Journal of Quantitatiûe Criminology, Vol. 18, No. 3, September 2002 (© 2002)

Identifying Unit-Dependency and Time-Specificity in Longitudinal Analysis: A Graphical Methodology

Laura Dugan[1]

Longitudinal analysis in criminology and other social sciences has become an important research tool because it allows us to draw conclusions from observing how multiple units change over time. Unfortunately, its results are more vulnerable to potential influences of unusual observational units or periods of time. Current leverage diagnostics are designed for cross-sectional analysis and are fallible when applied to longitudinal models. This article introduces a graphical diagnostic methodology to systematically examine the sensitivity of longitudinal results to extreme observational units and periods of time—unit-dependency and time-specificity. Further the article illustrates its use with an example testing policy effects on black and white female victimization of intimate partner homicide. Results are displayed in an easily understood graph that provides a snapshot of the results' time-specific patterns and robustness to unit-dependency. Currently, comparable tests for panel analysis are tedious and cumbersome. With this new illuminating methodology, researchers and policy-makers can easily decide whether a time-specific or unit-dependent pattern is consequential.

KEY WORDS: longitudinal analysis; observation dependency; outliers; spousal homicide; time specific effects; graphical diagnostics.

1. INTRODUCTION

Analysis of panel data has contributed greatly to the advancement of research in the field of criminology. Researchers have come to realize that it is more sophisticated than analysis of cross-sectional data because it allows them to use the same subjects repeatedly, and potentially examine the effects of changes in characteristics on changes in outcomes. With longitudinal data, we can create estimates from variation over time as well as across units. Including a time dimension is especially important because crime is an event, not a condition. By reducing our models to one period, we constrain incidents to be fixed conditions thus losing information about

[1]Department of Criminology and Criminal Justice, University of Maryland, College Park, Maryland, and the National Consortium on Violence Research, Carnegie Mellon University, Pittsburgh, Pennsylvania. Email: ldugan@crim.umd.edu

the dynamic nature of human behavior and criminal episodes. As researchers who are interested in understanding crime, we often want to identify the effects of changing conditions on the probability or frequency of criminal activity.

A special issue of this journal is dedicated to issues of longitudinal analysis in criminological research (McCord, 2000).[2] In it, Haapasalo, Tremblay, Boulerice, and Vitraro (2000) use person- and variable-based longitudinal approaches to better understand the association between kindergarten teachers' behavior ratings and three types of preadolescent adjustment problems. Earlier work by Levitt (1996) uses annual state-level observations from 1971 through 1993 to calculate the reduction in crime that is marginally associated with each increase in the prison population. Kposowa, Singh, and Breault (1994) use the Longitudinal Mortality Study to determine the relationship between social isolation and homicide victimization. Conaway and Lohr (1994) examine how previous criminal victimization experiences and subsequent interactions with the criminal justice system relate to a victim's current decision to report a crime to the police. Dugan, Nagin, and Rosenfeld (1999) examine the effects of changes in domestic violence resources and other factors in large U.S. cities on changes in rates of intimate partner homicide (see Paternoster, 1987 for a discussion comparing longitudinal and cross-sectional research in the deterrence literature).

Unfortunately, with the added sophistication of longitudinal analysis comes an added complexity that makes it difficult to identify potentially misleading anomalies within the data. This article addresses two problems of this sort, unit-dependency and time-specificity. The first problem occurs when the inclusion of a repeated outlying unit (e.g., a peculiar geographic region or person) alters the results of the analysis. Unit-dependency is also an issue in cross-sectional analysis. However, because each unit only appears in one row of data, conventional diagnostic tests, such as analyzing the residuals and hat-matrix, easily reveal influential observations (see Belsley, Kuh, and Welsch, 1980, for a thorough description of these diagnostics).

The influence of a *repeated* unusual observation is less clear because each row only represents a portion of the time that a unit is observed (i.e., unit i in panel t). By examining only one row of data, the remaining tA1 rows containing information on unit i will likely provide too much leverage to signal a problem. For example, in a study that measures crime rates in U.S. cities annually for ten years, each city is repeated ten times. If by including New York City (an urban region with unusually high levels of

[2]Readers should refer to Volume 16, number 2 of the *Journal of Quantitatiûe Criminology*.

crime) parameter estimates change, evidence of the city's leverage would be obscured in standard diagnostics because no single row reporting on New York City is singularly influential. Even with residual plots, clusters of ten are hardly obvious outliers.

Without systematic methods to clearly identify influential units in longitudinal analysis, statistical artifacts could give more credence to interventions than warranted. For example, research by Lott and Mustard (1997) examines criminal behavior in thousands of counties from 1977 to 1992, and concludes that laws granting the right-to-carry concealed handguns deter crime. When Black and Nagin (1998) reanalyzed their data, they found that without counties from the state of Florida, evidence of deterrence is, at best, ambiguous (see, Lott, 1998; Ludwig, 1998; and Duggan, 2000, for more discussion of this finding). Although the method that Black and Nagin used to identify Florida was effective, it is limited in its ability to identify abnormalities among *many* observational units. They interacted the law indicator with each state to test the hypothesis that it equally influences crime in all states. They did not test the hypothesis that the law equally affects the more than 3,000 counties included in the original analysis. That test would require more than 3,000 additional degrees of freedom reducing the model's power substantially. Furthermore, analyses are rarely used to identify the effects of only one factor. Interacting all units with all factors will likely be impossible because the number of parameters could exceed the degrees of freedom. A more systematic method to analyze panel data of all sizes for unit-dependency can help prevent misguided generalizations.

Just as geographic panel analysis is vulnerable to unusual units, individual-level panel studies are vulnerable to persons with atypical personality traits. For example, longitudinal examination of individuals' developmental stages and tendency towards antisocial behavior could fail to accurately represent the general population if an unusually deviant youth's leverage on the results is unnoticed. Work by Moffit (1993) and others (Nagin and Land, 1993; Nagin, Farrington, and Moffitt, 1995; Nagin and Tremblay, 1999; Maughan, Pickles, Rowe, Costello, and Angold, 2000) have identified unusually persistent offenders often referred to as life-course-persistent deviants or high-level chronics. These groups comprise a relatively small percentage of the population. For instance, Nagin and Temblay (1999) found that about 3% of their sample were chronic in their tendency towards aggression. When such individuals are included among other, more typical, persons for longitudinal analysis, outcomes could be altered, thus erroneously generalized to the entire population of interest. This implies that a systematic method to identify unusual individuals in longitudinal analysis can contribute greatly to the study of criminological behavior.

The second problem with longitudinal analysis, time-specificity, refers to unusually influential periods of time that can be obscured in the estimate of an overall effect. That is, the effectiveness of explanatory variables may not be constant over the entire time span. This could result from unmeasured changes during this period that are jointly related to the response and explanatory variables creating discontinuity of effect. For example, gun control laws that were ignored in the 1970s may be more effective in reducing gun crime as public perception of gun violence grows. Yet, with three decades of data in the analysis, the effectiveness in the latter years could be washed away with the null associations of the earlier years, thus hiding important information. A similar risk may be found when studying individuals' activity. Biological or social influences on antisocial behavior can change as young persons age (for a description of the developmental processes of criminality, see Loeber and Le Blanc, 1990). With visual displays of effectiveness over time, researchers can see patterns of association throughout the entire examination period (see Maltz and Mullany, 2000, who demonstrate how graphs can be used to analyze patterns in life course trajectories).

This article introduces a systematic graphical methodology to test the dependency of results from longitudinal analysis on observational units and to identify unusually influential periods of time. In it, I discuss the dependency and specificity problems that are obscured by the added time dimension and then show how the methodology addresses each problem by systematically deleting subsets of rows defined by the same observational unit, range of time, or both. The resulting product is a graphical display that shows the robustness of the relationship between each explanatory variable and the response variable. With it, we can easily identify results that are vulnerable to atypical units of analysis and examine temporal changes in the pattern of effectiveness.

The remainder of this article uses an example that tests the effectiveness of domestic violence resources on reducing the rates of female victimization of intimate partner homicide by race in 48 large U.S. cities (hereafter referred to as ''femicide'').[3] The data were assembled as part of a larger study that examines the effects of domestic violence laws, policies, and services on rates of intimate partner homicide (Dugan, Nagin, and Rosenfeld, 2000). Because partner femicide is an infrequent event, the model is especially vulnerable to city-dependency and time-specificity. The cities are repeated in six three-year waves spanning the years 1978 through 1996.

[3]This term has been used by scholars who study female victimization of intimate partner and other types of homicide (McFarlane, Campbell, Wilt, Sachs, Urlich, Xu, 1999; Gartner and McCarthy, 1991; Avakame 1999; Radford and Russell, 1994).

2. UNIT-DEPENDENCY AND TIME-SPECIFICITY

Much has been written about the diagnostic tests used to identify problems related to influential observations. Belsley, Kuh, and Welsch (1980) discuss techniques of row deletion that allow researchers to identify observations with a demonstrably larger impact on the calculated values of the estimated coefficients, standard errors, t-statistics, and other parameters. They generalize these methods to include subsets of rows that together have a strong influence on the findings. Their strategy to identify the components of the subsets is to compile a list of the most influential observations resulting from single-row deletion diagnostics. However, this method would fail to identify clusters of observations that only show influence when all members are deleted—similar to the New York City example above. In this case, when only a single observation is deleted, the remaining cluster provides enough leverage to prevent large changes in the parameter estimates. The fallibility of subset identification based on single-row deletion techniques highlights the need for a more intuitive method of identifying influential row subsets. Longitudinal analysis provides natural subsets because observations are repeated. Multiple occurrences of the same unit, as well as ranges of time, form subsets that would be natural candidates to influence outcomes.

In the context of femicide, unit-dependency occurs when the significance of a resource effect is entirely dependent on the inclusion of one city. Such an ‘‘outlying’’ city could lead us to draw conclusions about the average effect of a resource on black or white femicide that is not representative of the remaining cities in the sample. For instance, in Dugan *et al.*’s (1999) similar longitudinal study, they found that as cities developed additional counseling programs for victims of domestic violence, more women are killed by their unmarried partners. However, after conducting a cumbersome series of robustness tests, they found that once San Francisco was removed, the ‘‘lethal’’ counseling effect disappeared. Had that dependency gone unnoticed, policy-makers may have been tempted to limit funding for counseling services.

The problem of time-specificity in the context of femicide occurs when a policy appears ineffective for all six waves of data, but is effective in a portion of the total range. It may be that as an unmeasured time-varying component increases—such as intolerance of the mistreatment of women—policy factors could have an enhanced impact on reducing homicides. By including all six waves in the model, the coefficient estimates measure the ‘‘average’’ impact of factors from all waves, potentially hiding time-specific effects. For example, in the larger examination of this data, it was found that the availability of legal advocacy services is unrelated to the number of men who were killed by their wives (Dugan *et. al.*, 2000). However, after examining the model limiting the wave range, legal advocacy was found to be related to saving the lives of married men in the late 1970s through the early 1980s. In the later years this effect disappeared. Knowledge of the changing patterns of effectiveness can provide important substantive insights to researchers and practitioners.

The examples described above demonstrate that identifying unit-dependency and time-specificity in longitudinal data has previously been done. However, the methods were highly cumbersome making it inefficient to re-test after each iteration of model specification. Also, while each test is important on its own, separately they fail to identify effects that are both unit-dependent and time-specific. For example, they would fail to indicate if the effectiveness of legal advocacy on husband homicide in waves one and two was dependent on the inclusion of Atlanta. Furthermore, a finding that fails the city dependency test in all waves may be robust during a smaller segment of the total period. The method described by this article combines each test, and then displays the results using a series of boxplots showing the degree of significance for all associations using t-statistics. I also show that these graphs can be easily altered to display other relevant estimates such as those for coefficient magnitudes. This strategy will allow researchers to be more confident of their findings knowing that they are robust to outlying units. Furthermore, researchers will have a visual understanding of the temporal patterns of association.

3. DATA AND MODEL

3.1. Data

The analysis is based upon a panel data set of 48 of the 50 largest U.S. cities using six waves spanning 1978 to 1996. Table I lists the original 50 cities and their numeric codes, which will be referred to later. The response variables are the number of black and white women who were killed by their intimate partners within a three-year period.[4] The explanatory variables represent domestic violence laws, policies, and services, demographic attributes of each city related to women's relative economic status and marital domesticity, and finally state benefit levels for a federally funded program directed at low income families with children, Aid to Families with Dependent Children (AFDC).

[4]The broader analysis from which these data were assembled examines all intimate partner homicides calculated separately by victim's gender, marital relationship to offender, and race (Dugan *et. al.*, 2000). For demonstrative purposes, this article only examines female victimization by race.

Table I. City Identifiers

Code: City	Code: City	Code: City	Code: City
1: Albuquerque	14: Detroit	27: Milwaukee	40: San Antonio
2: Atlanta	15: El Paso	28: Minneapolis	41: San Diego
3: Austin	16: Fresno	29: Nashville	42: San Francisco
4: Baltimore	17: Fort Worth	30: New Orleans	43: San Jose
5: Boston	18: Honolulu	31: New York*	44: Seattle
6: Buffalo	19: Houston	32: Oakland	45: Saint Louis
7: Charlotte*	20: Indianapolis	33: Oklahoma City	46: Toledo
8: Chicago	21: Jacksonville	34: Omaha	47: Tucson
9: Cincinnati	22: Kansas City	35: Philadelphia	48: Tulsa
10: Cleveland	23: Long Beach	36: Phoenix	49: Virginia Beach
11: Columbus	24: Los Angeles	37: Pittsburgh	50: Washington
12: Dallas	25: Memphis	38: Portland	
13: Denver	26: Miami	39: Sacramento	

*New York and Charlotte were dropped due to missing resource data.

3.1.1. Homicide Data

The homicide data were extracted from the Supplementary Homicide Reports (SHR) (Federal Bureau of Investigation, 1998). The SHR is an adjunct to Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program that compiles information on crimes recorded by local law enforcement agencies. The SHR augments the standard ''Return A'' report on the number of homicides with information on the incidents themselves, including the victim's sex and relationship to the perpetrator. Because of this feature, it is possible to aggregate the data up to the city-level for each year, and record the number of female homicide victims who were killed by their partners.[5]

Homicide frequencies were summed over the current and subsequent two years. When more than one of these years were missing, the case was deleted. When only one of these three years was missing, the summation was adjusted by a factor of 3/2 and then rounded to a whole number. Three-year summations were used because spouse homicides are rare events and annual counts are highly unstable. Summing over three years is a smoothing procedure (similar to averaging) that reduces the amount of ran- dom variation and preserves the discrete nature of the data.

To reduce dependence across observations, every third year is used. Because 20 years of homicide data are available,[6] two years were dropped

[5]Because participation in the SHR program is voluntary, some law enforcement agencies fail to report their homicide incidents each month. Underreporting was corrected with an adjustment factor based on the total number of homicides reported to UCR. See Dugan (1999) for a description of the adjustment procedure.

[6]All data are available for the years 1976 through 1996. However, because the policy variables are vulnerable to endogeneity problems, they are lagged by one year. Therefore, the earliest year that the homicide data are available is 1977.

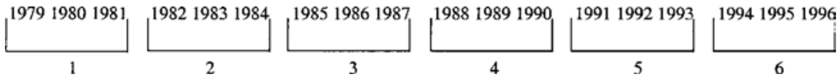

Fig. 1. Years in each wave.

to make the total divisible by three. The years 1977 and 1978 were excluded for two reasons. First, because the domestic violence resource data were collected by interviewing key informants, measurement error due to recollection bias is more likely to plague the earlier years than the later. Second, since the results are likely to be useful to decision-makers, the knowledge of the effects due to more recent policy changes are more informative to current policy decisions.[7] The responses to intervention in more recent years is more likely to accurately predict current responses. The wave compositions are illustrated in Fig. 1.

3.1.2. Domestic Violence Resources

Legal experts including members of the Pennsylvania Coalition Against Domestic Violence compiled the data on statutes from all 50 states. Practitioners from the Women's Center & Shelter of Greater Pittsburgh (WC&S) and a detective from the Pittsburgh Police Department collected information on local police and prosecution policies, and domestic violence services in 48 cities for 21 years. The information on local policy and services would normally be extremely difficult to collect because the primary intention of police, prosecutors, and service providers is to meet the daily demand of persons entering the system, not to document systematic changes over time. Therefore, the crux of the data collection strategy was to seek out informants within the local agencies and ask them to complete a survey instrument inventorying policies or activities by type and year of implementation. See Dugan (1999) for copies of the survey instruments.

Appendix A provides a listing of eleven resource variables that were used in the larger project. For demonstrative purposes, I reduced the dimensionality of that model using principal components factor analysis with varimax rotation. Three primary factors were retained. The first (*State Law*) has high loadings for all state law variables except for mandatory arrest. The second (*Arrest/Serûe*) combines the high loadings of mandatory arrest law with domestic violence services. The third (*Pros Policy*) loads highest for all prosecution variables. Measures on police policy failed to weigh heavily in any of the factors.

[7]Research by Dugan *et. al.* (2000) uses all twenty years of homicide data to incorporate an additional test for robustness.

3.1.3. Economic Measures and Domesticity

Although the focus of this example is primarily on the three resource factors described above, two measures of economic resources and four of marital domesticity were found to be related to intimate partner homicide in the broader study and, therefore, are used here. To measure relative economic status between genders, the models include race-specific ratios of the proportion of women age 25 years or older with at least 4 years of post-secondary education to the corresponding proportion for men. Both types of data were obtained from the U.S. Bureau of the Census for the decenial years 1970, 1980, and 1990, and were interpolated for the years between, then averaged over the appropriate three year periods (U.S. Bureau of the Census, 1993, 1981, 1973). I followed conventional practice in welfare analysis by measuring the generosity of AFDC benefits based on the benefit received by a family of four persons. All figures are adjusted to 1983 dollars using the consumer price index. Data on state AFDC benefits were obtained from annual versions of the ‘‘green book’’ compiled by the House Ways and Means Committee (1996).[8] The impact of domesticity on femicide is estimated with race specific marriage and divorce rates for each city and year. With eight control variables (including fixed effects), there are a total of fifteen explanatory variables. See Table II for a description of the data.

3.2. Model

Because the primary purpose of this article is to introduce and demonstrate a methodology that can be used with most likelihood functions, less attention is paid to the specification of the model chosen for this example. Readers interested in the factors related to intimate partner homicide should read Dugan *et al.* (2000) which provides a more thorough examination of all types of partner homicide.

The probability distribution of rare events like black or white intimate femicide is likely to be generated by a Poisson process (Maddala, 1983). While a Poisson regression is a common modeling choice, it assumes that the expected number of femicides is equal to its variance. When the variance is greater than the mean and the Poisson likelihood function is used, the resulting covariance matrix will be biased downwards, potentially overstating significance levels (Liao, 1994). In such cases, the more appropriate

[8]Data on 1995 AFDC benefit levels were missing. In cases where the 1994 benefit level was equal to the 1996 benefit level, 1995 benefit levels were assumed to be the same as the 1994 and 1996. Eight states had different benefits for 1994 and 1996—Colorado, Florida, Massachusetts, Maryland, New Mexico, Ohio, Oklahoma, and Texas. In these cases, to minimize measurement error, the average of the 1994 and 1996 levels was used for the 1995 value.

Table II. Variables in Model

Variable	Measure
Femicide (by Race)	Three year count of black or white women killed by their partners
Domestic Violence Resources	
State laws	A factor that heavily loads all state law variables described in Appendix A, except for mandatory arrest
Arrest and service	A factor that heavily loads the domestic violence service variables described in Appendix A, with mandatory arrest
Prosecution policy	A factor that heavily loads all prosecution variables described in Appendix A
Economic measures	
relative education (by Race)	The three-year average ratio of the percent of females to males, age 25 and older, who have at least four years of post-high school education
AFDC benefits	The three-year average of the yearly dollar amount given to a family of four, adjusted to 1983 dollars
Domesticity	
Marriage rate (by Race)	The three-year average percent of men or women over the age of 15 (14 for 1970) who are married
Divorce rate (by Race)	The three-year average percent of men or women over the age of 15 (14 for 1970) who are divorced or separated
Controls	
Homicide adjustment	The number of years within the three year homicide range that the were adjusted up due to low reporting months
Adult homicide	The three-year average rate of non-intimate adult homicides
Percent black	The three-year average percent of the population that is black
Wave 2	Indicator variable of wave 2
Wave 3	Indicator variable of wave 3
Wave 4	Indicator variable of wave 4
Wave 5	Indicator variable of wave 5
Wave 6	Indicator variable of wave 6

model is the Negative Binomial regression, which is a more general version of the Poisson regression that allows the variance to be overdispersed (McCullagh and Nelder, 1989). Because tests show overdispersion, the more appropriate model for this research is the Negative Binomial regression shown in equation one. If the model is expressed in terms of the expected number of homicides, λ_{it}, the following Negative Binomial regression is used with each observation weighted by the three-year average of the city's population:

$$\ln(\lambda_{it}) = \ln(n_{it}) + \sum_{k=0}^{K} \beta_k x_{itk} + \sum_{t=1}^{6} \phi w_t + u_{it}, \qquad (1)$$

where n is the total number of black or white women—the population at risk of homicide—and x_{itk} represents the data value for each explanatory variable k in city i during wave t. Time-fixed effects (ϕ) for each wave (w) are also included to control for the temporal data trends.[9] Also, $\exp(u_{it})$ follows a Gamma distribution with mean 1 and variance α, the overdispersion parameter (Long, 1997).

3.3. Findings from Negative Binomial Regressions

The results of model (1) for both black and white femicide victimization using all 48 cities and six waves are displayed in Table III. For illustrative purposes I will focus primarily on the results for the domestic violence factors. Of the six possible associations, only two are significant and negatively related to femicide. The table shows that, all else being equal, cities with stronger state laws, on average, have fewer killings of black women by their partners. White women seem to be better protected by stronger arrest legislation and services provided by domestic violence agencies. Fewer white women living in cities with strong support systems for domestic violence victims are killed by their partners. Finally, the strength of prosecution policy appears unrelated to the number of black or white female victims of femicide.

4. GRAPHICAL RE-ANALYSIS

4.1. Graphical Methodology

This methodology combines two types of sensitivity analyses and jointly displays the results, allowing researchers to easily identify cases of unit-dependency and time-specificity. The first robustness test is used to identify unusually influential units through a series of row subset deletions based on observational units (cities). Since the cities are repeated six times, their influence may not be obvious with conventional diagnostic tests that rely on single-row deletion statistics or residual analysis. The second test identifies sub-portions within the entire wave range that may have stronger associations than those generated from the whole. This is done by deleting partitioned rows of data based on wave ranges instead of observational

[9] Place fixed effects were not used because by including both time and city fixed effects the only variation is found in the changes within each city over time, thus masking important across city information. Levels of resources are hypothesized to be important predictors of homicidal outcomes. Furthermore, the fixed effect for any city without a homicide in a partial range of waves would be undefined in the Negative Binomial likelihood function because the natural logarithm of zero is infinite. See Dugan (1999) for a different approach to addressing this problem.

Table III. Negative Binomial Coefficients on Spouse Homicide Victimization Using all 48 Cities and Six Waves

Variable	Coefficient (Standard error)	
	Black	White
Domestic ûiolence factors		
State law	–0.124**	–0.042
	(0.033)	(0.037)
Arrest and service	–0.057	–0.096**
	(0.032)	(0.033)
Prosecution policy	–0.014	–0.0002
	(0.028)	(0.029)
Economic and domesticity		
AFDC	–0.001**	–0.0005**
	(0.0002)	(0.0002)
Relative education (by Race)	0.112	0.260
	(0.067)	(0.326)
Percent married (by Race)	–0.011	0.007
	(0.008)	(0.007)
Percent divorced (by Race)	0.043*	0.064**
	(0.017)	(0.014)
Controls		
Adult homicide rate	0.006	0.021**
	(0.003)	(0.003)
Percent black	–0.008**	0.013**
	(0.002)	(0.003)
Adjustment	0.018	–0.058
	(0.044)	(0.051)
Wave 2	–0.167*	0.138
	(0.075)	(0.085)
Wave 3	–0.138	0.242*
	(0.094)	(0.108)
Wave 4	–0.197	0.164
	(0.110)	(0.126)
Wave 5	–0.345**	–0.089
	(0.119)	(0.126)
Wave 6	–0.676**	–0.128
	(0.137)	(0.177)

** G*p*F0.01, * G*p*F0.05, two tailed.

units. Because the purpose of the diagnostics is to identify robust associations within a given model specification, t-statistics provide sufficient information to draw such conclusions. However, it will be shown that these tests can be generalized to examine any parameter estimate.

4.2. Testing Unit-dependency and Time-specificity

The program designed to generate estimates after sequentially deleting subsets of rows defined by units and wave ranges is described in Appendix B. The resulting matrix of estimates form a single data set containing the *t*-statistics with 528 rows (48B11) and 9 columns for each parameter.

4.2.1. Graphical Output

Figure 2 presents an example of the graph that summarizes the robustness test for the significance of a parameter estimate using boxplots.[10] Each box plot represents 48 data points that document the resulting t-statistics after each city is omitted. Outlying data points are labeled with a number identifying the omitted city that generated that t-statistic. For example, 19 shows the value of the t-statistic resulting from data after Houston is removed. Each box plot represents one of eleven wave ranges that were tested. The horizontal axis shows which waves were included in the model

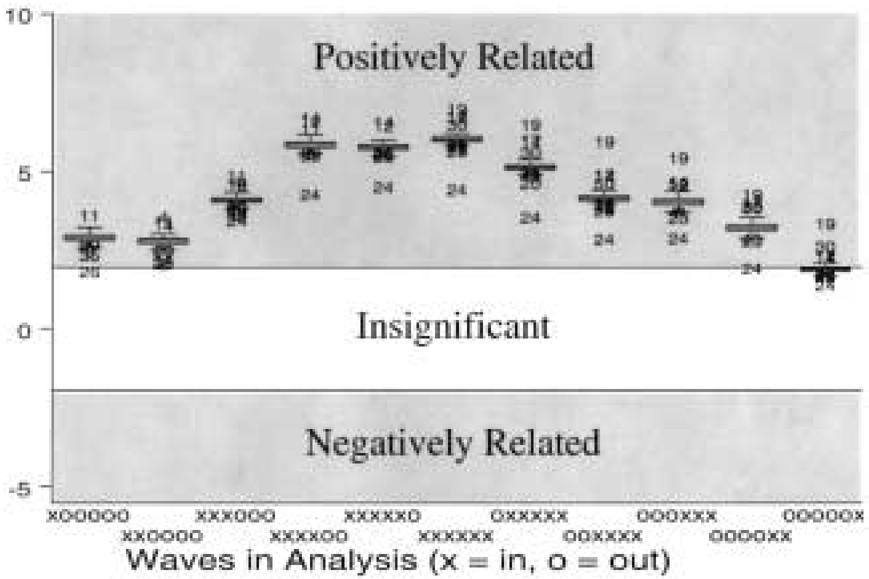

Fig. 2. Strength of association between the adult homicide rate and white femicide when a specific city is dropped during a specific wave range. For example, the valued marked by 24 shows the t–statistic when Los Angeles is omitted from the data.

[10]The graphs are generated using Stata 6.0.

for each box. For example, ‘‘xxxxoo’’ refers to a model that used waves one through four and omitting waves five and six. The range ‘‘xxxxxx’’ uses data from all six waves.

T-statistics for the coefficient estimate of interest are measured by the vertical axis. It is divided into three sections by two horizontal lines that mark the cut off points for two-tail significance tests at the 0.05 level.[11] Estimates that fall in the gray area above the two lines are significant and positively related to homicide. Those falling in the gray area below the two lines are significant and negatively related to homicide. Finally, those that fall in the clear area between the two lines have no significant association with homicide. This series of box plots shows the degree of significance in the association between the control variable *Adult Homicide* and *White Femicide*.[12] Since most boxes fall above the 0.05 significance threshold (upper gray area), we can conclude that *Adult Homicide* rate is positively and robustly related to *White Femicide* in all but the last wave range.

Not surprisingly, the model that uses all six waves (xxxxxx) has the most statistical power. Estimates that are generated from fewer waves—thus, using less data—will have less power, and therefore fall closer to the ‘‘insignificant zone.’’ Each box sequentially positioned away from the center contains one less wave. Therefore, boxes equal distance from ‘‘xxxxxx’’ have equal power. For this reason we would expect that boxes generated from factors with similar associations over time will fall symmetrically around the center box (as shown in Fig. 2).

Also note that city 24 falls lower than the other 48 cities. This shows that when city 24, Los Angeles, is removed from the data the association between *Adult Homicide* and *White Femicide* weakens. Had the t-test omitting Los Angeles fallen into the insignificant zone, this would have indicated that the significant relationship was dependent on Los Angeles.

Since most wave ranges are robust, it is also useful to examine how the magnitudes of the coefficient estimates change throughout the series of city and wave omissions described above. Figure 3 is similar to Fig. 2 except that the vertical axis now measures the coefficient estimates generated by each model. A horizontal line is placed where the estimate is equal to 0 indicating the boundary between positive and negative relationships.

All robust coefficients for *Adult Homicide* fall above 0.01. The stability of the central boxes suggests that the adult homicide rate has a fairly constant impact on white femicide through most of the waves. Because the

[11]This cut off point arbitrarily chosen for illustrative purposes.

[12]The adult homicide rate minus all those committed by intimate partners is a proxy for adult violence in general. It was included to control for changes in intimate partner homicide that are due to a more general change in adult violence. For this reason, it is expected to be positively related to white femicide.

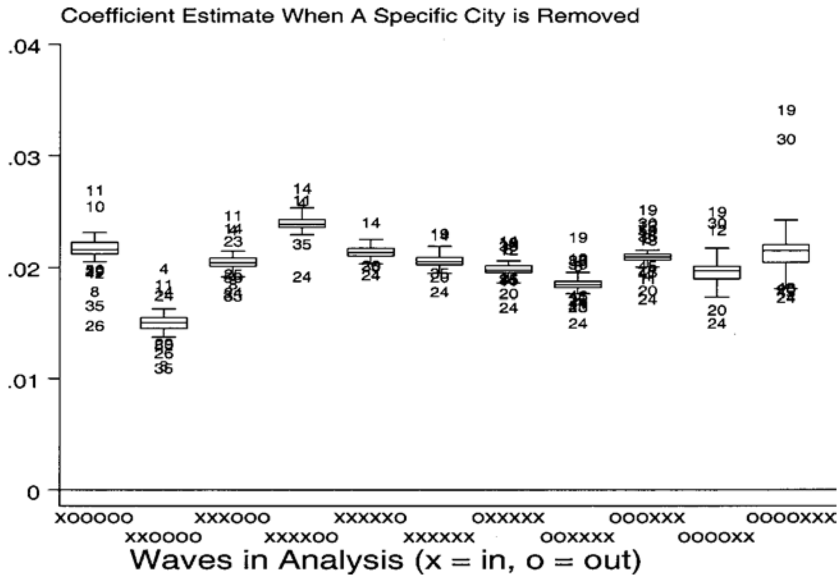

Fig. 3. Magnitudes of association between the adult homicide rate and white femicide when a specific city is dropped during a specific wave range. For example, the valued marked by 24 shows the coefficient estimate when Los Angeles is omitted from the data.

magnitudes in waves one and two (xxoooo) fall closer to zero than those beyond, the impact of *Adult Homicide* appears strongest between the years 1985 through 1996. Finally, recall that the end waves use fewer degrees of freedom resulting in wider confidence bounds to identify the true magnitude. For this reason, it is expected that the magnitude estimates displayed in the outer box plots will vary more than the center boxes.

5. RESULTS

5.1. Graphical Diagnostic Results

By using the graphical methodology described here, we can examine the robustness of the findings in Table III to influential cities across time periods. Graphical summaries of the relationship between the three domestic violence factors and black and white femicide victimization are displayed in Fig. 4. Also included are the graphs showing results for *AFDC* because they provide a good example of a finding that is both unit-dependent and time-specific.

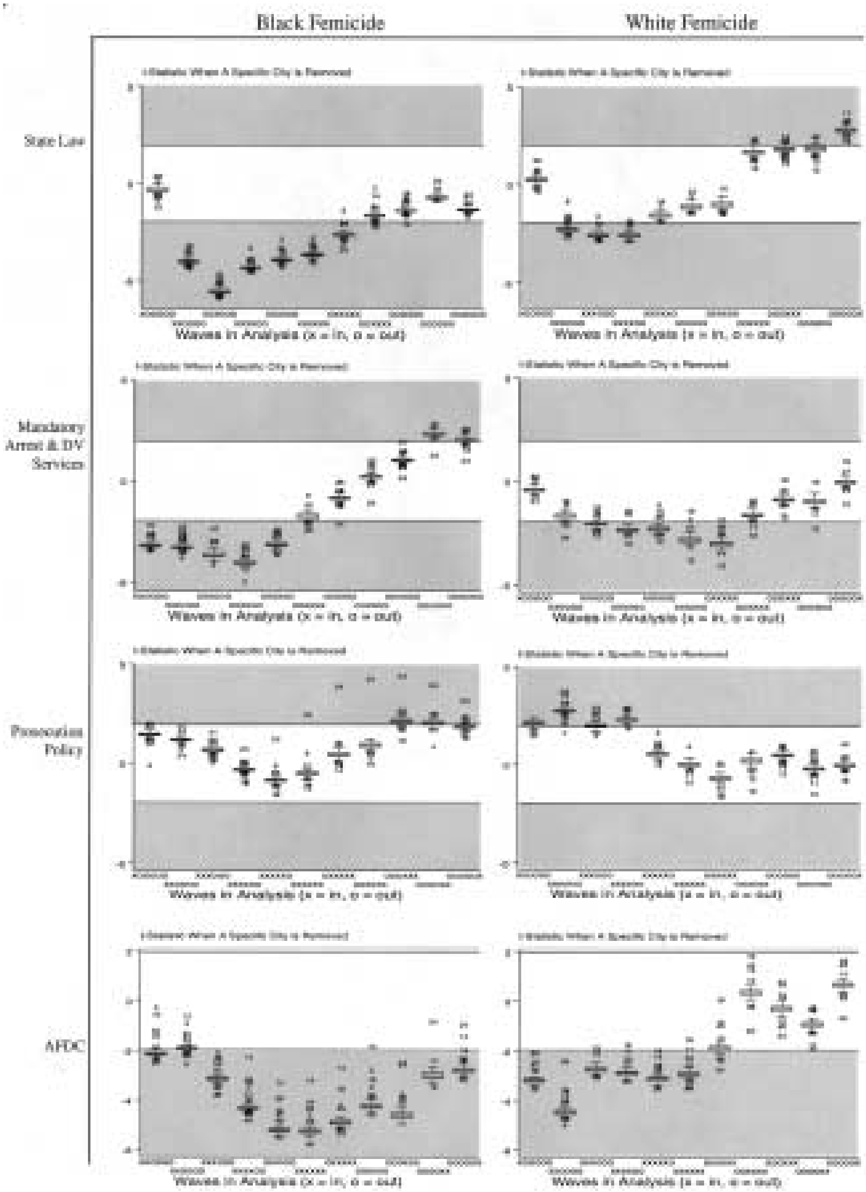

Fig. 4. Strengths of association when a specific city is dropped during a specific wave range. For example, the valued marked by 14 shows the t-statistic when Detroit is omitted from the data.

5.1.1. Robust Findings

Recall that *only two* of the six coefficient estimates of the resource effects displayed in Table III are negative and significantly related to femicide (*State Law* for black victims and *Arrest/Serûice* for white victims). Also, both findings for *AFDC* are negatively related to both black and white femicide victimizations. Diagnostic results testing whether these findings are dependent on a specific city are found in the box plots for waves one through six (xxxxxx) in Fig. 4. Of the four significant findings, only two pass the city-dependency test—*State Law* and *AFDC* for black female victims. Here, all 48 city data points fall in the lower gray area indicating significance at a 0.05 level. Furthermore, in the *State Law* graph the boxes representing the later waves rise into the ‘‘insignificant zone’’ quicker than those representing the earlier waves. This suggests that the robustness of the finding is primarily driven by the earlier years of the study. In contrast, the pattern of the box plots for the *AFDC* finding is nearly symmetrical around the center box, suggesting a time-consistent association. Despite the robustness of *AFDC*, note that city 14 (Detroit) strongly influences the significance of *AFDC* on *Black Femicide*. With Detroit, the t-statistic is −5.00 (as calculated from Table III). Once Detroit is omitted, it drops to −3.26.

5.1.2. City-dependency

Figure 4 shows that the second significant resource factor, *Arrest/Serûice* on *White Femicide*, fails the test for city-dependency. In wave range one through six (xxxxxx) this variable loses significance when cases from either city 19 (Houston) or city 8 (Chicago) are deleted. This is evident because both of those numbers fall in the clear ‘‘insignificant zone’’ between the two 0.05 cut-off points. The dependence of the *Arrest/Serûice* finding on Houston or Chicago suggests that further investigation of these cities is needed to better understand how mandatory arrest and domestic violence services might relate to female victimization of intimate partner homicide. Perhaps agencies in those cities are better equipped to implement policies and services.

5.1.3. Time-specificity

The graph for *Arrest/Serûice* on *Black Femicide* demonstrates how an association may only be evident during a portion of the total time range. Recall that in Table III *Arrest/Serûice* fails the t-test hypothesizing that it is related to *Black Femicide*. However, the graph in Fig. 4 reveals that it is negatively and significantly related to *Black Femicide* when the waves are limited to the earlier periods (xooooo, xxoooo, xxxooo, xxxxoo, and xxxxxo). This shows that cities with a mandatory arrest law and stronger domestic violence services have fewer killings of black women by their partners during the years 1979 through 1993. This finding urges the question of what happened beyond 1993 that would nullify this relationship. Perhaps something changed in the middle 1990s impeding prevention efforts that had earlier been successful.

5.1.4. City-dependency and Time-specificity

After examining the box plot for waves one through six (xxxxxx) in the graph relating *AFDC to White Femicide*, we could easily conclude that its significance is dependent on the inclusion of Houston (19) and Detroit (14), therefore, not robust. However, when only waves one and two are examined (xxoooo), the significance returns and is not dependent on any one city. The negative relationship between *AFDC* and *White Femicide* is specific to the years 1979 through 1984, suggesting that further investigation is needed to better understand how welfare support relates to the killing of white women by their partners during that period.

5.1.5. Null Findings and a City-dependent Robust Finding

The three remaining findings (*State Law* for white victims and *Prosecution Policy* for victims of both races) were null in Table III, and initially appear to remain null in Fig. 4. However, a closer investigation reveals that the omission of Los Angeles (24) dramatically alters the significance of *Prosecution Policy* on *Black Femicide*. Once that city is removed, the significance level changes from –0.50 to 2.36. This suggests that with the exception of Los Angeles more black women are killed in cities with strong prosecution policy, than in those with less. While an intriguing finding, it has yet to pass the tests for robustness. Figure 5 shows the graphical summary for the remaining 47 cities once Los Angeles is removed. As suspected, strong prosecution policy is robustly related to more killings of black women by their intimate partners beyond wave one. See Dugan, *et. al*. (2000) for a discussion of possible ‘‘retaliation effects’’ of domestic violence resources on intimate partner homicide.

5.1.6. Magnitude of Effect

Of the four resource and *AFDC* variables initially found to be significant, only three are robust—*State Law* and *AFDC* for *Black Femicide* and *AFDC* during waves one and two for *White Femicide*. Also, the null relationship between *Arrest/Serûice* and *Black Femicide* is negative, significant, and robust during the years 1979 through 1993 (xxxxxo). And finally,

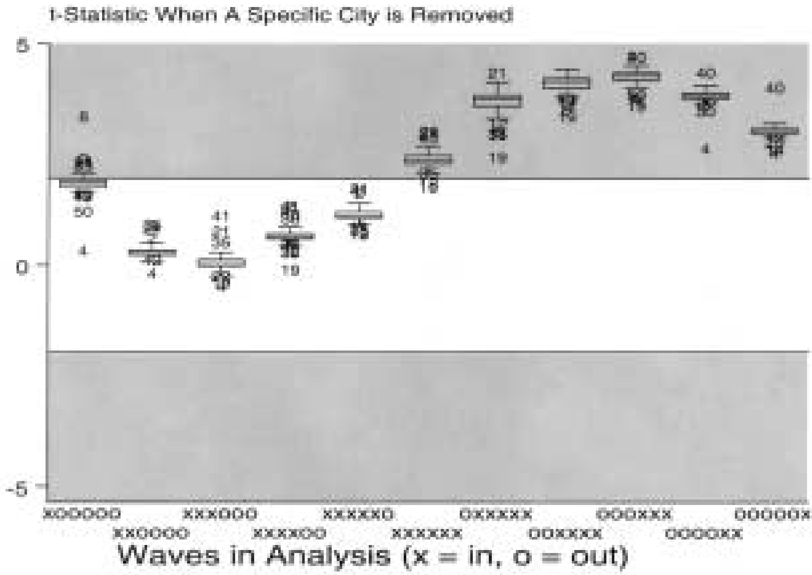

Fig. 5. Box plots testing the robustness of the association between prosecution policy and black femicide using all cities but Los Angeles.

tests reveal that in the absence of Los Angeles, prosecution policy is significant and positively related to *Black Femicide* from about 1982 to 1996.

While t-statistics are important to identify the degree of significance for each association, they fail to inform us on the magnitude or strength of association. To remedy this, similar graphical summaries are displayed in Fig. 6 summarizing the coefficient estimates for all eight relationships. Ovals enclose the estimates that were found to be robustly related to femicide (i.e., none of the t-statistics crossed into the insignificant area in Figs. 4 or 5). The graph showing how *Prosecution Policy* relates to *Black Femicide* generated from data omitting Los Angeles and therefore is comparable to Fig. 5. The vertical axes for the same factor are similarly scaled to easily compare magnitudes across victim types.

Most robust boxes show coefficient ranges that are relatively narrow and distant from zero when compared to the other box plots in the same graph. Further, two graphs show distinct temporal patterns. For black victims, the effects of arrest and services weakens as the more recent years are included in the analysis, suggesting that black women benefited most from these resources during the late 1970s to the early 1980s. Conversely, the ‘‘lethal’’ effects of prosecution resources on black women appear to

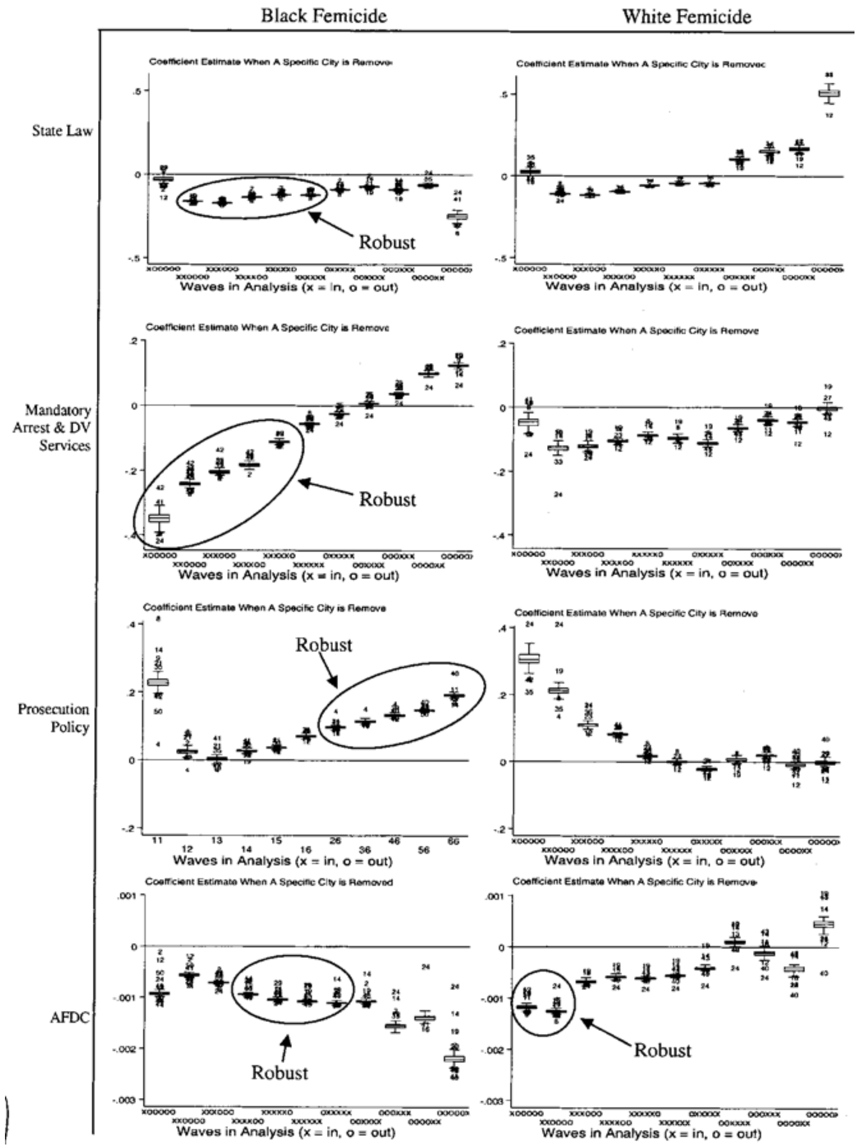

Fig. 6. Magnitudes for domestic violence factors when a specific city is dropped during a specific wave range. For example, the valued marked by 40 shows coefficient estimate when San Antonio is omitted from the data. Also, the prosecution policy graph for black femicide was created without Los Angeles.

strengthen over time as shown by the increasing magnitudes of the estimates. Of the three remaining robust findings, only *AFDC* on *Black Femicide* shows no obvious temporal relationship. The encircled box plots are relatively equidistant from zero and centered around ‘‘xxxxxx.’’

6. CONCLUSION

This article has introduced a systematic methodology to test the robustness of results generated from longitudinal analysis to influential units and periods of time. The graphs can reveal suspicious findings that are significant only when leveraged by a repeated observational unit. Further, the tests can also reveal robust results that are only apparent after the omission of portions of time, or leveraging units.

The example presented above, demonstrates that conclusions drawn from longitudinal data can be erroneous. Without conducting the systematic and rigorous series of robustness tests presented in this article, policy-makers would assume that existing domestic violence laws enhance the safety of black women while mandatory arrest laws and domestic violence services better serve white women. The failure of arrest laws and services to maintain its effect on white femicide as the composition of cities changes suggests that the original finding is suspicious, and that more investigation is needed. Also, the original regressions failed to expose the significance of those same resources on fewer killings of black women. Further investigation of how arrest and service provision has changed for black women over time could lead to innovative prevention strategies. Finally, without this type of graphical testing, policy-makers would be unaware of the potentially lethal affects of prosecution policy on black women involved in intimate relations.

As shown above, the graphical robustness tests can be an important tool to determine policy effects over long ranges of time. Temporally-dependent results could uncover important substantive insights when broader time-changing contextual factors are considered. For example, when examining life-course covariates to aggressive behavior, these graphs will allow researchers to visually examine age-specific associations that may otherwise be masked by behavior during more disciplined years. Furthermore, because this type of graphical exploration is easily generalizable to other longitudinal models, earlier findings can be reanalyzed for robustness and time-specificity. Such a tool can verify or expand the conclusions of others in past and future research. Researchers and policy-makers can now decide whether a time-specific or unit-dependent pattern is consequential with this powerful and illuminating methodology.

APPENDIX A

Original Resource Variables Included in the Factor Analysis

Variable	Measure	Possible values
Warrantless arrest	An indicator variable identifying states that have a warrantless arrest policy when protection orders are violated	0, 1
Mandatory arrest	An indicator variable identifying states that have a mandatory arrest policy when protection orders are violated	0, 1
Violation index	An index that sums the total number of the following consequences for violating a protection order: contempt (either civil or criminal), misdemeanor, or felony	0, 1, 2, 3
Exposure reduction index	An index that increases by one increment for each of the following statute provisions: no-contact order and custody relief, if married; and protection beyond cohabitation and no-contact order, if unmarried	0, 1, 2
Legal advocacy	Index that sums the number of agencies with a separate budget for legal advocacy with the number of agencies that have lawyers on staff, adjusted for the number of women over the age of 15 (14 for 1970) in the city	[0, S)
Hotlines	The total number of hotlines adjusted for the number of women over the age of 15 (14 for 1970) in the city	[0, S)
Police arrest index	An index totaling the number of the following arrest policies: pro-arrest for violation of a protection order, mandatory arrest for violation of a protection order, and mandatory arrest for domestic assault	0, 1, 2, 3
Police commitment index	An index that increases by one increment if the department has a domestic violence unit, and by one increment if it offers domestic violence in-service training to offices	0, 1, 2
DA willingness index	An index that increases by one increment if the prosecutor's office takes cases of protection order violation, and by another increment if the office has a written policy standardizing the prosecution of such cases	0, 1, 2
DA specialization index	An index that increases by one increment if the prosecutor's office has a domestic violence unit, and by one increment if the office has trained legal advocates on staff	0, 1, 2
No drop policy	An indicator variable that identify cities with prosecutors' offices that have no drop policies	0, 1

APPENDIX B

Combining Robustness Tests

The following program uses stata® language to combine tests of unit-dependency and time-specificity to generate a single data set containing the t-statistics generated from a changing series of deleting partitions of rows that rotate through cities and time ranges. The resulting data set will have 528 rows of data (48B11) with 9 columns for each parameter.

local *w* G1
while *w* ≤ 11 {
    local *i* G1
    while *i* ≤ 48 {
        $\mathrm{F}_{\mathrm{N,T}}^{(i,w)}$
        post (t-statistics for parameter estimates 1 through 9)
        local *i* G*i*C1 }
    local *w* G*w*C1 }

where each value of *w* represent a previously defined range of waves that will be omitted.[13] For example, let *w* G1 for wave 1; *w* G2 for waves 1 and 2; *w* G3 for waves 1, 2, and 3; . . . *w* G*T*A1 for waves 1 through *T*A1; *w* G *T* for only wave *T*; *w* G*T*C1 for waves *T*A1 and *T*; . . . *w* G2*T*A1 for waves 2 through *T*. Also, *i* represents each city, and $F_{\mathrm{N,T}}^{(i,w)}$ is the regression without city *i* and wave range *w*.

ACKNOWLEDGMENTS

This research was supported under award 97WTVX0004 from the National Institute of Justice, and grants from the National Consortium on Violence Research and Georgia State University. I would like to thank Daniel Nagin, Raymond Paternoster, Richard Rosenfeld, E. Michael Foster, Dean Dabney, Michael Maltz, Stephen Feinberg, William Vogt and an anonymous reviewer for their helpful comments. An earlier version of this article was presented at the 2000 meeting of the American Society of Criminology.

REFERENCES

Avakame, E. F. (1999). Females' labor force participation and intimate femicide: An empirical assessment of the backlash hypothesis. *Violence and Victims* 14: 277–292.

[13]In this example, I am interested in ranges that are contiguous in time and include either the beginning or ending wave. The selection of wave ranges can include noncontiguous waves if desired.

Belsley, D. A., Kuh, E., and Welsh, R. E. (1980). *Regression Diagnostics Identifying Influential Data and Sources of Collinearity,* Wiley, New York, NY.

Black, D. A., and Nagin, D. S. (1998). Do right-to-carry laws deter violent crime? *Journal of Legal Studies* 27: 209–219.

Conaway, M. R., and Lohr, S. L. (1994). A longitudinal analysis of factors associated with reporting violent crimes to the police. *Journal of Quantitatiûe Criminology* 10: 23–39.

Dugan, L. (1999). *The Impact of Policies, Programs, and Other Exposure Reducing Factors on Intimate Partner Homicide*. Doctoral Dissertation. Carnegie Mellon University.

Dugan, L., Nagin, D. S., and Rosenfeld, R. (2000). Exposure reduction or backlash? The effects of domestic violence resources on intimate partner homicide. *Report to the National Institute of Justice*.

Dugan, L., Nagin, D. S., and Rosenfeld, R. (1999). Explaining the decline in intimate partner homicide. *Homicide Studies* 3: 187–214.

Duggan, M. (2000). More Guns, More Crime, Working Paper 7967. *NBER Working Paper Series*. National Bureau of Economic Research.

Federal Bureau of Investigation. (1998). *Supplementary Homicide Reports 1976–1996*. Machine readable files and documentation obtained directly from the Uniform Crime Reporting program.

Gartner, R., and McCarthy, B. (1991). The social distribution of femicide in urban Canada, 1921–1988. *Law* & *Society Reûiew* 25: 287–311.

Haapasalo, J., Tremblay, R. E., Boulerice, B., and Vitaro, F. (2000). Relative advantages of person- and variable-based approaches for predicting problem behaviors for kindergarten assessments. *Journal of Quantitatiûe Criminology* 16: 145–168.

House Ways and Means Committee. (1996). *1996 Green Book,* US Government Printing Office, Washington, DC.

Kposowa, A. J., Singh, G. K., and Breault, K. D. (1994). The effects of marital status and social isolation on adult male homicides in the United States: Evidence from the National Longitudinal Mortality Study. *Journal of Quantitatiûe Criminology* 10: 277–289.

Levitt, S. D. (1996). The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. *The Quarterly Journal of Economics* 111: 319–351.

Liao, T. F. (1994). *Interpreting Probability Models Logit, Probit, and Other Generalized Linear Models*. Sage Publications, Thousand Oaks, CA.

Loeber, R., and Le Blanc, M. (1990). Toward a developmental criminology. *Crime and Justice* 12: 375–473.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks, CA.

Lott, J. R. (1998). The concealed-handgun debate. *Journal of Legal Studies* 27: 221–243.

Lott, J. R., and Mustard, D. B. (1997). Crime, deterrence, and right-to-carry concealed handguns. *Journal of Legal Studies* 26: 1–68.

Ludwig, J. (1998). Concealed-gun-carrying laws and violent crime: Evidence from State Panel Data. *International Reûiew of Law and Economics* 18: 239–254.

Maddala, G. S. (1983). *Limited-Dependent and Qualitatiûe Variables in Econometrics*. Cambridge Press, New York.

Maltz, M. D., and Mullany, J. M. (2000). Visualizing lives: New pathways for analyzing life course trajectories. *Journal of Quantitatiûe Criminology* 16: 255–281.

Maughan, B., Pickles, A., Rowe, R., Costello, E. J., and Angold, A. (2000). Developmental trajectories of aggressive and non-aggressive conduct problems. *Journal of Quantitatiûe Criminology* 16: 199–221.

McCord J. (2000). Longitudinal Analysis: An introduction to the special issue. *Journal of Quantitatiûe Criminology* 16: 113–115.

McCullagh P., and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman and Hall, London.

McFarland, J. M., Campbell, J. C., Wilt, S., Sachs, C. J., Ulrich, Y., and Xu, X. (1999). *Stalking and Intimate Partner Femicide*. Homicide Studies 3: 300–316.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Reûiew* 100: 674–701.

Nagin, D. S., Farrington, D. P., and Moffitt, T. E. (1995). Life-course trajectories of different types of offenders. *Criminology* 33: 111–139.

Nagin, D. S., and Land, K. C. (1993). Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson Model. *Criminology* 31: 327–362.

Nagin, D. S., and Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Deûelopment* 70: 1181–1196.

Paternoster, R. (1987). The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues. *Justice Quarterly* 4: 173–217.

Radford, J., and Russell, D. (1994). Femicide: The politics of woman killing. *The Australian and New Zealand Journal of Criminology* 27: 210.

US Bureau of the Census (1993). 1990 Census of the Population. Social and Economic Characteristics. USGPO, Washington, DC. (State volumes.)

US Bureau of the Census (1981). 1980 Census of the Population. Volume 1, Characteristics of the Population. USGPO, Washington, DC. (State volumes.)

US Bureau of the Census (1973). Census of the Population, 1970. Volume 1, Characteristics of the Population. USGPO, Washington, DC. (State volumes.)