become increasingly popular in the social sciences. For example, Adolph, Butler, and Wilson (2005) found that the number of political science articles in journals indexed in JSTOR using "time-series-cross-section" terminology increased explosively in the late 1980s, and that roughly two hundred studies published between 1996 and 2000 used time-series-cross-sectional data. Similarly, we find in EBSCO that the number of scholarly (peer-reviewed) journal articles that include the term "time-series-cross-section" in their abstract increased from four in the 1980s to fourteen in the 1990s and ninety-nine between 2000 and 2014.

Pooled TSCS data consist of "repeated observations on fixed units" (Beck and Katz 1995, 634). Thus, the total number of observations equals the number of cross-sections (I) multiplied by the number of time points (T). For example, our data from Messner et al. (2011) include 2,240 observations (i.e., IxT country-years), covering forty countries (I) for fifty-six years (T) between 1950 and 2005.

An Ordinary Least Squares (OLS) regression is not appropriate for this type of data because time-series observations are clustered within countries, inducing correlation among observations (Snijders and Bosker 2011). This violates the assumption of independence of observations, which is required for unbiased estimation of variances and standard errors in OLS regression.

The past several decades have witnessed several approaches to the correlated observations problem in analysis of pooled TSCS data.[1] Researchers, however, have to decide which approaches are appropriate for their research by checking whether the underlying assumptions are appropriate for their theories and data. Otherwise, they risk invalid parameter estimates, incorrect standard errors, and/or wrong type-I and type-II error rates. In other words, their findings may simply be wrong.

In this paper, we present a sequence of nested models to make explicit and test the assumptions that underlie each

1 Adolph, Butler, and Wilson (2005) compare: (1) a pooled regression by least squares (Beck and Katz 1995), (2) the Beck-Katz method with fixed effects, (3) an instrument of the lagged level of the dependent variable (Anderson and Hsiao 1981, 1982), and (4) Generalized Methods of Moments (GMM) estimation (Wawro 2002). See also Snijders and Bosker (2011, 197–202).

model. We chose the models based on theory, prior research, and the structure of the data. We start the sequence with a baseline model, the simplest model with the most restrictive assumptions. We specify subsequent models by relaxing these restrictions, and testing whether the restrictions are supported by the data. In this manner, we demonstrate a step-by-step approach to the analysis of TSCS data and illustrate a methodology for exploiting the properties of this data structure. In the first section of this paper, we describe and discuss several TSCS models, their similarities and differences, and what information each can contribute to help answer substantive research questions. In the second section, we illustrate these models with data on cross-national homicide trends from Messner et al. (2011). In the last section, we summarize our analysis and discuss methodological and theoretical implications for the analysis of pooled TSCS data.

1. Models

Several models are available for the analysis of pooled TSCS data. These include completely pooled models in which all observations – all cross-sections and all repeated observations – are assumed to be equivalent. That is, the pooled data are assumed to be a random sample from a population observed over time; and the data-generating process is assumed to be the same for all cross-sections. Fixed-effects models acknowledge that cross-sections and/or time periods may differ in unknown ways. This is known as unobserved variable bias. In pooled data this may also result in unequal variances, or *heteroskedasticity*, at the cross-section and/or time level. Fixed-effects models incorporate these departures from randomness by including a dummy variable for each cross-section and/or each time period, and Random-Effects models account for between-cross-section and/or between-time differences using parameters of a probability distribution.

Multilevel/Hierarchical Linear Models (MLM/HLM) model the nesting structure of the pooled data, whether time-periods are nested within cross-sections or cross-sections are nested within time periods. This is a fundamental ambiguity of time-series cross-section data and models. We can conceptualize them (1) as *i* cross-sections observed at each of *t* time-periods, or (2) as *t* time-periods observed for each of *i* cross-sections. The former are typically referred to as repeated cross-sections, when surveys on random samples are repeated over time; the latter as time-series cross-section. But these terms describe the perceptions and decisions of the researcher, rather than inherent properties of the data. Our interest and focus is on time – how the past affects the future. Throughout this paper we conceptualize the processes we model as *t* time-periods observed for each of *i* cross-sections. Our focus is on time-series within cross-sections, the parameters of the time-series, and similarities and differences of those parameters between cross-sections.

1.1. Completely Pooled Model

A completely pooled model can be expressed as:

$$(1)\ Y_{ti} = \alpha + \Sigma_k \beta_k X_{kti} + \varepsilon_{ti}$$

where $i = 1, 2, 3, \ldots$ I indexes cross-section; $t = 1, 2, 3, \ldots$ T indexes time; and $k = 0, 1, 2, 3 \ldots$ K indexes independent variables. Y_{ti} is a vector of the dependent variable that varies over cross-section and time; X_{kti} are the *k* independent variables that vary over cross-section and time; β_k are the coefficients on the *k* independent variables; and ε_{ti} are the stochastic errors that vary over cross-section and time.

There are several important aspects of this model. The total number of observations is N=IxT: one observation for each cross-section for each time unit. These are pooled into one homogenous data matrix with no structural distinctions with respect to cross-section or time. The data matrix can be transposed without affecting its statistical properties. β_k will be the same whether the data represent differences over time or between cross-sections (Beck and Katz 2004, fn. 4). As with individual-level survey data, observations are assumed to be equivalent and can thus be combined to estimate the effects of X on Y. The data are assumed to be homogenous, but this assumption is based on sampling design: If all observations are randomly sampled from the same population, they are in fact equivalent. Thus, Beck and Katz (2004) note the importance of homogeneity in the decision to pool. This is a critical assumption if the completely pooled model is used to make inferences about the population of cross-sections over time.

1.2. Fixed-/Random-Effects Model

Whether or not random samples of cross-sections over time are feasible depends on the substantive issues being investigated. Random sampling of observations of elementary schools over several years or even decades poses no insurmountable problems; but similar sampling of observations on market democracies does. The population of such countries is small; and those with consistent over-time observations are smaller still. Hence, we cannot rely on randomization to eliminate cross-national differences, and cross-sections are not equivalent. Time units may not be either, because historical events (such as recession, war, and international trade conventions) are unique and differentiate some historical periods from others. Therefore, we can expect both cross-sectional and over-time heterogeneity in the pooled data.

Furthermore, non-experimental research cannot control all factors that might impact the substantive issues under investigation. These unobserved and/or unmeasured variables are included in the stochastic error ε_{ti} and, if they are correlated with any independent variable, will induce a correlation between the error and the independent variable. This induced correlation will bias all of the parameter estimates. Fixed- and random-effects models statistically control for unobserved/unmeasured differences between cross-sections and/or over time. A generic fixed- and/or random-effects model is written as:

(2) $Y_{ti} = \alpha_i + \delta_t + \Sigma_k \beta_k X_{kti} + \varepsilon_{ti}$

where α_i is the cross-section effects, a vector of dummy variables indicating cross-section *i* (fixed effects), or a draw from a probability distribution (random effects); δ_t is a vector of dummy variables indicating time *t* (fixed effects) or a draw from a probability distribution (random effects); X_{kti} are the *k* independent variables that vary over cross-section and time; β_k are the respective coefficients indicating the effect of X_k on Y; and ε_{ti} are the stochastic errors that vary over both cross-section and time. It is important to note that, although model parameters may vary over cross-sections and/or time, they may be fixed or random. This is a choice the analyst makes. Fixed effects are estimated as fixed values – for example, a separate intercept for each cross-section or time period. Random effects are estimated as moments of a probability distribution (typically the normal distribution). Estimation of the former uses I-1 (or T-1) degrees of freedom. Estimation of the latter, if (as is typical) a normal distribution is assumed, uses two degrees of freedom (one each for the mean and the standard deviation). We prefer to follow Longford (1993) and call parameters that vary over cross-section and/or time variable-parameters, which may be fixed or random.[2]

The model of equation 2 estimates the effect of X_{ti} on Y_{ti} net of α_i and δ_t, that is, net of the effects of X_i on Y_i and X_t on Y_t. The only variation remaining in these data are cross-section effects that differ over time or time effects that are different for different cross-sections. Thus, the only effects X can have on Y are X_i on Y_j ($i \neq j$) and X_t on Y_s ($t \neq s$). All variation among cross-sections, regardless of functional form, is absorbed by α_i; and all variation over time, regardless of functional form, is absorbed by δ_t. The only way to model these differences within the context of this equation 2 is by including the interaction of α_i with δ_t. But this interaction absorbs all variation in the data and can provide no useful substantive information: All data points are fitted. Thus, one may not include the interaction between cross-section fixed effects and time fixed effects.

Substantively, the previous paragraph means that X_k can affect Y only if X_k varies across both cross-section and time, and that the over-time effect of X_k must then differ among cross-sections (or the cross-section effect must differ over time). But this cannot be accomplished within the context of equation 2. What equation 2 can accomplish is simply to test the following hypothesis: Is there an effect of X_k on Y net of stable cross-sectional differences in Y and net of temporal differences in Y that are constant among cross-sections? H_0: $\beta_k = 0$. Rejection of this null hypothesis tells us that X_k does affect Y *and* that this effect differs among cross-sections. Rejection of this null, how-

2 Estimating both cross-section and time fixed effects uses (I-1)+(T-1) degrees of freedom. Estimating both cross-section and time random effects uses four degrees of freedom.

ever, tells us neither what these differences are nor why they occur.

1.3. Multilevel/Hierarchical Linear Model (MLM/HLM)

How the MLM/HLM is written depends on the nesting structure of the data, whether time-periods are nested within cross-sections or cross-sections are nested within time periods. As noted, our interest is in the former, in line with virtually all of the pooled time-series cross-section literature (for an exception, see DiPrete and Grusky 1990).

In general, there are i=1, 2, 3, … I cross-sections. Each cross-section contains data for t=1, 2, 3, … T time periods. Interest is on the relationship between a set of independent variables X_{kti} and a dependent variable Y_{ti}. Suppose that relationship can be written for the i^{th} cross-section:

(3) $Y_{ti} = \alpha_i + \delta_t + \Sigma_k \beta_{ki} X_{kti} + u_{ti}$

Except for the additional *'i'* subscript on β_k, equation 3 is identical to equation 2. Equation 2 is a restricted or constrained version of equation 3 where β_k is constant across cross-sections. That is, the effect of X_k on Y is the same for all cross-sections. This constraint is assumed by all completely pooled and most fixed- and random-effect models. MLM/HLM has merit in that a wider range of models can be estimated, and that a richer set of data generating processes can be tested.[3]

The typical presentations of these models focus on the clustering of observations – on level 1 – within some larger units (geographical, organizational, social, etc.), frequently called *contexts* – or level 2 (see Snijders and Bosker 2011; Raudenbush and Bryk 2002). This implies that observations within contexts tend to be more similar to one another than observations from different contexts. Therefore, if observations are assumed to be independent (as statistical theory does), all variance estimates will be wrong. The impetus for modeling this non-independence is to obtain the correct estimates of variances and covariances. This process is generally applicable to pooled time-series cross-section (where time-series observations are nested within cross-sections) and to repeated cross-sections (where cross-section observations are nested within time periods). We think it more useful to focus on the modeling of the substantive process, rather than the statistical consequences of clustering. The development of MLM/HLM with time clustered within cross-section is the same as the development of MLM/HLM with cross-sections clustered within time (except for the issues involving the difference between time-series and cross-sections). At level 1:

$Y_{ti} = \alpha + \beta X_{ti} + u_{ti}$

This relationship holds for each cross-section *i*, but the model parameters may differ among cross-sections. Thus, the level 2 equations are:[4]

$\alpha = \alpha_i$

$\beta = \beta_i$

Their substitution gives the two-level model with time within cross-section:

$Y_{ti} = \alpha_i + \beta_i X_{ti} + u_{ti}$

This model controls for unobserved variables at level 2 by including the cross-section intercept α_i, which can vary among cross-sections. Cross-section differences in slope β_i are also modeled: The effect of X on Y differs among cross-sections. Here there are two varying parameters. In this case they are fixed effects. There are no stochastic components.

More complex models, such as the random intercept: $\alpha = \alpha_i + \varepsilon_i$ or $\alpha = \alpha_0 + \alpha_1 Z_i + \varepsilon_i$, can, of course, be written. These are random effects because of the stochastic component ε_i. In the second case, Z_i is a level 2 variable that varies among

3 At the limit, each independent variable for each fixed unit has a different effect (β_{ki} is different for every *i*). Beck and Katz (2007) refer to these as unit-specific models: a (perhaps different) model for each cross-section. If unit-specific models include the likely cross-equation correlation structure, these are Seemingly Unrelated Regression models described in econometrics (Judge et al. 1982; Wooldridge 2002).

4 Notation varies among authors. It is important to understand the models and not be wedded to some notation system.

cross-sections, but not time. The cross-section effect in this case is random (because of the stochastic component ε_i) and is a linear function of Z_i.

Consider the following abbreviated model from Messner et al. (2011):

$$\text{(4a) } \mathbf{D(Homrate)}_{ti} = \alpha_i + \delta_t + \beta_1 \mathbf{D(Homrate)}_{t-1,i} + \beta_2 \mathbf{D(Div)}_{ti} + \beta_3 \mathbf{D(LnGDPpc)}_{ti} + u_{ti}$$

This is a fixed-effects model with fixed intercepts for country (α_i) and time (δ_t). These two coefficients control for all variables that vary *only* between countries and between time periods. $\mathbf{D(Homrate)}_{t-1,i}$ is the annual change in the homicide rate of country *i* in year *t-1* (the previous year); $\mathbf{D(Div)}_{ti}$ is the annual change in the divorce rate; and $\mathbf{D(LnGDPpc)}_{ti}$ is the annual change in the log of per-capita income. All variables are measured as annual change because the levels are not stationary (see Raffalovich 1994). The annual changes in divorce rate and per-capita income in country *i* in year *t* have effects β_2 and β_3, respectively. Prediction error for country *i* in year *t* is u_{ti}. If one objects to divorce and/or income affecting homicide contemporaneously, one or both can be lagged by one year (or several) so that they will appear in equation 4a as $\mathbf{D(Div)}_{t-1,i}$ and $\mathbf{(LnGDPpc)}_{t-1,i}$. Note that the effects of change in divorce rate and per-capita income are modeled as constant across countries. In other words, the effect of change in the divorce rate on change in the homicide rate is constrained to be the same for all sampled countries, as is the effect of change in per-capita income. These assumptions of both the completely pooled model (equation 1) and the fixed-/random-effects model (equation 2) may or may not be reasonable, depending on substantive theory and prior research. In any case, they can be tested in the context of MLM/HLM. To do so, estimate equation 4a and also the more general model:

$$\text{(4b) } \mathbf{D(Homrate)}_{ti} = \alpha_i + \delta_t + \beta_{1i} \mathbf{D(Homrate)}_{t-1,i} + \beta_{2i} \mathbf{D(Div)}_{ti} + \beta_{3i} \mathbf{D(LnGDPpc)}_{ti} + u_{ti}$$

Equation 4b differs from equation 4a by the inclusion of the between-country differences in the effects of the lagged homicide rate, the divorce rate, and per-capita income on the country's homicide rate. Twice the difference in the log-likelihood (standard output from MLM/HLM software) has a chi-square distribution with degrees of freedom equal to the difference in the number of model parameters. In contrast to the fixed-/random-effects model (equation 2), the MLM/HLM offers insight into the data generating processes through which exogenous variables affect dependent variables of interest, as well as potential heterogeneity between countries and over time. However, there can be a very large number of parameters to be estimated in equation 4b; and thus some constraints are typically necessary in practice, for example, homogeneity for subsets of countries. Tests for these constraints are widely available (for example, Snijders and Bosker 2011). In the following, we will illustrate these observations using data from Messner et al. (2011).

2. Analysis

All data are from Messner et al. (2011). For the examples below, we excerpted the following variables:

- The national homicide rate $\mathbf{Homrate}_{ti}$, its one-year lag $\mathbf{Homrate}_{t-1,i}$, and the annual change $\mathbf{D(Homrate)}_{ti} = (\mathbf{Homrate}_{ti} - \mathbf{Homrate}_{t-1,i})$;
- Annual change in the homicide rate the previous year $\mathbf{D(Homrate)}_{t-1,i}$;
- Annual change in the national divorce rate $\mathbf{D(Div)}_{ti}$;
- Annual change in the log of national per-capita income $\mathbf{D(LnGDPpc)}_{ti}$.

We then used these data to estimate equations 1 through 4b in models 1 through 4d (see Tables 1 and 2). All estimates were produced using EViews-8, a widely used econometrics program. Other software will produce comparable results.

All models have the following form:

$$\text{(4a) } \mathbf{D(Homrate)}_{ti} = \alpha_i + \delta_t + \beta_1 \mathbf{D(Homrate)}_{t-1,i} + \beta_2 \mathbf{D(Div)}_{ti} + \beta_3 \mathbf{D(LnGDPpc)}_{ti} + u_{ti}$$

The various models are distinguished from one another by the application or removal of constraints on model parameters. For example, model 1, the completely pooled model, constrains $\alpha_i = \alpha$ (no unmeasured time-constant country effects) and $\delta_t = \delta$ (no unmeasured country-constant time effects), and also constrains the effects of prior change in the homicide rate, annual change in divorce rate, and annual change in per-capita income on the annual change in the homicide rate to be the same for all countries. Equation 4b removes the constraint that $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$, and $\beta_{3i} = \beta_3$.

$$\text{(4b) } \mathbf{D(Homrate)}_{ti} = \alpha_i + \delta_t + \beta_{1i}\, \mathbf{D(Homrate)}_{t\text{-}1,i} + \beta_{2i}\, \mathbf{D(Div)}_{ti} + \beta_{3i}\, \mathbf{D(LnGDPpc)}_{ti} + u_{ti}$$

There are N=1,478 country-year observations in these data.[5] The data are *unbalanced*, that is, the number of within-country observations varies among the forty countries. Not surprisingly, more observations are available for the United States (55) and Western Europe (median=53), than for Latin America and the Caribbean (median=44), Eastern and Southern Europe (median=34), and Asia/Other (median=25). We use a Feasible Generalized Least Squares (FGLS) estimator, with country weights and panel-corrected standard errors (PCSE). This estimator weights cases by the inverse of country-specific error variance. More precise estimates are weighted more heavily. The countries with more valid and reliable data are thus weighted more heavily. This might bias parameter estimates towards the more *developed* countries, as opposed to Latin America, Asia, and Africa. We test for regional differences later in the analysis.

Table 1: Pooled GLS estimations (cross-section weights, PCSE)

Dependent variable: D(Homrate)							
Independent variable	1	2a	2b	2c	3a	3b	3c
Common Effects							
Constant	0.0024	−0.0273 ***					
	(.008)	(.008)					
$D(Homrate)_{t-1}$	−0.2587 ***	−0.2696 ***	−0.2703 ***	−0.2803 ***	CSSE[a]	−0.2764 ***	−0.2770 ***
	(.029)	(.030)	(.030)	(.030)		(.030)	(.030)
D(Div)	0.1046 **	0.1035 **	0.0899 **	0.0896 **	0.0916 **	CSSE[a]	0.0843 *
	(.032)	(.032)	(.034)	(.034)	(.034)		(.034)
D(lnGDPpc)	−0.5761 **	−0.5198 **	−0.7996 **	−0.7153 **	−0.7000 **	−0.7805 **	CSSE[a]
	(.190)	(.201)	(.245)	(.262)	(.252)	(.251)	
Cross-section fixed effects							
Constant		no print		no print			
$D(Homrate)_{t-1}$					no print		
D(Div)						no print	
D(lnGDPpc)							no print
Time fixed effects							
Constant			no print	no print	no print	no print	no print
N	1,478	1,478	1,478	1,478	1,478	1,478	1,478
R^2	0.0847	0.1055	0.1442	0.1666	0.2046	0.1743	0.1763
Log likelihood	−617.8251	−600.8200	−575.5559	−555.9083	−522.3551	−550.3864	−546.1144
Log likelihood ratio test							
Restricted model		1	1	1	2b	2b	2b
$-2 \times (LL_R - LL_U)$		34.0102	84.5384	123.8336	106.4016	50.3390	58.8830
Df		39	52	91	39	39	39
P-value		0.6966	0.0029	0.0126	0.0000	0.1055	0.0214

[a] The variable is included as cross-section fixed effects. For the test of between-country differences, see the log likelihood ratio test in the bottom rows.
*** $p < .001$; ** $p < .01$; * $p < .05$

5 Messner et al. (2011) report N = 1,129~1,285, depending on the specific models. We excerpted fewer variables from these models, and thus lost fewer cases to listwise deletion.

When examining estimates, we should keep in mind that the completely pooled model (model 1) assumes that countries are identical in all unmeasured respects. Estimates are presented in the first column of Table 1. Except for the intercept (where the null is not rejected), the null hypothesis of no effect can be rejected at p<.01 for lagged change in the homicide rate, change in the divorce rate, and change in per-capita income. The intercept in a model of change represents the rate of change of the dependent variable (in this case the homicide rate) when the independent variables are zero. Here, that means the rate of change of the homicide rate when there is no change in prior homicide rate, divorce, or per-capita income. Model 1 tells us that if these variables do not change, the homicide rate does not change, except for random variation in the error term, because the intercept is not significantly different from zero.

Prior change in the homicide rate represents the accumulation of the effects of all determinants of current change through *t-1* (the past history of the process). It is included in these models because, following Messner et al. (2011), homicide rates are a function of historical patterns, rather than independently distributed through time. Interpretation of the coefficients of lagged dependent variables depends on how their past is believed to affect their future. We interpret neither the sign nor the magnitude of these effects, and refer to them as the effects of *history*.[6]

The effects of the changes in divorce rate and per-capita income are expected (Messner et al. 2011). There may be other variables – stable between-country differences and/or global changes that affect all countries – that we are unaware of or unable to measure. The independent variables in model 1 vary over both country and time, so may be correlated with these unmeasured variables. If so, the results of model 1 are biased and our inferences may therefore be wrong. Models 2a through 2c control for these unmeasured variables. Results are presented in the second, third, and fourth columns of Table 1. Model 2a in the second column includes country fixed effects; model 2b in the third column includes time fixed effects; and model 2c in the fourth column includes both country and time fixed effects. Our interest is not in the fixed-effects per se, but in the impact of their inclusion on the effects of the independent variables. Also, because there are so many fixed effects (forty countries and fifty-three years) we do not report them in this paper.[7] We do report the results of likelihood ratio tests in the bottom panel of Table 1.This test compares the log-likelihoods (LL) of two nested models: the unrestricted model and the restricted one. The models are identical except that the restricted model imposes a set of restrictions on parameters of the unrestricted model. Model 2a, for example, includes all parameters from model 1, as well as forty cross-section fixed effects (country-specific intercepts). In this example, model 2a is unrestricted. Model 1 imposes the restriction that thirty-nine of these effects are zero (and the remaining effect is the one intercept, which is not restricted). The restricted model is indicated in the bottom panel of Table 1, along with the log-likelihoods of both models, the chi-square statistic of -2 times the difference of log-likelihoods, and the degrees of freedom (the number of restrictions). In the second column we see that the restricted model is model 1, and that minus twice the difference in log-likelihoods is 34.0102, distributed as chi-square with 39 degrees of freedom and associated probability of .6966. We therefore fail to reject the null hypothesis that country fixed effects are jointly zero. Intercepts are the same for all countries.[8]

Time fixed effects (model 2b in the third column), on the other hand, are significantly different from one another: Intercepts vary significantly over time (p=.0029). Comparing the coefficients of lagged homicide, divorce, and per-capita income in models 1 and 2b, we see that they are the same sign, but slightly different in magnitude. These differences, however, are small relative to their standard errors. Substantively, findings from model 2b are the same as from model 1.

6 Estimates of models with lagged dependent variables will be biased if error terms are autocorrelated. Correlograms and Ljung-Box test statistics (Granger and Newbold 1986) indicate no significant autocorrelation.

7 These estimates are available on request.

8 We also estimated random between-country effects. Consistent with our findings for fixed-effects, random effects were zero.

The results show that the independent variables, which vary over both country and time, have significant effects net of country and time. This implies that the association between independent and dependent variables involves country-time interaction. We cannot estimate a country-time interaction within the context of model 2 because this $\alpha_i\delta_t$ interaction consumes all degrees of freedom and perfectly fits all data points. We can, however, fit a model where the over-time effect differs among countries. This cannot be done within the context of model 2 because just as with model 1, model 2 constrains model parameters to be the same for all countries. We have noted that model 2b has many parameters (53 just for the time fixed effects), and model 2c has an additional 40, preventing inclusion of either in Table 1. If we relax the assumption that the effects of independent variables are the same for all countries, the number of parameters to be estimated increases. For each of the three variables in the model, we would need an additional 39 coefficient estimates and estimated standard errors. Are all of these necessary to adequately represent the data generating processes? For current purposes, we define "adequate" as the absence of both redundant and omitted variables. A variable is redundant if the null hypothesis of no effect is not rejected. A variable is omitted if, despite theoretical and/or empirical evidence of its importance, it is not included in the estimated model. So redundancy is data-based and omission is theory-based. Fixed-effects were included, for example, because both the theoretical and statistical cases for inclusion were strong. To evaluate redundancy, we estimated a model with fixed effects (models 2a to 2c) and without them (model 1), and compared the likelihood ratios. We found that country fixed effects were redundant, but time fixed-effects were not.

Model 3 relaxes the assumption that the effects of the lagged change in homicide rate, change in the divorce rate, and change in per-capita income are the same for all countries. These results are presented in column 5 for lagged homicide (model 3a), column 6 for divorce (model 3b), and column 7 for per-capita income (model 3c). We do not present estimates for each country:[9] there are 40 different coefficients for each variable. We do present the results of hypothesis tests in the lower panel of Table 1. Like the models with fixed-country and fixed-year effects, these are also likelihood ratio tests. The null hypothesis is that the 40 country-specific coefficients are redundant; the alternative is that some (one or more) are not redundant. The variable being tested is indicated by "CSSE" (Cross-Section Specific Estimate) in the top panel. These are the between-country differences in this effect. The same null hypotheses are tested for between-country differences in the effect of change in divorce and in per-capita income. For these two variables, the null of no difference is not rejected: divorce at the .1 level and per-capita income at the .01 level.[10]

9 These are available on request.

10 The null would be rejected at the .05 level; and we would conclude that there were between-country differences in the effect of the change in per-capita income on the change in homicide. But consider that we are testing many hypotheses. At the .05 significance level, one out of twenty null hypotheses would be wrongly rejected (see Raffalovich et al. 2008). Therefore, we employ a more stringent threshold.

Table 2: Pooled GLS estimations (cross-section weights, PCSE) with regional dummies and interactions

Dependent variable: D(Homrate)					
Independent variable	2b	4a	4b	4c	4d
Common Effects					
D(Homrate)$_{t-1}$	−0.2703 ***	−0.2705 ***	−0.4142 ***	−0.2683 ***	−0.2698 ***
	(.030)	(.030)	(.049)	(.030)	(.030)
D(Div)	0.0899 **	0.0903 **	0.0951 **	0.0124	0.0916 **
	(.034)	(.034)	(.034)	(.057)	(.034)
D(lnGDPpc)	−0.7996 **	−0.7085 **	−0.6707 **	−0.6959 **	−0.0961
	(.245)	(.257)	(.254)	(.262)	(.458)
Northern Europe		*Reference*	*Reference*	*Reference*	*Reference*
Anglo America, UK, Oceania		0.0095	0.0074	0.0061	0.0101
		(.017)	(.017)	(.017)	(.023)
x D(Homrate)$_{t-1}$			0.1356		
			(.075)		
x D(Div)				0.0815	
				(.074)	
x D(lnGDPpc)					−0.0223
					(.737)
Latin America and Caribbean		−0.0243	−0.0270	−0.0284	−0.0114
		(.063)	(.063)	(.064)	(.067)
x D(Homrate)$_{t-1}$			0.1575 *		
			(.070)		
x D(Div)				0.1128	
				(.288)	
x D(lnGDPpc)					−0.5968
					(1.452)
E/S Europe		0.0064	0.0056	−0.0004	0.0297
		(.014)	(.014)	(.014)	(.019)
x D(Homrate)$_{t-1}$			0.2467 ***		
			(.068)		
x D(Div)				0.2054 *	
				(.096)	
x D(lnGDPpc)					−1.0075
					(.580)
Asia and Other		−0.0142	−0.0093	−0.0203	0.0149
		(.018)	(.017)	(.019)	(.025)
x D(Homrate)$_{t-1}$			0.3922 ***		
			(.110)		
x D(Div)				0.2044	
				(.254)	
x D(lnGDPpc)					−0.9905
					(.589)
Time fixed effects					
Constant	no print	no print	no print	no print	no print
N	1,478	1,478	1,478	1,478	1,478
R^2	0.1442	0.1450	0.1584	0.1484	0.1478
Log likelihood	−575.5559	−574.3539	−563.9780	−571.7798	−571.4492
Log likelihood ratio test					
Restricted model		2b	2b	2b	2b
$-2 \times (LL_R - LL_U)$		2.4040	23.1558	7.5522	8.2134
Df		4	8	8	8
P-value		0.6619	0.0032	0.4784	0.4129

*** $p < .001$; ** $p < .01$; * $p < .05$

Note: For the regions and countries, see Appendix.

Ideally, we would want the estimates of the between-country differences (in divorce, for example) so that we could investigate the reasons for these differences, obtain the measures of potential explanations, and test the hypotheses regarding these explanations. For each variable, however, there are 40 parameters to be estimated; for three variables, there are 120; period fixed effects add another 53. Estimating 173 parameters with almost 1,500 cases is not an insurmountable problem. Interpreting those estimates may well be, however, unless they describe a very simple pattern (for example monotonic). To reduce the complexity of this process, we aggregated the forty countries into the five regions defined by Messner et al. (2011). Thus, we further test in models 4a through 4d the regional differences, because regional differences may be more important than between-country differences.

Regional differences in the effects of lagged change in homicide, change in divorce, and change in per-capita income are estimated in Table 2 (see models 4a to 4d). Model 4a tests whether there are regional differences in the rate of change of homicide, net of divorce rates and per-capita income. Earlier, we found no between-country differences; so the finding of no regional differences (Chi-square with 4 df = 2.404, p>.5) is not surprising. Models 3a to 3c tested between-country differences in the effects of change in lagged homicide, divorce, and per-capita income on homicide change. Only the lagged dependent variable – *history* – was found to differ in effect among countries (model 3a, p<.01). Because there were forty countries in these data, and thus forty coefficients for the effect of the lagged dependent variable, specific country differences were hard to interpret, especially in the absence of strong theory and specific hypotheses. Model 4 is a little easier to interpret. First, as we expect from model 3a, only the effect of lagged homicide in model 4b displays regional differences (p<.001). Second, those differences are between Northern Europe (the reference region) and both Eastern/Southern Europe and Asia/Other. Why the historical patterns of homicide would have differential regional impact on annual change in contemporary homicide rates is a topic for future research.

3. Discussion

In this paper we have presented and discussed several models for the analysis of pooled time-series cross-section data, then illustrated these models with data from published research on homicide rates in a sample of forty countries over an average of more than forty years per country. Throughout, our focus has been on within-country time-series and between-country differences in time-series parameters.

The models we discuss range from completely pooled to regionally disaggregated. Completely pooled models require that sampled cross-sections be drawn from a population of equivalent cross-sections so that parameters do not vary among cross-sections and data can be combined to yield more precise estimates of common coefficients. This applies to measured cross-sectional differences, unmeasured cross-sectional differences (error variances and covariances), and to time-series processes within cross-sections. The advantage of pooling is this combining of information. More cross-sections in a sample means larger sample sizes; larger samples have smaller sampling error; and smaller sampling error means more precise parameter estimates. The major threats to validity are that the pooled cross-sections are not from the same population and that causal processes differ among cross-sections. The ability to avoid these threats depends, of course, on sampling design (thus attention to methodology must be emphasized); but frequently researchers rely on secondary data in which case sampling design is not under their control. In the absence of random sampling, a difficult achievement in many research contexts, cross-section and time homogeneity should not simply be assumed. Instead, researchers should test these assumptions. We suggest the likelihood-ratio test within the context of a fixed-effects multilevel statistical model as one useful method for testing heterogeneity in pooled models (see Snijders and Bosker 2011; Raudenbush and Bryk 2002). We do not discuss random effects models other than to note that they are an alternative to fixed-effects models in controlling for unmeasured heterogeneity. We are skeptical of these models because inference is to the population from which the data are sampled, whereas inference in fixed-effects models is conditional on the data in the sample. With TSCS data, the population is vaguely defined and sampling is typically by convenience. Also, random-effects models estimate moments of probability distributions, and this requires comparatively large samples to obtain reasonable estimates (but see Beck and Katz 2007). Fixed-effects models are more appropriate for the TSCS data analyzed here (Beck 2001).

Using the data from Messner et al. (2011), we first estimate a simple model (model 1) that assumes homogeneity with respect to cross-section and time. To rule out cross-sectional or over-time heterogeneity in the pooled data, which may be correlated with unobserved variables, we test this assumption. Thus, in model 2a through 2c, we relax this homogeneity assumption and include country fixed-effects (model 2a), time fixed-effects (model 2b), and both country and time fixed-effects (model 2c). The likelihood-ratio tests (between model 1 and models 2a to 2c) show that the time fixed-effects significantly improve the model. Thus, the later models (models 3a to 4d) include time fixed-effects, with model 2b serving as the restricted model for their likelihood-ratio tests. Next, we test the assumption that the effects of the predetermined variables are the same across the countries. The likelihood-ratio tests (between model 2b and models 3a to 3c) show that the country-specific effects of the lagged dependent variable differ among countries, whereas the effects of the exogenous variables are the same. Therefore, model 3a is the most appropriate of models 1 through 3c.

Model 3a tells us that countries differ in the impact of historical patterns of homicide on current patterns. Interpretation of these differences is problematic because of the large number of estimated parameters, and thus the large number of comparisons that need to be made. We simplify this task by aggregating countries into geographic regions, then testing for regional differences (model 4a to 4d). We find that, like models 3a to 3c, only the impact of historical patterns of homicide differs among regions. Specifically, the Eastern and Southern European region differs from the Anglo-American and Northern European regions, as does Asia and Other in this respect. The Latin America and Caribbean region does not differ from the reference region.

The sequence of models presented and illustrated suggest three important conclusions. First, country effects are redundant, net of change in divorce rate, change in per-capita income, and historical patterns of homicide rate change: The sample is homogenous with respect to stable between-country characteristics. Second, net of those same variables, time-effects are not redundant: The sample is not homogenous with respect to stable over-time differences, and statistical analysis of the pooled data must control for over-time heterogeneity. Third, the significance of independent variables that vary over both country and time implies country-time interaction. The effects of one or more independent variables must differ among countries. The sample is not homogenous with respect to causal processes. Statistical analysis of the pooled data must control for heterogeneity of causal processes.

The analysis of the homicide data shows that change in the divorce rate has a positive impact on the change in homicide rate, and that change in per-capita income has a negative effect. These findings are consistent with the research literature on homicide (for a literature review, see Messner et al. 2011). In addition, we find that these effects are constant across the countries in our data. This substantively important information is not obtainable from the analysis of the completely pooled or fixed-effects models of equation 1 or 2. Also unobtainable from completely pooled or fixed-effect models is the substantively important finding that the impact of history is not constant but varies among countries and regions. Methodologically, this information regarding homogeneity and heterogeneity is critically important to recognize and counter the threats that heterogeneity poses to validity. Heterogeneity is inherent in TSCS data, but not self-evident. Researchers must be diligent and test it.

References

Adolph, Christopher, Daniel M. Butler, and Sven E. Wilson. 2005. Which Time-Series Cross-Section Estimator Should I Use Now? Guidance from Monte Carlo Experiments. Paper presented at the 2005 Annual Meeting of the American Political Science Association, Washington, D.C.

Anderson, Theodore W., and Cheng Hsiao. 1982. Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics* 18 (1): 47–82.

Anderson, Theodore W., and Cheng Hsiao. 1981. Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association* 76 (375): 598–606.

Beck, Nathaniel, and Jonathan N. Katz. 1995. What to Do (and Not to Do) with Time-Series Cross-Section Data. *American Political Science Review* 89:634–47.

Beck, Nathaniel. 2001. Time-Series–Cross-Section Data: What Have We Learned in the Past Few Years? *Annual Review of Political Science* 4:271–93.

Beck, Nathaniel, and Jonathan N. Katz. 2004. Time-Series–Cross-Section Issues: Dynamics, 2004: Draft of July 24, 2004. Paper presented at the Annual Meeting of the Society for Political Methodology, Stanford University.

Beck, Nathaniel, and Jonathan N. Katz. 2007. Random Coefficient Models for Time-Series–Cross-Section Data: Monte Carlo Experiments. *Political Analysis* 15:182–95.

DiPrete, Thomas A., and David B. Grusky. 1990. Structure and Trend in the Process of Stratification for American Men and Women. *American Journal of Sociology* 96 (1): 107–43.

Granger, Clive W. J., and Paul Newbold. 1986. *Forecasting Economic Time Series*, 2d ed. Orlando, FL: Florida Academic Press.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee. 1982. *Introduction to the Theory and Practice of Econometrics*. New York: Wiley.

Longford, Nicholas T. 1993. *Random Coefficient Models*. Oxford: Clarendon.

Messner, Steven F., Benjamin Pearson-Nelson, Lawrence E. Raffalovich, and Zachary Miner. 2011. Cross-National Homicide Trends in the Latter Decades of the 20th Century: Losses and Gains in Institutional Control? In *Control of Violence: Historical and International Perspectives on Violence in Modern Societies*, ed. Wilhelm Heitmeyer, Heinz-Gerhard Haupt, Stefan Malthaner, and Andrea Kirschner, 65–89, New York: Springer.

Raffalovich, Lawrence E. 1994. Detrending Time Series: A Cautionary Note. *Sociological Methods and Research* 22:492–519.

Raffalovich, Lawrence E., Glenn D. Deane, David Armstrong, and Hui-Shien Tsao. 2008. Model Selection Procedures in Social Research: Monte-Carlo Simulation Results. *Journal of Applied Statistics* 35:1093–1114.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Advanced Quantitative Techniques in the Social Sciences Series 1, 2d ed. Thousand Oaks, CA: Sage.

Snijders, Tom A. B., and Roel J. Bosker. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2d ed. London: Sage.

Wawro, Gregory. 2002. Estimating Dynamic Panel Models in Political Science. *Political Analysis* 10 (1): 25–48.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: MIT Press.

Appendix: Regions and Countries

1. Anglo-America/U.K./Oceania (5)

Canada, United States, United Kingdom, Australia, New Zealand

2. Latin America and Caribbean (9)

Costa Rica, Mexico, Trinidad and Tobago, Uruguay, Venezuela, Dominican Republic, El Salvador, Nicaragua, Panama

3. Eastern/Southern Europe (12)

Austria, Bulgaria, France, Greece, Hungary, Italy, Poland, Portugal, Spain, Switzerland, Czech Republic, Estonia

4. Northern Europe (9)

Denmark, Finland, Germany, Iceland, Netherlands, Norway, Sweden, Belgium, Luxembour

5. Asia and Other (5)

Israel, Mauritius, Singapore, Japan, Thailand

Lawrence E. Raffalovich
lraffalovich@albany.edu

Rakkoo Chung
rchung@albany.edu

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014249
IJCV: Vol. 8 (2) 2014

The Analysis of Non-Stationary Pooled Time Series Cross-Section Data

Christoph Birkel, Federal Criminal Police Office (BKA), Wiesbaden, Germany

Vol. 8 (2) 2014

ISSN: 1864–1385

The Analysis of Non-Stationary Pooled Time Series Cross-Section Data

Christoph Birkel, Federal Criminal Police Office (BKA), Wiesbaden, Germany

It is common in macro-level research on violent crime to analyze datasets combining a cross-section (N units) with a time-series (T periods) dimension. A large body of methodological literature accumulated since the 1990s raises questions regarding the validity of conventional models for such Pooled Time Series Cross-Section (PTCS) data in the presence of non-stationarity (stochastic trends). Extant research shows that conventional techniques lead to consistent estimates only under specific conditions, and standard procedures for statistical inference do not apply. The approaches proposed in the literature to test for stochastic trends and cointegration (see the introduction to this issue) are reviewed, as well as methods for estimation and inference in the non-stationary PTCS context. A host of procedures have been developed, including methods to take cross-section dependence and/or structural breaks simultaneously into account. Thus all the tools needed for valid analyses of non-stationary PTCS data are now available, although many of them need large samples to perform well. The general approach to the analysis of non-stationary PTCS data is illustrated using a data set with robbery rates for eleven West German federal states 1971–2004. Several meaningful long-run relationships are identified and estimated.

It is quite common in macro-level research on violent crime to analyze datasets combining a cross-section (N units) with a time-series (T periods) dimension, where the number of observation periods is approximately as large as or even larger than the number of units. Examples include analyses of the effect of imprisonment rates on violent crime rates at the level of the US federal states (Vieraitis, Kovandzic, and Marvell 2007) and studies on the effect of economic inequality on murder rates among developed countries (Jacobs and Richardson 2008).[1] While such a design is, in principle, quite powerful, a large body of methodological research accumulated since the 1990s raises questions regarding the validity of conventional models for such Pooled Time-Series Cross-Section (PTCS) data in the presence of stochastic trends.[2] At the same time, it is known from single time-series analyses that violent crime rates are, in fact, often non-stationary (for example Hale 1998). This issue is often dealt with by estimating models in first differences (for example, Entorf and Winker 2005). The cost of this approach is a loss of information regarding long-run relationships. But there are some exceptions: One example is a paper by Entorf and Spengler (2000), who analyze crime rates for eleven German federal states. The most advanced application of non-stationary PTCS methods is the recent analysis of Swedish crime rates at county level by Blomquist and Westerlund (2014).

1 To be sure, micro-level data-sets might also have the property that T N, so that the discussion given here would also apply to them. But more often, N is much larger than T for micro data, and it is reasonable to treat T as fixed. In this case, different analytic results apply (Bond, Nauges, and Windmeijer 2005), which are not the topic of the present paper. Suffice to say – with respect to the popular Arellano-Bond dynamic panel models – that the first-differenced GMM estimator performs poorly under non-stationarity, which is not necessarily (but sometimes) the case for system-GMM estimators (Blundell and Bond 1998; Han and Phillips 2010; Binder, Hsiao, and Pesaran 2003). Furthermore, non-stationarity might also be an issue in the analysis of count data as well as binary and categorical variables (and, more generally, every kind of longitudinal analysis), but this is a question far beyond the scope of this paper.

2 An overview is given in Breitung and Pesaran (2008). In what follows, the term “panel” is often used synonymously for PTCS data, as is done in the econometric literature.

The purpose of this paper is to describe the consequences of non-stationarity for conventional PTCS analyses and to give a non-technical overview of approaches to estimation and inference for non-stationary PTCS data. Since research has shown that non-stationary PTCS methods are sensitive with respect to cross-section-dependence (which might be due to common shocks, common latent factors driving the individual series, or spatial autocorrelation) and structural breaks (such as shifts in the mean of the series), these problems are also dealt with. In section 1, I review analytic results and evidence from simulation studies regarding the behavior of conventional estimators for PTCS data in the presence of non-stationarity. Next, testing for unit roots (2.) and cointegration (3.) is discussed. In section 4, approaches to the estimation of long-run relationships are presented. An example using PTCS data for the West German federal states for 1971–2004 follows (5.). A brief discussion concludes (6.). Throughout the paper it is assumed that the reader has studied the introduction to cointegration and error-correction modelling by Helmut Thome in this issue.

1. Properties of OLS and Fixed-Effect Regressions with Non-Stationary Panel Data

First, I would like to summarize some analytic results that are corroborated by simulation evidence (Entorf 1997; Kao and Chiang 2000; Chen, McCoskey, and Kao 1999; Coakley, Fuertes, and Smith 2001; Urbain and Westerlund 2011) regarding standard estimators in this situation. The behavior of panel estimators under non-stationarity depends on the cointegration properties of the variables and the degree of homogeneity of long-run relationships. Several cases have been studied:

1.1. No Cointegration

In the first scenario, there is no cointegration between the left-hand side variable and the regressors: this is the classical “spurious regression” case. In contrast to single time-series analysis, in the PTCS context, there might be a long-run relationship between the variables that can be consistently estimated even when the residuals are non-stationary: this is what Phillips and Moon (1999) call the “long-run average regression coefficient”. This result is due to the fact that the pooling of series for several units attenuates noise, which restores the consistency of standard estimators. Thus, if there is no long-run relationship between two random walks, the panel estimator will converge to zero as T and N approach infinity in sequence. This consistency property is shown by Phillips and Moon (1999, 2000) for simple OLS regressions with driftless random walks without intercepts and heterogeneous, randomly varying long-run relationships.

This finding for heterogeneous (that is, unit-specific) long-run relationships also holds for regressions with detrended data.[3] Furthermore, it also pertains to the fixed-effect (FE) estimator when applied to pure random walks (Phillips and Moon 1999, 1090). An exception is the case of cross-section dependencies in both left-hand side and right-hand side variables due to common non-stationary factors (see 2.2.1. below) that are cointegrated across units (Urbain and Westerlund 2011, 124). Here, the OLS estimator behaves as in the classical “spurious regression” case in the analysis of time series for a single unit. Additionally, for FE regressions with drifting random walks and homogeneous coefficients, results mirror those obtained in the single time-series case, that is, the coefficient obtained is a consistent estimate of the ratio of the drift parameters (Entorf 1997).

For each of the situations where “spurious regression” does not occur, the distribution of the estimator is normal, but the variance depends on the specific scenario and cannot be estimated by the usual formulae. Thus, conventional t-tests can be highly misleading, as shown by the simulations of Kao (1999), for example.

1.2. The Cointegrated Case

If a cointegration relationship exists for all units, standard OLS and FE estimators are also consistent estimators of the long-run average relationship (which is not the average of

3 That is, they have been purged from a deterministic time trend by regressing them on a time index.

the cointegration parameters, if they vary across units).[4] Nonetheless, small-sample biases exist if the regressors are not strongly exogenous (in the sense that the error term of the regression is not correlated with contemporaneous or past random shocks on the independent variable), which is often the case; further biases arise if there is residual serial correlation which varies across units (Pedroni 2000; Phillips and Moon 1999; Phillips and Moon 2000; Kao and Chiang 2000). These results presume cross-section independence; if there are common stationary or non-stationary factors in the residuals, the FE estimator for homogeneous cointegration parameters is biased (Bai and Kao 2006; Bai, Kao, and Ng 2009). If there are cross-section dependencies due to common non-stationary factors in the dependent as well as the independent variables, on the other hand, the (homogeneous) cointegration parameters can be consistently estimated via OLS or FE regression (Urbain and Westerlund 2011).

The variance of the estimators depends on several features: the degree of homogeneity of (true) long-run relationships, the specification of deterministic components (unit-specific intercepts and/or time trends), and – in case of a common factor structure – the factor loadings (Phillips and Moon 1999; Kao and Chiang 2000; Baltagi, Kao, and Chiang 2000; Urbain and Westerlund 2011; Bai and Kao 2006). In any case, standard errors cannot be estimated in the usual way.

The bottom line of the research reviewed here is threefold: First, when the data are non-stationary, standard OLS or FE regressions produce consistent parameter estimates only under very specific circumstances; second, the properties of estimators are dependent on the presence or absence of cointegration; third, conventional estimates for standard errors as routinely reported by standard statistical software do not allow valid inference under non-stationarity. Therefore, for proper inference it is necessary to ascertain if the series are non-stationary, and if so, if they are cointegrated. Depending on the results of these preliminary analyses, appropriate models have to be estimated.

2. Unit-Root Tests

2.1. Unit-Root Tests Requiring Cross-Section Independence

The first unit-root tests for PTCS data proposed in the literature presuppose that the observations for unit i are not correlated with those for unit j. If this assumption is violated, these tests often exhibit a rate of alpha errors far above the specified nominal level, as shown in Monte Carlo simulations (see below).

2.1.1. ADF-Type Tests

There are three well-known adaptations of the Augmented Dickey-Fuller (ADF) test for single time series to the PTCS context: those of Levin, Lin, and Chu (2002, in the following referred to as LL), Breitung (2000), and Im, Pesaran, and Shin (2003; IPS). All these tests are based on estimating the parameters of an equation of the form

(1)[5] $$\Delta y_{it} = \alpha_{i0} + \alpha_{1i} t + \rho_i y_{it-1} + \sum_{p=1}^{P} \omega_{ip} \Delta y_{it-p} + e_{it}$$

After running (1) using OLS, the null-hypothesis i = 0 is tested, which implies non-stationarity of the form of a stochastic trend. Inference on i is complicated by the PTCS structure of the data. To account for this, all tests employ complicated computations involving several steps, resulting in test statistics which asymptotically obey the standard normal distribution.

Besides computational details, which I will not discuss here, the three tests differ in the formulation of the alternative hypothesis: In the tests proposed by Levin, Lin, and

4 This is due to the fact that the former is the *ratio* of the expectation of the long-run covariance to the expectation of the long-run variance of the regressor, while the latter is the *expectation of the ratio* of the long-run covariance to the long-run variance. In general, $E(x)/E(y) \neq E(x/y)$.

5 α_{i0} and $\alpha_{1i}t$ are optional, depending on assumptions regarding the alternative hypothesis (see section 3 of the introduction to this issue). Lagged values of Δ_{yt} are added – if necessary – as regressors to ensure that the residuals are not serially correlated. The formulation of the test equation in first differences chosen here is equivalent to the formulation in levels presented in the introductory essay by Helmut Thome in this issue: This can be seen by rearranging the most simple form of (1), without intercept, time trend, and lagged values of Δ_{yt}. For this purpose, ρ_i in (1) is designated ρ^*_i here. Then $\Delta y_{it} = \rho^*_i y_{it-1} + e_{it} \rightarrow y_{it} - y_{it-1} = \rho^*_i y_{it-1} + e_{it} \rightarrow y_{it} = \rho^*_i y_{it-1} + e_{it} + y_{it-1} \rightarrow y_{it} = (1 + \rho^*_i) y_{it-1} + e_{it}$. The expression in brackets might be combined to ρ_i: $(1 + \rho^*_i) = \rho_i$; then the equation reads $y_{it} = \rho_i y_{it-1} + e_{it}$, the PTCS-analogue to (3) in the introduction. When $\rho_i = 1$, then $\rho^*_i = 1 - \rho_i = 1 – 1 = 0$. Thus, testing $\rho^*_i = 0$ in the original equation with Δ_{yt} on the left side is equivalent to testing $\rho_i = 1$ in the rearranged equation.

Chu (2002) and Breitung (2000), the alternative hypothesis holds that data for *all* units follow identical stationary ($\rho_{i=1} = \rho_{i=2} = \ldots = \rho_{i=N} < 0$) or trend-stationary ($\rho_{i=1} = \rho_{i=2} = \ldots = \rho_{i=N} < 0$, $\alpha_{1i} \neq 0$ for all i) autoregressive processes. But the null hypothesis might also be wrong in other cases: for example, if *some,* but not *all* units exhibit stationarity; or if all series are stationary, but follow heterogeneous autoregressive processes. In contrast, the alternative hypothesis of Im, Pesaran, and Shin (2003) explicitly allows for heterogeneity by requiring only at least one series to be stationary and assuming that the autoregressive properties of the stationary series might vary.

2.1.2. Combination Tests

Maddala and Wu (1999) proposed to combine the p-values of individual unit-root tests into a single test statistic using meta-analytic methods. They utilize the fact that such combinations follow a well-defined distribution. Specifically, the following test statistic is computed:

$$(2),\quad \lambda = -2\sum_{i=1}^{N} \ln(\pi_i)$$

where π_i is the significance level of a unit-root test for unit i. Any unit-root test for single time-series might be used for this procedure. For fixed N, λ follows asymptotically (as $T\rightarrow\infty$) a χ^2-distribution with 2N degrees of freedom. If the number of cross-sections is large, it is advisable to use modified test statistics (called P_m and Z) developed by Choi (2001), which are valid under an asymptotic theory which assumes that N also approaches infinity (but slower than T). The exact formulation of the null and the alternative hypothesis depends on the unit-root test chosen, but generally, the significance of the test statistic implies that at least one unit is stationary.

2.1.3. Testing the Null of Stationarity

There might be situations in which it is attractive to view stationarity or trend-stationarity as the null hypothesis. Hadri (2000) proposes such a test. Here, the series are decomposed into deterministic components, a random walk, and a stationary white noise error term.[6] When the series are stationary, the variance of the non-stationary component σ_u will be zero. Therefore, the ratio of the variances of the random-walk and stationary white noise components (σ_u and σ_e) will also be zero – this is the null hypothesis of Hadris tests, implying that the data for all units are generated by a stationary or trend-stationary process. The alternative hypothesis holds that the ratio is greater than zero, meaning that all units exhibit unit-root processes. Hadri proposes two test statistics that are constructed using estimates of σ^2_u and σ^2_e, one under the assumption of homogeneous variances and one for the more realistic case of heterogeneous variances. These are based on regressions of the series on an intercept or an intercept and a time trend. For the case of serially correlated errors in these regressions, Hadri suggests using non-parametric heteroscedasticity- and autocorrelation-consistent variance estimators.

2.2. Panel Unit-Root Tests for PTCS-Data Exhibiting Cross-Sectional Correlation

Cross-section dependence affects the size of panel unit-root tests (see below).[7] Subtracting the period-specific mean from the data before applying a unit-root test removes cross-section correlation only if it is due to one common component (such as a common shock) with exactly identical influence on all units, which is not very plausible in most cases. Thus, panel unit-root tests have to be modified to take cross-section dependence explicitly into account – these are the so-called "second generation unit-root tests" (Hurlin and Mignon 2004).

2.2.1. The Common Factor Approach

One approach assumes that a common factor structure is the source of cross-section dependence: Here, the data are assumed to be generated by the following process:

$$(3),\quad y_{it} = \alpha_i + \beta_i t + \boldsymbol{\lambda_i}'\mathbf{F}_t + e_{it}$$

6 With "white noise" I refer to a series free of autocorrelation; in other words the observation for period t is not correlated with prior observations.

7 With "size" I refer to the rate of alpha errors of a test. A test is said to have a correct size if the rate of alpha errors corresponds to the nominal significance level chosen. In other words the test should wrongly reject the null hypothesis in at most 5 percent of cases if a significance level of 5 percent is chosen. Otherwise, I speak of "size distortions." If the actual rate of alpha errors is higher, for example, a test is said to be "oversized."

where $\mathbf{F_t}$ is a vector of common factors and $\boldsymbol{\lambda}_i$ a vector of associated factor loadings; such a common factor might be a national trend driving regional rates of violent crime, for example.[8] Thus, the data consist of deterministic components (intercept and time trend), common factors, and an idiosyncratic part e_{it}. The common factors might be stationary or non-stationary.

The most simple approach for this set-up are extensions of Maddala-Wu-type tests (designated $C\tilde{P}$ and $C\tilde{Z}$) as well as the IPS test (CIPS and CIPS*), developed by M. Hashem Pesaran, where it is assumed that there is one common stationary factor. In a first-step "cross-sectionally augmented Dickey-Fuller-Test" (CADF) this common factor is approximated by the lagged period-specific cross-section mean (Pesaran 2007). There is also a proposal to extend the approach to the case of several common stationary factors (Pesaran, Smith, and Yamagata 2008). Here, lagged period-specific cross-section means of other non-stationary variables which contain the same common factors as the variable of interest are also entered in the test equation.

A more explicit modelling of one or more stationary or non-stationary common factors has been proposed by Jushan Bai and Serena Ng in their "Panel Analysis of Non-stationarity in Idiosyncratic and Common Components" (PANIC) approach (Bai and Ng 2004). The procedure of Moon and Perron (2004) is very similar. Here, the common factors and the factor loadings are estimated. These estimates are then used for different purposes: Bai and Ng develop methods for separately testing the estimated common factors and idiosyncratic parts for unit-roots, while Moon and Perron are interested only in the behavior of the individual-specific component, and therefore apply a unit-root test to the defactored data. Thus, only the approach of Bai and Ng is able to detect non-stationarity if it is due to common factors; the tests of Moon and Perron as well as those of Pesaran will wrongly reject the hypothesis of a unit-root here, because the non-stationary common factor is removed before applying the test.

2.2.2. The Bootstrapping Approach

A more general alternative which does not presume a specific source of cross-section dependence is to conduct statistical inference based on empirical critical values obtained via bootstrapping. Chang (2004), for example, proposes a resampling scheme that preserves the cross-section correlation structure as well as the autoregressive properties of the residuals. Chang's approach is based on the assumption that the cross-section correlation is due to spatial dependencies in stationary components of the series.

Palm, Smeekes, and Urbain (2011), in contrast, develop a very general bootstrapping approach for simplified versions of the LL and IPS tests, which assumes a data-generating process as in (3), where the factors might be stationary or have a unit-root. Also, the approach also applies if there are (non-contemporaneous) dynamic dependencies between the idiosyncratic components e_{it}, if there is a correlation between e_{it} and e_{jt-1}, for example.

2.3. Structural Breaks

As in single time-series analysis, the power of panel unit-root tests is reduced by breaks (Im, Lee, and Tieslau 2005; Sethrapramote 2004).

There are several proposals for panel unit-root tests that take structural shifts into account.[9] Im, Lee, and Tieslau (2005), for example, assume a break in the form of a one-time shift of the mean, possibly at a different date for each unit. The advantage of the proposed test is that its distribution is invariant to the break date. The authors also consider a procedure to identify the break dates if they are not known – with the drawback that the distribution of the test statistic is then no longer invariant to the break date (Sethrapramote 2004, 42). In reaction to the latter problem, several authors have modified this approach (Tam 2006; Westerlund 2006a), suggesting alternative procedures for the determination of the break dates and bootstrapping procedures for the case of cross-section correlation. Tam also considers the case of a possible shift of the time trend

8 Blomquist and Westerlund (2014) report evidence for such a nationwide trend in Swedish county-level rates of property crime.

9 Besides the contributions reviewed here, see also Murray and Papell (2000); Jönsson (2005); Breitung and Candelon (2005); Harris, Leybourne, and McCabe (2005).

(if the data are assumed to have a time trend under the alternative hypothesis).

There are also suggestions for testing the null hypothesis of trend stationarity, possibly with a break: The approaches of Carrion-i-Silvestre, Barrio-Castro, and López-Bazo (2005) and Hadri and Rao (2008) are extensions of Hadri's stationarity test, augmenting the test equation with break-dummies to account for shifts in the mean and/or the time trend. In both papers, analytic results allowing the computation of the asymptotic expectations and variances of the test statistics – which are dependent on the break date here – are provided, so that it is possible to implement an appropriate test. If the break points are not known, they are determined empirically using methods similar to those suggested by Tam. Furthermore, Hadri and Rao propose a modified test based on the sum of two test statistics computed before and after the break, the distribution of which is not dependent on the breaks. Here it is assumed that the break consists of a mean shift only or a combined shift of the mean and the trend slope. For the case of cross-section dependence, the authors suggest bootstrapping procedures.

2.4. Small Sample Properties

Numerous simulation studies on the finite sample behavior of panel unit-root tests have been published (see the papers cited above and Hlouskova and Wagner 2006; Banerjee, Marcellino, and Osbat n.d.; O'Connell 1998; Maddala and Wu 1999; Gengenbach, Palm, and Urbain 2004; Gutierrez n.d.; Baltagi, Bresson, and Pirotte 2007; Sethrapramote 2004; Westerlund and Breitung 2013), but their results are, due to the variety of setups used, difficult to compare. Nonetheless it emerges that the size of panel unit-root tests is sensitive to first-order moving-average errors,[10] and tends to be distorted if the number of cross-sections is large compared to the time dimension. For small samples, for example with 10 cross-sections and 25 periods, power is generally modest, especially in the specification with deterministic trend. Among the first generation tests, the LL test and Breitung's test often, but by no means uniformly, perform best – especially if the autoregressive behavior of the data is homogeneous across units, as assumed under the alternative hypothesis of these tests. If the latter is not the case, the IPS and Maddala-Wu tests often outperform LL – which might, on the other hand, have more power if applied to nearly non-stationary data (i.e. if ρ_i in (1) is very close to 0) (Westerlund and Breitung 2013). According to the findings of Hlouskova and Wagner (2006), furthermore, Hadri's stationarity test is badly oversized as soon as the residuals are not totally free from serial correlation.

It is interesting to know how the first generation tests behave under cross-section dependency. This seems to depend on the specific strength and type of contemporaneous correlation: O'Connell (1998) reports large upward size distortions for the LL test: for average cross-section correlations of 0.9, he often finds rejection rates of more than 50 percent at a nominal level of 5 percent, especially if the number of cross-sections is large. Similarly, Banerjee, Marcellino, and Osbat (n.d.). find considerable size distortions if there is cointegration across units, that is, if y_{it} and y_{jt} are cointegrated (which is the case if there is a non-stationary common factor while the idiosyncratic part in (3), e_{it}, is stationary).

On the other hand, Hlouskova and Wagner (2006) as well as Maddala and Wu (1999) find only small increases in size for moderate cross-section correlations (of up to 0.6 in the case of Hlouskova and Wagner). In this situation, the relative ranking of the tests does not change. Baltagi, Bresson, and Pirotte (2007), who study the behavior of several tests for three models of spatial autocorrelation (in none of which common factors appear), report considerable size distortions (rejection rates up to 20 percent) only for strong spatial autoregressive processes, while they are weak for spatial moving-average and – especially – spatial error-component models. The performance of the conventional tests studied – LL, Breitung, IPS, Maddala-Wu

10 A first order moving-average-process is a form of autocorrelation where the observed value at time t is affected by the random shock on the series at time t-1.

and Choi – is very similar. Thus, it seems that cross-section dependence does not universally affect the size of panel unit-root tests.

Among second generation tests, unit-root tests assuming a factor structure sometimes show size distortions. This is the case for the CIPS test and Bai and Ng's approach if the time dimension is small, while bootstrapping procedures minimize them. Moon and Perron's tests perform best in terms of power, but tests assuming common factors generally need large samples to achieve satisfactory power. This finding applies also to bootstrapping procedures and approaches that take breaks into account.

Thus, the main conclusion to be drawn from the simulation evidence is that the performance of panel unit-root tests is moderate for the relatively small data sets which are employed often in comparative social research, especially if the data are subject to cross-section dependence and/or breaks.

2.5. Issues in the Application of Panel Unit Root Tests

When applying panel unit-root tests, one has to decide on the inclusion of an intercept and/or a time trend in the test equation. Furthermore, one has to select the number of lagged differences (p_i in (1)) to be included as regressors if a parametric correction for serial correlation is used. Regarding the first issue, the considerations pointed out in the contribution by Helmut Thome in this issue apply. With respect to the selection of p_i, it is common to use a general-to-specific-approach. Here, the test equation is computed sequentially, starting with a maximum p_i, p_{max}, which is chosen depending on T, and to step-wise reduce p_i until the t-Test for coefficient of the highest lag is significant. For the determination of p_{max}, the following formula performs well (Hayashi 2000, 594f.):

$$(4) \quad p_{max} = \mathrm{int}\left[12\left(\frac{T}{100}\right)^{1/4}\right]$$

Alternatively, information criteria can be used. Their drawback is that they tend to select a too small p_i, especially the Schwarz information criterion (SIC), if there is a specific form of serial correlation in the errors (negative moving-average processes) (Im, Pesaran, and Shin 2003, 68). Ng and Perron suggested modified information criteria which perform better in selecting p_i (Ng and Perron 2001).

2.6. Remarks Regarding the Choice of a Test and the Interpretation of Unit-Root Tests

Which of the conventional unit-root tests reviewed here should be chosen by the applied researcher? The answer is, first, dependent on the asymptotic theory deemed to be plausible: If the cross-sections can be viewed as a finite universe, the fixed-N, $T \rightarrow \infty$ asymptotics of Maddala and Wu are appropriate; generally, also the sequential limit theories of the other tests are valid in fixed-N situations, but this does not apply vice versa. Second, one should check if it is plausible to assume homogeneous autocorrelation properties across units under the alternative hypothesis: if this is the case, either the LL or Breitung's test might be applied; otherwise, IPS and the Maddala/Wu/Choi tests might be considered. To my view, it is difficult to imagine situations where the units exert literally identical dynamic behavior, as assumed by the former tests. Nonetheless, in view of the simulation results reviewed above, there might be situations where the tests of LL or Breitung are an option: these seem to outperform IPS in terms of power if the series exhibit strong autocorrelation. Regarding the stationarity test of Hadri, the simulation studies show that it tends to exhibit strongly inflated rates of alpha errors as soon as there is some autocorrelation in the data, making it useless in practical applications.

But which test should one choose if cross-section correlation is present? Here, the nature of cross-section correlation is crucial: if it is plausible to assume that it is due to common factors, factor analytic methods might be applied. Among these, the PANIC approach is conceptually most convincing, because both common as well as idiosyncratic components are tested. The drawback is that it needs large samples (especially large T) to perform well. For moderately-sized samples with a not too small time dimension ($T \geq 20$, say), CIPS* might be considered, although the presumption of one stationary common factor is somewhat restrictive. If the common factor model is not plausible, bootstrapping might be considered, although one has to be willing to assume that the source of cross-section correlation is stationary when using Chang's test, while the more general test of Palm and others has yet to be fully developed. The bootstrapping approach might also be chosen if the sample size is small, because – according to simulation studies – tests based on the common factor approach tend to over-reject in small samples.

Finally, in analyses of PTCS data, breaks in the series might be an issue. In PTCS analyses of crime statistics (crime rates, imprisonment rates, etc.) at the sub-national level, break dates are known in most cases, because they are due to changes in legislation and/or registration procedures. Here, the original approach of Im and coauthors might be a choice. In cross-national research, however, breaks at an unknown date are a realistic possibility, because it is difficult to gather comprehensive information on changes in legislation and recording procedures for every country in the sample. Thus, the procedures suggested by Tam and Westerlund might be considered, which also allow cross-section correlation to be taken into account.

I would like to close the section on unit-root tests with a caveat: Much care is needed in interpreting the results of panel unit-root tests, because the rejection of the null hypothesis suggests only that *at least one panel member* is stationary. But the test does not tell us for *how many units* the series are stationary. It would be important to know if, in fact, *all* series are stationary, or if there is a mixture of stationary and non-stationary processes. One way to check this would be to test also the null hypothesis that all units are stationary. But, as pointed out above, the only widely available panel stationarity test by Hadri (2000) performs extremely badly. So this is only a theoretical possibility. There are proposals for testing fractions of the units sequentially (Smeekes 2010) and approaches to unit-by-unit testing which control the size of the test (which would otherwise be inflated due to multiple testing) (Moon and Perron 2012; Hanck 2009). But, these procedures do not perform well in identifying the stationary units in PTCS data as long as T is not large (< 100) (Smeekes 2010). Thus, their usefulness is questionable for many situations. It remains only to exert caution if there are indications that non-stationarity is an issue for at least a part of the units, even if a formal test rejects the unit-root-hypothesis.

3. Testing for Cointegration

3.1. Testing for Cointegration in the Absence of Cross-Section Dependence and Breaks

The cointegration tests proposed by Pedroni (1999, 2004) are widely used.[11] These tests are based on the residuals of a regression of the following form:

$$y_{it} = \alpha_i + \delta_i t + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + \ldots + \beta_{Mi} x_{Mit} + e_{it} \qquad (5)$$

If there is cointegration between y and x, these residuals (e_{it}) should be stationary; thus, testing e_{it} for a unit-root amounts to testing for cointegration. Generally, Pedroni's tests allow the cointegration coefficients and the variances of the series to be heterogeneous and the regressors to be endogenous, although they should not be cointegrated.

He considers two basic approaches for the construction of panel cointegration tests: For the first it is assumed that the residuals of the individual static cointegration regressions follow identical autoregressive processes under the alternative hypothesis. These "panel" statistics are constructed based on pooled regressions. The second approach allows for heterogeneous serial correlation properties, therefore, the "group" tests are based on averages of individual test statistics. For each of these two types, panel cointegration tests based on the ADF test and the semi-parametric cointegration tests studied by Phillips and Ouliaris (1990) are developed.

Westerlund (2007) argued that there might be gains in power if the cointegration test is carried out as a test on the so-called error correction parameters γ_i in the following panel-error correction model:

11 In some situations it might be of interest to test the null hypothesis that two series are cointegrated; tests of this kind are not reviewed here; see McCoskey and Kao (1998) for such a test.

$$(6)\quad \Delta \mathbf{y}_{it} = \alpha_{1i} + \alpha_{2i}t + \gamma_i(\mathbf{y}_{i,t-1} - \vartheta_1{}'\mathbf{X}_{it}) + \sum_{j=1}^{p_i}\beta_{ij}\Delta \mathbf{y}_{i,t-j} + \sum_{j=0}^{p_i}\boldsymbol{\delta}'_{ij}\Delta \mathbf{X}_{i,t-j} + \varepsilon_{it}$$

Here, θ_i is a vector with unit-specific cointegration parameters. In the case of cointegration, the γ_i have to be negative ($\gamma_i < 0$). Thus, testing the hypothesis $\gamma_i = 0$ amounts to a cointegration test. The advantage of this approach is that the restriction of the long- and short-run dynamics to be identical, which is implicit in residual-based cointegration tests, can be avoided. He develops two tests: a test on the error correction parameter, and one based on the product of the error correction term and the number of time periods. Both come in a "panel" (P_ϑ, P_α) and a "group" (G_ϑ, G_α) version, so that four different tests result. A drawback of these tests is that strict exogeneity is required, which might be relaxed to weak exogeneity by adding leads of the first differences of the regressors.[12]

3.2. Dealing with Cross-Section Dependence and Breaks

The tests reviewed so far assume cross-section independence, an assumption which might be violated in many cases. Structural breaks also affect the performance of cointegration tests. According to simulations, at least Pedroni's t-tests lose power if there are breaks (Banerjee and Carrion-i-Silvestre 2006; Gutierrez 2005). Many suggestions for dealing with cross-section dependence also consider the issue of breaks.[13] The types of breaks studied are level shifts (implying a change of the intercept in estimation equations), changes of the time trend (change of the coefficient for a time index), and changes of the cointegration relationship (change of the cointegration parameter). None of the approaches suggested in the literature considers all possible combinations of these types of breaks, but they are all tailored to specific situations.

3.2.1. Allowing for Breaks in Absence of Cross-Section Dependence

To deal with breaks in the mean and/or the cointegration parameters in absence of cross-section correlation, there are two proposals to extend the approach of Gregory and Hansen (1996) for single time-series: Westerlund (2006b) constructs four test statistics which are cross-sectional sums of Gregory/Hansen-type test statistics for a mean-shift. Gutierrez (2005), in contrast, bases his tests on standardized sums of the *p-values* for Gregory/Hansen tests.

3.2.2. Testing for Cointegration in the Presence of Breaks and Cross-Section Dependence

Di Iorio and Fachin (2007) also adapt the Gregory/Hansen approach to the PTCS context in a test that uses the mean or the median of individual test statistics. They suggest computing critical values using a bootstrapping procedure, which also accounts for cross-section dependence.

Banerjee and Carrion-i-Silvestre (2006) consider every possible combination of single breaks in the mean, the slope, and the time trend of the cointegration relation, except a change restricted to the time trend and a simultaneous break in all three parameters. Specifically, they adapt two of Pedroni's parametric test statistics in order to take the breaks into account. For the case of cross-section dependence, Banerjee and Carrion-i-Silvestre use Bai and Ng's PANIC approach. Here, it is necessary to assume a common break date for all units; if it is not known, it has to be estimated. A similar test, which also follows Bai and Ng, has been proposed by Westerlund and Edgerton (2008), who also develop a procedure for the determination of the date of the break. But their test presumes that the common factors are stationary, and allows only for shifts in the intercept or in the intercept and the cointegration parameter.

3.3. Small Sample Behavior

The results from various Monte-Carlo studies (Banerjee, Marcellino, and Osbat 2004; Banerjee and Carrion-i-Silvestre 2006; Gengenbach, Palm, and Urbain 2005; Gutierrez 2003; Gutierrez 2005; Örsal 2007; Pedroni 2004; Wagner and Hlouskova 2007; Westerlund 2006b; Westerlund 2007; Westerlund and Basher 2008) are anything but clear-cut, but some general observations can be made: First, most tests show size distortions in the presence of residual serial correlation, especially Pedroni's parametric tests. One of Westerlund's tests on the error correction parameters, P_α, rejects too often when the regressors are endo-

12 For the concept of weak exogeneity see Enders (2004, 368).

13 One exception is the bootstrapping procedure developed by Westerlund (2007) to be applied with his error-correction tests in the case of cross-section correlation.

genous. The effect of cross-section dependence on size seems to be moderate in most cases. With respect to power, Pedroni's parametric statistics often perform best, but even for these, power is often quite low for small T (< 25), sometimes of the same magnitude as size. Similar results apply to other tests which I do not present here. Cointegration tests that accommodate breaks are generally well-sized, but need large T (= 100 or even 200) to achieve reasonable power, with the exception of Di Iorio and Fachin (2007).

3.4. Practical Considerations

When implementing residual-based cointegration tests, one has to specify the deterministic components (intercepts and time trends) of the static regression equation (5) as well as the number of lagged differences of the residuals to include in the ADF-type equation for testing the residuals for a unit-root. Here, similar considerations apply as in the case of unit-root tests. When serial correlation is accounted for non-parametrically, one has to choose a lag length for the band width of the kernel estimator; Pedroni suggests to do this in dependence on T according to the formula $\mathrm{int}[K=4(T/100)^{2/9}]$ (Pedroni 2004, 608).

3.5. Remarks Regarding the Choice of Test and the Interpretation of Cointegration Tests

How should one proceed in applied research, especially macro-level criminological research? Here, the difficulty arises that the time dimension of the data sets at hand – often between 20 and 50 periods – lies in the region where, according to the simulation results reviewed above, the power of cointegration tests is small in most cases. Thus, there is a real risk to miss substantively interesting long-run relationships. On the other hand, the implications of wrongly rejecting the null hypothesis are less serious than in the case of unit-root testing: the long-run parameter estimated in the next step will not reach significance if the model estimated is correctly specified, leading to the correct conclusion that the long-run effect is nil. One might therefore consider being more liberal with respect to alpha errors in cointegration testing than in unit-root testing, and putting more weight on power properties. Therefore, Pedroni's parametric tests, which often perform best in terms of power (but less so in terms of size), might be a choice if there are no indications of strong cross-section dependence (recall that the effects of cross-section correlation on cointegration tests are generally moderate). If the latter is the case, one might apply the error-correction tests (but not P_α, in view of the simulation results) with the bootstrapping procedure proposed by Westerlund. If there are breaks in the series, there is no satisfying solution yet, because appropriate methods need larger samples than available in most cases to achieve power. Besides that, none of the procedures mentioned above is implemented in standard statistical software. An ad-hoc approach might be to first adjust the series to the breaks (by regressing them on appropriate dummy-variables, for example) and then to apply conventional tests, although the results will only be roughly indicative then.

Regarding the interpretation of cointegration tests, finally, a caveat analogous to that with respect to panel unit-root tests applies here: The rejection of the hypothesis of no cointegration does *not* imply that *all* units are cointegrated. This has to be kept in mind in view of the fact that estimators for cointegration parameters assume that, indeed, cointegration holds for all units.

4. Estimating Cointegration Relationships in Non-Stationary PTCS Data

4.1. Fully Modified OLS and Dynamic OLS

First, there are several approaches to "fully modifying" OLS (FM-OLS) for cointegrated panels by using non-parametric corrections to (5) (Kao and Chiang 2000; Chiang, Kao, and Lo 2007; Pedroni 2000) for serial correlation and endogeneity.[14] They differ in the degree of homogeneity of variances and serial correlation properties assumed, while they presume generally homogeneous cointegration parameters. For each of them, analogues based on adding

14 In the case of cointegration it is also possible that several cointegration relationships ("cointegration vectors") exist if the estimation equation has more than one variable on its right-hand side. But since system approaches allowing the estimation of multiple cointegration vectors need huge data sets to perform well, I only cover the case of a single cointegration relationship. The interested reader is referred to the discussion in Wagner and Hlouskova (2007) and the references given there. I also do not treat the estimation of the "long-run average relationship" in the absence of cointegration, an issue that has received little attention so far (but see Sun 2004, Fuertes 2008, and Trapani 2012).

leads and lags of the first differences of the regressors to (5) have been proposed. This so-called Dynamic OLS approach (DOLS) is asymptotically equivalent to FM-OLS (Kao and Chiang 2000).

All these estimators converge to a normal distribution, the variance of which depends on the deterministic specification of the equation and the long-run covariance. To conduct statistical inference, kernel-density estimates of the long-run variance-covariance matrix are needed.

4.2. Error-Correction Models

An alternative way to deal with serial correlation is to model short-run dynamics explicitly using an error-correction model like (6). The disadvantage of this strategy in the panel context, however, is that it will be subject to the so-called Nickell bias due to the presence of the lagged dependent variable on the right-hand side of the equation if individual-specific intercepts are allowed for. But this bias usually vanishes fast with an increasing time-dimension (Judson and Owen 1999). Therefore, the error-correction model might nonetheless be useful for panels with a large enough time dimension. This approach has been favored by Pesaran, Shin, and Smith (1999), who consider three variants of the error correction model, varying in the degree of homogeneity assumed for long- and short-run parameters:

The "Dynamic Fixed Effects" (DFE) model is intended for situations in which it is reasonable to assume that short- and long-run parameters are identical for all cross-sections. Therefore, the following equation is estimated:[15]

$$(8),\quad \Delta\mathbf{y}_i = \varphi\, \mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^{p-1} \lambda_j^{*}\Delta\mathbf{y}_{i,-j} + \sum_{j=0}^{q-1} \Delta\mathbf{X}_{i,-j}\boldsymbol{\delta}_j^{*} + \alpha_i\mathbf{d}_i + \boldsymbol{\zeta}_i$$

where $\mathbf{X}_i$ is a matrix with the observations of the independent variables and $\mathbf{d}_i$ a matrix containing a unit-specific intercept and – if specified – a time trend. The cointegration parameters $\boldsymbol{\theta}$ can be computed as

$$\frac{-\boldsymbol{\beta}}{\varphi} = \vartheta.$$

If the cointegration coefficients are, in fact, heterogeneous, DFE will produce an inconsistent estimate of the average cointegration parameter.

For the case that it is plausible that the cointegration parameters $\boldsymbol{\theta}$ are homogeneous, but the short-run dynamics differ across units, Pesaran et al. propose the "Pooled Mean Group" (PMG) estimator (note the two subscripts of the $\boldsymbol{\lambda}^*$- and $\boldsymbol{\delta}^*$-parameter vectors):

$$(9)\quad \Delta\mathbf{y}_i = \varphi_i\mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta}_i + \sum_{j=1}^{p-1} \boldsymbol{\lambda}_{ij}^{*}\Delta\mathbf{y}_{i,-1} + \sum_{j=0}^{q-1} \Delta\mathbf{X}_{i,-j}\boldsymbol{\delta}_{ij}^{*} + \alpha_i\mathbf{d}_i + \boldsymbol{\zeta}_i$$

with the restriction

$$\frac{-\boldsymbol{\beta}_i}{\varphi_i} = \vartheta_i = \vartheta,\ i = 1,2,\ldots,N.$$

For the estimation of (9), a maximum likelihood algorithm is described by Pesaran, Shin, and Smith.

In the Mean Group (MG) estimator the assumption of homogeneous cointegration parameters is also relaxed. Here, an estimate for the average cointegration parameter θ_{MG} is computed as the arithmetic mean of the individual θ_i after running (9) for each unit via OLS without restrictions on $\boldsymbol{\beta}_i$ and φ_i. A consistent estimate of the variance of θ_{MG} can be obtained as follows (Pesaran, Smith, and Im Kyong So 1996, 157):

$$(10)\quad V_{\vartheta MG} = \frac{1}{N(N-1)}\sum_{i=0}^{N}(\hat{\vartheta}_i - \hat{\vartheta}_{MG})^2$$

Pesaran et al. suggest choosing among the three estimators based on a Hausman test of the null hypothesis that the difference between the MG and the PMG (or DFE) estimator is zero. If this hypothesis is rejected, the MG estimator is appropriate; otherwise, the more efficient pooled estimator is to be preferred.

15 The equation contains the so-called "Bårdsen transformation" of the error-correction model (6).

4.3. Estimation of Cointegration Parameters under Cross-Section Dependence

Bai and Kao suggest adapting the FM-OLS estimator to the case when there is cross-section correlation due to common stationary factors (Bai and Kao 2006). Here it is assumed that the cointegration parameters are the same for all units. In this approach, principal component analysis is used to extract the common factors from the residuals of a first-step OLS regression and to estimate the factor loadings. These are used to construct correction terms in a modified estimation formula, which is used to produce second-round parameters. The residuals of this second estimation step are used to construct new correction terms, and so on. This procedure is continued until convergence is achieved, resulting in the "continuously-updated and fully-modified estimator" (CUP-FM). In a later paper, the approach is extended to the case of common non-stationary factors. Furthermore, Bai, Kao, and Ng also consider an estimator where the correction is applied only in the last iteration, the so-called "continuously-updated and bias-corrected" estimator (CupBC) (Bai, Kao, and Ng 2009). Their approach is also valid when there is a mixture of stationary and non-stationary common factors, or a mixture of stationary and non-stationary regressors.

The computation of the "continuously-updated" estimators of Bai, Kao, and Ng is fairly involved. A simpler approach has been suggested by Pesaran and others, who transfer the logic of the "cross-sectionally augmented" unit-root tests to the estimation of cointegration regressions (Kapetanios, Pesaran, and Yamagata 2011): Common factors are simply accounted for by adding the cross-section averages of the dependent variable and the regressors to the left-hand side variables of the estimation equation. They consider static "common correlated effects" mean group estimators,[16] called CCEMG, which allow the cointegration parameters to vary across units, as well as a static pooled FE estimator, where the cointegration coefficients (but not the parameters for the cross-section averages) are assumed to be homogeneous, which they designate CCEP (Common Correlated Effects Pooled). Hypothesis tests on the average cointegration parameter can be conducted analogous to tests on MG. The estimation of the variance of coefficients estimated using CCEP, however, is more complicated and requires the computation of the CCEMG estimator.[17] The CCE estimators are consistent under various types of cross-section dependence, including single or multiple common factors, which might be stationary or non-stationary, and even when the idiosyncratic errors are cross-sectionally correlated (Kapetanios, Pesaran, and Yamagata 2011; Pesaran and Tosetti 2011).

4.4. Small Sample Properties

From several simulation studies (Breitung 2005; Pedroni 2000; Kao and Chiang 2000; Wagner and Hlouskova 2007; Bai and Kao 2006; Eberhardt and Bond 2009; Kapetanios, Pesaran, and Yamagata 2011), the following picture emerges: FM-OLS does not work very well in the reduction of bias, especially with respect to t-tests. DOLS performs better than FM-OLS, but even here estimates are biased and significance tests oversized in very small samples. This is especially a feature of an estimator proposed by Kao and Chiang (2000) which allows for heterogeneous variances and dynamics. Correlation of the residuals between cross-sections induces small increases in the bias of estimates and size distortions in tests for significance – but these are negative for DOLS. Cross-unit cointegration makes the biases of conventional estimates a bit worse. Unfortunately, there is no simulation evidence on the performance of the error correction models discussed above (MG, DFE, PMG).

Among the estimators for cointegrated PTCS data with cross-section correlation, the CUP-FM and especially the CupBC estimators perform well if there is one common factor (stationary or non-stationary), but there are large size distortions in significance tests if there are several common factors. The CCE estimators, on the other hand, seem to have good properties, although large samples (~ N = T = 100) are needed to achieve good power for significance tests.

16 With "static" I refer to estimation equations which contain no lagged values of the dependent variable or the regressors (as error-correction models do, for example) to model dynamic relationships.

17 The appropriate formula can be found in Kapetanios, Pesaran, and Yamagata (2011, 330f.).

4.5. Comments on the Estimation of Long-Run Relationships in Applied Research

If cross-section residual correlation is not a concern, and cross-unit cointegration is not plausible, the choice of the estimator depends on the degree of cross-section homogeneity with respect to parameters, residual variances, and (in case of static estimators) residual serial correlation one is willing to assume. If homogeneity of parameters is a reasonable assumption, DOLS is an option (although not in the variant due to Kao mentioned above), because it outperforms FM-OLS and avoids possible problems due to the Nickell bias which might occur with ECM estimators (such as DFE) with small T. The latter risk might be outweighed by the higher costs (in terms of bias) arising from wrongly imposing homogeneous parameters. Thus, if parameter heterogeneity is to be expected, one might turn to the MG or PMG estimators. The tenability of assumptions regarding parameter homogeneity can be tested within the ECM approach.[18] If cross-section dependence is an issue, one should turn to the CCE approach, which can be implemented in Stata and R (see below). The drawback of this method is the low power of significance tests when the CCE approach is applied to data sets of the size usually encountered in macro-level criminological research. One would have to accept this, due to the costs of ignoring cross-section dependence. Besides that, one should keep in mind that in this way, one can only estimate relationships within the idiosyncratic part of the data (that is, the portion which is not driven by common factors); thus, if a parameter turns out to be non-significant, this does not preclude the possibility that there are long-run relationships between factors common to all units.

5. An Example: Robbery Rates in the West German Federal States

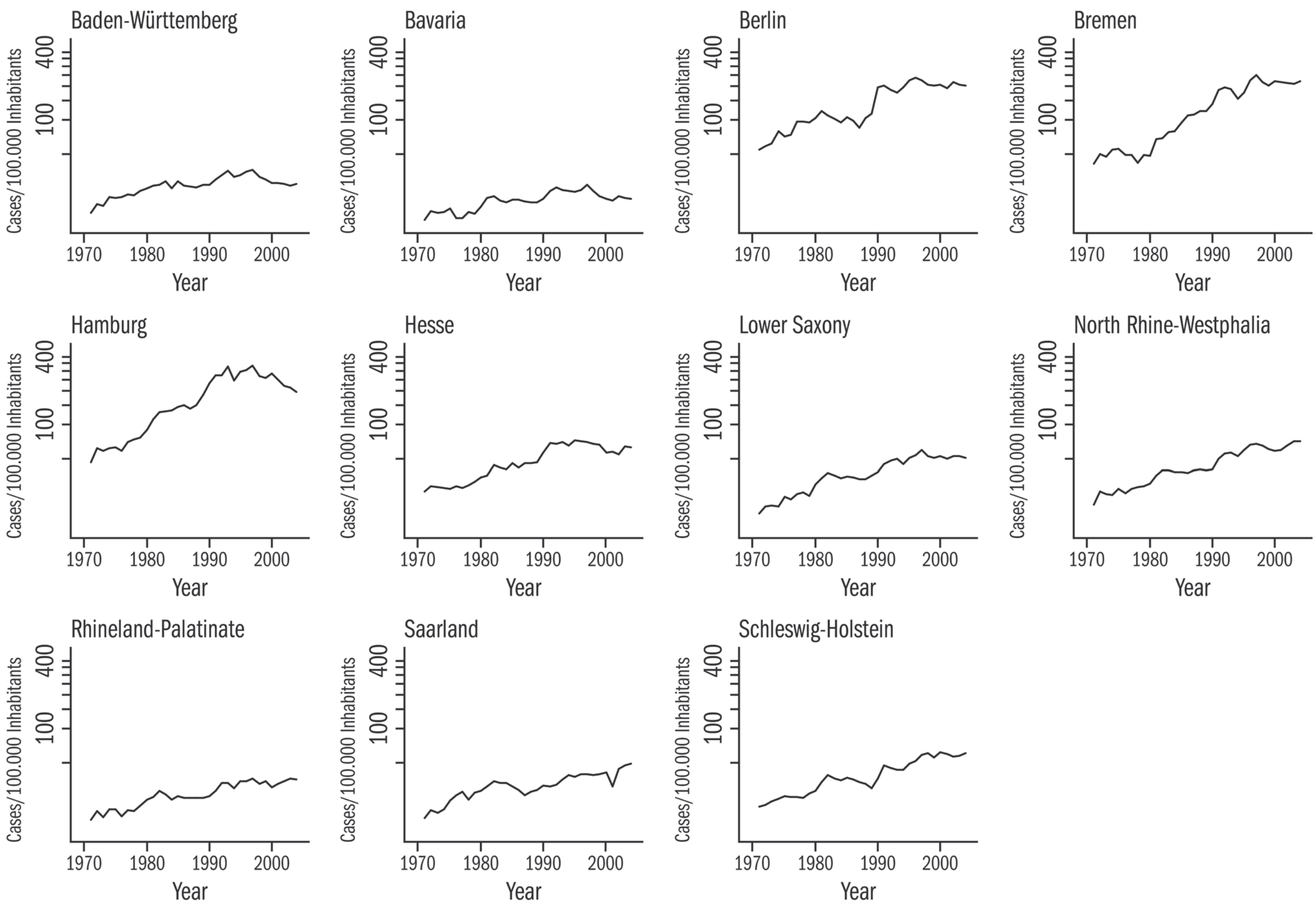
To illustrate the general approach to the analysis of non-stationary PTCS data described here, I use a data set with (completed) robbery rates for the eleven West German federal states for 1971–2004.[19] It includes also data for several plausible explanatory variables, namely per capita real disposable income, per capita consumption, and the clearance rate for robbery. It also contains the percentage of inhabitants aged 65 or older, as well as the percentage of foreigners (in the legal sense); these latter two variables serve as controls for changes in the demographic composition of the population.[20]

18 Although there will be cases where it is not possible to conduct the appropriate Hausman test, because there is no guarantee that the matrix in the denominator of the test statistic will be – as required – positive-definite.

19 Data for the five federal states on the territory of the former GDR are not used, because robbery rates for them are only available from 1993.

20 For data sources and the motivation for the selection of the variables, see Birkel (2015).

Figure 1: Robbery rates for the West German Federal States, 1971–2004



Source: Federal Criminal Police Office, Criminal Police Offices of the Federal States.

5.1. Unit-Root Tests

A panel of the robbery rates is shown in Figure 1 (note the logarithmic scale of the y-axes). The rates for most states exhibit a clear upward tendency until around 2000. Thus, the robbery rates may contain a unit-root. The graphs for the other variables – which are not shown here – suggest also stochastic trends. Therefore, the hypothesis of a unit-root was tested formally, using the IPS test as well as the Maddala-Wu Test based on individual ADF tests. These were conducted with unit-specific intercepts and a unit-specific time trend in the test equation, because at least for some federal states stationarity around a deterministic trend seemed to be a plausible alternative hypothesis. For lag-length selection, the modified AIC (MAIC) of Ng and Perron (see above) was used. The results are shown in Table 1.[21]

21 The unit-root test for the percentage of inhabitants aged 65 or older was applied to the first-differenced series, because preliminary analyses suggested that it might be a second-order integrated process.

Table 1: Unit-Root Tests

Variable	Period	Test	Test statistic	p
Robbery rate	1971–2004	IPS	0.79	0.788
		Maddal-Wu-ADF	16.42	0.795
Clearance rate robbery	1971–2004	IPS	–1.26	0.104
		Maddal-Wu-ADF	25.55	0.272
Real per capita disposable income	1971–2004	IPS	2.129	0.983
		Maddal-Wu-ADF	10.452	0.982
Real per capita consumption	1971–2004	IPS	2.417	0.992
		Maddal-Wu-ADF	8.237	0.997
Percentage foreigners	1971–2004	IPS	1.670	0.953
		Maddal-Wu-ADF	10.185	0.985
Δ Percentage 65+	1968–2004	IPS	5.013	1.000
		Maddal-Wu-ADF	0.827	1.000

For none of the variables can the null hypothesis of a unit-root be rejected. To check roughly if the results might be affected by cross-section correlation, the averages of the absolute values of the pairwise correlation coefficients between the first-differenced series for the individual federal states (|r|) were computed (Table 2).

Table 2: Average absolute value of the pairwise correlation coefficient and CIPS*-Unit-Root Tests

Variable	|r|	CIPS* statistic
Robbery rate	0.382	-
Clearance rate robbery	0.162	-
Real per capita disposable income	0.634	–2.019
Real per capita consumption	0.623	–1.598
Percentage foreigners	0.692	–1.207
Δ Percentage 65+	0.743	–3.003**

* $p < 0.10$ ** $p < 0.05$ (CIPS*-statistic only)

For all explanatory variables except the clearance rate, |r| exceeds the threshold of 0.6, above which – according to Hlouskova and Wagner (see above) – conventional unit-root tests are affected. Therefore, for these variables the CIPS* test by Pesaran was also computed (Table 2, third column). Pesaran's test was selected because the assumption of one common factor seemed a plausible possibility, while there was no reason to suspect that there might be several common factors. The CIPS* version was chosen because it seems to perform slightly better than CIPS in terms of size if T is small, according to the simulation results reported by Pesaran (2007). The results of the CIPS* test corroborate the findings of the conventional tests, with the exception of the first differences of the percentage of older people, for which the unit-root hypothesis is rejected. But it has to be kept in mind that this result applies only to the unit-specific component; it might nonetheless be that there is a common component with a unit-root, inducing non-stationarity in the observed series. Therefore, this variable was treated as non-stationary in further analyses despite the result of the CIPS* test.[22]

22 This should not lead to erroneous findings regarding long-term effects if this variable is, in fact, stationary: the residuals of a regression of the robbery rates on it will be non-stationary, and the cointegration test in the second step of the analysis will not reject the null hypothesis of no cointegration. If the latter happens, one could, in principle, miss a long-run effect of the (not-differenced) percentage of older people on the robbery rates, however. But if the null hypothesis of no cointegration is rejected (as here), it is fairly safe to conclude that the changes in the percentage of people aged 65 or older are, indeed, random walks (otherwise the regression residuals would not be stationary) and that they exert a long-run effect on the robbery rates.

5.2. Cointegration Tests

In a next step, it was assessed whether the supposed explanatory variables are cointegrated with the robbery rates. This was done using Pedroni's parametric t-statistics, which have comparatively good power in small samples. These were computed in the "group" as well as in the "panel" version, because none of the two seems to outperform the other under all possible circumstances. For lag selection the AIC was used, and besides a unit-specific intercept also a unit-specific time trend was specified. The tests were applied to the natural logs of the series, because a non-linear relationship is theoretically plausible.

Table 3: Parametric Pedroni cointegration tests

Independent variable (logged)	p-value panel-t-statistic	p-value group-t-statistic
Clearance rate robbery	0.066	**0.015**
Real per capita disposable income	0.339	**0.001**
Real per capita consumption	0.465	**0.012**
Percentage foreigners	**0.043**	**0.017**
Δ Percentage 65+	**0.002**	**0.000**

p-values < 0.05 are printed bold

The results are not totally clear-cut, because for three of the series the panel t statistics are not significant at the 5 percent level. I suspect that the lack of significance of some of the panel tests is due to the fact that they are constructed under the assumption that the residuals of the cointegration regression follow identical autoregressive processes under the alternative hypothesis, which might be too restrictive here. Therefore, I retain the group statistic, which allows the autoregressive behavior of the residuals to be heterogeneous across units, and conclude that all variables are cointegrated with the robbery rates. I did not apply a cointegration test constructed for the case of cross-section dependence because the effects of cross-section correlation on cointegration tests seem to be mild (see 4.4 above).[23]

5.3. Estimation of the Long-Run Parameters

To estimate the cointegration parameters, the MG and PMG estimators were computed and a Hausman test applied to determine if the more efficient PMG estimator is appropriate. Since complete parameter homogeneity seemed not realistic for the data at hand, I did not consider the DFE estimator.[24] All variables were entered in their natural logs. Besides the aforementioned variables, a time trend was entered into the estimation equation. The number of lagged first differences of the independent variables in the equation was determined using the Schwartz-Bayes Information Criterion (SBIC) after setting the maximum number to one (due to the limited number of observation periods). The resulting estimates for the average cointegration coefficients and the average error correction parameters are shown in Table 4.

Table 4: Error correction models for the logged robbery rate

Independent variable (logged)	PMG	MG
Real per capita disposable income	1.41**	2.74**
Real per capita consumption	–1.27**	–2.78**
Clearance rate robbery	–0.49**	–0.41*
Percentage foreigners	0.36**	0.29
Δ Percentage 65+	–4.70**	–5.05**
Lag specification	**1,1,1,1,1**	**0,0,0,0,0**
Average error correction parameter	–0.45	–0.56
Hausman-χ^2-statistic	3.23	
$\|r\|$	0.23	0.30
Number of observations	363	363

* p < 0.10 ** p < 0.05

23 There were no indications of breaks in the robbery rates. Therefore, there was no need to apply one of the cointegration tests (see 3.2 above) that account for breaks.

24 For the same reason, I decided against DOLS.

The Hausman test is not significant, so the hypothesis that the cointegration parameters are homogeneous cannot be rejected. Therefore, the PMG-specification is valid.

Besides that, the results obtained by the MG and the PMG estimators have the same sign, although the absolute values of the parameters differ, sometimes remarkably. Furthermore, the MG estimate for the coefficient of the percentage of foreigners among the population does not reach significance – probably because the MG estimator is less efficient than the PMG specification. It can be concluded that, in the long-run, the robbery rate drops if the clearance rate – an indicator for the probability of apprehension – for this crime rises, which is in line with the economic theory of crime (Becker 1968; Ehrlich 1973). Furthermore, the incidence of robbery also declines if real per capita consumption – which can be interpreted as a proxy for the supply of potential loot (the expected returns of criminal acts) – increases, while it grows with rising disposable income, a measure of legal income opportunities. These two findings are contrary to the predictions of economic theory; a detailed discussion of their implications is beyond the scope of the present article. The latter applies also to the findings regarding the percentage of foreigners and the change in the proportion of elderly people, which primarily served as control variables here.

Finally, there is only moderate cross-section correlation between the residuals for the individual federal states (see $|r|$ in the second last row in Table 4). In view of this finding, and because the consequences of cross-section correlation seem to be mild according to the simulation results mentioned above, I did not apply estimators for cross-sectionally correlated data.

6. Conclusion

In recent years, much effort has been spent on the study of non-stationarity in PTCS data. It emerges that non-stationarity has the potential to invalidate conventional approaches to the analysis of such data. Therefore, unit-root tests are mandatory. A variety of such tests have been proposed, some of which are also appropriate if the data are subject to cross-section dependence and/or structural breaks. Furthermore, if the data are, in fact, non-stationary, the appropriate method of estimation of long-run relationships depends on whether the variables are cointegrated or not. This can be determined using one of the cointegration tests reviewed here. After establishing cointegration, long-run relationships can be estimated; for this purpose, a number of approaches have been developed. For each of the three steps, many of the procedures proposed in the literature can be implemented using standard software packages.[25] The general approach to the analysis of non-stationary PTCS data was illustrated using data for the West German federal states 1971–2004. The results regarding cointegration relationships with robbery rates only partially support the economic theory of crime.

For most situations there are now procedures available that show good performance with sufficient sample size. The latter qualifier (sufficient sample size), however, can make it difficult in applied work to simultaneously account for all possible complications mentioned here, because the data sets actually available often have a modest size. There is no general rule of thumb as to which problem might be ignored without seriously jeopardizing the validity of the results in such a situation. Throughout the paper, advice was given with respect to the issues to be considered when deciding how to proceed. In addition, it is advisable to carefully explore the properties of the spe-

25 EViews 8 offers a full suite of “first-generation” methods, including the LL, IPS and Maddala-Wu panel unit-root tests, Pedronis cointegration tests, and FM-OLS as well as DOLS estimation. Stata 13 offers panel unit-root tests (LL, IPS, Maddala-Wu and others). In addition, there are user-written ado-files which allow the implementation of Pesaran’s CIPS* test (-pescadf-), Pedroni’s residual-based cointegration tests (-xtpedroni-), Westerlund’s ECM-based cointegration test (-xtwest-), DOLS (-xtdolshm-), panel error correction models (DFE, PMG, MG; -xtpmg-), and the CCEMG estimator (-xtmg-). Similarly, there are user-written codes available for RATS (panel unit-root tests (IPS, LL), Pedroni’s panel cointegration tests, DOLS and FM-OLS). For R, there is a package plm, by which several first generation unit-root tests (LL, IPS, Maddala-Wu, Hadri), Pesaran’s CIPS* test, as well as the CCEMG and CCEP estimators can be implemented; furthermore, the factor-analytic approach to unit-root testing of Bai and Ng (2004) can be applied via the package PANICr. The PANEL procedure in SAS/ETS 13.2 also provides limited facilities for non-stationary PTCS data (unit-root tests: LL, IPS, Maddala-Wu, Hadri, and a further test not discussed here due to Harris and Tzavalis [1999]).

cific data at hand and gather contextual information regarding relevant events which might induce breaks. If it emanates from this investigation that the observations are only moderately correlated cross-sectionally, and that there are no breaks in the series, one might apply "first generation" methods. If the results of the exploration are less reassuring, "second generation" approaches might be used, but one should be aware that there is some uncertainty with respect to the validity of the results, due to their limited performance when applied to samples of moderate size.

Besides that, further problems remain: For example, there is the issue of figuring out for which units the data are stationary if the hypothesis of a unit-root is rejected. Similarly, it would be useful to have a procedure available to find out if there is cointegration for all units if a cointegration test is significant. Regarding the latter problem, it is to be hoped that estimation methods which are valid even if there is a mixture of cointegrated and not-cointegrated units will be developed in the future,[26] which would obviate the need to partition the sample into cointegrated and not-cointegrated units.

References

Bai, Jushan, and Chihwah Kao. 2006. On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In *Panel Data Econometrics: Theoretical Contributions and Empirical Applications,* ed. Badi H. Baltagi, 3–30. Amsterdam: Elsevier.

Bai, Jushan, Chihwa Kao, and Serena Ng. 2009. Panel Cointegration with Global Stochastic Trends. *Journal of Econometrics* 149: 82–99.

Bai, Jushan, and Serena Ng. 2004. A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72 (4): 1127–77.

Baltagi, Badi H., Georges Bresson, and Alain Pirotte. 2007. Panel Unit Root Tests and Spatial Dependence. *Journal of Applied Econometrics* 22: 339–60.

Baltagi, Badi H., Chihwa Kao, and Min-Hsien Chiang. 2000. Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 7–51. Amsterdam: Elsevier.

Banerjee, Anindya, and Josep Lluis Carrion-i-Silvestre. 2006. *Cointegration in Panel Data with Breaks and Cross-Section Dependence,* Working Paper Series 591. Frankfurt am Main: Europäische Zentralbank.

Banerjee, Anindya, Massimiliano Marcellino, and Chiara Osbat. 2004. Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data. *Econometrics Journal* 7:322–40.

Banerjee, Anindya, Massimiliano Marcellino, and Chiara Osbat. n.d. *Testing for PPP: Should We Use Panel Methods?* Florence: European University Institute.

Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.

Binder, Michael, Cheng Hsiao, and M. Hashem Pesaran. 2003. *Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration.* Frankfurt am Main and Los Angeles: Johann Wolfgang Goethe-Universität and University of Southern California.

Birkel, Christoph. 2015. *Die Entwicklung der Gewaltkriminalität in Deutschland: Theoretische Erklärungsansätze im empirischen Vergleich.* Wiesbaden: Springer.

Blomquist, Johan, and Joakim Westerlund. 2014. A Non-Stationary Panel Data Investigation of the Unemployment-Crime Relationship. *Social Science Research* 44:114–25.

Blundell, Richard, and Stephen R. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87:115–43.

Bond, Stephen R., Céline Nauges, and Frank Windmeijer. 2005. *Unit Roots: Identification and Testing in Micro Panels,* Cemmap Working Paper CWP07/05. Oxford: Institute of Fiscal Studies.

Breitung, Jörg. 2000. The Local Power of Some Unit Root Tests for Panel Data. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 161–77. Amsterdam: Elsevier Science.

Breitung, Jörg. 2005. A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data. *Econometric Reviews* 24 (2): 151–73.

Breitung, Jörg, and Bertrand Candelon. 2005. Purchasing Power Parity during Currency Crises: A Panel Unit Root Test under Structural Breaks. *Review of World Economics* 141 (1): 124–40.

Breitung, Jörg, and Hashem Pesaran. 2008. Unit Roots and Cointegration in Panels. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice,* ed. László Mátyás and Patrick Sevestre, 279–322. Dordrecht: Kluwer.

Carrion-i-Silvestre, Josep Lluís, Tomás del Barrio-Castro, and Enrique López-Bazo. 2005. Breaking the Panels: An Application to the GDP per Capita. *Econometrics Journal* 8:159–75.

Chang, Yoosoon. 2004. Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics* 120 (2): 263–93.

Chen, Bangtian, Suzanne K. McCoskey, and Chiwah Kao. 1999. Estimation and Inference of a Cointegrated Regression in Panel Data: A Monte Carlo Study. *American Journal of Mathematical and Management Sciences* 19 (1–2): 75–114.

Chiang, Min-Hsien, Chihwa Kao, and Chih-Hsien Lo. 2007. The Estimation and Inference of a Panel Cointegration Model with a Time Trend. *Communications in Statistics – Theory and Methods* 36 (5–8): 1233–50.

Choi, In. 2001. Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20:249–72.

Coakley, Jerry, Ana-Maria Fuertes, and Ron Smith. 2001. *Small Sample Properties of Panel Time-Series Estimators with I(1) Errors.* Cambridge: Birkbeck College.

Di Iorio, Francesca, and Stefano Fachin. 2007. *Cointegration Testing in Dependent Panels with Breaks.* University of Rome La Sapienza.

Eberhardt, Markus, and Stephen R. Bond. 2009. *Cross-section Dependence in Nonstationary Panel Models: A Novel Estimator.* Oxford: Department of Economics, University of Oxford and Institute of Fiscal Studies.

26 See the proposal of Hu (2006) for cointegration testing and estimation in this situation.

Ehrlich, Isaac. 1973. Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *Journal of Political Economy* 81: 521–65.
Enders, Walter. 2004. *Applied Econometric Time Series,* 2d ed. Hoboken: Wiley.
Entorf, Horst. 1997. Random Walks with Drifts: Nonsense Regression and Spurious Fixed-Effect Estimation. *Journal of Econometrics* 80:287–96.
Entorf, Horst, and Hannes Spengler. 2000. Socioeconomic and Demographic Factors of Crime in Germany: Evidence From Panel Data of the German States. *International Review of Law and Economics* 75:75–106.
Entorf, Horst, and Peter Winker. 2005. *Investigating the Drugs-Crime Channel in Economics of Crime Models: Empirical Evidence from Panel Data of the German States,* Darmstadt Discussion Papers in Economics 165. Technische Universität Darmstadt.
Fuertes, Ana-Maria. 2008. Sieve Bootstrap t-Tests on Long-Run Average Parameters. *Computational Statistics and Data Analysis* 52:3354–70.
Gengenbach, Christian, Franz C. Palm, and Jean-Pierre Urbain. 2004. *Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling.* Maastricht: Department of Quantitative Economics, Universiteit Maastricht.
Gengenbach, Christian, Franz C. Palm, and Jean-Pierre Urbain. 2005. *Panel Cointegration Testing in the Presence of Common Factors.* Maastricht: Department of Quantitative Economics, Universiteit Maastricht.
Gregory, Allan W., and Bruce E. Hansen. 1996. Residual-Based Tests for Cointegration in Models with Regime Shifts. *Journal of Econometrics* 70:99–126.
Gutierrez, Luciano. 2003. On the Power Of Panel Cointegration Tests: A Monte Carlo Comparison. *Economics Letters* 80:105–11.
Gutierrez, Luciano. 2005. *Tests for Cointegration in Panels with Regime Shifts.* Sassari: Department of Agricultural Economics, University of Sassari.
Gutierrez, Luciano. n.d. Panel Unit Root Tests for Cross-Sectionally Correlated Panels: A Monte Carlo Comparison. Sassari: Department of Agricultural Economics, University of Sassari.
Hadri, Kaddour. 2000. Testing for Stationarity in Heterogenous Panel Data. *Econometrics Journal* 3:148–61.
Hadri, Kaddour, and Yao Rao. 2008. Panel Stationarity Test with Structural Breaks. *Oxford Bulletin of Economics and Statistics* 70 (2): 245–69.
Hale, Chris. 1998. Crime and the Business Cycle in Post-War Britain Revisited. *British Journal of Criminology* 38 (4): 681–98.
Han, Chirok, and Peter C.B. Phillips. 2010. *GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity,* Cowles Foundation Paper 1290. New Haven: Cowles Foundation for Research in Economics.
Hanck, Christoph. 2009. For Which Countries Did PPP Hold? A Multiple Testing Approach. *Empirical Economics* 37:93–103.
Harris, David, Stephen Leybourne, and Brendan McCabe. 2005. Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. *Journal of Business and Economic Statistics* 23 (4): 395–409.
Harris, Richard D. F., and Elias Tzavalis. 1999. Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed. *Journal of Econometrics* 91 (2):201–26.
Hayashi, Fumio. 2000. *Econometrics.* Princeton: Princeton University Press.
Hlouskova, Jaroslava, and Martin Wagner. 2006. The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. *Econometric Reviews* 25 (1): 85–116.
Hu, Ling. 2006. *A Simple Panel Cointegration Test and Estimation in I(0) / I(1) Mixed Panel Data.* Columbus: Department of Economics, The Ohio State University.
Hurlin, Christophe, and Valérie Mignon. 2004. *Second Generation Panel Unit Root Tests.* Paris: THEMA-CNRS, University of Paris X.
Im, Kyong-So, Junsoo Lee, and Margie Tieslau. 2005. Panel LM Unit-root Tests with Level Shifts. *Oxford Bulletin of Economics and Statistics* 67 (3): 393–419.
Im, Kyong So, M. Hashem Pesaran, and Yongcheol Shin. 2003. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115 (1): 53–74.
Jacobs, David, and Amber M. Richardson. 2008. Economic Inequality and Homicide in the Developed Nations From 1975 to 1995. *Homicide Studies* 12 (1): 28–45.
Jönsson, Kristian. 2005. Using Panel Data to Increase the Power of Modified Unit Root Tests in the Presence of Structural Breaks. *Applied Mathematics and Computation* 171:832–42.
Judson, Ruth A., and Ann L. Owen. 1999. Estimating Dynamic Panel Data Models: A Guide for Macroeconomists. *Economics Letters* 65 (1): 9–15.
Kao, Chihwa. 1999. Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics* 90:1–44.
Kao, Chihwa, and Min-Hsien Chiang. 2000. On the Estimation and Inference of a Cointegrated Regression in Panel Data. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 179–222. Amsterdam: Elsevier.
Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata. 2011. Panels With Nonstationary Multifactor Error Structures. *Journal of Econometrics* 160 (2): 326–48.
Levin, Andrew, Chie-Fu Lin, and Chia-Shang James Chu. 2002. Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties. *Journal of Econometrics* 108:1–24.
Maddala, Gangadharrao Soundalyarao, and Shaowen Wu. 1999. A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. In *Testing for Unit Roots and Cointegration Using Panel Data: Theory and Applications,* ed. Amindya Banerjee, 631–52. Oxford: Blackwell.
McCoskey, Suzanne, and Chihwa Kao. 1998. A Residual-Based Test of the Null of Cointegration in Panel Data. *Econometric Reviews* 17:57–84.
Moon, Hyungsik Roger, and Benoit Perron. 2004. Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* 122 (1): 81–126.
Moon, Hyungsik Roger, and Benoit Perron. 2012. Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Nonstationarity Properties of Individual Series in a Panel. *Journal of Econometrics* 169 (1): 29–33.
Murray, Christian J., and David H. Papell. 2000. Testing for Unit Roots in Panels in the Presence of Structural Change with an Application to OECD Unemployment. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 223–38. Amsterdam: Elsevier.
Ng, Serena, and Pierre Perron. 2001. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica* 69 (6): 1519–54.
O'Connell, Paul G. J. 1998. The Overvaluation of Purchasing Power Parity. *Journal of International Economics* 44 (1): 1–19.
Örsal, Deniz Dilan Karaman. 2007. *Comparison of Panel Cointegration Tests,* SFB 649 Discussion Paper 2007–029. Humboldt-Universität zu Berlin.
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.
Pedroni, Peter Louis. 1999. Critical Values for Cointegration Tests in Heterogenous Panels with Multiple Regressors. In *Testing for Unit Roots and Cointegration Using Panel Data: Theory and Applications,* ed. Amindya Banerjee, 653–70. Oxford: Blackwell.
Pedroni, Peter Louis. 2000. Fully Modified OLS for Heterogenous Cointegrated Panels. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 93–130. Amsterdam: Elsevier.
Pedroni, Peter Louis. 2004. Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the Purchasing Power Parity Hypothesis. *Econometric Theory* 20: 597-625.
Pesaran, M. Hashem. 2007. A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence. *Journal of Applied Econometrics* 22:265–312.
Pesaran, Hashem, Ron Smith, and Im Kyong So. 1996. Dynamic Linear Models for Heterogenous Panels. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications,* ed. László Mátyás, and Sevestre, Patrick, 145–95. Dordrecht: Kluwer.
Pesaran, M. Hashem, and Elisa Tosetti. 2011. Large Panels with Common Factors and Spatial Correlation. *Journal of Econometrics* 161:182–202.
Pesaran, M. Hashem, L. Vanessa Smith, and Takashi Yamagata. 2008. *Panel Unit Root Tests in the Presence of a Multifactor Structure.* Discussion Papers in Economics No. 2008/03. Heslington: Department of Economics and Related Studies, University of York.

Pesaran, M. Hashem, Yongcheol Shin, and Ron Smith. 1999. Pooled Mean Group Estimation of Dynamic Heterogenous Panels. *Journal of the American Statistical Association* 94 (446): 621–34.
Phillips, Peter C. B., and Hyungsik Roger Moon. 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67 (5): 1057–1111.
Phillips, Peter C. B., and Hyungsik Roger Moon. 2000. Nonstationary Panel Data Analysis: An Overview of Some Recent Developments. *Econometric Reviews* 19 (3): 263–86.
Phillips, Peter C. B., and Sam Ouliaris. 1990. Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* 58: 165–93.
Sethrapramote, Yuthana. 2004. *Panel Unit Root Tests with Structural Breaks.* Coventry: Department of Economics, University of Warwick.
Smeekes, Stephan. 2010. *Bootstrap Sequential Tests to Determine the Stationary Units in a Panel,* METEOR Working paper RM/11/03. Maastricht: Department of Economics, Maastricht University.
Sun, Yixiao. 2004. Estimation of the Long-Run Average Relationship in Nonstationary Panel Time Series. *Econometric Theory* 20 (6): 1227–60.
Tam, Pui Sun. 2006. *Breaking Trend Panel Unit Root Tests.* Faculty of Business Administration, University of Macau.
Trapani, Lorenzo. 2012. On the Asymptotic t-Test for Large Nonstationary Panel Models. *Computational Statistics and Data Analysis* 56: 3286–3306.
Urbain, Jean-Pierre, and Joakim Westerlund. 2011. Least Squares Asymptotics in Spurious and Cointegrated Panel Regressions with Common and Idiosyncratic Stochastic Trends. *Oxford Bulletin of Economics and Statistics* 73:119–39.
Vieraitis, Lynne M., Tomislav V. Kovandzic, and Thomas B. Marvell. 2007. The Criminogenic Effects of Imprisonment: Evidence from State Panel Data, 1974–2002. *Criminology and Public Policy* 6 (3): 589–622.
Wagner, Martin, and Jaroslava Hlouskova. 2007. *The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study,* Reihe Ökonomie/Economic Series 210. Vienna: Institut für Höhere Studien.
Westerlund, Joakim. 2006a. *Simple Unit Root Tests with Multiple Breaks.* Department of Economics, Lund University.
Westerlund, Joakim. 2006b. Testing for Panel Cointegration with a Level Break. *Economics Letters* 91:27–33.
Westerlund, Joakim. 2007. Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics* 69 (6): 709–48.
Westerlund, Joakim, and Syed A. Basher. 2008. Mixed Signals Among Tests for Panel Cointegration. *Economic Modelling* 25 (1): 128–36.
Westerlund, Joakim, and Jörg Breitung. 2013. Lessons from a Decade of IPS and LLC. *Econometric Review* 32 (5–6): 547–91.
Westerlund, Joakim, and David Edgerton. 2008. A Simple Test for Cointegration in Dependent Panels with Structural Breaks. Oxford *Bulletin of Economics and Statistics* 70 (5): 665–704.

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014272
IJCV: Vol. 8 (2) 2014

A Longitudinal Examination of the Effects of Social Support on Homicide Across European Regions

Kelly M. Thames, Department of Sociology, Appalachian State University, Boone, North Carolina
Patricia L. McCall, Department of Sociology and Anthropology, North Carolina State University, Raleigh, North Carolina

Vol. 8 (2) 2014

ISSN: 1864–1385

A Longitudinal Examination of the Effects of Social Support on Homicide Across European Regions

Kelly M. Thames, Department of Sociology, Appalachian State University, Boone, North Carolina
Patricia L. McCall, Department of Sociology and Anthropology, North Carolina State University, Raleigh, North Carolina

Since its introduction, social support theory has received generally consistent empirical support. Tests of social support theory have, however, mostly been cross-sectional and restricted to U.S. and Western European analyses. Measures of social support have tended to be inconsistent across studies and narrowly operationalized. The present project offers a longitudinal test of Cullen's (1994) social support theory using a more broadly defined measure of social support that is comparable across both Eastern and Western European countries. Using data gathered by Eurostat, this study applies "hybrid" regression panel analysis to test the effects of social support on homicide rates across European regions for 2000, 2005 and 2009. Results provide evidence for an effect of social support on homicide between Western and Eastern European regions and within those regions over time. The analyses also indicate that social support moderates the effect of economic deprivation on homicide across Western European regions, though not Eastern European regions.

In his presidential address to the Academy of Criminal Justice Sciences, Francis T. Cullen (1994) proposed a theory to provide an organizing framework for the field of criminology. While it has often been neglected by criminologists, the concept of *social support* has implicitly informed criminological theory since the early twentieth century, he argued. Cullen's ideas are drawn primarily from the work of the scholars of the Chicago School, who emphasized that "organized networks of human relations can assist people in meeting both expressive and instrumental needs" (Colvin, Cullen, and Vander Ven 2000, 24). While these traditional theories tend to focus on the deleterious effects of the breakdown of human relations networks (in other words, the negative phenomena that cause crime), Cullen shifts his focus to the forces that work to maintain, and even strengthen, these networks (the positive phenomena that work to prevent crime). Cullen conceptualizes these positive phenomena as social support, which, he argues, can explain variation in levels of social control, individual involvement in crime, and aggregate crime rates (Cullen 1994; Chamlin and Cochran 2003). Specifically, according to Cullen's theory, social support is hypothesized to be negatively associated with crime (Cullen 1994).[1]

The potential buffering effect of social support in the form of economic assistance – the most popular conceptualization of the concept – is of salient concern to criminological scholars interested in investigating the effects of global neoliberalization on cross-national rates of violent crime. Since the late 1970s, governments worldwide have adapted to growing post-industrial economic instability by way of instituting neo-liberal economic and social policies, which necessitate the retrenchment of social welfare programs (Harvey 2005; Esping-Andersen 1996). Following this worldwide neoliberal trend, the traditionally social democratic nations of Western Europe and the historically socialist nations of Eastern

This work was supported in part by the Harry F. Guggenheim Foundation. We would like to thank our anonymous reviewers for their constructive and insightful input.

1 Since Cullen's (1994) introduction of social support theory, the theory has been expanded to incorporate the theme of coercion. While Colvin, Cullen, and Vander Ven's (2002) differential social support and coercion theory provides a valuable theoretical expansion of Cullen's (1994) original formulation, the current project, along with much of the scholarly research investigating the effects of social support on crime, will focus exclusively on Cullen's social support paradigm.

Europe have been compelled to restructure social policy in an effort to maximize economic growth and competitiveness in the global economy (Esping-Andersen 1996). This restructuring has often involved the dissolution of the social and economic safety nets upon which the citizens of these countries have depended. Throughout this period of increasing austerity, European nations have seen growing levels of unemployment, poverty, and inequality (Esping-Andersen 1996; Standing 1996). During this same period, neo-liberal adaptation has weakened the institutionalized social support that, according to social support theory, should work to ease economic strain. For example, from 2001 to 2008, total unemployment benefits paid to citizens in the European Union decreased by approximately 0.4 percent (Eurostat 2014). The rate of change across countries, however, varies as some countries, particularly those of Eastern Europe, have seen much steeper decreases in expenditures on unemployment benefits. For instance, Poland has seen an 8.1 percent decrease in unemployment expenditures throughout this time period; Slovenia reports a 6.1 percent decrease in unemployment expenditures; and Romania has seen a 3.5 percent decrease (Eurostat 2014). Following the logic of social support theory, then, we should expect European crime rates to increase in association with the shrinking levels of social welfare across the continent. As such, the aim of the current project is to offer a longitudinal examination of social support theory in the European context.

An assessment of the body of literature examining Cullen's (1994) social support theory reveals that the theory and its underlying concepts have enjoyed generally consistent empirical support. While tests of theories related to social support theory (such as social disorganization, collective efficacy, social capital, social bonds, and institutional anomie) have provided partial support for social support theory, there have been relatively few direct tests of the theory (Kim and Pridemore 2005). To date, research by Chamlin and Cochran (1997), Chamlin, et al. (1999), Pratt and Godsey (2002, 2003), and Kim and Pridemore (2005) constitute the body of studies framed as direct empirical examinations of social support theory.

Although the majority of these studies offer evidence supportive of social support theory, further empirical examination of the theory is warranted. For example, the statistically null findings reported by Kim and Pridemore (2005) in their examination of social support in Russia highlight the need to further explore the effects of social support within transitional, unstable political and economic contexts (such as post-communist Eastern Europe) – a cross-national context not yet explored by scholars. What is more, these studies do not offer a consistent measure of social support and the measures used to test the theory tend to be rather narrowly conceptualized. And perhaps most importantly, extant tests of the theory employ cross-sectional data, which fail to capture the dynamic nature of the relationship between social support and crime over time.

In light of these limitations, the present project contributes to this body of research in a number of ways. Beyond testing social support theory among European countries, the present study also contributes methodologically to extant analyses of the relationship between social support and homicide. First, it offers a test of Cullen's (1994) original formulation of social support theory based on a more inclusive measure of the concept that comprises both public and private contributions and, therefore, incorporates dimensions of social support generally not considered in prior research. The measure of social support employed herein is also comparable across Eastern and Western Europe. Second, this study offers a cross-national test of social support theory at a level of aggregation lower than the country-level. Specifically, this study utilizes data for European regions, which allows one to take advantage of intra-country variation in levels of social support and crime, thereby extending cross-national studies of crime beyond the country-level (which currently dominates cross-national homicide research). This allows the researcher to assess the robustness of extant studies using country-level data to determine whether country-level findings apply to lower levels of analysis. Third, the present study offers a cross-national analysis of twenty-three European countries – moving beyond Western European countries typically represented in European studies of crime, to include Eastern European countries, as well. Fourth, the present study utilizes historical data, which allow for examination of the dynamic nature of changing levels of social support on crime rates over time. Therefore, the present study examines the relationship between social support and homicide across 247 European regions at the time points: 2000, 2005 and 2009, representing a total of 605 region-years. Eurostat's data archive is a rich source for social and economic indicators for European regions used for these analyses as they provide information for various levels of aggregation at the region-level as well as for countries and cities.[2] Finally, as opposed to using a conventional panel model with a fixed or random effects regression design, we employ a "hybrid" regression model to estimate the unique effect of social support on homicide (1) across regions of Europe and (2) within those regions over time. The hybrid model allows for the estimation of both the "over time" effects of social support on homicide – that is, the within-region or region effect over time, and the effects of social support on homicide across regions – that is, the between-region, comparable to cross-sectional, effects (Allison 2005; Phillips 2006). The benefits related to these methodological issues are elaborated in related discussions below.

1. Social Support Theory

While the concept of social support is implicit in traditional theories and underlies a number of contemporary criminological theories, including institutional anomie (Messner and Rosenfeld 1993), collective efficacy (Sampson, Raudenbush, and Earls 1997), and general strain theory (Agnew 1992), Cullen offers the most precise interpretation of the concept and of the foundational assumptions of these theories. Although many theories following this tradition assume that social support works to alleviate crime, Cullen makes this assumption explicit. Simply stated, Cullen argues that social support – in any form – reduces crime rates at both the aggregate and individual levels.

Drawing from extant analyses of the concept (House 1981; Lin et al. 1986; Vaux 1988), Cullen (1994), quoting Lin, defines social support as "the perceived or actual instrumental and/or expressive provisions supplied by the community, social networks, and confiding partners" (Lin 1986:18). Following this definition, social support can exist at both micro and macro levels of society and may be delivered formally or informally. Micro-level social support can come from a variety of social relationships, including family and friendship and can provide both instrumental supports, such as financial support/advice, and expressive supports, such as companionship. Macro-level support, on the other hand, originates from social networks, communities, and/or larger ecological units (Cullen 1994), and can include expressive supports received through networks and communities, such as support groups or clubs created around common interests, and instrumental support received through private organizations and/or the government, for example, welfare payments or complimentary financial advising. Informal delivery of social support occurs through relationships with individuals not affiliated with any state/official agency, while formal social support is delivered through formal organizations, such as schools, government welfare programs, and even the criminal justice system.

The crux of Cullen's thesis (1994) is the hypothesis that all forms of social support are negatively related to criminal behavior. Cullen suggests that social support might reduce criminal involvement in a variety of ways, including: reducing criminogenic strains (also see Cullen and Wright 1997); fostering effective parenting and nurturing strong family units; supplying both the human and social capital required to desist from crime; creating opportunities for prosocial modeling; strengthening informal and formal social control; and reducing opportunities for victimization. In addition to the direct effects social support has on reducing criminal involvement, and more pointedly relevant to macrolevel analyses, Cullen (1994) and Chamlin and Cochran (1997) note that the relationship between economic deprivation (poverty, economic inequality) and crime should be more pronounced in areas with limited social support (Pratt and Godsey 2003). They explain that, in addition to the established criminogenic effects of econ-

2 The current Eurostat archive contains region-level data for thirty-five European countries between 1990 and 2013, drawing on widely available data from country statistical agencies such as population totals. Data are more readily available between 2000 and 2010 for indicators used in these analyses. Data for many of these regions are missing, particularly for the social support indicator and especially for many Eastern European countries for the years leading up to 2000. Data for some regions are not available until 2006.

omic deprivation, social support should diminish the deleterious effects of economic deprivation on crime; that is, areas with high levels of social support will inhibit the impact of deprivation on crime and areas with low levels of social support will amplify the influence of economic deprivation on crime (Chamlin and Cochran 1997; Cullen 1994; Pratt and Godsey 2003). Therefore, the theoretical mechanisms outlined by Cullen imply both a direct relationship between social support and crime and a moderating relationship through the capability of social support to reduce the impact of criminogenic strain.

The aspects of Cullen's theory upon which the present study focuses include macro-level instrumental social support delivered by both formal and informal means. These institutionalized social supports are typically manifested in government welfare programs such as assistance to the unemployed, elderly, disabled, and family dependents. Basic healthcare also protects residents from financial hardship and poverty when costly medical treatment is required. Government-subsidized daycare supports single-parent households and households requiring two sources of income. Agencies often provide opportunities to acquire subsidized housing, and benefits are sometimes available to immigrant populations who are at risk of social exclusion and isolation. These benefits are provided by national and local government agencies, as well as by private organizations seated at both the local/community and nation levels, the level of development of which may indicate the extent to which the philosophy of social support has been institutionalized. As such, we are interested in the social supports available to individuals through government programs and both public and private community-level agencies, which work to reduce economic strains and provide individuals with human and social capital. The existence of programs and agencies responsible for providing social benefits allows individuals to anticipate assistance during times of economic downturn, and stress associated with financial hardship can be moderated by these systems of institutionalized social support.

1.1. Empirical Tests of Social Support

Relatively few studies have offered direct empirical tests of social support theory. Among economic indicators examined as explanations of crime rates, however, social support has received the most consistent theoretical support (Stamatel 2009). With the exception of the work of Chamlin et al. (1999), who found a positive relationship between social support and U.S. violent crime rates, and the work of Kim and Pridemore (2005), who found no association between social support and homicide in Russia, the results of these studies are consistent with the expectations of social support theory. Regardless of conceptualization and measurement, social support has been found to be statistically significant and negatively related to homicide rates (DeFronzo 1983, 1997; Messner and Rosenfeld 1997; DeFronzo and Hannon 1998; Savolainen 2000; Pratt and Godsey 2003).

Consistent with Cullen's social support theory, Messner and Rosenfeld (1997) demonstrated how levels of government social support were negatively related to homicide rates among a sample of countries using 1990 data. According to Messner and Rosenfeld's (1997) institutional anomie theory (IAT), the American economic institution dominates social life in such a way that it limits the ability of other institutions to insulate individuals from the pressure to achieve economic success by any means. In their cross-national test of IAT, the decommodification index, a measure of the ability of governments to insulate citizens from deleterious market forces, is negatively related to homicide rates among forty-five countries. Messner and Rosenfeld attempted to incorporate Esping-Anderson's concept of decommodification into their index, which includes three general dimensions of social support: (1) absolute and relative levels of expenditure for social support programs; (2) the sources of funding for those programs; and (3) the distribution of funding across types of social support programs (for instance, unemployment expenditures, family/dependents expenditures, workers' compensation, etc.). These dimensions are operationalized by way of an index comprised of social welfare expenditures as a percentage of GDP, annual benefits payments per capita, and the percentage of expenditures allocated to employment injuries. Similarly, in a re-examination of Messner and Rosenfeld's data and test of institutional anomie theory, Savolainen (2000) reported a significant negative relationship between homicide and welfare as it interacts with inequality.

Pratt and Godsey (2003, 621) confirm these earlier findings, revealing in a more comprehensive examination of forty-six countries that the percentage of total GDP spent on healthcare – a measure argued to represent the value placed on social institutions that may work against the criminogenic effects of "certain social-structural arrangements" – is negatively related to country-level homicide rates. Pratt and Godsey's measure of social support represents (1) the financial relief upon which a citizen can rely from their government when a family member requires medical attention, and (2) the extent to which the government allocates a proportion of the country's GDP to welfare benefits for its citizens. The former relates to the individual impact social support has on recipients and the latter represents the relative importance in governmental spending patterns. Pratt and Godsey also find empirical evidence for the moderating influence of social support as it acts to relieve the deleterious effect of economic inequality on homicide rates.

These generally consistent findings at different points in time and across various levels of analysis lend confidence to the validity of social support theory as a social force affecting crime rates (both directly and indirectly). Nevertheless, an examination of the theory in an even wider variety of political and economic environments and using a more generalized measure of social support is warranted. As explained below, such exploration will allow for the investigation of social support theory's generalizability across time and social environments.

2. Dynamic Effects of Social Support across Europe

As outlined above, the present study investigates the effects of social support on crime rates across regions within Europe – including European Union members, candidate countries, and members of the European Free Trade Association. The countries investigated in both Western European and post-communist Eastern European states, which is significant due to the differences in their economic and political conditions before and since the fall of communism in 1989. The transition from socialism to a democratic market economy was severely disruptive, as the economic transformation led to mass unemployment, rising mortality, and alarming increases in poverty and inequality (Kim and Pridemore 2005; Stamatel 2009; Standing 1996). Following a global trend of neoliberalization, Western European countries have also experienced a turbulent economic and social policy transitions (Esping-Andersen 1996; Harvey 2005). However, unlike Western European countries, which have been able to rely on institutionalized welfare programs (despite rising unemployment and austerity measures that have reduced welfare support), significantly weakened Eastern European governments have been unable to quell intensifying economic deprivation (Esping-Andersen 1996).

Social support theory should explain variation in crime rates across these varied political and economic contexts. Although all of the European countries included in the current analysis are facing economic and political challenges, the degrees to which their economic prosperity and welfare policies are strained by the changes vary. This variation provides an excellent opportunity to test the effectiveness of social support to reduce crime rates in a variety of economic climates. Moreover, if social support theory is to be upheld, regional levels of social support should also explain the variation in crime rates across time; changes in levels of social support should be negatively associated with changes in rates of crime. Therefore, the present study examines the effects of social support across three time points – 2000, 2005, and 2009 – among a sample of Eastern and Western European regions. The current analysis is restricted from examining more recent time points because of limited data availability for the 2010 time period (at the time analyses were conducted, data were not available for 2010).

3. Hypotheses

While social support theory applies to both individual and higher levels of aggregation, this analysis restricts itself to instrumental social support applied at the macro-level and delivered by government and private agencies. The following hypotheses are derived from the conceptual discussion:

> *H1: The association between region social support and crime will be negative. This refers to the direct relationship between social support and crime across regions.*

H*2: The association between intra-region change in social support and change in crime will be negative. This refers to the direct relationship between social support and crime over time, within regions.*

H*3: Between regions, social support will moderate the relationship between economic deprivation and crime: the effect of economic deprivation on crime will be less pronounced in regions with high levels of social support. This refers to the interaction between social support and economic deprivation across regions.*

H*4: Within regions, social support will moderate the relationship between economic deprivation and crime: as regional levels of social support increase, the effect of economic deprivation on crime will become less pronounced. This refers to the interaction between social support and economic deprivation over time, within regions.*

Hypotheses H_1 and H_3 concern the universality of the relationship between social support and homicide rates (across the varied political climates of European countries). Hypotheses H_2 and H_4 specify the effects of social support on homicide rates over the time frame (2000, 2005, and 2009).

4. Data and Methods

4.1. Data Source and Sample

All data included in this analysis are from Eurostat. As far as possible, Eurostat's data are standardized across countries (Eurostat 2014). One of the great advantages of Eurostat is the availability of data at sub-national levels of aggregation, which enables a cross-national test of social support theory at the region level. This allows the researcher to take advantage of variation in both the independent and dependent variables across these regions – that is otherwise masked in country-level measures. The units of analysis for this study are therefore regional areas of European Union member and candidate nations and EFTA countries.[3]

In addition to the availability of data for subnational levels of aggregation, yet another advantage to Eurostat data is the availability of data from Eastern European nations. While Pratt and Godsey's (2002, 2003) cross-national analyses included several nations outside of Europe, their sample did not include any Eastern European nations. Kim and Pridemore (2005) offered an analysis of the effects of social support on homicide rates in Russian regions but did not examine social support theory in any other post-communist contexts. Fortunately, Eurostat currently offers data from many Eastern European nations. Although the limited availability of comparable data necessitates the omission of much of the former Soviet Bloc, the countries included in this analysis represent a variety of economic and political climates.

While the availability of regularly updated data from both Western and Eastern European countries allows for an investigation of the effects of social support across a variety of political and economic contexts, the data available through the Eurostat archives are by no means complete. Therefore, the sample of regions included in the present study has been significantly restricted by the limits of Eurostat data (particularly at lower levels of aggregation).[4] Furthermore, because Western and Eastern European countries have distinct political and economic histories, the sample of European regions is divided according to a Western/Eastern categorization and examined separately. Preliminary analyses employed a dichotomous measure for Eastern European regions, but this measure was omitted in the final analysis (in favor of the split sample) due to its collinearity with the social support measure, GDP per capita, and the percent of the population aged 65 and over. After accounting for listwise deletion of cases and omitting influential outliers, the two samples include 197 Western European regions with 487 region-years and 50 Eastern European regions with 118 region-years across the three time points.

Table 1 offers an account of the number of regions in each country for each time point that are included in the analyses. Of the 35 countries in Eurostat's archives reporting population data, Austria, the Czech Republic, Germany, the Netherlands, Poland, Spain and Sweden provide more

3 Eurostat regional statistics are organized under the "Nomenclature of Statistical Territorial Units" (NUTS) classification system. The current project utilizes statistics documented for NUTS level 2 regions, hereafter referred to simply as "regions" (Eurostat 2014).

4 Region-level homicide rate indicators are available for thirty countries starting circa 1995 and ending 2009, and there are twenty-seven countries represented in the region-level data for the social benefits measure.

complete representation of region-level data (at least 80 percent) for our indicators of interest across all study time periods. Twelve countries represented in the Eurostat data holdings are omitted from our analyses because of a lack of complete data across the study years. These are Croatia, Cyprus, Denmark, Iceland, Lichtenstein, Luxembourg, Macedonia/Yugoslavia, Malta, Montenegro, Slovenia, Switzerland and Turkey.

Table 1: European regions (NUTS Level 2) represented in analyses and total number of regions

	2000	2005	2009
Austria	7/9	9/9	9/9
Belgium	4/11	11/11	11/11
Bulgaria*	0/6	6/6	4/6
Croatia*	0/3	0/3	0/3
Czech Republic*	7/8	8/8	7/8
Cyprus	0/1	0/1	0/1
Denmark	0/5	0/5	0/5
Estonia*	0/1	0/1	1/1
Finland	1/5	1/5	4/5
France	20/26	21/26	21/26
Germany	32/39	36/39	36/39
Greece	0/13	0/13	13/13
Hungary*	0/7	7/7	7/7
Iceland	0/1	0/1	0/1
Ireland	0/2	0/2	2/2
Italy	11/21	0/21	18/21
Latvia*	0/1	0/1	1/1
Lichtenstein*	0/1	0/1	0/1
Lichtenstein*	0/1	0/1	1/1
Luxembourg	0/1	0/1	0/1
Macedonia*	0/1	0/1	0/1
Malta	0/1	0/1	0/1
Montenegro	0/1	0/1	0/1
Netherlands	12/12	12/12	12/12
Norway	0/7	0/7	7/7
Poland*	15/16	15/16	15/16
Portugal	2/7	5/7	6/7
Romania*	0/8	6/8	6/8
Slovakia*	0/4	4/4	3/4
Slovenia*	0/2	0/2	0/2
Spain	16/19	18/19	18/19
Sweden	6/8	8/8	8/8
Switzerland	0/7	0/7	0/7
Turkey*	0/26	0/26	0/26
United Kingdom	28/32	31/32	30/32

* Eastern European countries (plus Turkey)

While Table 1 clearly illustrates the limitations of the Eurostat data holdings for the purposes of this study, the final sample remains substantial and represents countries characterized by widely varying political and economic characteristics. The study sample consists of 247 regions within twenty-three countries (fourteen Western and nine Eastern European). 162 regions are included for the year 2000; 200 regions for the year 2005; and 243 regions for 2009, representing a total of 605 region-years. Fortunately, nine of the twelve Eastern European countries remain in the sample: Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovakia. The remaining regions are located in fourteen Western European countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom. Whereas the findings from this study may not be generalizable across all European countries, these regions provide a good sample of regions across Europe.

4.2. Dependent Variable

The primary focus of this study is on the effects of social support on rates of homicide across regions within European countries. Because homicide is defined most similarly across countries, it is considered to be the most appropriate measure of violent crime for cross-national studies (LaFree 1999). Nevertheless, there are some differences across European countries regarding police recording practices. Therefore, the number of homicide victims obtained from cause of death statistics will serve as the measure for homicide.

Eurostat provides homicide statistics in the form of cause of death data, which are classified according to the *International Classification of Diseases* codes published by the World Health Organization (Eurostat 2014). Consequently, Eurostat data are equivalent in quality to those of the World Health Organization, the database widely considered to be the most reliable and valid source of data for cross-national studies of homicide (Kalish 1988; LaFree 1999). Eurostat provides age-standardized homicide rates for three-year averages, and pertinent to our study, circa the three time periods: for 2000 (averaging 1999 to 2001 rates), 2005 (2004 to 2006) and 2009 (2008 to 2010). Our time frame for analysis is truncated to 2009 as it is the last year included in that series. Three-year averaging avoids the overly inflated and/or deflated rates that result from extreme yearly fluctuations not uncommon among rare events such as homicide. Furthermore, age-standardization allows for the comparability of homicide rates across countries, as the measure acts as a control for each country's age structure.

Detailed descriptive statistics for homicide and all predictor variables are presented in Table 2 for Western and Eastern regions for each time point. Over this ten-year period, homicide rates across Western regions averaged .7 homicides per 100,000 population with a standard deviation of .5, slightly decreasing over the study period. In Eastern regions, the homicide rate was 2.5 times higher, with an average across the time period of almost 1.9 homicides per 100,000 population, decreasing slightly between 2005 and 2009. The standard deviation averaged approximately 1.3 for the Eastern regions. The covariates comprising our model specification are described below.

Table 2. Descriptive Statistics for European Regional Homicide Rates and Predictor Variables

	2000		2005		2009	
	Western Europe	Eastern Europe	Western Europe	Eastern Europe	Western Europe	Eastern Europe
Variables in models						
Standardized homicide rate (3-year average)	.77 [.46] (.2, 3.2)	1.96 [.76] (.9, 3.4)	.71 [.59] (.1, 4.7)	1.97 [1.41] (.4, 10.0)	.67 [.48] (0, 2.6)	1.73 [1.41] (.4, 7.1)
Logged homicide rate	.19 [.31] (-.4, 1.3)	.86 [.31] (.3, 1.4)	.11 [.38] (-.5, 1.6)	.81 [.40] (-.1, 2.4)	.09 [.36] (-.7, 1.1)	.68 [.46] (-.1, 2.0)
Social benefits[a] per capita (t-3) (in thousands of euros)	3.400 [.956] (.959, 5.380)	.409 [.051] (.334, .482)	4.377 [1.072] (1.549, 6.242)	.677 [.315] (.131, 1.124)	5.169 [1.332] (2.293, 8.854)	1.010 [.406] (.3, 1.7)
GDP per capitaa (in thousands of euros)	19.352 [5.581] (9.987, 58.370)	9.207 [3.777] (5.617, 23.912)	24.793 [7.181] (14.040, 78.001)	11.82 [6.07] (6.0, 37.3)	27.318 [8.297] (16.057, 87.797)	15.852 [6.268] (8.476, 46.428)
Unemployment rate	6.82 [4.04] (1.5, 20.1)	12.50 [5.09] (3.6, 21.0)	7.42 [3.74] (2.9, 21.7)	11.55 [5.30] (3.1, 22.4)	8.21 [4.05] (1.9, 25.6)	9.07 [3.74] (3.0, 20.9)
Sex ratio	.96 [.02] (.88, 1.01)	.94 [.02] (.90, .96)	.96 [.02] (.91, 1.02)	.94 [.02] (.85, .96)	.96 [.03] (.90, 1.03)	.94 [.03] (.85, .98)
Percentage aged 65 years and over	16.32 [2.30] (8.8, 25.0)	12.56 [1.58] (10.3, 16.3)	17.16 [2.54] (8.7, 23.2)	14.29 [2.00] (10.6, 21.3)	18.00 [3.04] (8.8, 27.1)	14.96 [2.10] (11.0, 22.1)
Variables not in models						
Social benefits[a] per capita (t) (in thousands)	3.886 [.990] (1.284, 5.699)	.693 [.068] (.591, .842)	4.986 [1.101] (2.033, 6.856)	.892 [.378] (.264, 1.566)	6.008 [1.349] (3.007, 10.922)	1.492 [.508] (.633, 2.227)
Population size (in thousands)	2139.0 [1615.9] (268, 11,020)	2014.4 [1162.0] (1007, 5113)	1907.7 [1523.8] (65, 11,442)	1866.4 [964.9] (593, 5146)	1908.3 [1680.6] (73, 11,728)	1900.8 [961.1] (882, 5204)
n (listwise)	139	23	152	48	196	47

Note: [a] In constant 2005 Euros. GDP multiplied by negative one is the of measure of economic disadvantage for model estimation.

4.3. Independent Variables

The independent variables are measured at three time points – 2000, 2005 and 2009 – except for the key concept of interest, social support. It is measured as social benefits expenditures per capita and has been entered as a three-year lagged measure for each time point – that is, for 2000 (1997), for 2005 (2002) and for 2009 (2006).[5] This measurement specification is informed by McCall and Brauer (2014, 94, 101), who provide evidence that the effects of social support may have a lingering rather than an immediate or contemporaneous influence on homicide rates (Messner and Rosenfeld 1997). Therefore we estimate a series of alternative lag specification models with contemporaneous as well as one-, two-, and three-year lagged social support measures. Appendix A displays the substantive differences across these alternative models, which are discussed below.

As social support theory does not explicitly suggest a particular operationalization of social support, previous studies testing social support theory have offered a variety of measures representing the concept.[6] Scholars have typically measured social support in the form of support provided by the government as described above. While, as a whole, the measures of social support employed by these scholars are somewhat diverse, taken independently, the measures are fairly limited in their operationalization of the concept. Most studies offer only one aspect of the variety of support that can be institutionalized in a society, such as healthcare and education expenditures (Pratt and Godsey 2003; Kim and Pridemore 2005). The measure of social support provided by Eurostat allows for a broader operationalization of the concept – that is, a standardized measure of the total annual social benefits expenditures per capita (reported in thousands of euros), which is defined as "all interventions from public and private organizations to relieve households and individuals of the burden of a defined set of risks or needs" (Eurostat 2008, 9). These risks/needs include: sickness/health care, disability, old age, survivors, family/children, unemployment, housing, and "social exclusion not elsewhere classified" (Eurostat 2008, 9). This measure allows the present analysis to reliably account for a wide range of sources of social support in each region, which include supports provided by both national and subnational public and private organizations.[7] To further control for inflation across time periods, the social support measure employed herein has been transformed to reflect constant 2005 euros.[8] Across the three time points, social benefits averaged 4,417 per capita in Western regions and 758 per capita in Eastern European regions. Refer to Table 2 for details across each time period.

Following previous cross-national studies of homicide and tests of social support theory, classic structural covariates of homicide are included in the analyses (Chamlin and Cochran 1997; Kim and Pridemore 2005; LaFree 1999; Pratt and Godsey 2003, 2002). These variables include indicators of economic prosperity and economic strain – measured in the present study using Gross Domestic Product (GDP) Purchasing Power Standard per capita and the percent of males aged 15 and over who are unemployed, respectively; the sex ratio (total males to total females), and percentage of total population aged 65 years and over. The average population size of all sampled regions was

5 Data for the social support measure is available for the majority of Western European countries beginning in 2000 but not available for some of the Eastern European countries until 2006, thereby accounting for a good deal of our missing cases. Note that the current Eurostat data holdings no longer provide data that were available for earlier years in the time series; therefore, we have retrieved data available from an earlier version of the Eurostat archive for 1995 social benefit spending and use it to interpolate social benefit data for 1997 through 1999 for the present analyses.

6 Studies of the effects of economic deprivation on homicide have included welfare support as another indicator of the economic needs of an area. Although theoretical rationale makes this assumption plausible, the current project aims to control for the conceptualization of economic deprivation through the inclusion of two measures which are both negatively correlated with the measure of social support.

7 Cullen's ideas about the macro effects of social support may be realized from the very existence of government (and private) programs and agencies which provide benefits in times of need. Consistent with that logic, a region rich in institutionalized social support available to various demographic groups is one in which residents can anticipate assistance when the need arises, thereby reducing general levels of stress as well as strain related to economic hardship. Eurostat's measure only offers an overall total measure of support and does not offer information by type; therefore, we are not able to include specific types of support which may seem more obviously connected to reducing homicides, such as unemployment and social exclusion.

8 The European Council uses the Harmonized Indices of Consumer Prices (HICP-CPI), which is comparable to our Consumer Price Index http://ec.europa.eu/eurostat/cache/metadata/en/prc_hicp_esms.htm.

1,900,000, and ranged from about 65,000 to 11,700,000 in Western European regions and from almost 60,000 to over 5,200,000 in the Eastern regions. These provide a wide variation in populations represented across these regions.

Scholars have struggled to incorporate valid indicators of economic deprivation or impoverishment in cross-national research (Messner et al. 2010). Hence, due to data limitations and collinearity problems characteristic of cross-national measures of absolute and relative deprivation, cross-national studies have most often included indicators of overall economic development (such as GDP and/or the human development index) and/or measures of relative deprivation (such as the Gini index) (Pridemore 2008; Messner et al. 2010). While Messner et al. (2010) find that measures of relative deprivation better predict cross-national rates of homicide, Eurostat does not supply the income-based data necessary to compile such measures at the regional level. Therefore, we are not able to capture this aspect of economic strain as a predictor of homicide rates in these analyses. As sufficient region-level measures of income inequality or poverty are unavailable from Eurostat, GDP per capita and male unemployment are included as traditional, cross-national measures of economic prosperity and economic strain.[9] In the present study, GDP is multiplied by -1 – henceforth, referenced as "negative GDP" – and represents the economic disadvantage of a European region. This is done to create an indicator consistent in sign with the posited direction of the relationship between economic deprivation and crime, and also eases the interpretation of the findings.

To test the moderating influence of social support on economic disadvantage as posited in hypotheses H_3 and H_4, an interaction term, using the product of social benefits per capita and negative GDP per capita, is incorporated into the analyses.[10] According to the conceptual discussion, support for this moderating mechanism of social support will be demonstrated with a negative, statistically significant coefficient for this interaction term. That is, the positive relationship between economic disadvantage and homicide will be diminished in regions with high levels of social support.

Finally, the percentage of the population aged 65 and over is included to control for growing aging populations that are likely to have great social support needs. As a reflection of the unique needs of elderly individuals, those countries with larger populations of individuals falling within the elderly age groups may have higher demands and, thereby, offer higher levels of social support.

4.4. Preliminary Analyses

Indications of both heteroskedasticity and collinearity among variables included in the analyses led to concerns over model specification and data transformations. An examination of residuals plotted against fitted values generated using ordinary least squares regression at each cross-section (2000, 2005, and 2009) led to the detection of heteroskedasticity, the correction for which involved the log transformation of the homicide rate (a common transformation in aggregate-level studies of homicide). Residual analysis conducted after log transformations indicated no patterns of unequal error variance. Additionally, inspection of bivariate correlation matrices (available upon request) indicates moderately high correlations among some of the study variables. One would anticipate that regions with high levels of social need (including high rates of poverty) are likely to exhibit high levels of social support.[11] Not surprisingly, strong correlations are found among these variables (especially between social support and negative GDP per capita). Even though the highest bivariate correlation is only .55, results of models presented herein are interpreted with caution to ensure the unique effects of predictor variables are identified and not masked by the effects of other highly correlated predictors. An analysis of variance inflation factor values (VIF) estimated for each time period

9 Eurostat's household income per capita indicator was considered for our measure of economic deprivation, but was too highly correlated with the other more conventional measures of economic wellbeing.

10 GDP per capita is used as a measure of economic prosperity, but for conceptual consistency with social support theory, GDP is multiplied by –1 to represent economic disadvantage for the region and, as such, serves as the component of the interaction term. This was chosen over using unemployment for the interaction measure because GDP is arguably a more reliable measure than unemployment.

11 Negative bivariate correlations between the homicide rate and social support provides initial support for the hypotheses and also indicates that social support is not an indicator of a region's economic deprivation, which would be positively correlated with homicide.

indicated that no VIF value exceeded 4, suggesting that multicollinearity is not an issue. Cook's distance values were examined and influential outliers were identified only in the Eastern region sample; therefore, cases with Cook's D values greater than the cutoff (4/n) were omitted from the Eastern European region analyses.[12]

4.5. Statistical Technique

In order to test the hypothesized relationships between social support and homicide, a series of panel models (also referred to as pooled time series), or more specifically, "hybrid" panel analysis regression models, were estimated. The Eurostat data provide for measures of change over the five-year period between 2000 and 2005 and for the four-year period between 2005 and 2009; recall that the available homicide rate data limited the time series for our analyses. Although greater detail and variation over time is afforded with annual time series analyses, which would capture the more nuanced covariation of trends between social support and homicide rates, the panel model allows the researcher to estimate change among regressors and avoids statistical challenges associated with annual time-series analyses, such as meeting assumptions of stationarity and serial independence (Ostrom 1990). In addition, limitations of data availability in cross-national research make the panel design attractive, as despite the absence of annual measures of social and economic indicators, researchers are able to model change over time. Researchers using a panel model design are nevertheless faced with issues related to assumptions of independence of error terms and omitted variable bias.

Fixed effects and random effects regression models are the two more commonly used methods for panel studies, or the analysis of cross-sectional time-series data – that is, data characterized by multiple measures of units over time (Allison 2005; Phillips and Greenburg 2008). Each of these models, however, suffers from significant limitations. Fixed effects models allow only for estimation of the within-region over-time effects of social support on homicide – treating the between-region effects as fixed and estimable. One benefit of the fixed effects model is its ability to control for unobserved (stable) traits. Here, dummy measures for each case (minus one) are used to control for unobserved, stable traits and can serve as a substitute for omitted variables, hence relieving problems associated with omitted variable bias.

Random effects models treat the between-region effects as independent and randomly distributed, estimating parameters that represent the combined effects of between- and within-region components (Phillips 2006). One condition of random effects models that is difficult to satisfy is that the error term is not correlated with any of the independent variables in the model (omitted variable bias). Therefore, many researchers opt for using the fixed effects model design. Yet, neither fixed effects nor random effects models allow the researcher to estimate the unique between-region *and* within-region over-time effects of regressors (Phillips 2006, 956–57).

In order to bypass the limitations of fixed effects and random effects regression models and following extant criminological literature, the present study employs a "hybrid model" (Allison 2005; Horney, Osgood, and Marshall 1995; Phillips 2006; Ousey and Wilcox 2007). The hybrid model allows for the estimation of parameter coefficients that are equivalent to those yielded by the fixed effects model (within-unit over time estimates, which are net of the effects of time-invariant characteristics of regions) and, unlike the random effects model, allows for the separation of these within-region effects from between-region effects. The hybrid model, then, takes the following form:

$$y_{jt} = \alpha + \beta X_j + \eta(x_{jt} - X_j) + \nu_j + \varepsilon_{jt}$$

The dependent variable y_{jt} represents the logged, age-standardized homicide rate for region *j* and year *t*, where signifies the intercept, β indicates the parameter estimates for the between-region component, X_j represents the mean

12 Eastern region influential outliers which were excluded from the analyses are, for 2000: LT00, LV00, PL31; for 2005: EE00, LT00, RO21, RO32; and for 2009: BG32, PL51, RO21, RO32, SI01, SI02, SK01. See Eurostat (2014) for region names associated with these region codes.

values over time for the predictors for region *j*, η represents the parameter estimates for the effects of the within-region component, and x_{jt} represents the predictor for region *j* at time *t* (Bryk and Raudenbush 1992; Johnston and DiNardo 1997; Judge et al. 1985; Phillips 2006). The region-specific error term is represented by v_j, while the ε_{jt} denotes the model error term that contains the random variation within regions over time. The inclusion of v_j in the model acts as a control for unique, region-specific characteristics, such as war or other political and/or economic transitions and also acts to correct for omitted variable bias as mentioned earlier.

In order to employ the hybrid model approach, the time-varying predictors must be separated into their respective between-region and within-region components. The between-region component of each predictor is acquired by calculating a mean score for each region – regional scores are averaged over the three study years (denoted X_j). This between-region component offers an examination of the effect of predictors across place; in other words, the between-region component is comparable to a cross-sectional analysis. The within-region component of each predictor is computed by calculating the difference between the value of the predictor at each time point and the mean score of the predictor for each region over the three time points (denoted $x_{jt} - X_j$). Distinct from the between-region component, the within-region component of the hybrid model offers an estimation of the effect of explanatory variables across time. Both the between-region and within-region components are included in a random-intercept regression model predicting the logged, age-standardized homicide rate. Additionally, in order to better control for possible year effects, dummy variables representing 2005 and 2009 are included in the models (2000 is omitted as reference year). Tests of hypotheses H_1 and H_3 are made possible through the between-region components of this model, as estimates indicate the effect of social support across regions. The within-region component of this hybrid model provides the tests for hypotheses H_2 and H_4, as estimates indicate the effects of social support within regions over time. Stata/SE 12.0's xtreg procedure is used with robust standard errors to estimate the coefficients and statistical tests for our ordinary least square random-effects regression analyses. The findings from the hybrid regression models are discussed below.

5. Results

In an effort to test the hypothesized relationships between social support and homicide between regions and within regions over time, to determine the optimal lag specification for social benefits per capita, and to explore the robustness of the findings – including the posited relationship of social support acting as a moderating influence on negative GDP per capita – a series of four hybrid models was estimated for each lag model specification: contemporaneous, one-, two-, and three-year lagged social benefits measures. After carefully examining the findings, the three-year lag model specification seems to be the most appropriate to capture the temporal effect of social support on homicide (recall, social benefits per capita measured in 1997 with all other predictor variables measured in 2000). Appendix A shows the regression coefficients and robust standard errors for social benefits per capita and for the interaction term (social benefits multiplied by negative GDP per capita) for all four lag specification models. Support for the hypotheses is reflected in the statistically significant negative regression coefficient for the social benefits per capita measure and the significant positive coefficient for the interaction term. Reviewing these findings from the contemporaneous through the three-year lag model specification, the numbers of statistically significant effects supporting the hypotheses increase across the models. These findings are consistent with McCall and Brauer's (2014) cross-national, longitudinal study of European homicide trends. The analyses were also conducted using the more conventional fixed-effects regression technique, with the comparable substantive findings denoted in bold in Appendix A.[13] More consistent findings appear among the three-

13 Of the twenty-four coefficients shown in the "Within" column (comparable estimates using fixed-effects regression), twenty (83 percent) are substantively comparable to the hybrid method findings. The robust findings across the lag specification provide support for the hypothesized relationships between social support and homicide.

year lag model. We interpret the three-year lag model because the relationship between social support and homicide appears to be strongest with this lag structure and because there is greater comparability between the results from the hybrid and the fixed effects regression techniques.

The results of the hybrid regression analyses used to test the hypotheses are presented in Table 3 with between-region effects in the top half of the table and within-region effects in the lower half. R-square values for the between- and within-region components of the models are also presented.

Table 3. Hybrid random intercept regression three-year lag specification panel models predicting homicide rates in European regions for 2000, 2005, and 2009

	Western Europe		Eastern Europe	
	Model 1	Model 2	Model 3	Model 4
Between-region predictors				
Social benefits per capita[a]	-.041* (.021)	-.149** (.046)	-.771** (.134)	-.953** (.260)
GDP per capita[a] (multiplied by –1)	-.005 (.005)	.018** (.009)	.013* (.006)	.025† (.016)
Unemployment	.034** (.008)	.033** (.008)	-.002 (.011)	.003 (.011)
Percent 65 years and over	-.018* (.009)	-.019** (.009)	.013 (.022)	.006 (.024)
Sex Ratio	-.363 (.943)	-.698 (.929)	–11.542** (1.580)	–11.608** (1.646)
Social benefits-GDP interaction term	--------	-.004** (.002)	--------	-.014 (.015)
Within-region predictors				
Social benefits per capita[a]	-.181** (.034)	-.119** (.050)	-.108 (.178)	-.263 (.327)
GDP per capita[a] (multiplied by –1)	.003 (.004)	-.004 (.008)	.018† (.012)	.029† (.020)
Unemployment	-.001 (.004)	-.001 (.004)	.003 (.006)	.005 (.007)
Percent 65 years and over	-.032** (.013)	-.025* (.013)	-.004 (.061)	-.009 (.063)
Sex Ratio	–6.725** (1.617)	–6.352** (1.596)	–1.252 (3.846)	–3.034 (5.203)
Social benefit-negative GDP interaction term	--------	.001 (.001)	---------	-.007 (.010)
2005	.126* (.054)	.069 (.063)	-.128 (.130)	-.070 (.166)
2009	.274** (.088)	.173* (.100)	-.183 (.238)	-.074 (.317)
Intercept	.485 (.998)	1.394 (1.018)	12.296** (1.750)	12.506** 1.838
R^2 (overall/within/between)	.18/.32/.19	.20/.33/.23	.60/.68/.62	.60/.68/.63
N (regions/region-years)	197/487	197/487	50/118	50/118

Note: **p<.01; *p<.05; †p<.10 (one-tailed test if in hypothesized direction).
[a] In constant 2005 euros.

The regression results displayed in models 1 and 3 represent tests for H_1 and H_2, for Western and Eastern European regions, respectively. These two models display the regression coefficients and effects of social support and the other predictors across the three time points: 2000, 2005, and 2009. Consistent with hypothesis H_1, the results of model 1 indicate that, net of the controls, social support is found to be statistically significant and negatively related to the homicide rate between Western European regions. Male unemployment and percent of the population aged 65 and over are also found to be statistically significantly related to homicide between regions in the theoretically predicted directions. Additionally, according to within-region effects of predictors presented in model 1 and supporting H_2 for Western Europe, changes over the three time periods in levels of social support are negatively related to changes in homicide rates and statistically significant.[14] Even with relatively limited change in social support over the ten-year time span, we find evidence that changes in social support are linked to changes in homicide rates in Western European regions. Changes in the percent of the population aged 65 and over and the sex ratio are also related to changes in homicide rates. On the other hand, changes in negative GDP and the percent unemployed males are not significantly related to changes in homicide rates. Model 3 shows the results for Eastern European regions and the related hypothesis tests of the direct effects of social support. The effect of social benefits per capita is also significant across regions (H_1), but not over time (H_2). Negative GDP per capita and the sex ratio are also significant between regions in this model, but none of the other regressors attain statistical significance. Accordingly, these results confirm both H_1 and H_2 in the Western model as social support explains variation in between-region homicide and within-region homicide, and provides support for H_1 in the Eastern model as the effect of social support is found in the between-region measure.

Focusing on the interaction terms added in Models 2 and 4, limited support for the moderating influence of social support on homicide is found in both Western and Eastern European samples. In fact, support for a moderating effect of social support is found only between Western European regions (H_3), as the interaction term is not statistically significant in either the Eastern European sample or the within-region, over-time components of the Western and Eastern European samples (H_4). These models show limited evidence for the moderating impact of social support on the economic influence of negative GDP on homicide rates.

6. Discussion and Conclusion

This project has presented a test of Cullen's (1994) social support theory that not only allowed for the broadening of the operationalization of social support but also for an investigation of the effects of social support over time and across a group of European regions characterized by varied political and economic contexts. Extant literature examining social support has been limited in both scope and measurement, whereas the present study provides a more comprehensive measure of institutionalized support characterizing these regions. In addition to testing the direct effect of social support on homicide, we examine the moderating influence of social support on strain produced by economic deprivation, which is also related to criminal offending. Hybrid regression panel techniques simultaneously provide estimates of both the variance explained in homicide rates across European regions and variance explained in homicide trends within regions over time. Results from the analyses of the time periods – 2000, 2005, and 2009 – offer mixed support for the research hypotheses. The findings suggest that, when controlling for the effects of economic deprivation, sex ratio, and the percent of the population aged 65 and over, social support is systematically related to homicide between and within regions in the manner consistent with Cullen's theory – statistically significant and negative. The robust support for the direct effects of social support is not matched by the evidence for its moderating effect. The interaction term measuring the moderating influence of social support on economic disadvantage measured with negative GDP per capita is demonstrated only between Western

14 Although the between-region component of the hybrid model is subject to the same potential biases as traditional random effects models, as between-region predictors are assumed not to be correlated with the error term (Allison 2005; Phillips 2006), within-region estimation is not affected by this assumption. Therefore, we can have confidence in the reported within-region estimates.

European regions. The negative coefficient indicates that the crime-inducing effect of economic disadvantage on homicide is lessened in areas with higher levels of social support.[15] This relationship is also proposed by Messner and Rosenfeld's (1997) institutional anomie theory and by Agnew's macrolevel general strain theory (1999).

The present study's findings are consistent with previous cross-sectional tests of the theory, as the between-region component of the hybrid panel model is, in essence, equivalent to a cross-sectional analysis (for example DeFronzo 1983, 1997; Messner and Rosenfeld 1997; DeFronzo and Hannon 1998; Savolainen 2000; Pratt and Godsey 2003). What is more, the present study offers robust evidence at the regional level (further substantiating research that has found support at the country level) and evidence for the direct effect of social support over time within regions, from 2000 to 2009.

Cullen's theory suggests that social support may act as a buffer against the deleterious effects of economic deprivation on crime – a proposition that receives support in our study across Western European regions, but not among Eastern regions and not over time. Whereas one might expect to find support among Eastern European regions as they lack the social support enjoyed by their Western counterparts, failure to achieve statistical significance could be attributed to the relatively small sample of Eastern European regions in our study; but perhaps the fundamental difference between ours and earlier support is related to our measure of economic disadvantage. No measure of economic inequality was available for regions in Eurostat's archive, and a test of Cullen's causal argument would benefit from such a measure. Even without that measure, our findings are consistent with the evidence Pratt and Godsey (2003) present in their cross-national, cross-sectional analysis of forty-six countries (which excluded Eastern European countries).

Although the present study does not offer longitudinal support for the theory across all models, as evidence is found only for Western European regions over time, the explanatory power of social support theory to account for longitudinal variation in crime rates cannot be wholly discounted. The varying results across the between-region and within-region component of the hybrid model may be attributed to the distinction between the effects of explanatory variables across place as opposed to over time – the stock vs. the flow effects of a predictor (see Phillips 2006). Scholars have noted differences in the stock effects of explanatory variables captured via cross-sectional analyses and the flow effects of explanatory variables most often captured via time-series analyses (Koreman and Miller 1997; Teachman, Paasch, Day, and Carver 1997; Phillips 2006). Alternatively, the absence of support for the effects over time may be the result of limited variation in homicide rates over the study period. While regional homicide rates are generally decreasing for this sample of regions, the magnitude of change may not offer a great deal of variation to explain. As future waves of data become available, additional data points and greater variation in homicide rates between regions and over time may reveal the deleterious effect that welfare retrenchment has on changes in homicide.

In addition to the limited range of values among variables, the limited availability of key covariates of homicide, and the limited range of longitudinal data points for regions over time, a number of further limitations suggest that the present study's results should be interpreted with caution. As previously discussed, while the study sample includes European regions characterized by varying social and political contexts, the limited number of countries represented in the analysis does not allow for the results to be generalizable across all European countries. While a number of Eastern European countries are represented, the bulk of the regions included in the analyses are located within Western Europe. Given the turbulent social, political, and economic histories of Eastern European countries, it seems plausible that social support may behave differently in these societies than in those of Western Europe. A more complete representation of regions (particularly in post-communist Eastern European countries) would allow for a more thorough investigation of the universality of social support theory and the mechanisms through which social support works to sup-

15 The results for the two-year lag specification model shows additional support for the interaction term for Eastern regions and should be interpreted with caution. We choose to interpret the evidence from the three-year lag model as its findings are more consistent with the fixed effects analyses.

press crime. Limited numbers of regions representing Eastern European countries also restricts the power of the Eastern region analyses and, thereby, merits caution in interpreting these findings with this caveat in mind.

In spite of data limitations, the results of the across-region test presented herein offer support for Cullen's social support theory, thereby warranting the attention of future research. The Eurostat archives have the potential to offer an invaluable resource for criminological scholars, especially as more complete data for a larger number of European regions and a greater number of time points become available. Scholars should take advantage of future data expansions.

References

Agnew, Robert. 1992. Foundation for a General Strain Theory of Crime and Delinquency. *Criminology* 30:47–87.
Agnew, Robert. 1999. A General Strain Theory of Community Differences in Crime Rates. *Journal of Research in Crime and Delinquency* 36:123–55.
Allison, Paul. 2005. *Fixed Effects Regression for Longitudinal Data Using the SAS System*. Cary, NC: SAS Institute.
Bryk, Stephen S., and Anthony W. Raudenbush. 1982. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park: Sage.
Chamlin, Mitchell B., and John K. Cochran. 1997. Social Altruism and Crime. *Criminology* 35:203–28.
Chamlin, Mitchell B., Kenneth J. Novak, Christopher T. Lowenkamp, and John K. Cochran. 1999. Social Altruism, Tax Policy, and Crime: A Cautionary Tale. *Criminal Justice Policy Review* 10:429–46.
Colvin, Mark, Francis T. Cullen, and Thomas Vander Ven. 2000. Coercion, Social Support, and Crime: An Emerging Theoretical Consensus. *Criminology* 40:19–22.
Cullen, Francis T. 1994. Social Support as an Organizing Concept for Criminology: Presidential Address to the Academy of Criminal Justice Sciences. *Justice Quarterly* 11:527–59.
DeFronzo, James. 1983. Economic Assistance to Impoverished Americans: Relationship to Incidence of Crime. *Criminology* 21:119–36.
DeFronzo, James. 1997. Welfare and Homicide. *Journal of Research in Crime and Delinquency* 34:395–406.
DeFronzo, James, and Lance Hannon. 1998. Welfare Assistance Levels and Homicide Rates. *Homicide Studies* 2:31–45.
Esping-Andersen, Gosta. 1996. After the Golden Age? Welfare State Dilemmas in Global Economy. In *Welfare States in Transition: National Adaptations in Global Economies*, ed. Gosta Esping-Andersen, 1–31. London: Sage.
Eurostat. 2008. *ESSPROS Manual: The European System of Integrated Social Protection Statistics (ESSPROS)*. Luxembourg: Office for Official Publications of the European Communities.
Eurostat. 2014. *Eurostat Regional Yearbook 2014*. Luxembourg: Office for Official Publications of the European Communities.
Harvey, David. 2005. *A Brief History of Neoliberalism*. New York: Oxford University Press.
Horney, Julie, D. Wayne Osgood, and Ineke H. Marshall. 1995. Criminal Careers in the Short Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances. *American Sociological Review* 60:655–73.
House, James S. 1981. *Work Stress and Social Support*. Reading, MA: Addison-Wesley.
Johnston, Jack, and John DiNardo. 1997. *Analysis of Panel Data*. Cambridge: Cambridge University Press.
Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*. New York: Wiley.
Kim, Sang-Weon, and William A. Pridemore. 2005. Social Support and Homicide in Transitional Russia. *Journal of Criminal Justice* 33:561–72.
LaFree, Gary. 1999. A Summary and Review of Cross-National Comparative Studies of Homicide. In *Homicide: A Sourcebook of Social Research*, ed. M. Dwayne Smith and Margaret A. Zahn, 125–45. Thousand Oaks: Sage.
McCall, Patricia L., and Jonathan R. Brauer. 2014. Social Welfare Support and Homicide: Longitudinal Analyses of European Countries from 1994 to 2010. *Social Science Research* 48:90–107.
Messner, Steven F., Lawrence E. Raffalovich, and Gretchen M. Sutton. 2010. Poverty, Infant Mortality, and Homicide Rates in Cross-National Perspective: Assessments of Criterion Validity and Construct Validity. *Criminology* 48:509–37.
Messner, Steven F., and Richard Rosenfeld. 1993. *Crime and the American Dream*. Belmont, CA: Wadsworth.
Messner, Steven F., and Richard Rosenfeld. 1997. Political Restraint of the Market and Levels of Criminal Homicide: A Test of Institutional-Anomie Theory. *Social Forces* 75:1393–1416.
Ostrom, Charles W. 1990. *Time Series Regression: Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage.
Ousey, Graham C., and Pamela Wilcox. 2007. Interactions between Antisocial Propensity and Life-Course Varying Correlates of Delinquent Behavior: Differences by Method of Estimation and Implications for Theory. *Criminology* 45:401–42.
Phillips, Julie A. 2006. Explaining Discrepant Findings in Cross-Sectional and Longitudinal Analyses: An Application to U.S. Homicide Rates. *Social Science Research* 35:948–74.
Phillips, Julie A., and David F. Greenberg. 2008. A Comparison of Methods for Analyzing Criminological Panel Data. *Journal of Quantitative Criminology* 24:51–72.
Pratt, Travis C., and Timothy W. Godsey. 2002. Social Support and Homicide: A Cross-national Test of an Emerging Criminological Theory. *Journal of Criminal Justice* 30:589–601.
Pratt, Travis C., and Timothy W. Godsey. 2003. Social Support, Inequality, and Homicide: A Cross-National Test of an Integrated Theoretical Model. *Criminology* 41:611–43.
Pridemore, William A. 2008. A Methodological Addition to the Cross-National Empirical Literature on Social Structure and Homicide: A First Test of the Poverty-Homicide Thesis. *Criminology* 46:133–54.
Sampson, Robert, Stephen Raudenbush, and Felton Earls. 1997. Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science* 277:918–24.
Savolainen, Jukka. 2000. Inequality, Welfare State, and Homicide: Further Support for the Institutional Anomie Theory. *Criminology* 38:1021–42.
Stamatel, Janet. 2009. Correlates of National-Level Homicide Variation in Post-Communist East-Central Europe. *Social Forces* 87:1424–48.
Standing, Guy. 1996. Social Protection in Central and Eastern Europe: A Tale of Slipping Anchors and Torn Safety Nets. In *Welfare States in Transition: National Adaptations in Global Economies*, ed. Gosta. Esping-Andersen, 225–55. London: Sage.
Teachman, Jay D., Kathleen M. Paasch, Randal D. Day, and Karen P. Carver. 1997. Poverty During Adolescence and Subsequent Educational Attainment. In *Consequences of Growing Up Poor*, ed. Greg Duncan and Jeanne Brooks-Gunn, 382–418. New York: Russell Sage Foundation.
Vaux, Alan. 1988. *Social Support: Theory, Research, and Intervention*. New York: Praeger.

Appendix A. Unstandardized regression coefficients (and robust standard errors) using various social support lag specifications: Eastern and Western Europe

	Hybrid Components		N Regions, Region-years
	Between	Within[a]	
Contemporaneous social support (t)			
Social support (Western Europe)	-.014 (.024)	.018 (.027)	201, 500
Social support (Eastern Europe)	-.387 (.089)*	**.027 (.079)**	51, 134
1-year lagged social support (t-1)			
Social support (Western Europe)	.007 (.024)	-.050 (.035)†	201, 491
Social support (Eastern Europe)	-.290 (.076)*	**-.003 (.112)**	51, 118
2-year lagged social support (t-2)			
Social support (Western Europe)	-.024 (.021)	**-.166 (.037)***	200, 490
Social support (EasternEurope)	-.436 (.107)*	**-.080 (.135)**	50, 117
3-year lagged social support (t-3)			
Social support (Western Europe)	-.041 (.021)*	**-.181 (.034)***	197, 487
Social support (Eastern Europe)	-.771 (.134)*	**-.108 (.178)**	50, 118
Contemporaneous social support (t)			
Social support (Western Europe)	-.157 (.049)*	**.086 (.040)***	201, 500
Social support*-GDP (Western Europe)	-.005 (.001)*	**.002 (.001)**	
Social support (Eastern Europe)	-.745 (.220)*	**.078 (.106)**	51, 134
Social support*-GDP (Eastern Europe)	-.028 (.014)*	**.003 (.003)**	
1-year lagged social support (t-1)			
Social support (Western Europe)	-.143 (.049)*	**-.042 (.048)**	201, 491
Social support*-GDP (Western Europe)	-.004 (.001)*	.002 (.001)*	
Social support (Eastern Europe)	-.521 (.200)*	**.095 (.151)**	51, 118
Social support*-GDP (Eastern Europe)	-.014 (.010)†	**.004 (.005)**	
2-year lagged social support (t-2)			
Social support (Western Europe)	-.153 (.054)*	-.076 (.054)†	200, 490
Social support*-GDP (Western Europe)	-.005 (.002)*	**.002 (.001)**	
Social support (Eastern Europe)	-.790 (.233)*	**-.118 (.226)**	50, 117
Social support*-GDP (Eastern Europe)	-.021 (.011)*	**-.003 (.007)**	
3-year lagged social support (t-3)			
Social support (Western Europe)	-.149 (.046)*	**-.119 (.050)***	197, 487
Social support*-GDP (Western Europe)	-.004 (.002)*	**-.001 (.001)**	
Social support (Eastern Europe)	-.952 (.260)*	**-.263 (.327)**	50, 118
Social support*-GDP (Eastern Europe)	-.014 (.015)	**-.007 (.010)**	

Note: * p<.05, † p<.10 (one-tailed test of significance if sign in predicted direction).
[a] Bolded values substantively consistent with fixed effects estimates.

Kelly M. Thames
kmthames@ncsu.edu

Patricia L. McCall
patty_mccall@ncsu.edu

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014297
IJCV: Vol. 8 (2) 2014

Does the Magnitude of the Link between Unemployment and Crime Depend on the Crime Level? A Quantile Regression Approach

Horst Entorf, Department of Applied Econometrics and International Economic Policy, Goethe University, Frankfurt am Main, Germany
Philip Sieger, Department of Applied Econometrics and International Economic Policy, Goethe University, Frankfurt am Main, Germany

Vol. 8 (2) 2014

ISSN: 1864–1385

Does the Magnitude of the Link between Unemployment and Crime Depend on the Crime Level? A Quantile Regression Approach

Horst Entorf, Department of Applied Econometrics and International Economic Policy, Goethe University, Frankfurt am Main, Germany
Philip Sieger, Department of Applied Econometrics and International Economic Policy, Goethe University, Frankfurt am Main, Germany

Two alternative hypotheses – referred to as opportunity- and stigma-based behavior – suggest that the magnitude of the link between unemployment and crime also depends on preexisting local crime levels. In order to analyze conjectured nonlinearities between both variables, we use quantile regressions applied to German district panel data. While both conventional OLS and quantile regressions confirm the positive link between unemployment and crime for property crimes, results for assault differ with respect to the method of estimation. Whereas conventional mean regressions do not show any significant effect (which would confirm the usual result found for violent crimes in the literature), quantile regression reveals that size and importance of the relationship are conditional on the crime rate. The partial effect is significantly positive for moderately low and median quantiles of local assault rates.

Introduction

According to an annual survey on the fears of German citizens (Ängste der Deutschen, conducted by insurer *R+V Versicherung*) the fear of becoming a victim of a criminal offense regularly ranks high on the list. In 2014, 26 percent of respondents stated that they were afraid of becoming a victim of a criminal offense. Although there was considerable variation across states in that year (from 38 percent in Schleswig-Holstein and Hamburg to 21 percent in Rheinland-Pfalz and Saarland), there is remarkably little variation in the national figure over the years of this century: in 2013 an all-time low was reached with 24 percent, while the highest value was 33 percent in 2002. This is in line with the observation that in Germany crime rates themselves were stable (or rather declined slightly) between 2003 and 2013 (Polizeiliche Kriminalstatistik [PKS] 2003, 2013) but show considerable variation across states. In addition to being one of the major fears, criminal activity is associated with large costs. According to Entorf and Spengler (2002), estimates of costs of crime range between 4 and 7 percent of GDP in most industrialized countries. Another fear which generally ranks among the top five is rising unemployment. This fear was expressed by roughly 50 percent of respondents in the early 2000s, increased to 68 percent in 2005 (when unemployment was particularly high in Germany with about five million registered unemployed), vanished from the top seven fears in 2007 and 2008 but was again expressed by more than 60 percent in 2009 and 2010 (when the financial crisis was expected to hit the German labor market). As with the fear of victimization, there is also large cross sectional variation in the fear of rising unemployment and in the unemployment rate itself.

In this paper we reconsider the complex link between unemployment and crime using Germany district panel data. The economic rationale why such a link might exist is the following: Declining labor market opportunities (manifested in an increasing unemployment rate) worsen legal income opportunities and therefore make crime more attractive. This idea was first formulized by Becker (1968). The many other studies focusing on the unemployment-crime relationship include Cantor and Land (1985), Young (1993), Levitt (2001, 2004), Raphael and Winter-Ebmer (2001), Gould et al. (2002), Edmark (2005), Öster and Agnell (2007), Lin (2008), Phillips and Land (2012), Lastauskas and Tatsi (2013), and Sieger (2014). These studies differ with respect to various aspects: estimation methods used, period and country under consideration, and conclusions drawn with respect to the magnitude of the effect of unemployment on crime. Lin (2008, 414) summarizes the results: "In terms of empirical evidence, recent studies reach consensus that unemployment does have a positive, significant but only small effect on property crime, and no effect on violent crime."

We depart from existing studies by applying quantile regression methods, which allow the identification of nonlinear crime-unemployment relationships (for example, a high impact of unemployment on crime for low-crime regions and a low impact for high-crime regions). That particular pattern would be consistent with a hypothesis of opportunity-based behavior: Those who become unemployed in a low-crime area have higher incentives to commit a crime than those in high-crime regions, because they would face less effective prevention by potential victims and lower competition from other criminals. However, there could also be an opposite nonlinear pattern, which we call the stigma-based hypothesis. This predicts low marginal effects from increasing unemployment rates in low-crime areas, because here any potential detection bears a higher risk of stigma than in regions where criminal behavior is more common. These examples show that there are good reasons to take a closer look at the unemployment-crime relationship using quantile regressions. Surprisingly, there is little research based on this technique in the criminological literature. To the best knowledge of the authors, the only contribution is Bandyopadhyay et al. (2015). Based on time-series evidence from six crime categories and forty-three police force areas, they confirm not only that unemployment does increase crime but that it does so more in high-crime areas. Moreover, they find that the crime-reducing effect of higher detection rates is stronger in low-crime areas.

The quantile analysis conducted in this paper is based on a panel data set covering about four hundred German *Landkreise* (districts) and urban municipalities (*kreisfreie Städte*) for the years 2005 to 2009 in Germany. The same source (German districts and urban municipalities) has recently also been used by Messner et al. (2013) and Lastauskas and Tatsi (2013).

1. Factors of Crime

1.1. Economic Factors

Legal income opportunities represent an important factor of crime. Following Becker (1968), higher legal income should decrease criminal activity, because legal income represents part of the opportunity costs of conviction. Higher legal income prevents a potential offender from committing a crime because they fear losing it. All other things being equal (probability of detection and conviction and illegal income opportunities) higher legal income is expected to decrease criminal activity. However, if one switches from a micro to a macro perspective, there is another channel through which legal income affects crime. If average legal income in a certain region (as a German district) increases, the potential offender is on the one hand more likely to have a higher legal income, and hence less likely to commit a crime. On the other hand, a higher average legal income might also increase illegal income opportunities, since now there is more income or wealth to steal from. At least for property crime, a higher legal average income could therefore also increase criminal activity. Mobile criminals from other regions might also be attracted. This would increase the utility of committing a crime and, in turn, also the likelihood of rising local crime rates (note that crime rates are registered in the city or district where the crime is committed). The effect of disposable income is therefore ambiguous, since it influences the decision to commit a crime (or not) through different channels.

The potential channel through which unemployment affects the crime rate has already been briefly mentioned above: Declining labor market opportunities (manifested in an increasing unemployment rate) worsen legal income opportunities and therefore make crime more attractive. In their influential paper, Raphael and Winter-Ebmer (2001, 262) express this idea as follows: "Conceptualizing criminal activity as a form of employment that requires time and generates income, a 'rational offender' should compare returns to time use in legal and illegal activities and make decisions accordingly. Holding all else equal, the decrease in income and potential earnings associated with involuntary unemployment increases the relative returns to illegal activity." The idea of time allocation between legal and illegal activities and its influence on the decision to participate in criminal activities was formalized in a theoretical framework by Grogger (1998). As Raphael and Winter-Ebmer (2001) lay out, Grogger's model implies four different employment-crime situations which can be used to predict how unemployment affects criminal activity. For individuals who engage in both criminal activity and job market activity, the model predicts that unemployment increases time allocated to crime. For individuals who do not work in the regular job market but only commit crimes, an unemployment spell cannot affect the time allocated to criminal activity. For workers not committing crimes, the effect of unemployment depends on whether the return to the first hour of criminal activity exceeds the reservation wage. Individuals whose reservation wage is high are unlikely to be pushed into crime by an unemployment spell. Individuals with comparably low reservation wages are more likely to be influenced by unemployment and might try to offset lost income by engaging in criminal activity. Thus, Grogger's model predicts that for two out of four situations an unemployment spell will increase time allocated to criminal activity (and thus increase the crime rate), while for the remaining two cases, there is no response to an unemployment spell. Applying the model to regional data, theory would predict that responses to changing unemployment rates should be smaller in regions with already high crime rates than in regions where crime rates are low (given that reservation wages are not prohibitively high).

1.2. Demographics, Education, and Urbanity

Becker's (1968) seminal economic model of crime abstracts from some important features of the criminal's decision problem. Several other determinants of crime have been discussed in the literature besides deterrence variables (probability of conviction or severity of punishment). One of these is the age structure of society. As outlined by Farrington (1986), who focuses on the United Kingdom and the United States, the age-crime curve usually peaks in teenage years and declines afterwards. Grogger also provides evidence for this phenomenon: "Thirty five percent of all Philadelphia males born in 1945 were arrested before the age of 18, and one-third of all Californian men born in 1956 were arrested between the ages of 18 and 30. The 1990 census counted 1.1 million persons in jail, the vast majority of whom were men in their twenties and thirties." (1998, 756). Similar patters can be observed for Germany. Those aged 6 to 20 make up 26.1 percent of all crime suspects but only 13.7 percent of the population, while those aged 40 and above make up 32.4 percent of all crime suspects but 56.9 percent of the population (PKS 2009). Given the descriptive evidence and the mostly accepted empirical evidence from other studies (for example Freeman 1996),[1] it seems imperative to include age structure as a further control variable when it comes to explaining crime. One would expect the proportion of people of crime-prone age in the population to have a positive influence on criminal activity. Younger people are also victimized more often (PKS 2009, table 91), so a larger proportion of young people might therefore foster criminal activities in two ways: it increases both the supply of criminals and the supply of victims.

Data from the German police statistics (PKS) show that non-German crime suspects make up 21.1 percent of all crime suspects, although contributing only 8.7 percent of the total population (PKS 2009).[2] Possible reasons for this huge overrepresentation are discussed in Albrecht (1997). He mentions, among other things, deprivation and control theories, which focus on problems of social integration and reduced opportunities to develop ties to mainstream society. The reasons for the apparent overrepresentation of foreigners in criminal activity will not be discussed in detail here, but the numbers indicate the need to control for the composition of the regional population.

Overrepresentation of crime suspects can be observed in yet another demographic group: men. Inspection of the

1 Levitt (1999, 2004) argues that the age structure alone has only a limited influence on the evolution of crime rates, because the decline in crime rates during the time period from 1995 to 2004 in the United States was at odds with a rising share of the most crime-prone demographic age group of young males.

2 Even after excluding those offenses which can only be committed by non-Germans (such as offenses against asylum law), the numbers only go down to 19 percent (2003) and 19.4 percent (2009) respectively (PKS 2009, 105).

raw numbers tells the following story: in 2009, out of the 2.19 million crime suspects, 1.64 million were male (75 percent). Controlling for the gender composition of the respective district hence seems to be as important as controlling for the demographic variables discussed above.

Another determinant of criminal behavior is education. Unfortunately, there is no comprehensive data on educational attainment of the German population at the district level. The only variable that covers education at the district level is the proportion of workers subject to social security contributions who have not completed vocational training *(sozialversicherungspflichtig Beschäftigte ohne abgeschlossene Berufsausbildung)*. This variable only covers the education of a certain group, namely those who are subject to social security contributions. The predicted influence of this variable on crime is therefore hard to determine: on the one hand, less educated people are expected to commit more crimes. One could therefore expect a positive influence of this variable on crime. On the other hand, a high proportion of workers subject to social security contribution not having completed vocational training means that there are good labor market opportunities even for unskilled workers. Under this interpretation, a higher proportion of such workers would have a negative effect on crime. Empirical evidence for this can be found in Gould et al. (2002).

The last determinant discussed in this section is population density. There are several theories why population density might be an important determinant of crime. On the one hand, densely populated areas (usually large cities) feature a weaker net of social control (Glaeser and Sacerdote 1999). The anonymity of the city makes it easier for individuals to commit crimes, since the potential stigma involved in being caught is less. In addition, similar to the argument applied above to age composition, a high population density makes a "match" between criminal and victim more likely. Criminals may also have greater access to the wealthy in urban areas. Glaeser and Sacerdote (1999, 227) also mention the possibilities that criminals face a lower probability of detection and arrest in urban areas and that urban areas themselves attract (or create) crime-prone individuals. These theoretical considerations are confirmed for the data set used in this analysis. The bivariate correlation between overall crime rates and population density is remarkably high, with 0.63. One would therefore expect a positive impact of population density on crime rates.

2. Data Used

The empirical analysis is based on data covering districts *(Landkreise)* and urban municipalities *(kreisfreie Städte)* in Germany. *Landkreise* usually include one or more moderate-sized towns, as well as villages and rural areas, whereas municipalities are organized as stand-alone communities *(kreisfreie Stadt)*. In the following, both urban municipalities and rural counties will be referred to as "the districts."[3] This section introduces the variables included and presents detailed summary statistics. Crime data (number of offenses and clearance rates) are collected by regional state offices of the German Federal Criminal Police Office *(Bundeskriminalamt)* and are published in *Polizeiliche Kriminalstatistik* (PKS, police criminal statistics). Covariates come from two sources: unemployment and employment data are gathered by the German Federal Employment Agency *(Bundesagentur für Arbeit)*, whereas demographics and income data are obtained from the Federal Statistical Office *(Statistisches Bundesamt)*.

2.1. The Dependent Variables

The dependent variables used in this study are the crime rates in each district. Before defining the term "crime rates" we describe which offenses are included. These are burglary, auto theft, and assault. The offenses are defined as follows in the German penal code (*Strafgesetzbuch*, StGB).[4]

3 Messner et al. (2013) prefer to use the German word " Kreise," because they differ from counties or districts in the United States. For example, large city such as Houston may be within a district with other large cities; however, in Germany Houston would be a stand-alone community *kreisfreie Stadt*, i.e. counted as "Kreis".

4 The translation covers the most important points. German speaking readers are referred to the original source.

- Burglary (*Wohnungseinbruchsdiebstahl*, §244 Abs. 1 Nr. 3 StGB): entering a home by force or deception with the intention of stealing property.
- Auto theft (*Diebstahl in/aus Kraftfahrzeugen*, §242 StGB): Stealing a car or stealing property from a car.
- Assault (*Körperverletzung*, §223–227, 229, 231 StGB): Bodily injury, dangerous bodily injury, maltreatment of wards, serious bodily injury, bodily injury resulting in deaths, negligent bodily injury, participation in a brawl (see the official translation of the German Criminal Code: http://www.iuscomp.org/gla/statutes/StGB.htm#223)

"Crime rates" are defined as the frequency ratio *(Häufigkeitszahl)* from the German police statistics. This is the number of cases (of a given offense) reported to the police per 100,000 inhabitants in the district where the crime was committed. As is pointed out in the PKS (for example PKS 2003,14), the explanatory power of the frequency ratio is limited by the fact that only part of the committed crimes are reported to the police and by the fact that illegal aliens, foreign tourists and transients might also commit crimes but are not counted as inhabitants of Germany. However, the latter restriction is negligible: in 2009 out of the 2,187,217 crime suspects only 46,132 (or 2.11 percent) were illegal aliens and 6,739 (0.31 percent) were foreign tourists and transients, adding up to only 2.42 percent of all crime suspects. A slightly broader perspective, which also includes asylum seekers (22,137 or 1.01 percent) and stationed armed forces, including their family members (2,249 or 0.1 percent), produces a share of 3.53 percent of all crime suspects. The second problem of unreported crimes is more severe, though it can be mitigated by using fixed-effect models (see below).

Table 1 presents descriptive statistics for the frequency ratio for burglary, auto theft, and assault. It is apparent that there is huge variation in the respective crime rates. The overall distribution of the frequency ratio for auto theft is displayed in Figure 1, which nicely visualizes what can also be inferred from percentiles in Table 1. Although the maximum frequency ratio for auto theft is 2,437 (recorded in Bremen in 2007), the 95 percent percentile is only 878.5, with a median of only 246. The minimum is as small as 20, recorded in the district Forchheim (Bavaria) in 2008. Hence the distribution is heavily skewed. Moreover, the geographical distribution (Figure 2) shows a north-south pattern with higher frequency ratios in the north. Urban municipalities, at least in the south (Bavaria and Baden-Württemberg), do not stand out particularly on visual inspection of Figure 2.

Table 1: Frequency ratios for burglary, auto theft, and assault (descriptive statistics)

Percentile	Burglary	Auto theft	Assault
5%	22	66.5	325
25%	47	138.5	421
50%	84	246	538.5
75%	138.5	413	704
96%	275	878.5	1094.5
Minimum	3	20	202
Maximum	605	2437	2108
Mean	105.05	325.68	597.06
Standard deviation	79.19	274.18	242.45

Note: Statistics based on 3,020 pooled annual district and urban municipality data points for 2003 to 2009. Due to administrative reforms, the number of districts fell from 438 in 2003 to 412 in 2009. Frequency ratio is the number of reported offenses per 100,000 inhabitants.

Figure 1: Distribution of frequency ratio auto theft, 2003 to 2009

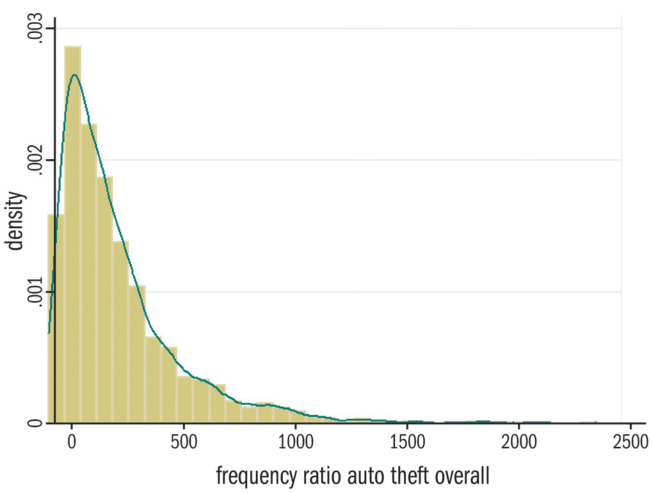
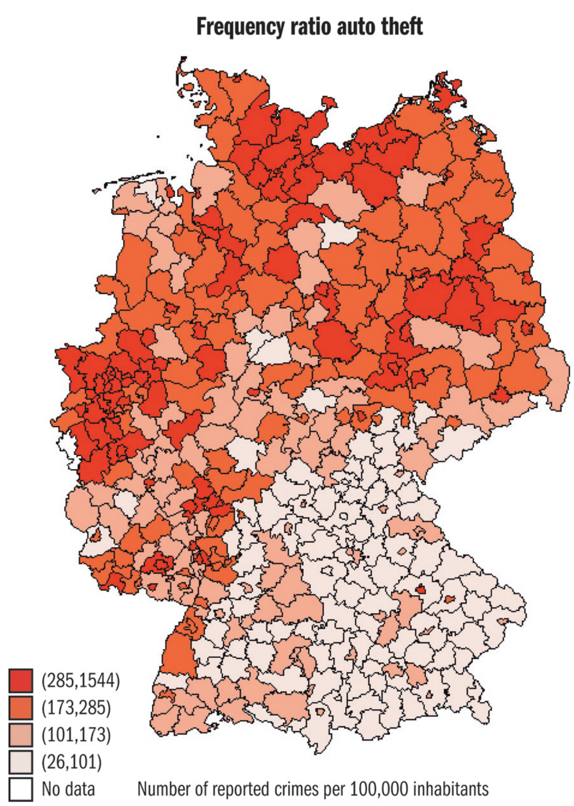
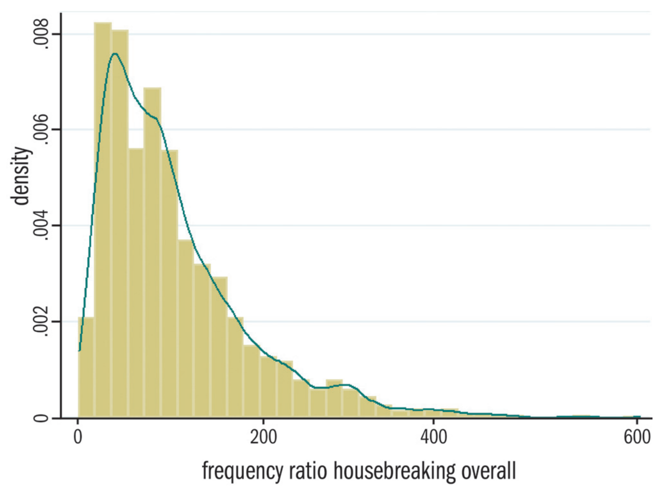


Figure 2: Regional distribution of frequency ratio auto theft, 2009



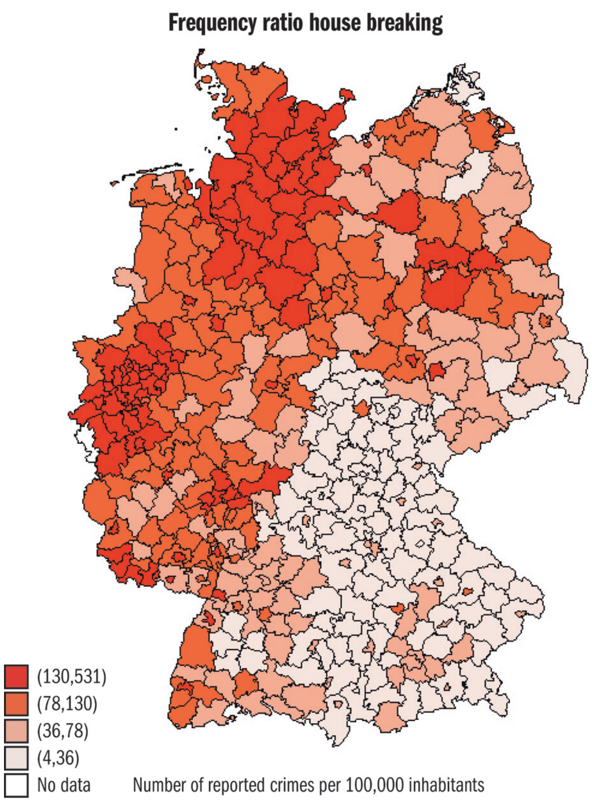
Similar patterns hold true for the frequency ratio for burglary. Here, too, we observe a heavily right-skewed distribution and enormous variation. The minimum frequency ratio for burglary is only 3 (Hildburghausen, Thuringia, 2008), with the 5 percent percentile as low as 22. In contrast, the maximum frequency ratio of 605 (Cologne, North Rhine-Westphalia, 2003) is about two hundred times the minimum. The distribution over the whole time period under consideration and the graphical visualization of the distribution in 2009 are displayed in Figure 3 and Figure 4, respectively. Noteworthy is the clustered appearance of burglaries in the north and west, while the south-west does not exhibit high frequency ratios even in the urban municipalities.

Figure 3: Distribution of frequency ratio burglary, 2003 to 2009



Assault, with a minimum frequency ratio of 202 (Enzkreis, Baden-Württemberg, 2003) and a maximum of 2,108 (Neumünster, Schleswig-Holstein, 2007), does not show as much variation as the other offenses. The ratio of minimum to maximum is lower (ten compared to one hundred for auto theft and two hundred for burglary). In addition, the distribution is more symmetrical than to the other distributions (Figure 5). Urban municipalities are among the most heavily affected districts for assault. They clearly stand out in the geographical distribution for 2009 (Figure 6). Besides the urban municipalities, the city states Berlin, Bremen, and Hamburg, the region around the city of Hannover, and the Rhine-Ruhr metropolitan region all show elevated frequency ratios. The contrast between the south and the north is less pronounced than it is for auto theft or burglary.

Figure 4: Regional distribution of frequency ratio burglary, 2009



One possible objection to using crime rates at district level is that criminals do not necessarily live in the district where they commit the crime. For the offenses under consideration, however, about 75 percent of criminals live in the district where they committed the crime (PKS 2009).

Figure 5: Distribution of frequency ratio assault, 2003 to 2009

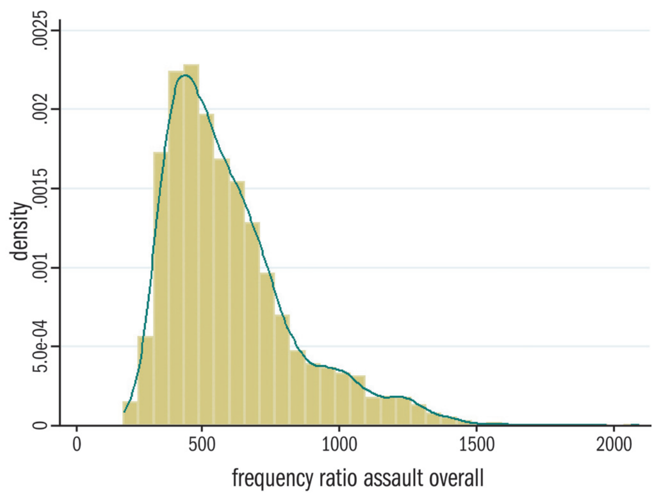
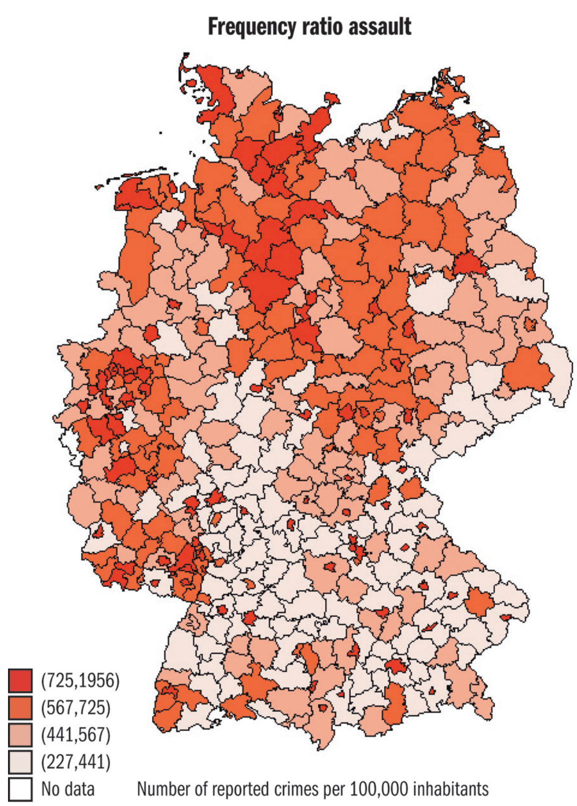


Figure 6: Regional distribution of frequency ratio assault, 2009



2.2. Economic Explanatory Variables

Table 2 shows descriptive statistics for the explanatory economic variables: unemployment rate and net household income. As unemployment is the major variable of interest, it is important to know its exact definition. The unemployment rate is defined as the number of unemployed persons divided by the total workforce. This raises the question who is counted as an unemployed person and which persons are considered to be in the workforce. According to the German Social Security Code (*Sozialgesetzbuch 3* [SGB 3], §16 Abs. 2), a person is to be considered unemployed if he or she

1. is temporarily not in an employment relationship or works less than 15 hours per week,
2. is looking for employment subject to social security contributions,
3. at the disposal of the employment agency,
4. has registered as unemployed at the employment agency.

The workforce consists of all persons in dependent civilian employment plus all self-employed persons and helping family members.

The unemployment rate varies considerably across German districts. The district with the lowest unemployment rate during the period under consideration is Eichstätt (Bavaria) (1.6 percent, 2008). The district with the highest rate is Uecker-Randow (Mecklenburg-Western Pomerania) (29.3 percent, 2004). Figure 7 shows the distribution of unemployment rates during the period under consideration. It is right-skewed with a peak at about 8 percent. A considerable number of districts (more than 5 percent) experienced unemployment rates exceeding 20 percent. The geographical distribution of unemployment rates in (Figure 8) shows that even nineteen years after Reunification, the new German states still lag behind in terms of labor market success. Districts with unemployment rates higher than 10 percent are almost exclusively located in eastern Germany (along with a few urban municipalities in the west, especially in the Rhine-Ruhr metropolitan area), where very few districts have a rate smaller than 7 percent. The vast majority of districts with rates below 5 percent are found in the south (Bavaria and Baden-Württemberg), while in the rest of Germany rates range between 5 and 10 percent.

Table 2: Descriptive statistics for economic variables

Percentile	Unemployment rate	Net household income (euros)
5%	0.039	29,907
25%	0.061	34,747
50%	0.087	39,650
75%	0.130	43,527
95%	0.202	49,845
Minimum	0.016	24,545
Maximum	0.293	69,030
Mean	0.100	39,520
Standard deviation	0.051	6,289

Note: Statistics based on 3,020 pooled annual district and urban municipality data points from 2003 to 2009. Due to administrative reforms of geographical boundaries, the number of districts changed from 438 in 2003 to 412 in 2009.

Figure 7: Distribution of unemployment rate, 2003 to 2009

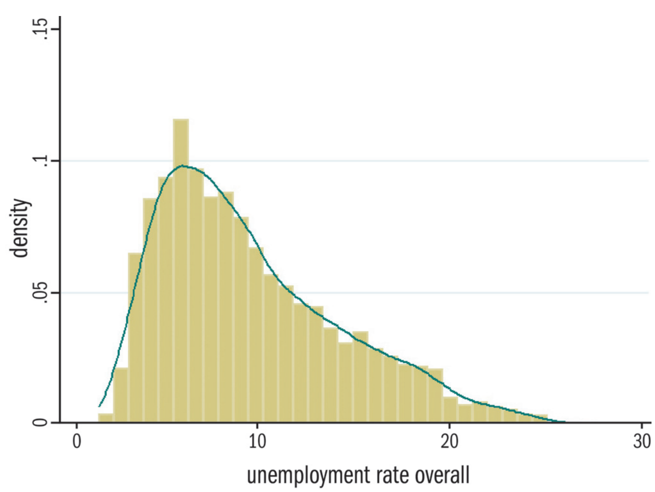
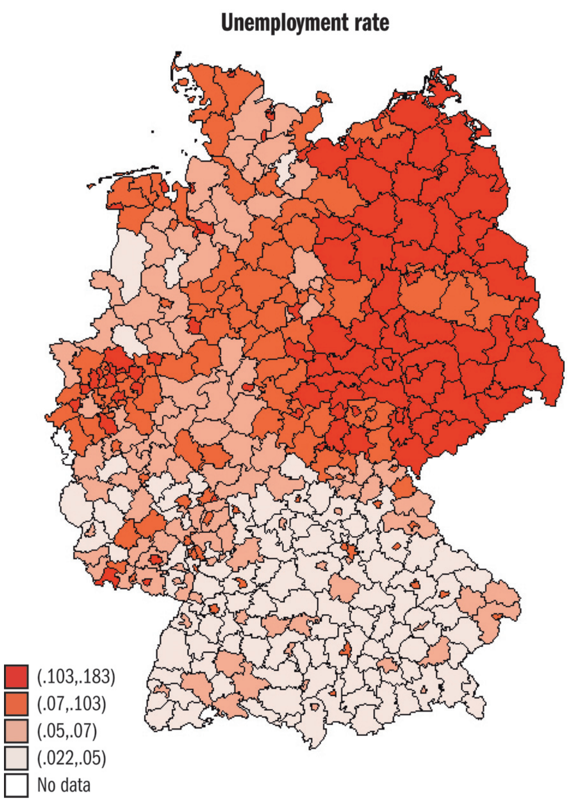


Figure 8: Regional distribution of unemployment rate, 2009



The variation in net household income is less pronounced than the variation in unemployment rates using the coefficient of variation as a measure of dispersion. The lowest average net income was reported in Leipzig (Saxony) in 2003 with 24,545, while the highest one was reported in Starnberg (on the periphery of Munich, Bavaria) in 2006. 50 percent of districts have an average net household income between 35,000 and 44,000. Only 5 percent have less than 30,000 and only 5 percent have more than 50,000. As already mentioned, only in very few districts is average household income above 50,000.

2.3. Other Explanatory Variables

The share of workers subject to social security contributions without completed vocational training (median = 15.6 percent, mean = 15.2 percent; from here on referred to as the share of unskilled workers in the workforce) has been introduced as factor representing the prevailing level of education. However, as described above, it is not entirely clear whether this variable truly captures education or rather job opportunities for unskilled workers. The share of unskilled workers in the workforce is low in the new federal states, high in the south-west, and rather mixed in the rest of Germany, suggesting that the share of unskilled workers in the workforce captures labor market opportunities for unskilled workers rather than educational attainment. For example, the small and medium-sized and manufacturing businesses concentrated in Baden-Württemberg seem to offer jobs to unskilled workers, whereas the economic situation in the new German states is generally less beneficial. The lowest share of unskilled workers in 2009 (7.3 percent) was reported in Greiz (Thuringia), the highest share (30.6 percent) in Tuttlingen (Baden-Württemberg).

Deterrence plays a crucial role in economic models of crime. The severity of the expected punishment and the probability of arrest are deterrence measures that influence the likelihood of committing a crime. The size of the police force (Levitt 1997, Lin 2008), the incarceration rate (Raphael and Winter-Ebmer 2001, Phillips and Land 2012), and the clearance rate (Entorf and Spengler 2000) are frequently used in empirical analyses. Although theory suggests that it is imperative to include a variable that measures some form of deterrence, some studies fail to do so (Öster and Agell 2007, or Yearwood and Koinis 2009). Based on data availability, we follow Entorf and Spengler and use the clearance rate as a measure of deterrence. The clearance rate is defined as the number of "solved" cases in a given year as a percentage of the total number of crimes recorded by the police in the same period (PKS 2009, 14). A case is considered to be "solved" when a suspect is identified and a charge is laid, regardless of whether the accused is convicted. As some cases reported in the previous year are solved in the current year, this might result in a clearance rate greater than 100 (which indeed happened in the period under consideration). Clearance rates significantly differ by type of crime. Whereas average clearance rates are rather low for burglary (26.0 percent, median 23.2 percent) and car theft (15.4 percent, median 12.8 percent), the rate for assault is 90.7 percent (median 91.2 percent).

The demographic structure plays a key role in explaining criminal behavior. The share of the population aged 15 to 24 (mean = 11.8 percent, median 11.7 percent; referred to as the youth population) ranges from 8.9 percent in Greiz (Thuringia, 2009) to 17.7 percent in the urban municipality of Greifswald (Mecklenburg-Western Pomerania, 2005) with. Turning to the share of the population aged 25 to 54 (referred to as the "active population"; mean 42.6 percent, median 42.4 percent) one can observe a close-to-normal distribution with a slightly more pronounced right tail (maximum 49.3 percent, in Heidelberg, minimum 37.4 percent, in Lüchow-Dannenberg (Lower Saxony).

Population density, the share of males, and the share foreigners complete the list of explanatory variables. In Germany, population density varies considerably from 38 per square kilometer in Mecklenburg-Kreilitz to 4,282 in Munich. The mean population density (512 inhabitants per square kilometer) does not convey much information about the "typical" district: more than 70 percent of all districts have a smaller population density than this mean, which is inflated by a small number of extremely densely populated areas (the median is 197). The district with the lowest share of males is Baden-Baden (Baden-Württemberg) with a share of 46.0 percent (recorded in 2004) while the district with the highest share is Aachen (North Rhine-Westphalia) with 51.5 percent (recorded in 2009). 95 percent of all districts have male shares less than 50 percent during the period under consideration, so in almost all districts there are more women than men (mean 49.0 percent, median 49.1 percent). However, it turned out that the variation of males across districts and over time is rather low and highly collinear with other factors of the population structure such that we had to omit it from the econometric analysis. The share of foreigners varies considerably across German districts. The lowest share of foreigners was recorded in Sömmerda (Thuringia) in 2007 with only 0.7 percent, while the highest share was recorded in Offenbach am Main (Hesse) in 2003 with more than 26 percent. A look at the percentiles shows that there are many districts with rather low shares of foreigners (50 percent have rates lower than 5.8 percent; mean 6.8 percent), while there are few districts with high shares of foreigners (5 percent of the districts have shares of foreigners higher than 15 percent).

3. Methodology

Ordinary least squares regressions determine the conditional mean of a response (dependent, endogenous) variable given values of explanatory (exogenous) variables. In this section we go beyond this standard method and also analyze the relationship between unemployment and crime using quantile regression (for example Koenker 2005). This technique has been proposed to discover relationships in cases with unequal impacts of explanatory variables for different ranges of the dependent variable. Hence, quantile regression allows identification of relationships even when there is no relationship or only a weak relationship between the *means* of such variables, but perhaps one at the median or in lower or upper parts of the distribution. Therefore, application of quantile regression seems to be promising in regional data sets with uneven distributions of the response variable, which is certainly the case for the heavily skewed distribution of crime rates across regions.

3.1. Mean Regression

The starting point for the empirical analysis is the following model specification for the dependence of crime on unemployment:

$$(1)\quad Crime_{i,t} = \beta\ Unemployment_{i,t} + \gamma' X_{i,t} + \theta_t + \varepsilon_{i,t}$$

The coefficient of interest is β which captures the effect of unemployment in year t in district i on crime in year t in district i. The vector of parameters γ captures the influence of other explanatory variables as demographic, economic, or deterrence variables. The θs are time-fixed effects and capture the influence of shocks on the crime rate which affect all districts in the same way. $\varepsilon_{i,t}$ denotes the error term.

This model specification suffers from unobserved heterogeneity (for example due to region-specific shares of unreported crimes) which would lead to inconsistent and biased OLS estimates. The problem can be tackled by utilizing the panel structure of the data. Panel data are superior to a pooled cross-section in that the former allows the researcher to consider unobserved effects (or individual fixed effects). They are able to capture time-invariant (or slowly changing) factors that influence the crime rate and are specific to a certain district (rural areas, for example, are fundamentally different from urban areas). These factors can all be lumped together in the fixed effects. Their inclusion can therefore help to mitigate the problem of omitted variables (Wooldridge 2002, 247). Note that the use of random-effects (RE) modelling does not provide a reasonable alternative to the fixed-effects approach used in this study. Consistency of RE panel data modelling requires that unobserved factors of unobserved heterogeneity (the α_is, see below) are uncorrelated with included regressors. This presumption seems rather unrealistic as observed factors such as local unemployment, income, or clearance rates are most likely related to unobserved factors such as the share of unreported crime in the region. Nevertheless, it should also be noted that including fixed effects is no cure-all against omitted variable biases, because factors such as the share of unreported crimes may change over time. They are only useful when included observed factors change much faster than excluded unobserved factors. We assume that this is plausible for the data under comparison.

By including these fixed effects, the resulting regression equation reads:

$$(2)\quad Crime_{i,t} = \alpha_i + \beta\ Unemployment_{i,t} + \gamma'\, X_{i,t} + \theta_t + u_{i,t}\,,$$

where α_i denotes the fixed effect for district i and $u_{i,t}$ is the new error term. Although the α_is are unobservable it is still possible to estimate the parameters of interest in equation (2) by subtracting the (over time) mean of each district from the respective observation (or, equivalently, by including district dummy variables). Denoting mean values by overlining, the regression equation reads:

$$(3)\quad Crime_{i,t} - \overline{Crime}_i = \beta(Unemp_{i,t} - \overline{Unemp}_i) + \gamma'(X_{i,t} - \overline{X}_i) + \theta_t + u_{i,t} - \bar{u}_i$$

Note that the unobserved effect no longer appears in equation (3), but the parameters are the same as in equation (2). It is hence possible to estimate the parameters of interest by applying OLS to equation (3).

This specification might still suffer from the potential problem that unemployment is not an exogenous variable in equation (3). Econometric endogeneity problems (inconsistency and biasedness of parameter estimates) arise when regressors are correlated with residuals of the statistical model. The major reason for endogeneity of unemployment can be suspected in a potential correlation between unemployment and unobserved factors in the error term, such as the degree of regional social disruption and social control. This shortcoming relates to the omitted variable bias discussed above. A further potential reason for endogeneity is simultaneity, which might for instance occur when high local crime rates have a reversal effect on corresponding labor markets. The potential endogeneity of unemployment in crime equations is beyond the scope of this article, but has been addressed at length elsewhere (Raphael and Winter-Ebmer [2001] and Lin [2008]; Lastauskas and Tatsi [2013] and Sieger [2014] consider German district data). Experience from previous research has shown that the likelihood of potentially biased parameters on unemployment is much smaller when panel data are used and time as well as district effects are included.

3.2. Quantile Regression

While mean regression delivers a single parameter estimate for the *average* partial effect of unemployment on crime, quantile regression allows different impacts of unemployment on crime depending on the level of criminal activity. This is useful for several reasons. For instance, one might find an insignificant effect of unemployment on crime in mean regressions, while there is in fact a negative (and significant) effect of unemployment on crime for low-crime areas and a positive (and significant) effect for high-crime areas. In mean regressions both effects would simply cancel out, leaving the researcher with the false conclusion that unemployment does not affect crime. Moreover, as suggested by the Grogger (1998) model, it makes a difference whether a certain percentage change Δu^* of the unemployment rate hits a region with few criminals, or a region with a comparatively large proportion of full-time criminals. Quantile regression can therefore be seen as a tool for deeper inspection of the results of the mean regression, a path that does not seem to have yet been pursued in the context of analyzing the relationship between crime and unemployment (with the notable exception of Bandyopadhyay et al. 2015).

Mean regression estimates the conditional mean function, given values of explanatory variables. That function describes how the mean of the dependent variable changes with the vector of explanatory variables. The underlying assumption is that the error term in the regression equation has the same distribution independent of the values of the explanatory variable. Instead of predicting the mean of the endogenous variable, quantile regressions aim at predicting the quantiles of the regession, i.e., the median (50 percent median), 25 percent, 75 percent etc. However, there is a possibility that the explanatory variables influence the conditional distribution of the dependent variable in many other ways: stretching one tail of the distribution, inducing multimodality, or expanding its dispersion (Koenker and Hallock 2001, 143). Investigating these other possibilities might offer a more detailed view on the relationship between the dependent and explanatory variables. In particular, it might shed light on the question whether the effect of unemployment on crime differs between different levels of crime.

There are (at least) two alternative crime-unemployment links that are imaginable from a theoretical point of view:

i) A declining crime-unemployment link, where the effect of unemployment on crime is high in low-crime areas and low in high-crime areas.

ii) An increasing crime-unemployment link, where the effect of unemployment on crime is low in low-crime areas and high in high-crime areas.

These two different crime-unemployment relationships correspond to two different interpretations of how criminals react to the level of criminal activity. A declining crime-unemployment relation would give rise to what we call opportunity-based behavior. It would also be in line with the Grogger (1998) model. When criminal activity is low, the supply of crime is highly elastic (that is, criminals show strong responsiveness to changing incentives). Hence, in such situations an increase in unemployment has a relatively large impact on crime: There are attractive and unprotected victims and only few competitors. If there are only a few drug dealers in the street, becoming a drug dealer is more profitable than if there are already many drug dealers around. If there are only a few burglars around, trying to break into a house is more profitable (maybe also because people do not invest so much in crime-preventing equipment such as alarm and warning devices). If crime is already high, that means the "crime market" is already rather saturated, and engaging in criminal activities after becoming unemployed is not as attractive anymore. Then the supply becomes inelastic and the effect of unemployment on crime would be lower or insignificant. At a first glance, the reasoning seems plausible for property crimes, but less so for violent crimes such as assault. However, it also makes a difference for violent crimes whether the marginal crime effect of a certain change of the unemployment rate Δu^* hits a neighborhood of less protected citizens in low-crime areas or a region of already high crime rates where further increases become unlikely, in particular because more and more people have taken precautions and avoid risky places.

An increasing crime-unemployment link, on the other hand, would follow from what we call stigma-based behavior. If criminal activity is low, being unmasked as a criminal creates a strong stigma, since the person concerned is one of only a few criminals. Funk (2004) describes stigma of potential detection as a crime deterrent. Higher unemployment rates would not necessarily push a person into criminal activity, since the fear of the stigma prevents the potential offender from doing so. However, if there is already a lot of criminal activity, there is less impediment to becoming a criminal, since even detection would not make the person a "black sheep." A rise in unemployment would hence more easily push the person into criminal activity.

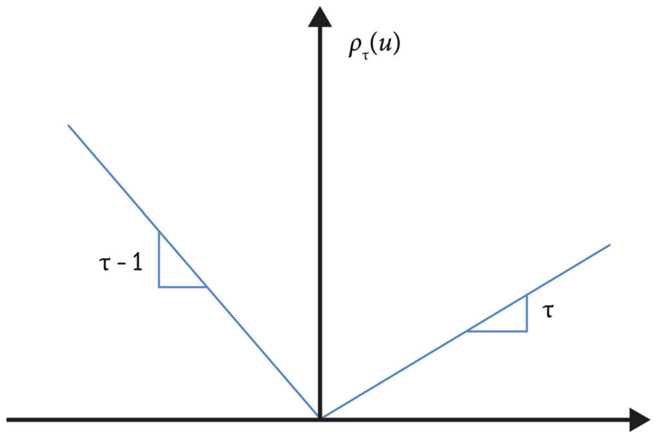
3.2.1. Ordinary Quantile Regression

It might come as a mild surprise that quantiles, although linked to the operations of ordering and sorting, can also be defined via a simple optimization problem (Koenker and Hallock 2001, 145). Similarly to OLS, where estimation is based on minimizing a sum of squared residuals u_{it}, quantile estimation is based on minimizing a sum of weighted absolute residuals $\rho_\tau(u_{it})$. More precisely, estimating the conditional quantile function for quantile τ is achieved by solving the following minimization problem

$$(4)\quad min_\beta \sum \rho_\tau (y_{i,t} - \xi(x_{i,t}, \beta))$$

where β is the parameter of interest, $\rho_\tau(u) = [\tau I\{u \geq 0\} + (1 - \tau)\, I\{u < 0\}]\, |u| = u(\tau - I\{u < 0\})$ is the asymmetric quantile loss function (visualized in Figure 9),[5] respectively weighting function of residuals u_{it}, and $\xi(x_{i,t}, \beta)$ is some parametric function of explanatory covariates, which may include controls for time effects (these are included in performed regressions but omitted here for notational convenience). In a first step, the parametric function will be a linear function of the explanatory variables and the parameters to be estimated, as the right hand side of regression equation (1), i.e. $\xi(x'_{i,t}, \beta) = x_{i,t}\beta$ (in vector notation). This approach suffers from the same deficiencies described above, in particular that it is not fully exploiting the panel structure by using fixed effects and taking unobserved heterogeneity into account. This feature will be added in the next subsection. The interpretation of quantile regression coefficients follows the interpretation of ordinary regression coefficients, with the important difference that reported parameter estimates only affect the quantile in question (instead of the mean). Thus, in the median regression the constant is the median of the sample while in the .75 quantile regression the constant is the 75th percentile for the sample, etc.

Figure 9: Quantile loss function



Note: See Koenker and Hallock (2001) for a similar illustration.

3.2.2. Quantile Regression with Fixed Effects

The approach presented above might suffer from the problem of unobserved heterogeneity. Following Koenker (2004) we consider the following model for the conditional quantile functions of the dependent variable of individual i at time t:

$$(5)\quad Q_{y_{i,t}}(\tau | x_{i,t}) = \alpha_i + x'_{i,t}\,\beta(\tau)$$

Where the a_i again denote the individual fixed effect, $x_{i,t}$ is a vector of explanatory variables and the τ-dependent vector β is the vector of parameters to be estimated. In such models the fixed effects a_i imply a pure location shift on the conditional quantiles of the response. Thus, the effects of the covariates are permitted to depend on the quantile, τ, whereas the effects a_i do not, but they are still useful to control for unobserved heterogeneity and can be interpreted in the way discussed above. In order to estimate model (5) for several quantiles simultaneously, Koenker (2004) proposes solving the following:

$$(6)\quad \min_{(\alpha,\beta)} \sum_{k=1}^{q} \sum_{t=1}^{T} \sum_{i=1}^{n} \omega_k \rho_{\tau_k} \left(y_{i,t} - \alpha_i - x'_{i,t}\,\beta(\tau_k) \right)$$

5 I denotes the indicator function taking the value 1 if the expression in the cambered brackets is true and 0 otherwise.

or, if the number of individuals is large relative to the number of time periods, a penalized version of (6), which reads as

$$(7)\quad \min_{(\alpha,\beta)} \sum_{k=1}^{q}\sum_{t=1}^{T}\sum_{i=1}^{n} \omega_k \rho_{\tau_k}\left(y_{i,t}-\alpha_i-x'_{i,t}\beta(\tau_k)\right)+\lambda\sum_{i=1}^{n}|\alpha_i|$$

where the ω_ks are weights which control the relative influence of the q quantiles $[\tau_1, \dots, \tau_q]$ on the estimation of the α_i parameters (Koenker 2004, 77), $\rho_\tau(\cdot)$ is again the quantile loss function and λ is a shrinkage parameter. For $\lambda \to 0$, one would obtain the fixed effect estimator based on optimizing (6), while for $\lambda \to \infty$ one would obtain an estimate of the model purged of the fixed effects (Koenker 2004, 78).[6] A routine that implements this estimator (and variants of it) has been provided by Roger Koenker and Stefan Bache and is available for the statistical software package *R*. More recent work on fixed effects quantile regressions also deals with potential endogeneity of explanatory variables. The approach outlined in Harding and Lamarche (2009) tries to overcome this problem by extending the work of Chernozhukov and Hansen (2008) and developing an estimation technique which is able to control for unobserved heterogeneity on the one hand, but is also able to incorporate the idea of instrumental variables.

4. Estimation Results

4.1. Mean Regressions

Table 3 shows the results for the two mean regressions applied in this study. The dependent variable is the logarithm of the frequency ratio of the respective offense. Besides the unemployment rate, OLS regressions also include the logarithm of the clearance rate for the respective offense lagged by one period,[7] the logarithm of disposable income, the logarithm of population density, the share of foreigners, the share of the young population (aged younger than 15), the share of the youth population (aged 15 to 24), the share of the "active population" (aged 25 to 55), the share of unskilled workers, and time dummies. The analysis covers the years from 2005 to 2009, although data are available from 2003 onwards. The reason for the choice of this time span is a major labor market reform (the so called "Hartz-Reform") implemented in 2005. This reform had led to a redefinition of unemployment: Most people who were receiving social welfare benefits *(Sozialhilfe)* before 2005 have been counted as "employable" thereafter and therefore unemployed after 2005 (instead of being out-of-the-labor force before). To avoid potentially biased results stemming from the changing definition of the unemployment rate, we restricted the time window to the years 2005 to 2009.

Table 3 displays the results of OLS and fixed effects mean regressions.[8] The estimated parameters are to be interpreted as semi-elasticities: an increase of the unemployment rate by one unit (one percentage point in this case) increases criminal activity by percent. Based on the standard OLS regression, unemployment has a positive and significant effect on crime for burglary (9.6 percent) and auto theft (10.3 percent), and a negative but insignificant effect on assault (-0.4 percent). These results are in line with previous findings from the literature: the unemployment rate usually has a significant positive effect on property crimes (here burglary and auto theft) while only small or insignificant effects on violent crime, here measured in terms of assault. This is consistent with the vast majority of cross-section findings in the literature, and

6 If the shrinkage parameter goes to infinity, the estimated fixed effects have to approach zero in order to find a minimum of equation (7).

7 Lagging the clearance rate by one period mitigates the problem of simultaneity.

8 Note that the maximal number of 398 districts used in the subsequent multivariate analysis differs from the one in Messner et al. (2013), who report results based on 413 districts. The difference might be explained by the way data are employed. As violent offenses can be rare events in less populated districts (contrary to large cities and the more urbanized areas), Messner et al. decided to use the average annual robbery and assault rates per 100,000 population for the three-year period 2005, 2006, and 2007. By contrast, our paper fully exploits the panel data structure of the five years period from 2005 to 2009, i.e. data are collected over time and over the same districts and then regressions (in form of panel econometric methods and quantile techniques) are run over these two dimensions. In turn, some district observations are lost due to redefinitions of geographical district boundaries which took place during 2007 and 2009 in the East German states Sachsen-Anhalt and Sachsen (see Wikipedia 2015, for details of boundary reforms in Germany). Further observations are lost due to missing data of explanatory variables. Lastauskas and Tatsi (2013), who estimate cross-sectional spatial models based on data from 2008 and 2009, report the use of 402 districts. In 2007 the total number of districts was still 439. This number fell (with time-variant boundaries) to 412 in 2009. As of 2015 there are 402 districts (295 *Landkreise* and 107 *kreisfreie Städte*).

even with findings on the influence of the contemporaneous unemployment rate on the assault rate in time series studies (Phillips and Land 2012). However, the OLS specification does not consider the panel structure, so regional peculiarities such as locally high or low shares of unreported crimes or unobserved factors of urbanity are not taken into account. The fixed effect regression (column FE in Table 3) does include district fixed effects and is therefore able to control for unobserved heterogeneity across districts.[9] Applying it does not change the insignificance of unemployment on assault, but parameters of the FE estimation on car theft and burglary differ substantially from the ones of OLS estimation. The effect on auto theft becomes insignificant, and the estimated parameter on the link between burglary and unemployment is been reduced to 4.4 (recall that the median district unemployment rate is about 9 percent, so a one percentage point fall would be equivalent to a -11.1 percent change experienced by a median district).[10]

Reported inference is based on cluster-robust standard errors. The employed Stata command is based on the work of White (1980, 1984) and Huber (1964, 1967), and allows the assumption of independently distributed residuals to be relaxed. The routine produces consistent standard errors if the residuals are correlated within, but uncorrelated between clusters (districts). In spatial models it may be rather optimistic to assume that the residuals are correlated within but uncorrelated between clustered regions. Beck and Katz (1995) suggested the application of panel-corrected standard errors (PCSEs) which correct for contemporaneous correlation between the clusters. However, their approach is based on large T-asymptotics (a large time-series dimension), while our approach is based on a large cross-sectional dimension N, with N » T. Hoechle (2007) points out that the PCSE estimate will be rather imprecise if the ratio T/N is small. Thus, we stick with White-Huber robust standard errors, which seems to be justified as we only consider five years of data and also correct for time and district fixed effects.

Table 3: Results from the mean regression for the effect of unemployment on crime

Offense	OLS	FE
Assault	–0.364 (0.289)	0.175 (0.382)
Burglary	9.634*** (0.909)	4.432*** (1.493)
Auto theft	0.319*** (0.911)	–1.053 (0.851)
Number of observations	1,947	1,947

Note: Cluster-robust standard errors in parentheses. Dependent variable: log (frequency ratio). OLS regressions include log(clearance rate) for the respective offense lagged by one period, log(disposable income), log of population density, share of foreigners, share of the young population, share of the youth population, share of the active population, share of unskilled workers and time dummies. FE regressions include log (clearance rate) for the respective offense lagged by one period, log (disposable income) and time dummies. All regressions are weighted using the size of the district population. Note that due to regional reorganizations some districts had to be excluded from the data set. Further note that panel data analysis requires at least two subsequent periods with identical regional boundaries and without missing data. This was the case for 383 districts with five-year time spans, one additional district for the four-year time span between 2006 and 2009, and an additional fourteen districts for the time span 2008/09, resulting in 1,947 observations. *, **, and *** denote significance at the 10%, 5% and 1% level respectively.

9 However, this comes at some costs. On inclusion of fixed effects all slowly varying or quasi time-invariant variables became highly collinear and completely insignificant. For this reason share of foreigners, share of unskilled workers, and all other variables representing the population structure have been omitted from the fixed-effect specification.

10 Sieger (2014) confirms significance (respectively insignificance) and sign of presented FE results using an IV approach.

4.2. Quantile Regressions

Table 4 gives an overview of the results for the ordinary quantile regression. Some interesting insights emerge from comparing the methods under consideration. While the estimated effect of an increase in the unemployment rate on the rate of assault was insignificant in the OLS regression (see Table 3), it is positive and significant in the ordinary quantile regression at least for low levels of crime (the 5 percent and 25 percent quantiles). In addition, the strength of the crime-unemployment link is slightly decreasing (Figure 10). The downward slope is even more pronounced for burglary and auto theft (Figures 11 and 12), giving rise to the interpretation that agents are committing crime when the "supply" of crime is rather low, and "tolerance" towards crime is still high (see also Ehrlich 1996, who argues that tolerance towards crime represents the demand side of a market of offenses). Moreover, OLS estimates are within the middle of the respective quantile regressions, supporting the apprehension that in the OLS regression the effects at different quantiles are simply averaged and do not reveal the full picture of the crime-unemployment relationship (but note that mean and median results differ due to the skewed crime distribution).

Table 4: Results from the ordinary quantile regression for the effect of unemployment on crime

Offense	Quantiles				
	0.05	0.25	0.5	0.75	0.95
Assault	1.126* (0.681)	0.947*** (0.307)	0.418 (0.333)	0.412 (0.392)	–0.715 (0.977)
Burglary	15.185*** (1.212)	9.718*** (0.885)	7.491*** (0.671)	7.038*** (0.960)	5.924*** (1.051)
Auto theft	12.789*** (1.619)	9.835*** (0.741)	8.829*** (0.689)	8.579*** (0.889)	7.921*** (1.631)

Note: Bootstrapped standard errors in parentheses. Dependent variable: log (frequency ratio). All regressions are based on 1,947 observations and include log(clearance rate) for the respective offense lagged by one period, log (disposable income), log (population density) share of foreigners, share of the young population, share of the youth population, share of the adult population, share of unskilled workers and time dummies. All regressions are weighted using the size of the district population. See Table 3 for details on data. *, **, and *** denote significance at the 10%, 5% and 1% level respectively.

Figure 10: Effect of an increase in unemployment on assault at different quantileshere

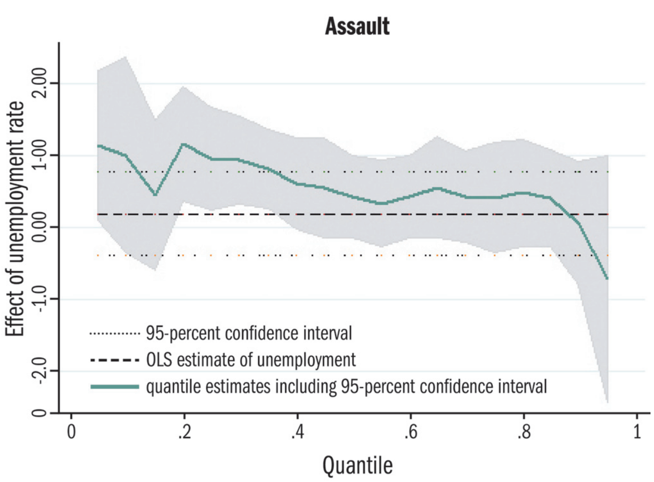


Figure 11: Effect of an increase in unemployment on burglary at different quantiles

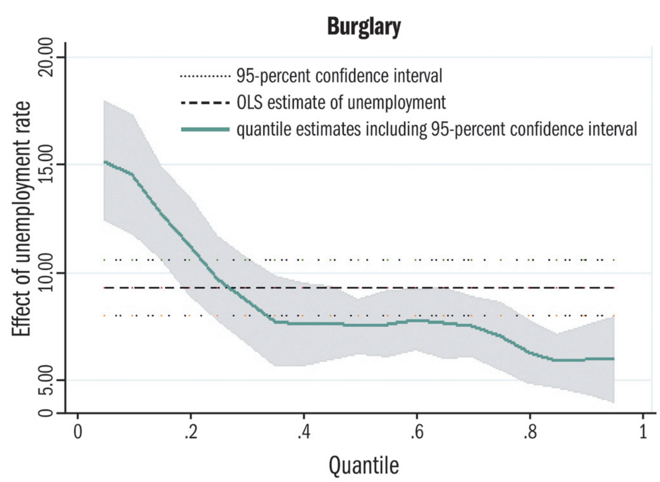


Figure 12: Effect of an increase in unemployment on auto theft at different quantiles

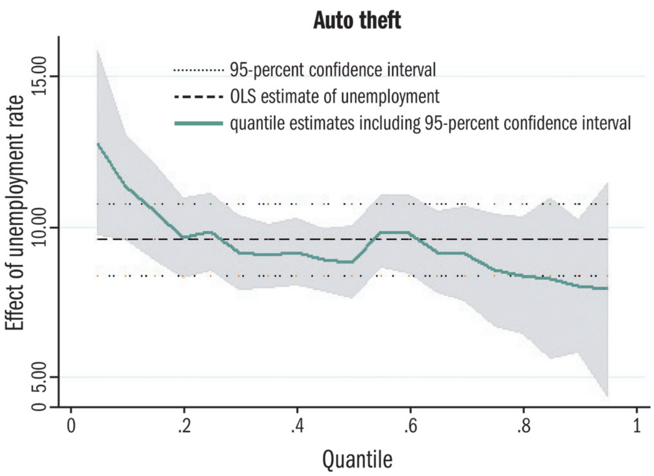


Table 5 displays the results from quantile regression with fixed effects. Note that the usual inclusion of time-fixed effects has caused indications for serious multicollinearity problems such that we deviate from previous specifications by including a linear time trend instead of time dummies. We consider quantile regressions with fixed effects as the most reliable and preferred, as they control for potential district-specific unobserved heterogeneity. The significance of parameter estimates is in line with the one obtained from ordinary quantile regression, but the pattern of the unemployment-crime link has changed, in particular for property crimes. The effect on burglary and auto theft is still significant for all quantiles, but results do not confirm the decreasing pattern in Table 4 (where estimates below the 50 percent quantile of the regional crime distribution are particularly high). Instead, quantile parameters exhibit a rather flat crime-unemployment profile, which is not indicative for or against stigma or opportunity-based behavior. The effect of unemployment on assault is significant for rather low-crime (25 percent-quantile) and median-crime regions (50 percent quantile) and insignificance is confirmed for quantiles above 50 percent. This lack of significance for high-crime areas is in line with opportunity-based criminal behavior, but the FE approach does not confirm the strictly downward effect found with ordinary quantile regressions.[11]

Table 5: Results from quantile regression with fixed effects for the effect of unemployment on crime

Offense	Quantiles				
	0.05	0.25	0.5	0.75	0.95
Assault	0.908 (0.570)	0.896** (0.378)	1.344*** (0.432)	0.980 (0.856)	0.388 (1.610)
Burglary	11.910*** (1.297)	10.990*** (1.383)	11.270*** (1.363)	11.318*** (1.596)	11.888*** (1.418)
Auto theft	12.416*** (1.505)	11.005*** (1.330)	11.706*** (1.079)	12.628*** (1.560)	14.434*** (1.943)

Note: Bootstrapped standard errors in parentheses. Dependent variable: log (frequency ratio). All regressions are based on 1,947 observations and include log (clearance rate) for the respective offense lagged by one period, log (disposable income) and a linear time trend. All regressions are weighted using the size of the district population. See Table 3 for details on data. *, **, and *** denote significance at the 10%, 5% and 1% level respectively.

11 Sieger (2014) also presents some preliminary quantile FE IV estimates. However, the results lack robustness and seem to be highly sensitive to the choice of specification such that we do not comment on them further in this paper.

4.3. Further Results

Tables 6 and 7 show results for the effects of other factors of the economics of crime model: clearance rate and net income. In line with theoretical expectations, the log of the lagged clearance rate is negative and significant for all but one model specification. The only exception is the fixed effects mean regression of assault. The effect of the clearance rate for property crimes ranges between -0.15 and -0.30. The much larger effect of a 1 percent change in the clearance rate for assault can be explained by its relatively high median clearance rate of 91 percent (compared to only 13 percent for auto theft, and 23 percent for burglary) and its low variation across districts. The quartiles at 25 percent and 75 percent are 89 percent and 93 percent, respectively, such that a change by 1 percent (for instance, from a 90-percentile down to the 89.1-percentile) already represents a substantial change, in particular given that assault rates – in contrast to auto theft and burglary – do not show strong variation across districts (see above).[12] So when interpreting and comparing parameter estimates it needs to be taken into account that increasing the clearance rate for assault by 1 percent would be more difficult, less likely, and perhaps also much more costly than increasing the clearance rate for property crimes by the same amount.

As regards the structure of the quantile regression estimates, Bandyopadhyay et al. (2015) report that the crime-reducing effect of higher detection rates is stronger in low-crime areas. This can be confirmed for assault and using ordinary quantile regressions (as also applied by Bandyopadhyay et al.) in Table 6, whereas for fixed-effects and property crimes there is no obvious quartile-specific pattern.

Table 6: Results from pooled OLS and ordinary quantile regression

Log(clearance rate), lag(-1)		Quantiles		
	POLS	0.25	0.5	0.75
Assault	2.762*** (0.275)	−2.305*** (0.367)	−1.685*** (0.342)	−1.590*** (0.465)
Burglary	−0.313*** (0.035)	−0.274*** (0.053)	−0.309*** (0.034)	−0.299*** (0.028)
Auto theft	−0.201*** (0.027)	−0.156*** (0.026)	−0.184*** (0.027)	−0.164*** (0.031)

Net income		Quantiles		
	POLS	0.25	0.5	0.75
Assault	−1.265 *** (0.079)	−1.137*** (0.089)	−1.187*** (0.113)	−1.118*** (0.099)
Burglary	−0.010 (0.212)	−0.729*** (0.239)	−0.613*** (0.212)	−0.269 (0.226)
Auto theft	−0.007 (0.233)	−0.830*** (0.256)	−0.717*** (0.271)	−0.289 (0.254)

Note: Cluster-robust (POLS) and bootstrapped standard errors in parentheses.
Dependent variable: log (frequency ratio). See Tables 4 and 5 for further details. *, **, and *** denote significance at the 10%, 5% and 1% level respectively.

12 The 25% quartiles of auto theft and burglary are 8.1% and 20%, respectively. The corresponding 75% quartiles are 15.6% and 33.9% (Sieger 2014).

Table 7: Results from fixed effects means and quantile FE regressions

Log(clearance rate), lag(-1)		Quantiles		
	FE	0.25	0.5	0.75
Assault	0.214 (0.182)	–3.919*** (0.393)	–4.158*** (0.376)	–4.437*** (0.586)
Burglary	–0.091*** (0.025)	–0.299*** (0.053)	–0.395*** (0.052)	–0.376*** (0.039)
Auto theft	–0.040*** (0.011)	–0.252*** (0.047)	–0.245*** (0.038)	–0.278*** (0.041)

Net income		Quantiles		
	FE	0.25	0.5	0.75
Assault	–0.430** (0.197)	–0.899*** (0.106)	–0.777*** (0.112)	–0.844*** (0.170)
Burglary	–0.810 (0.770)	0.593 (0.391)	0.220 (0.403)	0.141 (0.412)
Auto theft	0.219 (0.431)	0.423 (0.360)	0.315 (0.299)	0.261 (0.382)

Note: Cluster-robust (FE) and bootstrapped standard errors in parentheses.
Dependent variable: log (frequency ratio). See Tables 4 and 5 for further details. *, **, and *** denote significance at the 10%, 5% and 1% level respectively.

Results for net income (Table 7) are more heterogeneous. The clearest empirical evidence can be observed for assault, where all estimates confirm the hypothesis that better legal earning opportunities are associated with lower crime rates. The same effect occurs for the 25 percent and 50 percent percentiles (but not for the 75 percent percentile) of ordinary quantile regressions for auto theft and burglary. However, quantile fixed effects, pooled OLS, as well as standard mean FE modelling do not confirm this result so that we cannot reach a unanimous conclusion with respect to the effect on property crimes.

5. Summary and Conclusions

This paper uses regional panel data from about four hundred German districts and quantile regressions to study the effect of unemployment on crime. The main contribution is to test the hypothesis that size and significance of the effect of unemployment on crime may depend on the relative position of the prevailing regional crime level within the overall distribution of crime rates, i.e. whether the local crime rate is relatively low or large. We present two conjectures about the non-linear pattern of the relationship between unemployment and crime. First, there could be a downward sloping crime-unemployment link with a high marginal impact of unemployment on crime for low-crime regions. This pattern might arise when job losses imply high incentives and relatively large opportunities to become criminals. Likewise, potential criminals would face less crime prevention and precautions from potential victims than those in regions where crime is already more elevated. The opposite pattern might follow from the alternative stigma effect: If there are only a few criminals around, there are strong moral obstacles to becoming a criminal, since any detection would make the person a “black sheep.” This contrasts to a situation with many criminals in the neighborhood where acting illegally becomes more likely as many others or even peers already have criminal experience. Empirical results show that conventional mean regressions might indeed produce misleading results. For instance, while simple OLS and FE regressions depict insignificant results for *assault* (which would confirm the usual result for violent crime found in the literature), the preferred fixed quantile regressions reveal positive and significant effects for the districts representing the 50 percent and 25 percent percentiles of the crime distribution, i.e. for median- and low-crime regions, respectively. The analysis of *property crimes* illustrates that results based on quantile fixed effect modeling might substantially differ from those of ordinary quantile regressions. The latter seem to indicate behavior in line with the opportunity-based approach (the effect of unemployment on crime in regions with relatively low crime rates is stronger than in regions with relatively high crime rates), but this result cannot be confirmed when including fixed effects. As this technique has the advantage that it corrects for unobserved heterogeneity and is therefore a favored estimation strategy, we may conclude that the positive and significant effect on considered property crime categories is rather constant across quantiles. This indicates a conventional linear relationship between property crime and unemployment and corroborates standard theoretical explanations based on expected values of distributions and usual mean regressions. However, future work should also use individual data to identify and better understand the complexity of incentives and activities of potential criminals in high- and low-crime regions.

References

Albrecht, Hans-Jörg. 1997. Ethnic Minorities, Crime, and Criminal Justice in Germany. *Crime and Justice* 21:31–99.

Bandyopadhyay, Siddhartha, Samrat Bhattacharya, and Rudra Sensarma. 2015. An Analysis of the Factors Determining Crime in England and Wales: A Quantile Regression Approach. *Economics Bulletin* 35 (1): 665–79.

Beck, Nathaniel, and Jonathan. N. Katz. 1995. What to Do (and Not to Do) with Time-series Cross-section Data. *American Political Science Review* 89: 634–47.

Becker, Gary. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76 (2): 169–217.

Cantor, David, and Kenneth C. Land. 1985. Unemployment and Crime Rates in the Post-World War II United States: A Theoretical and Empirical Analysis. *American Sociological Review,* 50 (3): 317–32.

Chernozhukov, Victor, and Christian Hansen. 2008. Instrumental Variable Quantile Regression: A Robust Inference Approach. *Journal of Econometrics* 142 (1): 379–98.

Edmark, Karin. 2005. Unemployment and Crime: Is there a Connection? *Scandinavian Journal of Economics,* 107(2): 353-373.

Ehrlich, Isaac. 1996. Crime, Punishment, and the Market of Offenses. *Journal of Economic Perspectives* 10 (1): 43–67.

Entorf, Horst, and Hannes Spengler. 2000. Socioeconomic and Demographic Factors of Crime in Germany: Evidence from Panel Data of the German States. *International Review of Law and Economics* 20 (1): 75–106.

Entorf, Horst, and Hannes Spengler. 2002. *Crime in Europe*. Heidelberg and New York: Springer.

Freeman, Richard B. 1996. Why Do So Many Young American Men Commit Crimes and What Can We Do About It? *Journal of Economic Perspectives* 10 (1): 25–42.

Farrington, David P. 1986. Age and Crime. *Crime and Justice* 7:189–250.

Funk, Patricia. 2004. On the Effective Use of Stigma as a Crime-deterrent. *European Economic Review* 48:715–28

Glaeser, Edward L., and Bruce Sacerdote. 1999. Why Is There More Crime in Cities? *Journal of Political Economy* 107 (6): 225–58.

Grogger, Jeff. 1998. Market Wages and Youth Crime. *Journal of Labor Economics* 16 (4): 756–91.

Gould, Eric, Bruce Weinberg, and David Mustard. 2002. Crime Rate and Local Labor Market Opportunities in the U.S.: 1979–1997. *Review of Economics and Statistics* 84 (1): 45–61.

Harding, Matthew, and Carlos Lamarche. 2009. A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables. *Economics Letters* 104 (3): 133–35.

Hoechle, Daniel 2007. Robust Standard Errors for Panel Regressions with Cross-sectional Dependence. *Stata Journal* 7 (3): 281–312.

Huber, Peter J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35.

Huber, Peter J. 1967. The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 221–33. Berkeley, CA: University of California Press.

Koenker, Roger, and Kevin F. Hallock. 2001. Quantile Regression. *Journal of Economic Perspectives* 15 (4): 143–56.

Koenker, Roger. 2004. Quantile Regression for Longitudinal Data. *Journal of Multivariate Analysis* 91 (1): 74–89.

Koenker, Roger. 2005. *Quantile Regression*. Cambridge, Mas: Cambridge University Press

Lastauskas, Povilas, and Eirini Tatsi. 2013. Spatial Nexus in Crime and Unemployment at the Dawn of Crisis: Evidence from Germany. Cambridge Working Papers in Economics 1359, Faculty of Economics, University of Cambridge.

Levitt, Steven D. 1997. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review* 87 (3): 270–90.

Levitt, Steven D. 1999. The Limited Role of Changing Age Structure in Explaining Aggregate Crime Rates. *Criminology* 37 (3): 581–97.

Levitt, Steven D. 2001. Alternative Strategies for Identifying the Link between Unemployment and Crime. *Journal of Quantitative Criminology* 17 (4): 377–90.

Levitt, Steven D. 2004. Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not." *Journal of Economic Perspectives* 18 (1): 163–90.

Lin, Ming-Jen. 2008. Does Unemployment Increase Crime? *Journal of Human Resources* 43 (2): 413–36.

Messner, Steven F., Raymond H. C. Teske Jr., Robert D. Baller, and Helmut Thome. 2013. Structural Covariates of Violent Crime Rates in Germany: Exploratory Spatial Analyses of Kreise. *Justice Quarterly* 30 (6): 1015–41.

Öster, Anna, and Jonas Agell. 2007. Crime and Unemployment in Turbulent Times. *Journal of the European Economic Association* 5 (4): 752–75.

PKS (Polizeiliche Kriminalstatistik). Various issues from 2003 to 2009, 2013 Bundeskriminalamt. Wiesbaden, Germany. Website: http://www.bka.de/DE/Publikationen/PolizeilicheKriminalstatistik/pks__node.html.

R+V Versicherung Websites: https://www.ruv.de/de/presse/download/pdf/aengste-der-deutschen-2014/grafiken-bundeslander.pdf, https://www.ruv.de/de/presse/download/pdf/aengste-der-deutschen-2014/grafiken-bundesweit.pdf

Raphael, Steven, and Rudolf Winter-Ebmer. 2001. Identifying the Effect of Unemployment on Crime. *Journal of Law and Economics* 44 (1): 259–83.
Phillips, Julie, and Kenneth C. Land. 2012. The Link between Unemployment and Crime Rate Fluctuations: An Analysis at the County, State and National Level. *Social Science Research* 41 (3): 681–94.
Sieger, Philip. 2014. Essays in Microeconometrics: Establishing Causality in Observational Studies. Ph.D. dissertation, Goethe University, Frankfurt am Main.
Sozialgesetzbuch (SGB) Drittes Buch (III). Website: http://www.gesetze-im-internet.de/sgb_3/
Strafgesetzbuch (StGB). Website: http://www.gesetze-im-internet.de/stgb/
Wikipedia. 2015. Kreisreformen in der Bundesrepublik Deutschland nach 1990, Wikipedia, http://de.wikipedia.org/wiki/Kreisreformen_in_der_Bundesrepublik_Deutschland_nach _ 1990.
White, Halbert. 1980. A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica* 48: 817–38.
White, Halbert. 1984. *Asymptotic Theory for Econometricians*. Orlando, FL: Academic Press.
Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.
Yearwood, Douglas L., and Gerry Koinis. 2009. Revisiting Property Crime and Economic Conditions: An Exploratory Study to Identify Predictive Indicators beyond Unemployment Rates, Munich Personal RePEc Archive, MPRA Paper 16834.
Young, Thomas J. 1993. Unemployment and Property Crime: Not a Simple Relationship. *American Journal of Economics and Sociology* 52 (4): 413–16.

Horst Entorf
entorf@wiwi.uni-frankfurt.de

Philip Sieger
sieger@wiwi.uni-frankfurt.de

International Journal
of Conflict and Violence

urn:nbn:de:0070-ijcv-2014226
IJCV: Vol. 8 (2) 2014

Local Media in Global Conflict: Southeast Asian Newspapers and the Politics of Peace in Israel/Palestine

Yakubu Ozohu-Suleiman, Department of Mass Communication, Ahmadu Bello University, Zaria, Nigeria
Sidin Ahmad Ishak, Department of Media Studies, University of Malaya, Kuala Lumpur, Malaysia

Vol. 8 (2) 2014

ISSN: 1864–1385

Local Media in Global Conflict: Southeast Asian Newspapers and the Politics of Peace in Israel/Palestine

Yakubu Ozohu-Suleiman, Department of Mass Communication, Ahmadu Bello University, Zaria, Nigeria
Sidin Ahmad Ishak, Department of Media Studies, University of Malaya, Kuala Lumpur, Malaysia

It is often assumed that local media are a potential deescalating tool in global conflict. This study examines how four leading newspapers in Southeast Asia (*Star* of Malaysia, *Philstar* of the Philippines, *Jakarta Post* of Indonesia, and *The Nation* of Thailand) reported the Israeli-Palestinian conflict during the year after the 2009 Gaza War. A census of 536 reports was coded for tones (to detect alignment), frames (to detect characterization of the conflict), and sources (to examine correlation with coverage tones). The results show fragmented alignment of the newspapers with Palestine and Israel. Conflict frames on offensives, fighting, threats, military strategies, demonization, death, and destruction were most prevalent. Coverage tones were significantly correlated with sources, suggesting that the potential of local media to serve as deescalating tools in global conflicts is subject to the varying political contexts in which they operate in relation to specific conflicts.

Despite growing recognition of the importance of local media in shaping conflicts (Puddephatt 2006), existing evidence still upholds Blondel's observation that "much of the research on the role of the media in conflict has focused on international news organizations" (2004, 27). Very few studies have been conducted on the roles local media play in bringing news of international conflicts to the attention of local audiences. This evident research deficit implies that local media involvement in global diffusion of conflict is ignored. Of specific concern here is the growing assumption that local media are a potential de-escalation tool in global conflict, on which there has been little in the way of empirical research. Attempts to examine how local media function in conflict de-escalation have actually been based on local conflicts, and focused on specific peace projects – examples include Rwanda's Studio-Ijambo (Hagos 2001; Paluck 2007) and Bosnia's Open Broadcast Network (OBN) and Free Exchange Radio Network (FERN) (Bratic 2009, 21–22). What seems to be more common are studies focusing on how global media report local conflicts, example the CNN effect in Somalia (McSweeney 2011) and Aljazeera's and the BBC's framing of Darfur (Kinner 2005).

Surprisingly, the most neglected conflict in local media research is the most globally diffused and intractable one – the Israeli-Palestinian conflict, whose metamorphosing recurrence and growing chain of links to global terrorism has intensified security surveillance and restrictions on civil liberty all over the world. While this conflict seems central to the globally impacting revolution that swept through Middle East and North Africa, how local media around the world present the conflict to their audiences is an important question that has remained largely unanswered. This observation stems from negative antecedents, in which local media have often been found to be effective in pushing people to engage in conflict and mobilizing public support for war. Examples of this include the manipulation of media to justify use of armed force in the Chechen conflict (Baranovsky, 2012); the abuse of local media in facilitating conflict in former Yugoslavia (Bratic 2009) and modern Russia (Glukhov 2009); and the goading of ethnic genocide by Radio-Télévision-Libre-des-Milles-Collines in Rwanda in 1994 (Des Forges 2007; Paluck 2007).

The prevailing assumption seems to be that the "international media" is the most appropriate choice for assessing the media's role in "international conflict." In a sense, this can be defended considering that in international conflicts of global relevance such as the Israeli-Palestinian conflict, international media and news agencies sometimes serve as sources of information to local media outlets, which are often not financially equipped to source first-hand information. However, in terms of both global diffusion and de-escalation, local media operating within a country actually have greater potentials than international news media, as Blondel (2004) argues, depending on the role they choose or are able to play.

This study was conceived to examine how major newspapers in selected Southeast Asian countries – namely, *The Star* (Malaysia), *The Philstar* (the Philippines), *The Jakarta Post* (Indonesia), and *The Nation* (Thailand) reported the Israeli-Palestinian conflict during the year after the 2009 Gaza war. Of course that fighting, perhaps more than ever in the history of the Israeli-Palestinian conflict, attracted civic and news reactions from all over the world including Southeast Asia. The goal of this study is therefore to determine how the local press handled the one-year post-war period in terms of the dominant news frames used and their tones towards the conflict actors, Israel and Palestine. It is also of interest in this study to identify the major news sources used by those newspapers in reporting the conflict, and to establish whether there are correlations between the tone of their coverage and the sources they use. These goals are relevant to the need for increased understanding of the roles of local media as possible tools of de-escalation and peace building in global conflict.

1. Global Conflict in Local News

In a conflict of international or global dimensions, there always seems to be a nexus that explains coverage in local media. A local media outlet may report international conflict to serve its commercial interests, as the *Australian* did in relation to the Democratic Republic of Congo (Hawkins 2009). They may also rally round the flag in patriotism or acquiescence to a nation's interest, as the British media did during the 1982 Falklands War (Goddard, Robinson, and Parry 2008). In another sense, the local media may report an international conflict simply to meet its ethical responsibility to bring international news events to the attention of local audiences, as the British *Daily Mirror* did in its coverage of the US-led invasion of Iraq (see Goddard, Robinson, and Parry 2008).

Again, local media may report an international conflict to promote a specific local agenda and response based on ideological considerations, as Dutch media demonstrated in their stereotypical reporting of the Bosnian war and proposals for Dutch government action (Ruigrok 2008). Local media may also become involved in international conflict situations in order to contradict perceived opponents of the geopolitical interests and ideological values that define their existence and operations. This is journalism of attachment, in which the reporters are regarded as participants in the conflict they report (Ruigrok 2008). Local media attachment in international conflict sometimes occurs where the home country is directly involved in the conflict. In this case, the media often accompany their states to the war front in a rally-round-the-flag approach, and function as tool of government propaganda as Taylor (1992) observed. This was exemplified in the role played by the British media during the Falklands war, and by the American media during the Vietnam and Gulf wars (Hallin 1986; Bennet and Paletz 1994).

Where the home country is not directly involved in a conflict, shared transnational ideology, for instance democracy, capitalism, religion, or a complex combination of some or all of them, might connect local media to a global conflict. In recent years, religion seems to have become the most important transnational ideology affecting media coverage of global conflicts. It seems indeed correct to assert that since the demise of the global anti-communist propaganda of the cold war era, religion has emerged as the most important ideological influence in media coverage of global conflicts, ostensibly because religion is at the heart of current trends in global conflict. This trend appears to affect the Israeli-Palestinian conflict, in which Muslims and Jews and, in some places, Christians are perceived and cited as the conflict actors, whose global presence explains the global spread of the conflict. Islam is heavily present in civic life and government levels in Malaysia and Indonesia, where over 60 percent and over 80 percent of local populations respectively are Muslims (Hosen 2005). Christianity is heavily present in the Philippine where over 90 percent of the population are Christians (Miller n.d.). In Thailand, over 90 percent of the population practices Buddhism.[1]

A few attempts have been made in public discourse and research to examine the linkage between religion and the Israeli-Palestinian conflict. Some scholars, like Edward Luttwak and Shibley Telhami consider religion not to be the key determinant in the Israeli-Palestinian conflict.[2] Researchers like Slater (2006) and Roy (2004), who also share this view, believe that the political layers are more important than religion for understanding the Israeli-Palestinian conflict. There are, however, other researchers (Gopin 2002; Ranstorp 1996) who positively associate religion with the conflict, seeing it as an asset in the search for peace; and others still (Baumgartner et al. 2008), who find a strong association between religion and the global spread of the conflict. While these works are not directly related to mass media, lack of research into the roles of local media in global conflict makes it even more difficult to find scientific evidence on the linkage between religion and local media coverage of the Israeli-Palestinian conflict. In fact, religion has not been a widely used variable for explaining the sources of foreign policy attitudes, as Baumgartner et al (2008) observed. However, it has been established that an association exists between correspondents' demographic backgrounds and their coverage of international conflicts (El-Nawawy 2002).

An association between demographic background and a journalist's account of international conflict suggests that religion, as a demographic indicator, may be able to explain differences in local media coverage of a conflict between predominantly Islamic setting and predominantly Christian settings. On that basis, this study assumes that Southeast Asian newspapers operating in a predominantly Islamic setting (*Star* of Malaysia and *Jakarta Post* of Indonesia) are likely to report the Israeli-Palestinian conflict in favor of Palestinians. Similarly, Southeast Asian newspapers operating in a predominantly Christian setting (*Philstar* of the Philippine) are likely to report the conflict in favor of Israel. Southeast Asian newspapers operating under neither Islamic nor Christian religious predominance (*Nation* of Thailand) are likely to be non-aligned and therefore more neutral than inclined towards Israel or Palestine. To investigate these assumptions, this study questions the coverage tones in the newspapers thus:

> Research question 1: Does the tone of coverage of the Israeli-Palestinian conflict by Southeast Asian press reflect alignment with Israel and Palestine, and how does this vary in newspapers from different religious backgrounds?

To determine how the press characterized the conflict, the study asks:

> Research question 2: What is the dominant frame used by Southeast Asian presses in reporting the Israeli-Palestinian conflict, and how does this differ between newspapers from different religious backgrounds?

In developing a framework for understanding the influences around US media coverage of the Vietnam War, Hallin (1986) documented the sources used by the US media. Taking a cue from Hallin, this study goes on to ask:

> Research question 3: What are the major news sources used by Southeast Asian press to report the Israeli-Palestinian conflict, and how does this vary between newspapers from different religious backgrounds?

To find out if the tones of coverage (as dependent variable) relate to the sources used by the newspapers (as independent variable), the study followed up with:

> RQ4: Are there significant associations between the tones of coverage and the news sources from which local newspapers in Southeast Asia reported the Israeli-Palestinian conflict?

2. Method

This study is based on a content analysis of coverage of the Israeli-Palestinian conflict in the *Star* of Malaysia, Jakarta

1 Information regarding religion in Malaysia and Thailand from http://www.globalsecurity.org/military/world/malaysia/religion.htm and http://www.amazing-thailand.com/relig.html respectively (accessed June 20, 2013).

2 Edward Luttwak is Senior Associate of the Center for Strategic and International Studies, Washington, D.C.; Shibley Telhami is Anwar Sadat Chair and Professor for peace and development, University of Maryland.

Post of Indonesia, *Philstar* of the Philippines, and *The Nation* of Thailand during the year following the 2009 Gaza war. Although these newspapers are published in English only, they were selected on the strength of their circulation and popularity within their countries of origin. According to the Audit Bureau of Circulation (2011), Star is the largest English-language newspaper in Malaysia, with daily circulation close to 300,000. The newspaper is owned by an alliance of the Malaysian Chinese Association and Malaysia's ruling party, Berisan Nasional. The *Jakarta Post* is a leading Indonesian English daily, independently owned by Bina Media Tenggara but with a political orientation toward public office-holders (Eklof 2004). According to the Nielson Media Index (2011), Philstar is among the three largest newspapers in the Philippines. *The Nation* is one of Thailand's top English newspapers with daily circulation of between 60,000 and 80,000. The paper is considered nationalist, pro-royalist, and pro–elitist government in its editorial policies.[3]

The unit of analysis was the article. The texts of articles were obtained from the websites of the newspapers using the search terms "Israel Palestine," "Israeli Palestinian," and "Israel Palestinian". Stories published between November 22, 2009, and November 21, 2010, were analyzed. The newspapers published different volumes of relevant reporting during this period. The *Star* of Malaysia produced 230 related articles and the *Jakarta Post* of Indonesia 222, while the *Philstar* produced 49 and *The Nation* of Thailand 35. Due to the low output of the latter two newspapers, we conducted a census study in which all the articles were included in the study population. Overall, 536 articles were analyzed.

2.1 Categories and Measurement

For coding the characterization of the Israeli-Palestinian conflict by Southeast Asian Press, this study drew upon the most commonly used frames in media coverage of conflict identified by Semetko and Valkenburg (2000): "conflict," "human interest," "economic consequences," "morality," and "responsibility," but also created a "peace" frame to determine the newspaper's tendency to play a deescalating role. Semetko and Valkenburg's "economic consequences" was modified as "consequences," broadening its meaning to include non-economic consequences. The "conflict," "morality," and "responsibility" frames were retained as defined by the authors.

Coverage tones were analyzed using the "slant" category, coded into "favorable," "unfavorable," and "neutral" stories. The "sources" were coded into "news agencies" (mostly AP, Reuters, and Xinhua), "other media," "government" (former or active members of parliament, members of the executive, including the president), "civic bodies" (external individuals, human rights, civic and interest groups), and "independent" sources (editorials, columns, opinions, and analysis by internal staff). Articles that did not fit into any of these categories were coded in a residual category of "others." Drawing upon existing literature (Galtung 1998; McGoldrick and Lynch 2000; Semetko and Valkenburg 2000; Howard 2003), these frames and categories were defined as follows:

1. *Peace:* stories on or with emphasis on peace initiatives, events, or subjects.
2. *Conflict:* stories that emphasize offensives, fighting, threats, military strategies, death, and destruction.
3. Human interest: stories that stress suffering and distress in the conflict.
4. *Consequences:* stories that make salient the implications and likelihood of spread of conflict.
5. *Morality:* stories questioning or justifying the moral stand taken by conflict participants and mediators.
6. *Responsibility:* stories that provide background on causes and suggest remedial actions.

The frames, sources, and tones were measured as quantitative variables by identifying and coding articles in which their descriptors are present as "1," and others in which their descriptors are absent as "0." Cases of multiple frames occurring in a single article were resolved by initially recording each occurrence of a descriptor as "1," such that each article (unit of analysis) is coded for the most preva-

3 For more information on the circulation and editorial line of the Nation see http://en.wikipedia.org/wiki/List_of_newspapers_in_Thailand#Newspapers.

lent descriptor. Drawing on Lee and Maslog (2005), this was done to comply with the rule of coding each unit of analysis into only one category. Data was analyzed using SPSS16.0.

2.2 Inter-coder Reliability Test

An inter-coder reliability test was conducted using two experienced coders, who also received one month of specific training. A systematic random procedure was used to sample fifty-three articles, which constituted 10 percent of the content populations, for pilot coding. ReCal2 internet-based software was used to calculate inter-coder reliability. The test result shows a reliability coefficient of between 94 and 100 for percent agreement, and between .72 and 1.0 for Scott *Pi*, Cohen Kappa and Krippendorff's alpha. Reliability is substantial if it yields kappa coefficient ranging between .61 and .80 (Stemler 2001). This range of value is similarly held as substantially reliable in Scott *Pi* and Krippendorff's alpha.

3. Results

Each newspaper produced different search results, all of which were coded. Therefore, rather than base data reporting on mere numerical frequency, we focus on relative percentage occurrences to report the value of each coded variable. Chi square and correlation statistics were then used to answer the research questions.

3.1. RQ1: A Journalism of Fragmented Alignments

To discover whether the newspapers exhibit alignment in reporting the Israeli-Palestinian conflict, the incidence of neutral slanted stories was analyzed.

Table 1: Coverage Tone

	Slanted stories *n* (%)	Neutral stories *n* (%)	Total *n* (%)	Mean	Standard deviation
Star Malaysia	127 (55.2)	103 (44.8)	230 (100)	1.55	.498
Jakarta Post Indonesia	136 (61.3)	86 (38.7)	222 (100)	1.61	.488
Philstar Philippines	36 (73.5)	13 (26.5)	49 (100)	1.73	.466
Nation Thailand	20 (57.1)	15 (42.9)	35 (100)	1.57	.502
Total	319 (59.5)	217 (40.5)	536 (100)	1.60	.491

Table 2: Breakdown of slanted stories

	Favorable to		Unfavorable to		Total
	Israel *n* (%)	Palestine *n* (%)	Israel *n* (%)	Palestine *n* (%)	*n* (%)
Star Malaysia	8 (3.5)	35 (15.2)	80 (34.8)	4 (1.7)	127 (100)
Jakarta Post Indonesia	43 (19.4)	24 (10.8)	65 (29.3)	4 (1.8)	136 (100)
Philstar Philippines	27 (55.1)	3 (6.1)	2 (4.1)	4 (8.2)	36 (100)
Nation Thailand	0 (0.0)	4 (11.4)	11 (31.4)	5 (14.3)	20 (100)
Total	78 (14.6)	66 (12.3)	158 (29.5)	17 (3.1)	319 (100)

Overall, there were significantly more slanted stories than neutral stories $x^2(1, n=536) = 19.410$, $p<.001$ (see Table 1). This result implies that the Southeast Asian press shows a strong general alignment with the conflict actors.

3.1.1. Newspapers from Predominantly Muslim Environments

The *Star* of Malaysia and *Jakarta Post* of Indonesia publish in predominantly Muslim cultures. At the aggregate level, they produced 263 slanted stories and 189 neutral stories (Table 1). Further analysis shows that the amount of slanted stories produced by this group was significantly higher than the neutral stories it produced: $x^2(1, n=452) = 12.115$, $p<.001$. This means that Southeast Asian newspapers operating in predominantly Islamic contexts showed meaningful levels of alignment in reporting the Israeli-Palestinian conflict in the period under study. This alignment is revealed by the large amount of stories slanted against Israel produced by this group (Table 2).

At the level of individual newspapers, there was no significant difference between the amount of slanted and neutral stories in the Star of Malaysia $x^2(1, n=230) = 2.504$, $p=.114$, but in the Jakarta Post of Indonesia, evidence was found of a significant difference in the occurrences of slanted and neutral stories $x^2(1, n=222) = 11.261$, $p=.001$. Both papers were significantly sympathetic towards Palestine in terms of the amount of slanted content that favored and disfavored Palestine – *Star* of Malaysia: $x^2(1, n=39) = 24.641$, $p<0.001$; *Jakarta Post:* $x^2(1, n=28) =14.286$, $p<0.001$. Thus Southeast Asian newspapers from predominantly Islamic countries displayed strong alignment with Palestine in their reporting of the Israeli-Palestinian conflict during this period.

3.1.2. Newspapers from Predominantly Christian Environments

Philstar of the Philippine was the only newspaper included in this study that operates in a predominantly Christian context. Evidence was found of a significant difference between the amount of slanted and neutral stories produced by *Philstar:* $x^2(1, n=49) = 10.796$, $p=.001$. This means the *Philstar* displayed strong alignment in its reporting of the Israeli-Palestinian conflict. This alignment is reflected in the high volume of slanted stories produced by the paper, which significantly favored Israel $x^2(1, n=30) = 19.200$, $p<0.001$. It is noteworthy that *Philstar* produced the largest amount of pro-Israel content among the researched publications, with 55.1 percent of its coverage (Table 2).

3.1.3. Newspapers from Other Religious Environments

The Nation of Thailand is the only newspaper in the research sample from a context dominated by neither Islam nor Christianity. Thailand, as noted above, is a Buddhist culture where almost 95 percent of the population practices Buddhism. The assumption of this study is that *The Nation* of Thailand is "non-aligned." Aggregate analysis of coverage tone (Table 1) shows that there was no significant difference between the amounts of slanted and neutral stories produced by *The Nation:* $x^2(1, n=35) = 0.714$, $p=.398$. However, a breakdown of the slanted stories (Table 2) reveals the paper's unfavorable stance on Israel as against its sympathy for Palestine, thus negating the study assumption on *The Nation.*

3.2. RQ2: Conflict Frames Most Prevalent

The coding of identified frames was analyzed to answer the second research question. The analysis focused on determining the distribution of frames in overall and individual newspaper coverage, and identifying the most prevalent frame at each level. This supplied an understanding of how the press characterized the conflict.

Overall, the conflict frames were most prevalent with 19 percent. This was closely followed by the peace frames with 17.7 percent. The human-interest component of the conflict was the third most salient with 12.5 percent (see Table 3).

Examining the individual newspapers, *Philstar* produced the highest figure for "conflict" frames (40.8 percent of its coverage), followed by *Star* of Malaysia (22.6 percent). *The Nation* of Thailand turned out the lowest proportion of "conflict" frames (2.9 percent), but with 22.9 percent was second to *Jakarta Post* (25.7 percent) in producing "peace" frames.

In the "human interest" frame, the highest proportion (17.4 percent) was found in *Star* of Malaysia. Virtually all the articles published by *Star* in the "human interest" category portrayed Palestinians as victims of Israel. This line was closely shared by *Jakarta Post*, where 12.2 percent its stories sympathized with Palestine in the "human interest" frame. *The Nation* of Thailand was the least likely to report the Israeli-Palestinian conflict from a "human interest" angle. Instead, the paper focused greater attention on the "consequences" of the conflict (37.1 percent of articles), and questioning its "morality" (17.1 percent). *Jakarta Post* took the lead in the "responsibility" frame, while *Philstar* was least likely to consider the "consequences" or question the "morality'" and "responsibility" issues (Table 3). As the results show, individual Southeast Asian newspapers held a range of different perspectives on the Israeli-Palestinian conflict. There was no strict religious dimension to the use of identifiable frames, but at aggregate level, the "conflict" frames was the most prevalent. There were also important levels of attention to the "peace" and "human interest" perspectives, with the latter skewed largely in favor of Palestine.

Table 3: Frames employed by individual newspapers

	Conflict *n* (%)	Peace *n* (%)	Human interest *n* (%)	Consequence *n* (%)	Morality *n* (%)	Responsibility *n* (%)	Others *n* (%)	Total *n* (%)
Star Malaysia	52 (22.6)	25 (10.9)	40 (17.4)	21 (9.1)	21 (9.1)	4 (1.8)	67 (29.1)	230 (100)
Jakarta Post Indonesia	29 (13.0)	57 (25.7)	27 (12.2)	23 (10.4)	15 (6.7)	37 (16.7)	34 (15.3)	222 (100)
Philstar Philippines	20 (40.8)	5 (10.2)	0 (0.0)	2 (4.1)	2 (4.1)	2 (4.1)	18 (36.7)	49 (100)
Nation Thailand	1 (2.9)	8 (22.9)	0 (0.0)	13 (37.1)	6 (17.1)	2 (5.7)	5 (14.3)	35 (100)
Total	102 (19.0)	95 (17.7)	67 (12.5)	59 (11.0)	44 (8.3)	45 (8.4)	124 (23.1)	536 (100)

Table 4: Sources

	News agencies n (%)	Other media n (%)	Government n (%)	Civic bodies n (%)	Independent n (%)	Other n (%)	Total
Star Malaysia	31 (13.5)	9 (3.9)	65 (28.3)	33 (14.3)	89 (38.7)	3 (1.3)	230 (100)
Jakarta Post Indonesia	129 (58.1)	0 (0.0)	9 (4.1)	41 (18.5)	42 (18.9)	1 (0.4)	222 (100)
Philstar Philippines	30 (61.2)	0 (0.0)	0 (0.0)	4 (8.2)	10 (20.4)	5 (10.2)	49 (100)
Nation Thailand	0 (0.0)	1 (2.9)	1 (2.9)	21 (60)	12 (34.2)	0 (0.0)	35 (100)
Total	190 (35.4)	10 (1.9)	75 (14.0)	99 (18.5)	153 (28.5)	9 (1.7)	536 (100)

3.3. RQ3: Foreign News Agencies are Dominant Sources

To answer the third research question, the articles were coded for five common news sources: news agencies, other media, government, civic bodies, and independent. Stories that did not fit into any of these categories were coded "Others." News agencies were the dominant source for Southeast Asian press reporting of the Israeli-Palestinian conflict. Overall, 35.4 percent of stories were obtained from the Associated Press (United States), Reuters (United Kingdom) and Xinhua (China). Although independent sources followed closely, with 28.5 percent, and civic bodies with 18.5 percent, it is apparent that foreign news agencies were the most common sources used by the Southeast Asian Press for reporting the Israeli-Palestinian conflict within the period investigated.

The sources used by individual newspapers offer an insight that is not noticeable at the aggregate level. *Star* of Malaysia, which produced the largest volume of stories on the Israeli-Palestinian conflict, acquired them principally from independent sources (38.7 percent), government sources (28.3 percent), and civic bodies (14.3 percent). The *Jakarta Post,* on the other hand, relied mostly on news agencies (58.1 percent, principally Associated Press), civic bodies (18.5 percent), and independent sources (18.9 percent). *Philstar* also relied mostly on news agencies (61.2 percent) (principally Associated Press and Xinhua) and independent sources (20.4 percent). *The Nation* of Thailand sourced most of its stories on the conflict from civic bodies (60.0 percent) and independent sources (34.2 percent). Table 4 summarizes the general distribution of news sources.

3.4. RQ4: Coverage Tones Significantly Correlated with News Sources

At the aggregate level, a large majority (87.6 percent) of neutral stories were sourced from foreign news agencies (AP, Reuters, and Xinhua) while half the slanted stories (48 percent) were obtained from independent sources. Government sources (18.2 percent) and civic bodies (31 percent) also played an important role in sourcing slanted stories.

3.4.1. Tone-News Source Relationship: Star of Malaysia

The main sources of neutral stories produced by *Star* of Malaysia were news agencies (mainly Reuters) (30.1 percent) and government sources (61.2 percent), while most of the slanted stories came from civic bodies (26 percent) and independent sources (70.1 percent). Evidence of significant positive correlation was found between the tone adopted by *Star* of Malaysia towards Israel and Palestine and the sources from which it reported the conflict $r(228) = .859$, $p<.001$. This means that the sources from which Star of Malaysia reported the Israeli-Palestinian conflict were likely explain the tone adopted by the paper within the period investigated.

3.4.2. Tone-Sources Relationship: *Jakarta Post* of Indonesia

All of the neutral stories published in the *Jakarta Post* were sourced from news agencies (mainly AP). News agencies also contributed the largest proportion (31.6 percent) of its slanted stories, just ahead of independent sources (30.9 percent) and civic bodies (30.1 percent) . Evidence of a positive correlation was found between the tone adopted by *Jakarta Post* towards Israel and Palestine and the sources from which it reported the conflict $r(220) = .653$, $p<.001$. This also means that sources were likely to explain the tone of reporting of the Israeli-Palestinian conflict in Jakarta Post within the period investigated.

3.4.3. Source-Tone Relationship: *Philstar*

Like the *Jakarta Post,* all the neutral stories that appeared in *Philstar* came from news agencies (mainly AP), which also constituted an important source of its slanted stories (47.2 percent). The second source of slanted stories in *Philstar* was the independent sources (27.8 percent). A strong positive correlation was found between the tone adopted by *Philstar* towards Israel and Palestine and the sources used by the paper $r(47) = .467$, $p=.001$. Again, this means that sources in *Philstar* were likely to have influenced the paper's tone of coverage of the Israeli-Palestinian conflict this period.

3.4.4. Tone-News Sources Relationship: *The Nation* of Thailand

Most (86.7 percent) of the neutral stories published in *The Nation* of Thailand were sourced from civic bodies, while its slanted stories came mainly from independent sources (60.0 percent) and civic bodies (40.0 percent). A significant positive correlation was found between the tone adopted by *The Nation* towards Israel and Palestine and the sources it used $r(33) = .611$, $p<.001$. This similarly suggests that sources could explain the tone of *The Nation's* coverage of the Israeli-Palestinian conflict in the post-Gaza period.

4. Discussion

The local press in Southeast Asia, like its counterparts in other parts of the world, is faced with many challenges in reporting the Israeli-Palestinian conflict, including political environment and ideological attachment. The investigated newspapers were found to be divided in their alignments. In predominantly Muslim environments, the *Star* of Malaysia and *Jakarta Post* of Indonesia were aligned with Palestine and significantly disfavored Israel in their coverage. *Philstar* of the Philippines, in a largely Christian environment, was aligned with Israel. *The Nation* of Thailand, which operates in a context that is neither Islam-dominated nor Christian-dominated, was sympathetic to Palestine in its reporting of the conflict. It appears then, that religion might offer a useful paradigm for explaining the attitudes of the Southeast Asian press towards the Israeli-Palestinian conflict.

In terms of frames, conflict-focused language ("attacks," "hostilities," "hostages," "clashes," "escalation of violence," "risks," etc.) dominated the pages of these newspapers. Consciously or unconsciously, content highlighting peace featured less prominently. *Philstar* for example, which produced the largest amount of conflict frames, was clearly sympathetic to Israel with favorable stories representing 55.1 percent of its coverage of the conflict. Its articles made frequent reference to Israel, and predominantly offered defenses for Israel's positions. For example, the paper once reported:

> A day after the Arab League (AL) Committee on the Middle East peace process recommended to Palestinian President Mahmoud Abbas to decide on when to start direct peace talks with Israel, Gaza militants fired a long-range Russian-made rocket from the Gaza Strip at southern Israel. The Israeli army immediately responded to Friday's attack, during which the rocket hit a populated area in the southern coastal Israeli city of Ashkelon, causing some damages, but no injuries were reported … Hamas armed wing al-Qassam Brigades vowed to revenge [sic] … (*Philstar,* July 31, 2010).

This story and many others like it, which *Philstar* sourced from Xinhua and the Associated Press is mirrored in the Star of Malaysia, which relied predominantly on Reuters and devoted 34.8 percent of its articles to criticizing Israel's position. Words such as "bully," "goliath," and "criminal" were frequently associated with Israel in the *Star,* which also presented a human-interest picture of the conflict through frequent portrayal of Palestine as Israel's victim. *Jakarta Post* of Indonesia took a similar anti-Israeli position, but with slightly more pro-Israel content than the *Star. The Nation's* sympathy for Palestine was conveyed with stories that focused mostly on cross-border consequences and moral obligations in the conflict. For example, an editorial titled "Time We Grasped Palestinian Issue" (sic) categorically stated:

> The issue of Palestine and Palestinian statehood will eventually hit Bangkok's front door and it's best to take up the debate now and prepare our country for it. Along the way, the government could be genuinely scoring political points with the Malay Muslims in the South, instead of insulting their intelligence by bringing foreign clerics who know nothing about the historical sentiment of the region and the mistrust … As a member of the UN Human Rights Council, Bangkok should at least feel it has a moral obligation to the people in Israel and Palestine." (*The Nation,* June 14, 2010).

The portrayal of Israel as the bully and Palestinians as the victims by the *Star* of Malaysia, the defense of Israel by *Philstar,* the sympathy of the Thai *Nation* towards Palestinians, and the *Jakarta Post's* anti-Israel frames are clear indications of alignment in the Southeast Asian press. Obviously, this is an important challenge to the local media playing a deescalating role in global conflict.

Bina (2007) observed that Malaysian media cooperate closely with the government to support its policy of maintaining unity between the Muslim world and Malaysia. The situation is similar in Indonesia, which has the world's largest Muslim population and a constitution that stresses "Pancasila" – the principle of one supreme God. However, Indonesia's open-door media policy, which is considered a step in the right direction, might explain *Jakarta Post's* extensive reliance on America's Associated Press, which perhaps led the paper to produce pro-Israel content that ranked second to the highest in amount. The policy might also see the press balancing its views and becoming neutral. *The Nation* of Thailand probably feels no obligation to favor Muslims or Jews or Christians, but in an attempt to take a dispassionate look at the issues, it found itself softening towards Palestine. This suggests how difficult it can be for the media to be impartial in reporting an asymmetric conflict. The Philippine media are known to favor US views and policies (Bina 2007), in light of the country's strong business relationship with Israel. If this is considered from the viewpoint of US support for Israel, it then may well explain the pro-Israel position of *Philstar.*

6. Conclusion

The cultural proximity of regional media to its audience offers a good reason to be optimistic that it can serve as a deescalating tool in global conflict, but a complex combination of glocal political engagements of local media actors and the helpless dependence of local media on foreign news agencies makes it difficult to maintain this hope. This study reveals that the Southeast Asian press shares the global sentiment on the Israeli-Palestinian conflict, and is consciously aligned in reporting the conflict. This is particularly the case with newspapers in Islam-dominated and Christian-dominated political environments. Looking at the relationships between coverage tones and news sources at aggregate level, Southeast Asian reporting of the Israeli-Palestinian conflict reflects the broader orientation of foreign news agencies towards the conflict. Firstly, the majority of local news about the conflict was sourced from foreign news agencies. Secondly, at the level of individual newspapers, an important proportion of slanted stories were also sourced from foreign news agencies. This renders local media coverage of global conflict vulnerable to the remote influence of foreign news agencies.

At the aggregate level, government is not a major source of news about the Israeli-Palestinian conflict for the Southeast Asian press. In Malaysia, however, the *Star* newspaper framed an important proportion of its neutral stories on the conflict around government sources, portraying the Malaysian government's concerned neutrality in the Israeli-Palestinian conflict. This framing in reality reflects the close political cooperation between government and the media in Malaysia (see Bina 2007).

As the results of this study also show, stories from civic bodies were cleverly framed as popular opinions coming from individuals, human rights, and interest groups. This and independent sources (editorials, columns, opinions, and analysis by staff journalists) perhaps most explicitly reflected the Southeast Asian press alignment in the Israeli-Palestinian conflict, as they were found at aggregate and individual newspaper levels to be the major sources of the slanted stories. Drawing on these results, it is plausible that the potential of local media to serve as deescalating tools in global conflicts is subject to the varying political contexts in which they operate in relation to specific conflicts.

References

Audit Bureau of Circulations Malaysia. 2011. About the Star and Sunday Star. http://abcm.org.my (accessed June 22, 2013).

Baranovsky, A. Nikolaevna. 2012. Framing Chechnya: an Analysis of News Coverage of Russia's Chechen Wars. M.A. Thesis: Harvard University.

Baumgartner, Jody C., Peter L. Francia, and Jonathan S. Morris. 2008. A Clash of Civilizations? The Influence of Religion on Public Opinion of U.S. Foreign Policy in the Middle East. *Political Research Quarterly* 61 (2): 171–79.

Bennett, W. Lance., and David L. Paletz, eds. 1994. *Taken by Storm: The Media, Public Opinion and US Foreign Policy in the Gulf War.* Chicago: University of Chicago Press.

Bina, L. Rebekah. 2007. Conflict, Terrorism and the Media in Asia. *Federal Communication Law Journal* 59 (2): 445–56.

Blondel, I. Ylva. 2004. International Media Get the Most Attention, But Don't Ignore the Local Media in Diffusing Conflicts. In *Media, Conflict Prevention and Reconstruction,* ed. James Barry, 27–29. Paris: UNESCO.

Bratic, V. 2009. Examining Peace Oriented Media in Areas of Violent Conflict. http://www.sfcg.org/Documents/DME/bratic_peace_oriented_media.pdf (accessed March 31, 2010).

Des Forges, Alison. 2007. Call to Genocide: Radio in Rwanda, 1994 In *The Media and the Rwanda Genocide,* ed. A. Thompson, A. London and Kampala: Pluto, Fountain Publishers, and IDRC.

Eklof, Stefan. 2004. *Power and Political Culture in Suharto's Indonesia: The Indonesian Democratic Party and the Decline of the New Order (1986–1998).* Copenhagen: Nordic Institute of Asian Studies (NIAS) Press.

El-Nawawy, Mohammed. 2002. *The Israeli-Egyptian Peace Process in the Reporting of Western Journalists.* Westport, CT: Ablex.

Galtung, Johan. 1998. Peace Journalism: What, Why, Who, How, When, Where. Paper presented at the TRANSCEND workshop, "What Are Journalists For?" Taplow Court, UK, September 3–6.

Glukhov, P. Andrey. 2009. *Mass Media as a Resource of Mobilization in Ethnic Conflicts in Modern Russia.* http://home.hiroshimau.ac.jp/heiwa/Pub/E20/glukhov.pdf (accessed June 22, 2013).

Goddard, Peter., Piers Robinson, and Katy Parry. 2008. Patriotism Meets Plurality: Reporting the 2003 Iraq War in the British Press. *Media, War, and Conflict* 1 (1): 9–30.

Gopin, Marc. 2002. *Holy War, Holy Peace: How Religion Can Bring Peace to the Middle East.* Oxford: Oxford University Press.

Hagos, Asgede. 2001. *Media Intervention in Peace Building in Burundi – The Studio Ijambo Experience and Impact.* Washington, DC: MSI-USAID.

Hallin, C. Daniel. 1986. *The Uncensored War.* Berkeley: University of California Press.

Hosen, Nadirsyah. 2005. Religion and the Indonesian Constitution: A Recent Debate. *Journal of Southeast Asian Studies* 36 (3): 419–40.

Hawkins, Virgil. 2009: National Interest or Business Interest: Coverage of Conflict in the Democratic Republic of Congo in *The Australian* Newspaper. *Media, War, and Conflict* 2 (1): 67–84.

Howard, Ross. 2003. Conflict Sensitive Journalism. Copenhagen: IMS and IMPACS.

Kinner, Kelly. 2005. Aljazeera.net and BBC.CO.UK: Media Framing of the Darfur Humanitarian Crisis. http://www.colorado.edu/journalism/globalmedia/Darfur-Student-Papers/DARFUR-FINAL-Kinner.pdf (accessed April 19, 2012).

Lee, S. Ting, and Crispin C. Maslog. 2005. War or Peace Journalism? Asian Newspaper Coverage of Conflicts. *Journal of Communication* 5 (2): 311–29.

McGoldrick, Annabel., and Jake Lynch. 2000. *Peace Journalism—How to Do It.* http://www.transcend.org/pjmanual.htm (accessed March 13, 2013).

McSweeney, Daniel. 2011. The CNN Effect in Somalia. http://www.e-ir.info/2011/08/11/the-cnn-effect-and-somalia/ (accessed April 19, 2012).

Miller, Jack. n.d. *Religion in the Philippines.* http://asiasociety.org/religion-philippines (accessed June 21, 2013).

Paluck, E. Levy. 2007. Reducing Intergroup Prejudice and Conflict with the Mass Media: A Field Experiment in Rwanda. Ph.D. diss., Yale University.

PhilStar. 2010. Gaza Militants, Israel Escalate Violence to Thwart Direct Peace Talks. 31 July. http://www.philstar.com/breaking-news/598343/gaza-militants-israel-escalate-violence-thwart-direct-peace-talks (accessed June 21, 2013).

Puddephatt, Andrew. 2006. *Voices of War: Conflict and the Role of the Media.* N. p.: International Media Support.

Ranstorp, Magnus. 1996. Terrorism in the Name of Religion. *Journal of International* Affairs 50 (1): 41–63.

Roy, Sera. 2004. Religious Nationalism and the Palestinian-Israeli Conflict: Examining Hamas and the Possibility of Reform. *Chicago Journal of International Law* 5 (1): 251–70.

Ruigrok, Nel. 2008. Journalism of Attachment and Objectivity: Dutch Journalists and the Bosnian War. *Media, War, and Conflict* 1 (3): 293–313.

Semetko, A. Holli., and M. Patti Valkenburg. 2000. Framing European Politics: A Content Analysis of Press and Television News. *Journal of Communication* 50 (2): 93–109.

Slater, Jerome. 2006. Polemics in the Israeli-Palestinian Conflict. *Human Rights and Human Welfare* 6:105–17

Stemler, Steve. 2001. An Overview of Content Analysis. *Practical Assessment, Research and Evaluation* 7 (17). http://PAREonline.net/getvn.asp?v=7&n=17 (accessed September 10, 2010).

Taylor, M. Phillip. 1992. *War and the Media: Propaganda and Persuasion in the Gulf War.* Manchester: Manchester University Press

Warshel, Yael. 2007. "As Though There Is Peace": Opinions of Jewish-Israeli Children about Watching Rechov Sumsum/Shara'a Simsim Amidst Armed Political Conflict. In *Children and Media in Times of War and Conflict,* ed. Dafna Lemish and Maya Götz, 309–32. Cresskill, NJ: Hampton.

Yakubu Ozohu-Suleiman
lastbornee@yahoo.co.uk

Sidin Ahmad Ishak
mdsidin@um.edu.my

International Journal of Conflict and Violence

urn:nbn:de:0070-ijcv-2014217
IJCV: Vol. 8 (2) 2014

Social Cohesion Activities and Attitude Change in Cyprus

Direnç Kanol, International Relations Department, Cyprus International University, Nicosia, Cyprus

Vol. 8 (2) 2014

ISSN: 1864–1385

Social Cohesion Activities and Attitude Change in Cyprus

Direnç Kanol, International Relations Department, Cyprus International University, Nicosia, Cyprus

Do social cohesion activities change the attitudes of the participants? This paper uses intergroup contact theory to explore attitude change resulting from contact with out-group(s) in social cohesion activities. Results from a pre-test/post-test design with fifty-five participants in two bicommunal camps in Cyprus show how attitudes change at the immediate end of these activities; an analysis of fourteen participants' comments after one, thirteen, and twenty-five months provides a medium- to long-term assessment of attitude change. Not all participants were completely positive towards the other community before they took part, as assumed by some. There is clearly space for impact in terms of attitude change. Social cohesion activities represent indispensable tools for reducing prejudice and improving relationships between former enemies in post-conflict countries.

Can social cohesion activities alleviate the negative socio-psychological effects of the deliberate negative representation of the "other" group(s) in divided societies? We know that one catalyst of many conflicts is the lack of contact between groups (Webster 2005; Vasilara and Piaton 2007; Hadjipavlou 2007). Some authors argue that social cohesion activities such as bicommunal camps have a substantial effect on participants' attitudes, significantly increasing trust and understanding (Ungerleider 2001, 2006; Hadjipavlou and Kanol 2008). Loizos, for instance, asserts that: "In Cyprus, the strongest case for bicommunal initiatives might be to claim that without them the antagonism between nationalists on both sides could have been more intense, drawing in more waverers, with the possibility of further military conflict and loss of life" (2006, 181).

On the other hand, social cohesion activities have been subjected to various criticisms. According to Broome, bicommunal gatherings do not go "deep" enough (Broome 2005). Counter-socialization forces that preach hatred and work to create an image of the adversary community as the "other" curb the possible effects (if any) of these short-term initiatives, which can be effective only for a very short period (Paffenholz 2010). Loizos takes note of the following danger: "Bicommunal contacts are somewhat removed from concrete livelihood contexts. Once you leave the workshop, you can, if you choose to, forget the whole thing, especially if its resolutions or lessons cannot be realistically activated in your place of work or your home" (2006, 188). Similarly, the Cyprus Center for European and International Affairs suggests that after participants leave these activities, they "go back to their 'normal' lives where they are confronted with prejudice, social pressure, and a lack of understanding regarding bi-communal activities" (2011, 10). Moreover, participants in this type of activities are more likely to have a prior positive attitude towards the other community, and the activities do not succeed in reaching out to the wider population (Paffenholz 2010; Cyprus Center for European and International Affairs 2011; Hadjipavlou and Kanol 2008). This is quite problematic as the aim of such activities is to convert negative public attitudes into positive attitudes. Furthermore, some scholars emphasize the point that more extreme people are more likely to resist change (Eagly and Chaiken 1988). So, it is debatable if an activity that manages to have a positive effect on the attitudes of a relatively positive section of the population can have the same effect on a completely negative section of the population.

However, although there is much criticism of the effectiveness of social cohesion activities, they still seem to be embraced by many conflict resolution activists and scholars. Therefore, the effectiveness of social cohesion activities remains unclarified. In order to fill this gap in the literature, this paper strives to answer the following question: Do social cohesion activities change the attitudes of the participants? Intergroup contact theory is used to interpret attitude change engendered by contact with the out-group(s) in social cohesion activities. Initially, the results from a pre-test/post-test design study with fifty-five participants in two bicommunal camps in Cyprus will illuminate how attitudes change at the immediate end of these activities. Secondly, an analysis of fourteen participants' own comments after one, thirteen, and twenty-five months provides a medium to long-term assessment of attitude change.

1. Theoretical Framework

Why would participants in social cohesion activities change their attitude towards the other community? Possible answers to this question can be found in the growing work on intergroup contact theory. According to Allport (1954), contact with other groups can reduce prejudice against them. Allport argues that there are four preconditions for a contact to have an impact. Other scholars working on the theory added another precondition, which is currently accepted by leading scholars (Pettigrew 1998). First of all, group members who come into contact must have almost equal status in that situation. Secondly, those who come into contact must wish and strive for a common goal. The third precondition is that the groups work together to achieve this common goal without any intergroup competition. Allport's final precondition (1954) was the existence of an authority at the top encouraging these favorable conditions.

Others sought to expand this list by suggesting the need for active participation (Maoz 2005), a common language, voluntary contact, and a prosperous economy (Wagner and Machleit 1986). Some suggested that the group's views before coming into contact should not be very negative (Ben-Ari and Amir 1986; Yogev, Ben-Yeshoshua and Alper 1991) and that stereotype disconfirmation is crucial (Cook 1978). However, these expansions are criticized by Pettigrew as facilitating, rather than essential conditions (Pettigrew 1998). One exception (the fifth precondition) is the condition of the possibility to become friends in the situation of contact, which implies a circumstance of close interaction (Pettigrew 1998; Pettigrew and Tropp 2006).

Of course, this theory would not be so robust if the processes that lead to prejudice reduction could not be explained so thoroughly. Pettigrew (1998) greatly advanced our understanding of the causal mechanisms by summarizing four processes where contact may show its effect. The first process takes place when learning about the individual(s) from the other group. During this process, stereotypes and negative attitudes are challenged as individuals get to know counterparts from the out-group better (Stephan and Stephan 1984). Based on this understanding, one can expect the prejudices that in-group members have about out-group members to erode after participating in social cohesion activities. Consequently, they start to humanize the adversary group and notice commonalities with the out-group rather than differences.

Hypothesis 1: Participants are more likely to see the commonalities with the out-group at the end of social cohesion activities.

The second possible process has behavior exogenous to attitude: The individual changes his/her behavior vis-à-vis an out-group member; if the process is repeated, positive attitude change results over time (Aronson and Patnoe 1997; Jackman and Crane 1986). The third process concentrates on the role of strong affective ties, intimacy, friendship and empathy in explaining attitude change. As the bonds between members of the two groups strengthen during contact, in-group members become more affectionate and empathetic towards the members of the out-group (Wright, McLaughlin-Volpe, and Roppe 1997; Pettigrew 1997a, 1997b; Pettigrew and Meertens 1995; Hamberger and Hewstone 1997; Batson et al. 1997; Turner et al. 2007; Hewstone et al. 2006; Davies et al. 2011). Based on these arguments, one may expect the participants of social cohesion activities to sympathize and empathize more with the out-group members as a result of the intimacy and friendships created.

Hypothesis 2 Participants are more likely to express mutual concern rather than selfish concern at the end of social cohesion activities.

The final process is intergroup appraisal, which like the first process relies on the learning process to explain attitude change. The difference is that intergroup appraisal emphasizes on the impact of contact on revising attitudes about the in-group as well as the out-group. The individual who comes into contact with the out-group learns new perspectives and takes less pride in the culture and values of the in-group. The individual accepts that the in-group's way may be neither the only way, nor the best (Pettigrew 1998; Pettigrew et al. 2011). Based on this understanding, one could expect the participants of social cohesion activities to be more likely to make self-criticism of their in-group after these activities.

Hypothesis 3: Participants are more likely to express in-group self-criticism after social cohesion activities.

As a consequence of all these changes, it is plausible to expect contact to reduce prejudice toward the out-group. There is solid empirical evidence for the robustness of intergroup contact theory based on meta-analyses, reviews, and recent data (including longitudinal studies) showing that attitudes change toward different types of groups (Pettigrew and Tropp 2011; Christ and Wagner 2013; Hewstone et al. 2014). Pettigrew and Tropp (2011) also refine the discussion about the mechanisms discussed above, showing that positive findings are not restricted to a specific country or culture. Findings from all over the world seem to be encouraging for the supporters of contact to alleviate prejudice and conflict.

However, an important limitation is that the theory has not been extensively tested in the context of intractable conflict (Wagner and Hewstone 2012). Wagner and Hewstone (2012) distinguish three phases in regard to intergroup contact theory in an environment of intractable conflict: previolence phase, physical violence phase, and postviolence phase. Similar to Hewstone and colleagues (2008), who tested the intergroup contact theory with a longitudinal study of Catholics and Protestants in Northern Ireland, Wagner and Hewstone (2012) find support for the theory in the context of protracted conflict. The next section of this paper describes the current research measuring the effect of bicommunal camps in Cyprus. This contribution to the civil society and peacebuilding literature also provides evidence for the effects of intergroup contact in a postviolence phase of an intractable conflict.

2. Method

A pre-test/post-test research design was used to test the aforementioned hypotheses, comparing the attitudes of the participants before and after two bicommunal camps in Cyprus. Crossroads II Bicommunal Theatre Camp and Friendship for Cyprus Summer Camp for Teenagers aimed to promote the peace process by bringing together young people (aged between 15 and 18) from the Greek Cypriot and Turkish Cypriot communities. Previous research does not show any difference of effect between age groups (Pettigrew et al. 2011), which implies that this research might be generalizable to other age cohorts. Crossroads II Bicommunal Theatre Camp, which took place from July 15 to 24, 2011, attempted to accomplish this goal by creating a shared living place for the participants where they could learn and practice drama skills under the supervision of instructors with theatre experience. Cyprus Friendship Program Summer Camp for Teenagers was very similar, except without the focus on theatre skills. It took place from July 22 to 29, 2013. Participants in both camps were given the same opportunities and treated equally throughout. In Crossroads II Bicommunal Theatre Camp, the participants all aimed to learn theatre skills and create a play at the end of the camp working in mixed groups without any intergroup competition. In the Cyprus Friendship Program Summer Camp, the teenagers participated in various sports, fun, and educational activities. The facilitators acted as "soft" authority throughout the camps and the participants were able to develop close ties as a result of the intimate and intense period they shared.

The organizers of the two camps agreed to assist this study by administering a survey to the participants at both the beginning and the end of the respective camps. For the Crossroads II Bicommunal Theatre Camp, the first data was collected as soon as the participants arrived. The post-camp data was collected on the last day of the camp, when the participants were getting ready to leave. For the Cyprus Friendship Program Summer Camp for Teenagers, the author personally collected the data before the camp, during a meeting where the participants were given information about the logistics. The data after the camp was collected by the organizers on the last day of the camp. The short questionnaire aiming to capture attitude change was given to 55 participants, of whom 29 were Greek Cypriots and 26 Turkish Cypriots. The sample includes all of the 14 participants of the Crossroads II Bicommunal Theatre Camp, with a 100 percent return rate. The sample includes 41 participants from the Friendship for Cyprus Summer Camp for Teenagers, which had 44 participants. Here, three questionnaires were not returned.

The questionnaire used three questions to explore the attitudes of the Turkish Cypriot and Greek Cypriot participants towards each other. Perception of commonalities with the out-group (hypothesis 1) was measured by asking the participants to choose a response to the statement: "we have so much in common with the Turkish/Greek Cypriots" with the following possible answers: "strongly disagree" (coded as 0), "somewhat disagree" (coded as 1), "neither agree nor disagree" (coded as 2), "somewhat agree" (coded as 3), or "strongly agree" (coded as 4). Mutual concern (hypothesis 2) was measured by the statement: "The Cyprus problem must be solved on the basis of a mutually acceptable compromise". Participants were asked to respond using the same scale from "strongly disagree" (coded as 0), to "strongly agree" (coded as 4). Openness to self-criticism (hypothesis 3) was measured by asking the participants to comment on the statement: "I recognize that both communities have made mistakes in the past," again using the same five-point scale. Cronbach's alpha of 0.71 for the data taken before the camps began shows that the items have an acceptable level of internal consistency. The analysis compared the change in responses before and after the camp using a t-test.

Kelman argued that measurement of the effect of interactive problem-solving workshops should be conducted not only before and immediately after the workshop but also after a considerable period (2008, 47). To the author's knowledge, there is only one study (Malhotra and Liyanage 2005) that specifically measured the long-term effect of peace workshops with an experimental design. The present study combines the short-term analysis with a semi-structured questionnaire distributed on September 1, 2012, which was approximately thirteen months after Crossroads II. In order to increase the number of observations, the study also included the participants of Crossroads I Bicommunal Theatre Camp, which took place approximately twenty-five months before data collection, as well as Crossroads III which took place about one month before data collection. The sample included two participants from Crossroads I, five participants from Crossroads II and seven participants from Crossroads III. Qualitative analysis and quotations related to the hypotheses are included in the results section. The participants were directly asked if and why participating in the camp made them realize that they have more things in common with the other community (hypothesis 1), participation in the camp made them empathize more with the other community (hypothesis 2) and participating in the camp changed their views to make them more open to criticizing their own community (hypothesis 3). The author personally conducted this survey during a reunion of Crossroads Bicommunal Theatre Camp participants. This provided some findings on the question of whether the effect of social cohesion activities is long-lasting or not. The qualitative and quantitative questionnaires can be found in the appendix.

3. Results

The descriptive statistics (see Table 1) show that the differences between the means before and after the camps are significant in the expected direction. On the 0 to 4 scales, the mean before the camps is 3.40 with respect to the first hypothesis, 3.60 with respect to the second hypothesis, and 3.46 with respect to the third hypothesis. The respective figures after the camps are 3.87, 3.87 and 3.91. Examining the paired t-test results for the three hypotheses (Table 2), the means before and after the camps are significantly different at the 99 percent confidence level with respect to the first, second, and third hypotheses ($t= -4.8942$, $t= -2.6734$, and $t= -4.0379$ respectively). Therefore, the results obtained from the pre-test/post-test study provide empirical evidence for all three hypotheses. For visualizations of the difference of means before and after the camps for hypotheses 1, 2, and 3, see figures 1, 2 and 3 respectively.

Table 1: Descriptive statistics

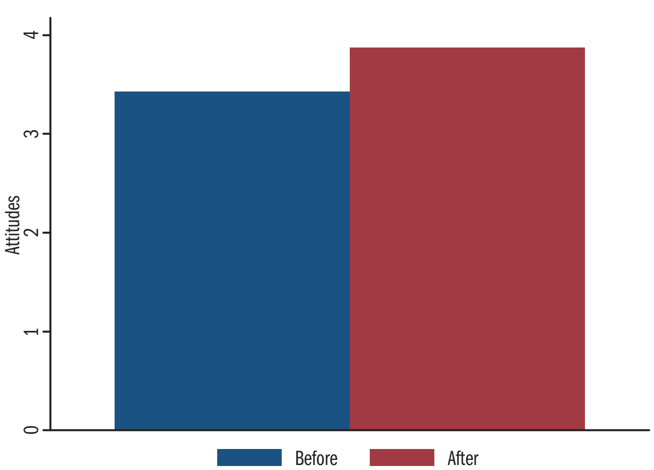
	N	*Mean*	*SD*	*Min*	*Max*
Commonality before	55	3.40	0.68	2	4
Commonality after	55	3.87	0.51	1	4
Mutual compromise before	55	3.60	0.60	2	4
Mutual compromise after	55	3.87	0.47	1	4
Past mistakes before	55	3.46	0.77	0	4
Past mistakes after	55	3.91	0.44	1	4

Table 2: Paired t-tests

	t-value	*p-value*
Commonalities (hypothesis 1)	–4.8942	0.01***
Mutual concern (hypothesis 2)	–2.6734	0.01***
Past mistakes (hypothesis 3)	–4.0379	0.01***

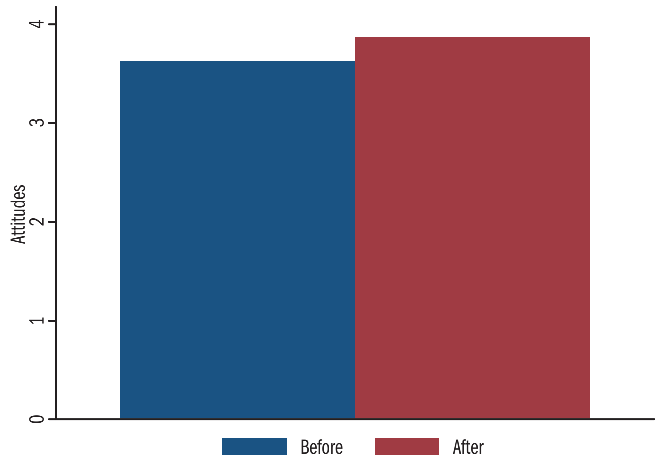
Note: All p-values measured as two-tailed.
*** significant at p<0.01.

Figure 1: Commonality with out-group before and after camp



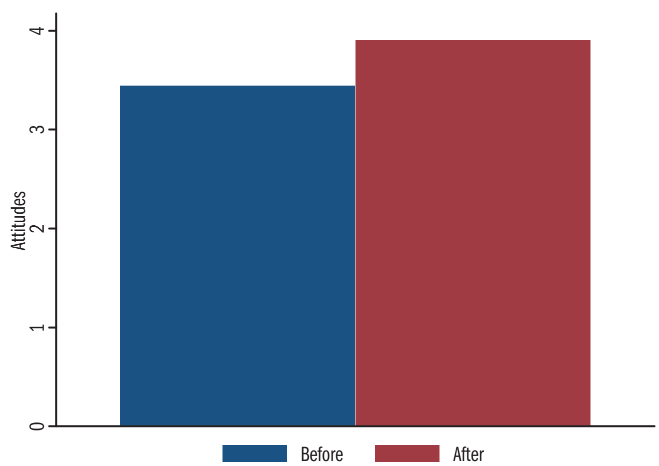
251658240

Figure 2: Mutual acceptability of a solution before and after camp



251658240

Figure 3: Acceptance of in-group's past mistakes before and after camp



251658240

Moving on to the survey of previous camp participants, thirteen out of fourteen reported a significant attitude change in the medium and long term after participating in one of the Crossroads camps, based on hand-coding of what they wrote on the semi-structured questionnaires. Where an answer was not clearly quantifiable, the respondent was consulted face-to-face for clarification. This quantitative finding suggests that the camp not only had an immediate effect, but also a long-term one. Almost all participants self-reported significant positive attitude change. The participants' comments suggest that the causal mechanisms are compatible with the intergroup contact theory. As the following quotes illustrate, participants are more likely to see commonalities with the out-group after having participated in social cohesion activities (hypothesis 1):

> "Through living together and talking about our everyday activities and interests, we came to the conclusion that we are more alike than different."

> "I realized that we have so much in common because we lived together for some time and this gave me a chance to get to know them and their way of life better."

The following two quotes suggest that the participants are more likely to have mutual concern rather than selfish concern after having participated in social cohesion activities (hypothesis 2):

> "The camp made me empathize with them more. They lost houses and relatives like us. I empathize because we have a lot in common."

> "I empathize more with the other community now because we all lost some important things and we all feel the same."

And the following two quotes suggest that the participants are more likely to express self-criticism of their in-group after having participated in social cohesion activities (hypothesis 3):

> "After living together in the camp, barriers seem to disappear and now I am more open to criticizing my own community."

> "Now I am more open to criticizing my own community because now I know that we are all the same and we are all in this thing together."

Where participants reported no significant long-term change in their attitudes, their argument was not that the effect of the camp faded with time but that they were already completely positive towards the other community:

> "I guess the camp didn't make me realize any commonalities I have with them that I didn't know. It just reminded me of the similarities that I have forgotten during the time I haven't seen many Greek Cypriots."

> "By participating in the camp I didn't start criticizing my own community more. I always criticize my own community."

4. Conclusion

Peacebuilding needs activities to overcome the negative socio-psychological effects caused by forces that may include education, media, and negative rhetoric of politicians or family members. Social cohesion activities aim to achieve just this but there are question marks over their effectiveness. The results of the study reported here confirm the intergroup contact theory suggesting that social cohesion activities can indeed be effective. Relying on this theory and using a pre-test/post-test study, this paper showed that the fifty-five participants in the Crossroads Bicommunal Theatre Camp II and Cyprus Friendship Program Summer Camp for Teenagers saw the commonalities with the out-group more, had more mutual concern compared to selfish in-group concern, and were more open to self-criticism of their in-group after the camps. Furthermore, statements made by fourteen participants in the Crossroads I, Crossroads II, and Crossroads III camps provided some evidence for long-term attitude change and further substantiated the finding that social cohesion activities can be effective. Thirteen of these fourteen participants self-reported significant positive change.

Researchers working on the endogeneity problem in regard to the question of tolerant people seeking contact or contact decreasing prejudice found important evidence for simultaneous causation working both ways (Binder et al. 2009; Sidanius et al. 2008) and in fact a stronger effect when contact is the independent variable (Pettigrew 1997a; Pettigrew and Tropp 2006; Powers and Ellison 1995; Wilson 1996; Van Dick et al. 2004). The sample reported here shows that not all participants in these camps were completely positive towards the other community before they took part, as is assumed by some. So, there was clearly space for impact in terms of attitude change. Social cohesion activities might be quite effective tools in achieving positive attitude change in post-conflict societies such as Cyprus.

One important shortcoming of this study was the lack of control groups. At the time of the camps, I followed the news during the periods the camps took place. There were no significant developments with respect to the track I level negotiations at the time of the camps. This is encouraging for the validity of the results presented. Nevertheless, like any pre-test/post-test study, the validity of the findings is much less robust when control groups are not present. Therefore, the findings in this paper should be cross-checked. Future research may use different measurement techniques with the presence of control groups in order to put the argument to a more stringent test.

Appendix

Quantitative items

1. We have much in common with the Turkish/Greek Cypriots.
2. The Cyprus problem must be solved on the basis of a mutually acceptable compromise.
3. I recognize that both communities have made mistakes in the past.

Response scale

Strongly agree (4)
Somewhat agree (3)
Neither agree nor disagree (2)
Somewhat disagree (1)
Strongly disagree (0)

Qualitative items

1. Has participating in the bicommunal theatre camp made you think that you have more in common with the other community than you thought before attending the camp? Why?
2. Has participating in the bicommunal theatre camp made you empathize more with the other community? Why?
3. Has participating in the bicommunal theatre camp changed your views to become more open to criticizing your own community more? Why?

References

Allport, Gordon W. 1954. *The Nature of Prejudice.* Reading, MA: Addison-Wesley.
Aronson, Elliot, and Shelley Patnoe. 1997. *The Jigsaw Classroom: Building Cooperation in the Classroom.* New York: Longman.
Batson, Daniel C., Marina P. Polycarpou, Eddie Harmon-Jones, Heidi J. Imhoff, Erin C. Mitchener, Lori L. Bednar, Tricia R. Klein, and Lori Highberger. 1997. Empathy and Attitudes: Can Feeling for a Member of a Stigmatized Group Improve Feelings toward the Group? *Journal of Personality and Social Psychology* 72 (1):105–18.
Ben-Ari, Rachel, and Yehuda Amir. 1986. Contact between Arab and Jewish Youth in Israel: Reality and Potential. In *Contact and Conflict in Inter-group Encounters,* ed. Miles Hewstone and Rupert Brown, 45–58. New York: Basil Blackwell.
Binder, Jens, Hanna Zagefka, Rupert Brown, Friedrich Funke, Thomas Kessler, Amelie Mummendey, Annemie Maquil, Stephanie Demoulin, and Jacques-Philippe Leyens. 2009. Does Contact Reduce Prejudice or Does Prejudice Reduce Contact? A Longitudinal Test of the Contact Hypothesis amongst Majority and Minority Groups in Three European Countries. *Personality and Social Psychology Bulletin* 96 (4):843–56.
Christ, Oliver, and Ulrich Wagner. 2013. Methodological Issues in the Study of Intergroup Contact. In *Advances in Intergroup Contact,* ed. Gordon Hodson and Miles Hewstone, 233–61. London: Psychology Press.
Cook, Stuart W. 1978. Interpersonal and Attitudinal Outcomes in Cooperating Interracial Groups. *Journal of Research and Development in Education* 12 (1): 97–113.
Cyprus Centre for European and International Affairs. 2011. *Enhancing Civil Society's Role in Conflict Prevention and Peace-building in Cyprus.* Nicosia: University of Nicosia.
Davies, Kristin, Linda R. Tropp, Arthur Aron, Thomas F. Pettigrew, and Stephen C. Wright. 2011. Cross-Group Friendships and Intergroup Attitudes: A Meta-analytic Review. *Personality and Social Psychology Review* 15 (4):332–51.
Eagly, Alice H., and Shelly Chaiken. 1998. Attitude Structure and Function. In *The Handbook of Social Psychology,* vol. 1, ed. Daniel Gilbert, Susan, T. Fiske, and Gardner Lindzey, 269–322. New York: McGraw-Hill.
Hadjipavlou, Maria. 2007. The Cyprus Conflict: Root Causes and Implications for Peacebuilding. *Journal of Peace Research* 44 (3):349–65.
Hadjipavlou, Maria, and Bülent Kanol 2008. *Cumulative Impact Case Study: The Impacts of Peacebuilding Work on the Cyprus Conflict.* N. p.: CDA Collaborative Learning Projects.
Hamberger, Jürgen, and Miles Hewstone. 1997. Inter-Ethnic Contact as a Predictor of Prejudice: Tests of a Model in Four West European Nations. *British Journal of Social Psychology* 36 (2):173–90.
Hewstone, Miles, Ed Cairns, Alberto Voci, Juergen Hamberger, and Ulrieke Niens. 2006. Intergroup Contact, Forgiveness, and Experience of "The Troubles" in Northern Ireland. *Journal of Social Issues* 62 (1):99–120.
Hewstone, Miles, Simon Lolliot, Elissa Myers, Ananthi Al Ramiah, Hermann Swart, Alberto Voci, and Ed Cairns. 2014. Intergroup Contact and Intergroup Conflict. *Peace and Conflict: Journal of Peace Psychology* 20 (1): 39–53.
Hewstone, Miles, Nicole Tausch, Joanne Hughes, and Ed Cairns. 2008. *Can Contact Promote Better Relations? Evidence from Mixed and Segregated Areas of Belfast,* Report to the Community Relations Unit, Northern Ireland. http://www.ofmdfmni.gov.uk/can_contact_promote_better_relations_evidence_from_mixed_and_segregated_areas_of_belfast_-_full_report.pdf
Jackman, Mary R., and Marie Crane. 1986. "Some of My Best Friends are Black…": Interracial Friendship and Whites' Racial Attitudes. *Public Opinion Quarterly* 50 (4):459–86.
Kelman, Herbert C. 2008. Evaluating the Contributions of Interactive Problem Solving to the Resolution of Ethnonational Conflicts. *Peace and Conflict: Journal of Peace Psychology* 14 (1):29–60.
Loizos, Peter. 2007. Bicommunal Initiatives and Their Contribution to Improved Relations between Turkish and Greek Cypriots. *South European Society and Politics* 11 (1):179–94.

Malhotra, Deepak, and Sumanasiri Liyanage. 2005. Long-term Effects of Peace Workshops in Protracted Conflicts. *Journal of Conflict Resolution* 49 (6):908–24.
Maoz, Ifat. 2005. Evaluating the Quality of Communication between Groups in Dispute: Equality in Contact Interventions between Jews and Arabs in Israel. *Negotiation Journal* 21 (1):131–46.
Paffenholz, Thania. 2010. What Civil Society can Contribute to Peacebuilding. In *Civil Society and Peacebuilding: A Critical Assessment,* ed. Thania Paffenholz, 381-404. Boulder: Lynne Rienner.
Pettigrew, Thomas F. 1997a. Generalized Intergroup Contact Effects on Prejudice. *Personality and Social Psychology* 23 (2):173–85.
Pettigrew, Thomas F. 1997b. The Affective Component of Prejudice: Empirical Support for the New View. In *Racial Attitudes in the 1990s: Continuity and Change,* ed. Steven A. Tuch and Jack K. Martin, 76–90. Westport, CT: Praeger.
Pettigrew, Thomas F. 1998. Intergroup Contact Theory. *Annual Review of Psychology* 49:65–85.
Pettigrew, Thomas F., and Roel W. Meertens. 1995. Subtle and Blatant Prejudice in Western Europe. *European Journal of Social Psychology* 25 (1):57-75.
Pettigrew, Thomas F., and Linda R. Tropp. 2006. A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology* 90 (5):751–83.
Pettigrew, Thomas F., and Linda R. Tropp. 2011. *When Groups Meet: The Dynamics of Intergroup Contact.* Philadelphia, PA: Psychology Press.
Pettigrew, Thomas F., Linda R. Tropp, Ulrich Wagner, and Oliver Christ. 2011. Recent Advances in Intergroup Contact Theory. *International Journal of Intercultural Relations* 35 (3):271–80.
Powers, Daniel A., and Christopher G. Ellison. 1995. Interracial Contact and Black Racial Attitudes: The Contact Hypothesis and Selectivity Bias. *Social Forces* 74 (1):205–26.
Sidanius, Jim, Shana Levin, Colette Van Laar, and David O. Sears. 2008. *The Diversity Challenge: Social Identity and Intergroup Relations on the College Campuses.* New York: Russell Sage Foundation.
Stephan, Walter G., and Cookie W. Stephan. 1984. The Role of Ignorance in Intergroup Relations. In *Groups in Contact: The Psychology of Desegregation,* ed. Miller Brewer and Marilyn B. Brewer, 229–56. Orlando, FL: Academic.
Turner, Rhiannon N., Miles Hewstone, Alberto Voci, Stefania Paolini, and Oliver Christ. 2007. Reducing Prejudice via Direct and Extended Cross-group Friendship. *European Review of Social Psychology* 18 (1):212–55.
Ungerleider, John. 2001. Bicommunal Youth Camps for Peace in Cyprus. *Peace Review* 13 (4):583–89.
Ungerleider, John. 2006. The Impact of Youth Peacebuilding Camps: Connectedness, Coping, and Collaboration. *Cyprus Review* 18 (1):137–47.
Van Dick, Rolf, Ulrich Wagner, Thomas F. Pettigrew, Oliver Christ, Carina Wolf, Thomas Petzel, Vanessa S. Castro, and James S. Jackson. 2004. The Role of Perceived Importance in Intergroup Contact. *Journal of Personality and Social Psychology* 87 (2):211–27.
Vasilara, Marina, and Gotelenne Piaton. 2007. The Role of Civil Society in Cyprus. *Cyprus Review* 19 (2):107–21.
Wagner, Ulrich, and Miles Hewstone. 2012. Intergroup Contact. In *The Oxford Handbook of Intergroup Conflict,* ed. Linda Tropp, 193–209. Oxford: Oxford University Press.
Wagner, Ulrich, and Uwe Machleit. 1986. "Gastarbeiter" in the Federal Republic of Germany: Contact between Germans and Migrant Populations. In *Contact and Conflict in Inter-group Encounters,* ed. Miles Hewstone and Rupert Brown, 59–78. New York: Basil Blackwell.
Webster, Craig. 2005. Optimism and Pessimism regarding Interacting with Turkish Cypriots: Greek Cypriot Public Opinion and Its Implications for Reconciliation. Paper given at international symposium "Cyprus Management in the Non-profit Sector: The Role of Civil Society Organizations in Achieving Participatory Democracy and Reconciliation," February 25–26, 2005, Nicosia, Cyprus.
Wilson, Thomas C. 1996. Prejudice Reduction or Self-selection? A Test of the Contact Hypothesis. *Sociological Spectrum: Mid-South Sociological Association* 16 (1):43–60.
Wright, Stephen C., Arthur Aron, Tracy McLaughlin-Volpe, and Stacy A. Ropp. 1997. The Extended Contact Effect: Knowledge of Cross-group Friendships and Prejudice. *Journal of Personality and Social Psychology* 73 (1):73–90.
Yogev, Abraham, Naama S. Ben-Yeshoshua, and Yael Alper. 1991. Determinants of Readiness for Contact with *Jewish Children among Young Arab Students in Israel.* Journal of Conflict Resolution 35 (3):547–62.

Direnç Kanol
direda@yahoo.co.uk

International Journal
of Conflict and Violence

urn:nbn:de:0070-ijcv-2014263
IJCV: Vol. 8 (2) 2014

Teen Dating Violence in French-speaking Switzerland: Attitudes and Experiences

Jacqueline De Puy, Institute of Legal Medicine, University of Lausanne, Switzerland
Sherry Hamby, Department of Psychology, The University of the South, Sewanee, United States
Caroline Lindemuth, Department of Psychology, The University of the South, Sewanee, United States

Vol. 8 (2) 2014

ISSN: 1864–1385

Teen Dating Violence in French-speaking Switzerland: Attitudes and Experiences

Jacqueline De Puy, Institute of Legal Medicine, University of Lausanne, Switzerland
Sherry Hamby, Department of Psychology, The University of the South, Sewanee, United States
Caroline Lindemuth, Department of Psychology, The University of the South, Sewanee, United States

Research on dating violence has tended to focus on North American college students. This study innovates with data collected in Switzerland from a sample of 132 school pupils and vocational education students aged 14 to 22 using a self-administered questionnaire. The study investigates relationships between attitudes and experiences about dating violence and the effect of gender. Biases against women were common in the sample. Females reported less endorsement of patriarchal attitudes about women's roles, but both genders reported similar levels of disparagement of women. Participants reported high rates of physical violence perpetration (41.9 percent) and victimization (48.8 percent). Pro-violence attitudes were related to psychological and physical perpetration as well as physical victimization. For female respondents, essentialist beliefs about women's innate abilities appear more persistent than beliefs about appropriate roles. Male participants endorsed both types of gender stereotypes at high rates. Male-perpetrated violence was perceived less favorably than female-perpetrated violence. Our data suggest that general attitudes toward violence are the most consistent predictor of physical and psychological aggression within dating relationships. More attention needs to be paid to subtypes among attitudes on women and violence, which past research assumed were monolithic. This study shows the need for prevention programs to address pro-violence attitudes.

Teen dating violence is increasingly recognized as a serious problem affecting many adolescents (Black et al. 2011; Hamby, Finkelhor, and Turner 2012) , but has received less attention outside of North America, with a particular lack of information on younger teens and the non-college population. Evidence is needed to document the extent of this public health problem in societies outside of North America. Attitudes and experiences relating to teen dating violence may differ in other sociocultural contexts. There is a need for more scientific evidence on this problem in countries outside North America, in order to contribute to awareness and shape prevention efforts. Attitudes, in the form of gender role stereotypes and concerning the acceptability of violence in relationships, are among the most commonly studied risk factors for teen dating violence (Foshee et al. 2000; Foshee et al. 1998; Simon et al. 2010). Few studies, however, have examined variations in types of attitudes and types of violence. Instead, negative attitudes toward women and favorable attitudes toward violence are usually assumed to be unitary constructs.

This study presents and analyzes the first data on dating violence attitudes and experiences among adolescents in French-speaking Switzerland. Existing data indicate that dating violence is common in Switzerland, as it is in most parts of the world (Chan et al. 2008), but there has been little research among younger adolescents and non-college students. In this respect, it is especially relevant in Switzerland to include students in vocational education and training, where about two-thirds of young people start this type of education in their early teens after basic schooling. Early adolescence is a prime risk period for the onset of teen dating violence and it is important, especially for prevention efforts, to know patterns of teen dating violence and identify risk factors that can be targeted by prevention and intervention programs. The data from our study have several unique characteristics. In addition to being, as far as we are aware, the first study of teen dating violence in Switzerland, it is also one of the first European studies to focus on vocational students (in contrast to university students). In addition to physical aggression these data also examine psychological aggression, and include three categories of attitude: toward gender role egalitarianism, disparagement of women and teen dating violence.

1. Background

1.1. Intimate Partner Violence and Dating Violence in Switzerland

A recent study of college students in Switzerland reported high prevalence rates of dating violence. Over 28 percent of males and 23 percent of females reported having perpetrated assaults, while a smaller percentage, 25.0 percent of males and 16.6 percent of females, reported having been a victim of violence (Chan et al. 2008). In that study, however, the mean age of the participants was relatively high, at 34.3 years of age, and it thus represents an even older sample than typically seen in U.S. college student surveys. The Swiss Optimus study found that among teenagers, sexual victimization was often perpetrated by dating partners or ex-partners, with 42 percent of victims reporting at least one incident of sexual contact victimization perpetrated by their partner or date (Averdijk, Müller-Johnson, and Eisner 2011). In a nationally representative study of adult Swiss women, those aged 18–24 were at the greatest risk (26 percent) of being victims of violence (Killias, Simonin, and De Puy 2005). Given the higher rates of violence among the young adult population, and recent data on the extent of sexual violence among teenagers, we anticipate that physical and psychological forms of dating violence are also prevalent among Swiss adolescents.

1.2. Attitudes Associated with Dating Violence

Favorable attitudes about violence have long been thought to be important antecedents to violent acts (DeWall, Anderson, and Bushman 2011) and have long been a primary focus of research on teen dating violence (Foshee et al. 1998). Two types of attitudes have received particular attention in research on teen dating violence: the extent to which youth endorse gender stereotypes and the extent to which they endorse dating violence under particular circumstances (Foshee et al. 1998; Simon et al. 2010). Despite the decades-long social movement promoting egalitarian attitudes about men and women, gender hostility (men having negative attitudes about women and vice versa) remains common. Nearly 50 percent of men and 60 percent of women in a multi-national college sample showed some degree of gender hostility, with 5 percent of men and 7 percent of women exhibiting extreme hostility (Dutton, Straus, and Medeiros 2006). Although trending down historically, favorable attitudes towards at least certain forms of violence remain common. It has also been noted that floor effects are common in certain attitude measures. For example, one research team evaluating a sexual assault prevention program found that most college students rejected most rape myths at pretest (Klaw et al. 2012). Yet they also noted that a couple of items on their rape myth scale were still endorsed at distressingly high rates. Even after the prevention program, 50 percent of male participants endorsed the statement that: “Men don’t usually intend to force sex on a woman, but sometimes they get too sexually carried away.” They note that a more specific approach to problematic attitudes may be warranted. Similarly, Simon and colleagues (2010) found that girls in general, as well as youth who had dated, were less accepting of male-perpetrated than female-perpetrated physical violence. Although most research still unfortunately sums all attitudes into a single score, these data suggest that attention to variation across attitudes is warranted and ought to be investigated in a teenage population. For instance, it is often assumed that younger generations will have more egalitarian attitudes about gender than their parents, because of progress in the status of women at the level of society. In Switzerland, equality between men and women was included in the Federal Constitution in 1995 and there have been particular efforts to promote egalitarian norms in the public education system. Similarly, the first campaign on intimate partner violence against women in Switzerland was launched in 1997 following the publication of the first scientific study on the problem (Gillioz, De Puy, and Ducret 1997). Since that ground-breaking study, multiple Swiss institutions and organizations have developed prevention programs.

1.3. Sociodemographic Characteristics and Teen Dating Violence

In U.S. samples, a number of sociodemographic characteristics have been found to increase the risk of teen dating violence and other forms of youth victimization. Teen dating violence increases with age as youth move through adolescence (Turner et al. 2013). Children who live in single-parent or other nontraditional households are also at elevated risk for most forms of youth victimization (Turner et al. 2013). Peer networks are increasingly recognized as an important risk factor (Swartout 2013), but less is known about how exposure to peer victims is associated with teen dating victimization and perpetration. Few of these factors have received extensive study outside the United States.

1.4. Purpose of the Study

Our first purpose was to assess patterns of gender stereotypes, favorable attitudes towards violence, and rates of teen dating violence perpetration and victimization in a sample of Swiss adolescents. This is not a randomly selected sample, and is not statistically representative of the whole teenage Swiss population: we were interested in identifying patterns that could indicate potential targets of future research and intervention. We also explored variations in attitudes about gender and violence and examined how these intersect with gender. Finally, we examined how attitudes, sociodemographic indicators, and relational characteristics are associated with teen dating violence perpetration and victimization in this Swiss sample. We expected higher levels of patriarchal and pro-violent attitudes to be associated with higher levels of involvement in teen dating violence.

2. Methods

2.1. Participants

The study surveyed 132 teenagers, 42 percent of whom were girls and 58 percent boys. The study was conducted in French-speaking Swiss towns, in several youth centers and one vocational education program. Participants ranged from 14 to 22 years in age with a mean of 17.75 years (SD 1.63). One- quarter (25.0 percent) of the sample were 14 to 16 years old, 19.5 percent were 17, 27.3 percent were 18, 15.6 percent were 19, and 12 percent were 20 to 22 years old. A majority of the participants (75.0 percent) reported having been in at least one dating relationship lasting a month or longer. A majority (61 percent) described themselves as Swiss citizens, although most of the non-Swiss teens reported having lived in Switzerland for a long period of time (mean=10.9 years, SD=5.4). When asked with whom they lived, 87 percent of participants reported living with their mother, while only 60 percent lived with their father (53 percent lived with both). Families were relatively small, with participants living with a mean of 1.4 siblings (SD=1.2). The sample was drawn predominantly from working class families. When asked to indicate their parents' education levels, the most common response was vocational education and training diplomas (48.7 percent of fathers and 35.6 percent of mothers). More than one fifth indicated that the highest level of education that their parents had completed was the middle school level (22.6 percent of fathers and 27.1 percent of mothers). Only 7 percent of both mothers and fathers had received education at the university level.

2.2. Procedure

Data were collected in French-speaking Switzerland at one vocational education center and two community youth centers. The questionnaire for this study was administered as part of a pilot study in preparation for an evaluation of the SEESR program.[1] The organizations offered the SEESR program (including participation in this study) as one of their activities. Following the usual procedure, documentation was sent to parents describing all of the center's activities, including the SEESR prevention program, and parents signed a permission form allowing their children to participate in the various activities and fill in the questionnaire anonymously. The form provided the option for parents to refuse permission for their children to engage in any specific activity on the list. The consent of the participants themselves was obtained at the beginning of the program by center staff. Data analyzed in this article were collected at the beginning of the first session, prior to the presentation of any prevention programming.

2.3. Measures

2.3.1. Egalitarian Social Norms

The questionnaire included ten questions about social norms, adapted from the First Swiss national survey on partner violence (Gillioz, De Puy, and Ducret 1997). The

1 Sortir Ensemble et Se Respecter is a nine-session program promoting healthy relationships and dating violence prevention among adolescents (De Puy, Monnier, and Hamby 2009), adapted from Safe Dates (Foshee and Langwick 1994).

ten items were presented as statements. Participants were given both verbal and visual response categories. This format was suggested by program facilitators and pre-test participants, in order to make it easier to understand and user-friendly. Two happy faces were described as "strongly agree," one happy face represented "agree," one sad face signified "disagree," and two sad faces indicated "strongly disagree." See Table 1 for a description of items.

In the original Swiss survey (Gillioz, De Puy, and Ducret 1997), these questions were analyzed at the item level. For data reduction purposes, a principal factors analysis with a promax rotation was conducted on these items. Two factors accounted for 36 percent of the variance, each with four items with loadings of .3 or higher. Two items were dropped because they did not load on either factor. One factor was endorsement of patriarchal attitudes, which included "It is good when men participate in housework." Such statements phrased in egalitarian terms were reverse-coded. The second factor was disparagement of females, as indicated by negative opinions about women's personal characteristics, and included the item "Women are by nature less talented at math than men." Items were summed to create two scores, with higher scores indicating more attitudes of gender hostility. Internal consistency was adequate for patriarchal attitudes (.62) and fair for disparagement of females (.48).

2.3.2. Attitudes towards Dating Violence

Eight items from the original Safe Dates evaluation (Foshee et al. 1998) were translated and back-translated by the authors. One example is: "It is OK for a boy to hit his girlfriend if she insulted him in front of his friends." Other items are listed in Table 1. Participants were given response options of strongly agree, agree somewhat, disagree somewhat, or strongly disagree, represented pictorially as described above. We coded following the same procedure as Foshee and colleagues, totaling all items to obtain an overall score of dating violence attitudes, with higher scores indicating more pro-violent attitudes. Cronbach's Alpha was .77.

2.3.3. Dating Violence: Perpetration and Victimization

The survey included eighteen questions from the Revised Conflict Tactics Scales (CTS2) (Straus et al. 1996). Statements examining dating negotiation included items such as "I showed my partner I cared even though we disagreed." Dating negotiation items were used only to facilitate disclosure and were not included in analyses. Statements of dating violence assessed a range from verbal insults ("My partner called me fat or ugly") to physical assault ("I threw something at my partner that could hurt"). The CTS2 has shown good internal consistency and construct validity in a number of studies (Straus, Hamby, and Warren 2003). All eighteen items asked how many times the event had occurred within the past twelve months as well as whether it had ever happened earlier. They were combined into a single lifetime score. These questions were only asked of respondents who reported having a dating history (n=86). Items were grouped into four scores: psychological perpetration, psychological victimization, physical assault perpetration, and physical assault victimization. Internal consistency was adequate, at .68 for both psychological perpetration and victimization, .76 for physical perpetration, and .77 for physical victimization.

2.3.4. Demographics

We asked seven demographic questions, including age, gender, country of origin, and length of residency in Switzerland. Participants were asked to indicate whether they were currently in a relationship, and if not, whether or not they had ever been in a relationship. Those who had been or were currently in a relationship were asked the length of the relationship. Two questions asked about the level of education completed by both the mother and father, with seven possible responses ranging from primary school to university.

2.3.5. Father in Home

Family structure was assessed by asking participants to indicate the members of their household. For the analyses we created a variable for presence of the father in the home. Two out of five participants (40 percent) did not live with their father.

2.3.6 Know Female Victims of Teen Dating Violence

Participants were asked to indicate how many victims of teen dating violence they knew personally among friends and family. More than half (55 percent) reported knowing at least one victim.

3. Results

3.1. Descriptive Statistics for Attitudes about Gender Roles and Teen Dating Violence

Descriptive analyses indicated that endorsement of patriarchal attitudes was fairly common, with 40.8 percent saying, for example, that it is better if a woman stays home. As can be seen in Table 1, chi-square analyses indicate significant gender differences for every item about patriarchal attitudes, with boys agreeing with patriarchal attitudes more often than girls. Endorsement of patriarchal attitudes did not, however, vary in connection with the presence or absence of a father in the home.

Endorsement of disparaging beliefs about women's personal characteristics was also fairly common, with more than half of the sample (53.5 percent) agreeing that "women are more easily influenced than men" and more one in four agreeing with statements that "a woman without children is unfulfilled" (28.9 percent) and "women are less talented at math than men" (25.6 percent). More than one in ten (11.5 percent) even agreed to the statement that a "wife must submit to sex with her husband." As also seen in Table 1, these attitudes did not vary significantly by gender. However, it was surprising that boys agreed with the statement that "males and females are equally courageous" more often than girls. For the most part, responses did not vary by family structure either, although respondents with no father in the home were somewhat more likely than those with a father in the home to endorse the statement that a wife must submit to sex with her husband (17 percent versus 6.6 percent, $p < .08$).

Table 1: Attitudes about Gender Roles and Teen Dating Violence

Item	Total (n=132)	Gender: Female (n=52)	Gender: Male (n=71)	χ^2
Endorsement of patriarchal attitudes				
Better if a woman stays home	40.8	26.0	50.7	7.43**
More women not needed in politics	27.9	17.6	34.3	4.12*
Wife should not have equal influence	15.5	5.8	20.0	5.04*
Men should not have to do housework	9.2	1.9	12.9	4.74*
Endorsement of disparaging beliefs about women's personal characteristics				
Women are more easily influenced than men	53.5	52.0	54.3	0.06
A woman without children is unfulfilled	28.9	34.7	23.9	1.65
Women are less talented at math than men	25.6	26.9	23.5	0.18
Wife must submit to sex by husband	11.5	9.6	14.3	0.60
Belief in male and female equality				
Males and females are equally courageous	33.3	17.6	47.8	11.74***
Equal confidence in male and female surgeons	21.7	30.8	15.9	3.76†
Attitudes about justifications for dating violence				
Okay in retaliation if boyfriend hits first	40.8	42.3	43.5	0.02
Boyfriends deserve to be hit sometimes	21.4	21.2	20.0	0.02
Okay if girlfriend makes boyfriend jealous	11.5	3.8	18.6	6.00*
Girlfriends deserve to be hit sometimes	8.4	1.9	14.3	5.56*
Okay if girlfriend insults boyfriend	6.1	1.9	10.0	3.18†
Okay if boyfriend needs to get control	3.9	3.8	4.3	0.02
Okay if girlfriend annoys boyfriend	1.5	0.0	2.9	1.51
Okay in retaliation if girlfriend hits first[a]	22.1	19.2	25.7	0.71

†$p < .08$; *$p < .05$; **$p < .01$; ***$p < .001$. n = 132 except for analyses by gender for which n = 123.
[a]This item double-loaded on both factors and was omitted from further analysis.

Many respondents also endorsed justifications for dating violence, including more than two out of five (40.8 percent) agreeing that it was okay for a girlfriend to hit her boyfriend in retaliation. More than one out of five (22.1 percent) said it was okay for a boyfriend to hit a girlfriend in retaliation. Gender differences were observed for the items about girlfriends making boyfriends jealous, girls deserving to be hit sometimes, and (in a statistical trend), girlfriends insulting their boyfriends, all of which were considered acceptable justifications for violence by more boys than girls (see Table 1). Family structure influenced attitudes about items relating to retaliation, which were both endorsed more frequently by respondents with no father in the home than by those who lived with their father.

3.2. Attitudes, Sociodemographic Characteristics, and Teen Dating Violence

We next examined how these attitudes contribute to actual perpetration and victimization. Four logistic regression analyses were conducted, with perpetration and victimization of physical assault and perpetration and victimization of psychological aggression as the dependent outcome variables. Gender, age, and presence of father in the home were demographic independent variables. We also included length of most recent relationship and personal knowledge of female victims of teen dating violence. Attitudes toward teen dating violence, patriarchal attitudes, and female disparagement were also entered as independent variables. These analyses are limited to the 71 percent of respondents who had been in at least one dating relationship. Bivariate correlations among these variables are shown in Table 2.

Table 2: Intercorrelations among variables used in regressions

	1. Psychological aggression perpetration	2. Psychological aggression victimization	3. Physical assault perpetration	4. Physical assault victimization	5. Relationship length	6. Gender	7. Age	8. Live with father	9. Knows female victims of teen dating violence	10. Gender role attitudes	11. Female disparagement	12. Dating violence attitudes
1. Psychological aggression perpetration	1	.602**	.555**	.451**	.209	-.284**	.102	-.221*	.152	-.125	-.110	.105
2. Psychological aggression victimization	.602**	1	.302**	.390**	.176	-.310**	.071	-.106	.047	-.076	-.084	.006
3. Physical assault perpetration	.555**	.302**	1	.680**	.193	-.161	.144	-.248*	.087	.024	-.076	.141
4. Physical assault victimization	.451**	.390**	.680**	1	.086	.017	.142	-.115	.101	.025	-.058	.196
5. Relationship length	.209	.176	.193	.086	1	-.212	.150	.066	-.011	-.025	-.152	-.128
6. Gender	-.284**	-.310**	-.161	.017	-.212	1	.076	.178	.091	.332**	-.074	.228*
7. Age	.102	.071	.144	.142	.150	.076	1	.148	-.047	-.002	.075	-.089
8. Live with father	-.221*	-.106	-.248*	-.115	.066	.178	.148	1	-.105	.071	-.172	-.201
9. Knows female victims of teen dating violence	.152	.047	.087	.101	-.011	.091	-.047	-.105	1	.096	-.120	.002
10. Gender role attitudes	-.125	-.076	.024	.025	-.025	.332**	-.002	.071	.096	1	.228*	.182
11. Female disparagement	-.110	-.084	-.076	-.058	-.152	-.074	.075	-.172	-.120	.228*	1	.373**
12. Dating violence attitudes	.105	.006	.141	.196	-.128	.228*	-.089	-.201	.002	.182	.373**	1

Table 3: Attitudinal and demographic predictors of physical and psychological teen dating violence

	Physical assault						Psychological aggression					
	Perpetration			Victimization			Perpetration			Victimization		
Predictor	*B*	OR	95% CE	*B*	OR	95% CE	*B*	OR	95% CE	*B*	OR	95% CE
Relationship length	0.04	1.04	(.99–1.09)	0.02	1.02	(.98–1.07)	0.08†	1.08	(1.00–1.17)	0.03	1.03	(.96–1.11)
Gender	–1.06†	0.35	(.10–1.22)	–0.25	0.78	(.25–2.44)	**–1.85***	**0.16**	**(.03-.73)**	**–1.93***	**0.15**	**(.03-.72)**
Age	0.32	1.37	(.14–1.37)	0.26	1.30	(.87–1.94)	0.52†	1.69	(1.00–2.85)	**0.70***	**2.00**	**(1.16–3.47)**
Father in home	–1.14†	0.32	(.10–1.04)	–0.48	0.62	(.21–1.84)	–1.08	0.34	(.08–1.39)	–0.76	0.47	(.12–1.86)
Know female TDV victim	0.07	1.07	(.82–1.39)	0.05	1.05	(.81–1.37)	0.27	1.31	(.83–2.05)	–0.10	0.91	(.68–1.23)
Female disparagement	–0.22	0.80	(.59–1.08)	–0.12	0.88	(.67–1.17)	**–0.41***	**0.66**	**(.45-.98)**	**–0.50***	**0.61**	**(.40-.92)**
Patriarchal attitudes	0.14	1.15	(.88–1.51)	0.08	1.08	(.84–1.40)	0.08	1.09	(.77–1.53)	0.17	1.19	(.87–1.64)
Violence attitudes	**0.07***	**1.18**	**(1.01–1.39)**	**0.17***	**1.18**	**(1.01–1.37)**	**0.32****	**1.38**	**(1.11–1.72)**	0.16	1.17	(.97–1.41)

Notes: n = 86 for participants who have been in a dating relationship. For attitude items, higher scores indicate greater endorsement of attitude.
Gender dummy-coded: female = 1, male = 2. *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .10$.

For both physical assault perpetration and victimization, only endorsement of pro-violence attitudes about dating violence was significantly associated with increased likelihood of occurrence. Being female and having an absent father both approached, but did not meet, statistical significance for perpetration of dating violence (not victimization). Pro-violent attitudes were associated with increased likelihood of psychological aggression perpetration. Female participants also reported higher rates of both psychological perpetration and victimization, although this could be a reporting artifact. Older participants reported greater vulnerability to psychological aggression. Counter to hypothesis, female disparagement was associated with lower rates of psychological perpetration and victimization (see Table 3).

4. Discussion

The key findings of this study are: 1) surprisingly high rates of bias against women and pro-violent attitudes, compared to other Swiss studies; 2) the identification of pro-violent attitudes as the most consistent associate of risk for perpetration (physical and psychological) and vulnerability to victimization of a range of attitudinal, relational, and demographic characteristics; and 3) unexpectedly, female disparagement was associated with lower rates for psychological aggression perpetration and victimization.

The survey found strong biases against women, which encompassed not only social expectations such as staying home to raise children, but also perceived limitations such as less talent in math. We were surprised to observe that substantial proportions of the sample thought women should not work outside the home or become more involved in politics. The female participants endorsed some patriarchal values, with 26 percent agreeing that it is better if women stay at home while men work. Among female participants, 5.8 percent agreed that men need not participate in housework, 17.6 percent thought more women are not needed in politics, and about one in six (15.5 percent) asserted that wives should not have equal influence in the home. For each of these items, male respondents endorsed the patriarchal position more strongly than females. Thus, while there are some indicators of movement towards more egalitarian norms, there is also evidence of persistent patriarchal attitudes, as is also the case in the United States (Foshee et al. 1998).

We found strong endorsement of disparaging attitudes towards women among female respondents. Almost one in ten (9.6 percent) agreed that a wife must submit to sex with her husband, while 34.7 percent agreed with the statement that a woman without children is unfulfilled.

Our sample reported rates of dating violence acceptance similar to those found in a 2010 study of American sixth-grade students (Simon et al. 2010). In both the American sample and our Swiss sample, girls were significantly less accepting of male-perpetrated violence while gender had no significant effect on acceptance of female-perpetrated violence. Both studies also found that youth who had never dated were significantly less accepting of female-perpetrated dating violence. Our results diverged from the American sample with regard to the significance of the effect of dating history on acceptance of male-perpetrated violence. In our sample, youth who had dated and youth who had never dated were similarly accepting of male-perpetrated violence, while in the American sample respondents with a dating history were significantly more accepting of male perpetration than those who had never dated (Simon et al. 2010).

4.1. The Connection between Attitudes and Behavior

We next examined reports of dating violence behaviors for the subset (71 percent) who had been involved in at least one dating relationship. More than two thirds of the sample reported being either a perpetrator (66.3 percent) or a victim (69.8 percent) of psychological aggression. Fewer indicated that they had either inflicted (41.9 percent) or endured (48.8 percent) physical assault, although both values are high. These rates are higher than found in college student samples in Switzerland (Chan et al. 2008; Straus 2008) and are at the high end of rates in North American adolescent samples (Reeves and Orpinas 2011). Our respondents reported committing and experiencing violence in greater numbers than they reported tolerance of violent behaviors.

Nonetheless, we did find that attitudes were significantly associated with violence, especially endorsement of rationalizations and justifications for using violence against a dating partner. We did not expect that high levels of disparagement would be associated with lower rates of violence, but that is what the data suggest. Future research is needed to see if such findings replicate, and if so, why disparagement of women relates to lower rates of violence. One possible explanation might be provided by recent social network analyses of bullying that suggests, contrary to some portrayals, that youth are more likely to bully those of similar social status (Faris and Felmlee 2011).

4.2. Limitations

This study was limited in several respects, due in part to it being the first research of its kind with a Swiss population. The study would have been strengthened by larger sample sizes, particularly since a significant portion of this young age group had never been in a dating relationship. Since our sample is not representative, it is unknown whether our findings can be generalized to other teenage populations. In a trade-off for making the survey of manageable length for a young population, we were limited in the different types of attitudes assessed. We experienced floor effects on the item level in our analyses, particularly with reports of male-perpetrated violence. The original Safe Dates scale was not balanced in terms of attitudes toward male and female perpetration, with six items on male perpetration and only two on female perpetration. We have recently become aware of an adaptation that provides gender balance in the items and would be worth considering for future research (Reeves and Orpinas 2011). Adding items addressing more socially acceptable acts might reveal a clearer picture of actual experiences. Our data showed a majority of respondents rejecting dating violence justifications, yet reporting higher levels of experience of dating violence than we expected. It is important for clinical providers to be aware of this disconnect and to not rely solely on breaking down justifications as a means of reducing dating violence.

4.3. Implications

Our findings suggest that general attitudes toward violence are a more consistent predictor of physical and psychological aggression within dating relationships than attitudinal factors. This has implications for the large body of research that often examines only gender-based attitudes without considering other risk factors that might be better predictors.

Our findings support the need for dating violence prevention among the adolescent population in Switzerland, including evaluations of the effects of such programs. The evaluations need to take into account the practical aspects of gathering the data, including getting the subject population to cooperate in filling out the questionnaires. Our data suggest that attitudes about violence may be more proximally related to actual violence than attitudes regarding women and gender roles. The U.S. Expect Respect Dating Violence Prevention Program is tailored to at-risk youth and provides one useful model for ways to address at-risk Swiss youth with pro-violent attitudes. In addition to a school-wide prevention program and anti-violence youth leadership training, Expect Respect places at-risk students in support groups to provide a healthy social network and peer group (Ball, Kerig, and Rosenbluth 2009). The Swiss SEESR program has already been successfully introduced at several institutions in French-speaking Switzerland, and its training program for facilitators is currently undergoing an evaluation by the Fondation Charlotte Olivier, with the support of the Optimus Foundation (Minore and Hofner 2013). The feasibility of introducing this program on a larger scale is also being evaluated.

Our results indicate that other individual characteristics warrant attention in prevention programming. Our findings confirm substantial gender differences in certain attitudes about violence, and attitudes about inherent differences between men and women (Simon et al. 2010; Straus 2008). More study of the intersections between gender, attitudes, and violence is warranted. A recent review of eight sexual assault prevention programs in American universities (Vladutiu, Martin, and Macy 2011) supports the idea that effective programs are those aimed at single-gender audiences. The influence of dating history should be a special consideration for adolescent prevention programs, a notion that has been supported by the findings of previous studies (Reeves and Orpinas 2011; Simon et al. 2010). All of these studies have found that adolescents tend to become more accepting of dating violence once they have been in a dating relationship.

Our results also indicate the need to pay more nuanced attention to differences across specific attitudes. Person-centered analyses, such as latent class analysis, that enable the examination of subgroups might be useful in future research with larger samples. Future research could also explore whether other items might better capture the specific attitudes of youth. Among female adolescents, we found that essentialist beliefs about women's innate abilities (or lack thereof) appear to be more persistent than beliefs about women's appropriate roles. Male adolescents continued to endorse both types of gender stereotypes at high rates. We also found, for both males and females, that male-perpetrated violence was perceived less favorably than female-perpetrated violence. In this sample, these attitudes interacted in complex ways with violent behavior. Despite this complexity, focusing on attitudes that are still endorsed at high rates may be a means to more accurately assess true attitudes about violence with less influence of social desirability. This may be useful both in terms of measurement and ability to detect change in prevention programs (Klaw et al. 2012) as well as for targeting the areas that are most in need of change in order to reduce future violent incidents.

References

Averdijk, Margrit, Katrin Müller-Johnson and Manuel Eisner. 2011. *Sexual Victimization of Children and Adolescents in Switzerland*. Zurich: UBS Optimus Foundation.

Ball, Barbara, Patricia K. Kerig, and Berri Rosenbluth. 2009. Like a Family but Better Because You Can Actually Trust Each Other: The Expect Respect Dating Violence Prevention Program for At-Risk Youth. *Health Promotion Practice* 10:45S–58S. DOI: 10.1177/1524839908322115.

Black, Michele L., Kathleen C. Basile, Matthew J. Breiding, Sharon G. Smith, Mikel L. Walters, Melissa T. Merrick, Jieru Chen, and Mark R. Stevens. 2011. *The National Intimate and Sexual Violence Survey: 010 Summary Report*. Atlanta: Centers for Disease Control and Prevention. DOI: 10.1016/j.jmwh.2008.04.016.

Chan, Ko Lin, Murray A. Straus, Douglas A. Brownridge, Agnes Tiwari, and Wai-Ching Leung. 2008. Prevalence of Dating Partner Violence and Suicidal Ideation among Male and Female University Students Worldwide. *Journal of Midwifery and Women's Health* 53 (6): 529–37.

De Puy, Jacqueline, Sylvie Monnier, and Sherry L. Hamby. 2009. *Sortir ensemble et se respecter: un outil de prévention des violences et de promotion des compétences relationnelles auprès des adolescent-e-s*. Geneva: Editions IES.

DeWall, Nathan C., Craig A. Anderson, and Brad J. Bushman. 2011. The General Aggression Model: Theoretical Extensions to Violence. *Psychology of Violence* 1 (3): 245–58.

Douglas, Emily M. 2006. Familial Violence Socialization in Childhood and Later Life Approval of Corporal Punishment: A Cross–Cultural Perspective. *American Journal of Orthopsychiatry* 76 (1): 23–30. DOI: 10.1037/a0023842.

Dutton, Leila B., Murray A. Straus, and Rose A. Medeiros. 2006. *Gender Equality and Gender Hostility among University Students: A Multi-National Analysis*. University of New Hampshire.

Faris, Robert, and Diane Felmlee. 2011. Status Struggles: Network Centrality and Gender Segregation in Same- and Cross-Gender Aggression. *American Sociological Review*.76:48–73. doi: 10.1177/0003122410396196.

Foshee, Vangie A., Karl E. Bauman, Wendy F. Greene, Gary G. Koch, George F. Linder, and James E. MacDougall. 2000. The Safe Dates Program: 1-Year Follow-up Results. *American Journal of Public Health* 90:1619–22.

Foshee, Vangie A., Karl E. Bauman, R. W. Helms, Gary G. Koch, and George F. Linder. 1998. An Evaluation of Safe Dates, an Adolescent Dating Violence Prevention Program. *American Journal of Public Health* 88:45–50.

Gillioz, Lucienne, Jacqueline De Puy, and Véronique Ducret. 1997. *Domination et violence envers la femme dans le couple*. Lausanne: Editions Payot Lausanne.

Hamby, Sherry L., David Finkelhor, and Heather Turner. 2012. Teen Dating Violence: Co-occurrence with Other Victimizations in the National Survey of Children's Exposure to Violence (NatSCEV). *Psychology of Violence* 2 (2): 111–24.

Hines, Denise. 2007. Predictors of Sexual Coercion Against Women and Men: A Multilevel, Multinational Study of University Students. *Archives of Sexual Behavior* 36 (3): 403–22. DOI: 10.1007/s10508-006-9141-4.

Killias, Martin, Mathieu Simonin, and Jacqueline De Puy. 2005. *Violence Experienced by Women in Switzerland over Their Lifespan*. Berne: Staempfli.

Klaw, Elena, Beth Lee, Britany R. Alarid, Daniel Miao, Angela Hickenbottom, and Serena Del Mundo. 2012. Rape Myths in the Residence Halls: An Intervention. Family and Intimate Partner Violence Quarterly 4 (3): 199–207.

Minore, Raphaela, and Marie-Claude Hofner. 2013. "Sortir ensemble et se respecter", Swiss Prevention Program: Feasibility Study Report. Fribourg: Fondation Charlotte Olivier.

Reeves, Patricia M., and Pamela Orpinas. 2011. Dating Norms and Dating Violence among Ninth Graders in Northeast Georgia: Reports from Student Surveys and Focus Groups. Journal of Interpersonal Violence 27 (9): 1677–98. DOI : 10.1002/ab.21441

Simon, Thomas R., Shari Miller, Deborah Gorman-Smith, Pamela Orpinas, and Terri Sullivan. 2010. Physical Dating Violence Norms and Behavior among Sixth-Grade Students from Four U.S. Sites. Journal of Early Adolescence 30 (3): 395–409. DOI: 10.1177/0272431609333301.

Straus, Murray A. 2008. Dominance and Symmetry in Partner Violence by Male and Female University Students in 32 Nations. Children and Youth Services Review 30:252–75.

Straus, Murray A., Sherry L. Hamby, Sue Boney-McCoy, and David Sugarman. 1996. The Revised Conflict Tactics Scales (CTS2): Development and Preliminary Psychometric Data. Journal of Family Issues 17 (3): 283–316.

Straus, Murray A., Sherry L. Hamby, and W. Louise Warren. 2003. The Conflict Tactics Scales Handbook. Los Angeles: Western Psychological Services.

Turner, Heather A., David Finkelhor, Anne Shattuck, and Sherry L. Hamby. 2013. Family Structure, Victimization, and Child Mental Health in a Nationally Representative Sample. Social Science and Medicine 87:39–51.

Vladutiu, Catherine J., Sandra L. Martin, and Rebecca J. Macy. 2011. College- or University-Based Sexual Assault Prevention Programs: A Review of Program Outcomes, Characteristics, and Recommendations. Trauma, Violence, and Abuse 12 (2): 67–86. DOI: 10.1177/1524838010390708.

Jacqueline De Puy
jacqueline.de-puy@hospvd.ch

Sherry Hamby
sherry.hamby@sewanee.edu

Caroline Lindemuth
lindecg0@gmail.com