Bulletin of Applied Economics, 2022, 10(1), 1-19
https://doi.org/10.47260/bae/101

Unemployment and Suicide in the United States: The Import of Addressing Cross-Sectional Dependence

Mitch Kunce[1]

Abstract

Recent reviews of the sociological and economic-based ecological studies of suicide find cyclical unemployment to be a key suicide risk factor, though the evidence presented is mixed at best. The ambiguity of the ecological associations appear to stem from faulty statistical methodologies. Panel treatments offer advantages over conventional time-series methods by exploiting cross-section variation. However, if the added cross-section units are cointegrating (dependent) and independence is presumed, incorrect statistical inference and inconsistent coefficient estimation can result. Herein, we fully address the import of cross-sectional dependence on the ecological relationship between U.S. unemployment rates and suicide rates using an 81-year panel of the 48 contiguous states and the District of Columbia. When proper allowances are made for cross-section dependence at each step of the examination, we find no significant statistical association, short-run or long-run, running from unemployment to suicide rates in the U.S. Results of this sequential analysis highlight potential sources of the ambiguity found in the literature.

JEL classification number: B55, C31, C49.
Keywords: Suicide rates, Cointegration, Cross-section dependence.

---

[1] DouglasMitchell Econometric Consulting Laramie, WY USA. DouglasMitchellConsulting@protonmail.com

Article Info: Received: October 15, 2022. Revised: October 28, 2022. Published online: November 2, 2022.

1 Introduction

A vast literature in the sociological study of suicide has been motivated by Emile Durkheim's (1897/1951) *Le suicide*. Durkheim's theory suggested that the rate of suicide is a function of societies' social relationships and these relationships vary according to their level of integration and moral regulation. In a digression, Fedden (1938) speculated that poverty, in developed societies, could also be an important suicide risk factor. Given the materialism fostering value formation in modern societies, it is reasonable to expect that variations in suicide rates are also related to economic factors and their fluctuations. In support, Henry and Short (1954) spend the better part of their book on the empirical correlates of suicide rates to the business cycle. They posit that economic improvement decreases frustration therefore aggression which is linked to self-harm.[2] Hamermesh and Soss (1974) developed a more formal economic theory based on the lifetime utility function of an individual. The present value of an individual's utility is shown to be increasing in net income and decreasing in age. This individual will take his own life when the total discounted lifetime utility equals zero. From a cohort of these individuals, they derived an age-specific suicide rate function that depends on tastes against suicide and the distribution of net incomes. The comparative static, from this cohort equation, of the suicide rate related to net income is signed negative. Empirically, Hamermesh-Soss modeled country-level suicide rates as a time-series and separate cross-section functions of real income and age. Expected income was allowed to deviate based on the economic cycle. They found a strong negative relation between suicide rates to income and business cycles.

To date, a quick Google Scholar search on the keywords, 'business cycle' 'suicide rates', returns over 280,000 results. The, mostly, seminal work cited above has spurred a sizable cross-discipline interest in the impacts of economic cycles on suicide occurrence (see Vandoros and Kawachi 2021 for a recent review). Conceivably, the concept of the 'business cycle' is overly broad and researchers generally focus on key indicators of economic fluctuations. Unemployment is one of (if not) the most important business cycle indicators. For example, Hatzius and Stehn (2012) referred to the unemployment rate as their "desert island economic indicator", the one they would choose if they had to choose only one indicator to provide information about the cyclical economy. Suicide researchers appear to agree, in a conceptual review and meta analysis of unemployment as a causal effect of suicide, Milner et al (2014) screened 10,258 articles from a search of four separate databases. More recently, Lin and Chen (2018) provided an extensive and thoughtful review of the ecological treatments. Interestingly, studies reviewed found positive, negative or no statistical association between unemployment and the rate of suicide. These ambiguities and lack of causal certainty demand further investigation.

Within the ecological strand of this literature reviewed above, most of the empirical specifications involve strictly time-series data. Only within the last 15 years or so have practitioners paid particular attention to variable integration and time-series cointegration. Of late, so called 'macro' panels, where the number of time-series observations T and grouped units N are relatively large, have received growing interest. So much so that limiting distributions of double indexed integrated processes had to be developed in the econometric literature (Phillips and Moon 1999). The prospect of N and T being large has split the thinking on model estimation into two factions. The first continues to be critical of pooling the data and promotes heterogeneous specifications for each grouped unit (see Baltagi 2021 for a review). Trend variables included in the right-hand-side (RHS) of these single time-series models provide inadequate controls for latent correlated factors that are time-varying within the grouped units. Another drawback of these single

[2] The authors' frustration-aggression theory also applies to the harm of others.

grouped unit specifications is that estimates critically rely on T being sufficiently large. The second faction champions applying time-series procedures to panels. Those in this camp are not resistant to pooling grouped unit data, extend as much heterogeneity as possible, pay close attention to non-stationarity and examine cointegration structures of all variable combinations (see Beenstock and Felsenstein 2019 for a review). The aim of pooled non-stationary panel analysis is to gain observations and statistical power from the added cross-sections. Moreover, estimators and test statistics obtained from using non-stationary panels benefit from having normal limiting distributions. This is in contrast to non-stationary time-series estimators where the limiting distributions are complex functions of Weiner processes.

The macro panel empirical literature is large, but generally ignores error cross-section dependence. A common problem in the estimation of panels with a large number of cross-sectional group units, N, is cointegrating relationships among the units. Presuming that errors are cross-sectionally independent can lead to incorrect inference and inconsistent estimates. Moreover, testing for variable non-stationarity and series cointegration are also impacted by cross-sectional dependence. The purpose of this examination is to fully dissect the import of cross-sectional dependence on ecological relationships between U.S. unemployment rates and suicide rates using an 81-year panel of the 48 contiguous states and the District of Columbia. This analysis, however, is purposed in statistical method and should not be considered a test of any specific suicide theory involving other covariates. In section 2 we will examine the underlying stochastic process generating each variable's series through unit root testing. Methods that address then ignore cross-sectional correlation are compared. Section 3 focuses on cointegration structures and tests for them. First, we assess the import of cross-section dependence along with structural stability on any cointegrating relationship. Comparisons are then made to Pedroni's (1999, 2004) workhorse approach. Section 4 presents four estimation scenarios (cases) intended to reinforce the impact of the misspecification error when cross-section dependence is ignored. Lastly, section 5 provides concluding remarks.

2 Panel Unit Root Testing

A natural first step in the analysis is to establish the non-stationary characteristics of each panel series. Panel unit root testing benefits from increased power through the exploitation of cross-sectional information. However, conventional panel unit root tests have been criticized, of late, for assuming that cross-section cointegrating relationships are not present (Westerlund and Breitung 2013). Requiring cross-sectional independence, when perhaps strong error dependence is in play, tends to distort the size of the estimated test statistics that reject the null of non-stationarity too often. Pesaran (2004, 2015) proposes a test statistic that can be adapted to test individual variable cross-section dependence,

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{i\,j}\right), \tag{1}$$

where, following Pesaran (2007), $\hat{\rho}_{i\,j}$ denotes the pair-wise correlation coefficient from the residuals of cross-sectioned ($i = 1, \ldots, N$; $t = 1, \ldots, T$) Augmented Dickey and Fuller (1979, 1981) ADF regressions.[3] This CD statistic is distributed, generally, asymptotically standard normal and the null hypothesis depends on the relative expansion rates of N and T. Following Pesaran (2015), the null is shown to be,

[3] See Cheng et al (2012) for a similar adaptation of Pesaran (2004). For convenience, we may duplicate the use of some equation variable symbols as the paper proceeds. While there will be some overlap, consider similar symbols and definitions equation specific.

$$0 \le \phi < (2-\varepsilon)/4, \tag{2}$$

where ϕ is the exponent of cross-sectional dependence derived in Bailey et al (2016) and ε is the degree to which T expands relative to N. In the case of large panels (both N and T tend to infinity at the same rate), the appropriate null is defined as weak cross-sectional dependence. In most panel estimations, only strong cross-sectional error dependence can pose deleterious effects on estimation and inference.

Table 1 describes, provides sources, shows descriptive statistics and depicts estimated *CD* statistics and p-values for the natural log series of each panel variable: suicide rate and unemployment rate.[4] Natural logarithmic transformations are used as a means to remove potential growth in variance over time. Cross-sectional test statistics shown are sufficient to reject the implicit null hypothesis of weak cross-sectional dependence at the < 5% level.

Table 1: Data descriptions, sources, descriptive statistics and CD tests

Suicide Rate. Age Standardized per 100,000 total state population. 48 contiguous states and D.C. included.[5] National Center for Health Statistics, National Vital Statistics System, 1940-2020. Year 2000 age standardization applies. Mean 13.39, STD 3.86. Natural Log Suicide Rate, Mean 2.56, STD 0.28. Pesaran 2007 CD, 1.97 (p-value 0.049).
Unemployment Rate. Average annual rates for 48 contiguous states and D.C., in percent of labor force. U.S. Bureau of Labor Statistics, 1940-2020.[6] Mean 5.32, STD 2.30. Natural Log Unemployment Rate, Mean 1.58, STD 0.42. Pesaran 2007 CD, 2.01 (p-value 0.044).

Faced with relatively strong cross-sectional dependence for each panel variable − we opt for Bai and Ng's (2004, 2010), Panel Analysis of Non-stationarity in Idiosyncratic and Common components (PANIC), method for panel unit root testing.[7] The PANIC unit root test is based on a factor model in which non-stationarity can arise from common factors, idiosyncratic components, or both. Consider the following stochastic process for a series S_{it},

$$S_{it} = D_{it} + \lambda_i' F_t + \eta_{it}, \tag{3}$$

where the series is the sum of a deterministic component D_{it}, a common component $\lambda_i' F_t$, and an error η_{it} that is idiosyncratic. The deterministic component is comprised of cross-section intercepts c_i and can include a linear trend $\beta_i t$. The r x 1 vector of common factors is denoted F_t where λ_i is an r x 1 vector of factor loadings. Factor estimation follows the information criteria proposed by Bai and Ng (2002) and is based on the method of principal components. PANIC avoids inconsistent estimation of the components by applying the method to the first-differenced data. Relative to the number of cross-sections (N), the number of common factors (r) are usually small.

---

[4] See the Appendix for a discussion of errors in variables and aggregation bias that plague suicide and unemployment rates.
[5] Data back to 1940 were not available, of course, for Alaska and Hawaii.
[6] Many of the older annual series were found in the *Book of the States* (published annually since 1935 by the Council of State Governments) where the Bureau of Labor Statistics was cited as the source.
[7] This approach is arguably the workhorse in panel unit root testing, however, can suffer from small sample distortion particularly when the number of cross-sections N is 'small'.

Multivariate common factors ($r > 1$) from equation (3) are tested, herein, using the modified version of the, more general, *Qc* test developed by Stock and Watson (1988) denoted *MQc* by Bai and Ng (2004).[8] The *MQc* test, which corrects for serial correlation of arbitrary form through non-parametric estimation, parallels the multivariate procedure suggested by Phillips (1987). Testing proceeds using a successive procedure for determining the number of stochastic trends underlying the r common factors. The null hypothesis states that r common factors have at most r common stochastic trends, against the alternative that they have less than r common trends. If the null is not rejected at the onset, we conclude that r common factors are non-stationary. If the null is rejected, decrement by 1 and repeat until we fail to reject the null or only one trend remains. If the null is rejected when one trend remains, we conclude that all r are stationary.

For each idiosyncratic component $\hat{\eta}_{it}$, the Augmented Dickey-Fuller test is applied to each cross-section. Accordingly, a pooled panel unit root statistic (distributed $\mathcal{N}(0,1)$) for the idiosyncratic terms can be constructed,

$$P_{\hat{\eta}} = \frac{-2\sum_{i=1}^{N}\ln\left(p_{\hat{\eta}_i}\right) - 2N}{2\sqrt{N}}, \tag{4}$$

where $p_{\hat{\eta}_i}$ denotes the probability values from the cross-sectioned Augmented Dickey-Fuller tests. Pooling of p-values has broad appeal and allows more heterogeneity in the cross-sections. The null hypothesis of this test is all cross-sections have a unit root (non-stationary). Note that the null holds only if *no* stationary combination of the single variable (S_{it}) cross-sections exists. As such, the pooled test mirrors a panel test for no cointegration among all cross-sections for each variable separately. If cross-section correlation can be represented by common factors, idiosyncratic error independence is then assumed in developing the pooled test even though univariate testing permits weak cross-sectional correlation.[9] Moreover, tests on the idiosyncratic components are asymptotically independent of tests on the common factors. Lastly, a series with a factor structure is non-stationary (unit root) if one or more of the common factors are non-stationary, or the idiosyncratic error is non-stationary, or both. Table 2 provides the PANIC results for the natural log series of each variable separately. Table 3 provides conventional panel unit root tests, requiring cross-section independence, for comparison.

Table 2: PANIC results

	LN Suicide		**LN Unemployment**	
	Lags	**Stat**	**Lags**	**Stat**
Alabama	2	0.030	5	-1.188
Arizona	4	-0.576	0	**-2.874*****
Arkansas	5	1.308	2	0.113
California	0	1.671	0	-0.609
Colorado	2	**-2.058****	1	**-2.439****
Connecticut	4	-0.593	1	**-2.014****
DC	1	-1.518	0	**-2.059****
Delaware	0	**-1.951****	4	-1.066
Florida	0	-1.037	4	**-2.092****
Georgia	1	-1.442	0	**-2.076****

[8] When $r = 1$, PANIC suggests an Augmented Dickey-Fuller test for unit root in the common factor.

[9] The behavior of this pooled test holds up against other PANIC residual based tests involving pooled autoregressive coefficients and sample moments (Bai and Ng 2010).

Idaho	4	-0.684	0	-1.505
Illinois	0	**-2.963*** **	0	**-2.078** **
Indiana	3	-0.960	0	-1.158
Iowa	0	**-2.622*** **	1	**-2.244** **
Kansas	1	**-3.477*** **	2	-0.115
Kentucky	3	-0.253	1	**-2.257** **
Louisiana	1	-0.419	0	**-2.791*** **
Maine	1	**-1.826***	0	-0.364
Maryland	2	1.086	0	**-2.556*** **
Massachusetts	1	**-1.875***	0	**-3.601*** **
Michigan	2	-1.388	0	-1.219
Minnesota	1	-1.252	0	-1.566
Mississippi	1	-0.265	4	-0.198
Missouri	2	-0.041	2	-1.138
Montana	3	-0.936	0	-0.886
Nebraska	3	0.069	4	-1.433
Nevada	2	1.384	0	-0.703
New Hampshire	1	-1.328	0	-1.124
New Jersey	2	0.281	2	**-2.230** **
New Mexico	0	**-2.381** **	0	-1.533
New York	0	**-2.629*** **	0	**-2.028** **
North Carolina	0	-1.568	2	-0.167
North Dakota	4	-0.064	0	-1.298
Ohio	2	-0.816	0	-1.553
Oklahoma	1	0.206	0	-1.573
Oregon	2	-1.423	2	-1.238
Pennsylvania	1	-1.560	0	-1.549
Rhode Island	1	-1.022	7	-0.707
South Carolina	1	**-2.109** **	1	-1.243
South Dakota	3	-0.763	5	-0.087
Tennessee	3	-1.553	1	-1.524
Texas	1	-1.268	5	-1.518
Utah	1	**-2.112** **	5	0.653
Vermont	1	-1.231	6	0.485
Virginia	3	0.598	2	-1.545
Washington	6	1.942	1	**-2.126** **
West Virginia	2	0.548	1	-1.338
Wisconsin	3	-0.508	5	-0.547
Wyoming	3	0.201	0	-1.445
Cross-Section Rejections		**11**		**15**
Common Factors	**8**	**60.29**	**7**	**31.68**
Pooled Idiosyncratic		**4.726*** **		**6.876*** **
Deterministics: Constant and Trend				
Significance (***) < 1%, (**) < 5%, (*) < 10%.				

Regarding the idiosyncratic components, individual cross-section unit root tests show null hypothesis (unit root) rejections in 11 states for suicide rates and in 15 states for unemployment rates. The last row of Table 2 shows the pooled idiosyncratic component test (equation (4)) for each panel variable. The pooled statistic rejects no cointegration among cross-sections for both suicide and unemployment rates separately.[10] As depicted near the bottom of Table 2, multiple common factors are determined for each panel variable; 8 in the suicide rate series and 7 in the unemployment rate series. For each variable, failure to reject the null of retaining the common factors indicates that *all* common factors are non-stationary. Overall, PANIC results are consistent with non-stationarity in U.S. suicide rates and unemployment rates, pervasive in the common factors and finite in the idiosyncratic components.[11]

Table 3: Conventional panel unit root tests

		LN Suicide	**LN Unemployment**
LLC	Constant	**-1.521***	**-3.043*****
LLC	Constant, Trend	**-2.247****	**-3.094*****
IPS	Constant	**-2.029****	**-6.593*****
IPS	Constant, Trend	**-6.358*****	**-8.055*****
LLC attributed to Levin et al (2002) IPS attributed to Im et al (2003)			
Significance (***) < 1%, (**) < 5%, (*) < 10%. Null: Unit Root			

Table 3 shows first-generation panel unit root tests for comparison. All test specifications reject the null of non-stationarity at conventional levels. As pointed out above, cross-section dependence can lead to deceptive inference in panel unit root testing. O'Connell (1998) first documented the nontrivial size distortions in panel unit root tests developed under the assumption of cross-section independence. As a result, the PANIC results appear more reliable.

3 Cointegration

Testing for cointegration is a necessary second step prior to estimating long-run variable relationships. As in the unit root analysis above, cross-section dependence plays an important role in cointegration testing. Popular testing methods such as Kao (1999) and Pedroni (1999, 2004) assume cross-section independence among the grouped units of panel data which is rare in most regional economic settings. Pesaran (2004, 2015) suggests a cross-section dependence test similar to equations (1) and (2) but based on the average of pair-wise correlation coefficients of the residuals from a pooled panel regression. To obtain these residuals, we regress log suicide rates on log unemployment rates in a simple panel least squares specification.[12] The resulting Pesaran CD test statistic is 2.12 with a p-value of 0.034 which rejects weak cross-sectional correlation in the residuals at the < 5% level (Pesaran 2015). This test result indicates that panel cointegration testing must account for the presence of cross-sectional dependence. In keeping with the factor-based unit root testing in section 2 above, cointegration testing will follow the companion factor-based procedure developed by Banerjee and Carrion-i-Silvestre (2015) which relaxes the assumption of cross-section independence and addresses the structural stability of any cointegrating relationship. Advantages of this method include the ability to paint a more complete picture of the stochastic properties of all variables and components impacting the model. Assessing the properties of the common factors is of particular import because it allows for cointegration among the variables of interest (suicide and unemployment) alone or helps specify whether any non-stationary common factors are needed to form a

[10] This does not imply that all idiosyncratic components are stationary. The rejection of the null implies a finite number of stationary components.
[11] All first differenced tests confirm I (0) when differenced, stationarity.
[12] Similar results can also be obtained with fixed and random effects specifications. However, the CD test based on panel least squares is more robust to slope and error-variance heterogeneity.

cointegrating relationship. Following the same set of assumptions of Bai and Ng (2004), define, $Y_{it} = (y_{it}, x'_{it})'$, a vector of stochastic processes including a dependent variable y_{it} with regressors x_{it} where,

$$Y_{it} = D_{it} + \pi_i \, F_t + u_{it} \,. \tag{5}$$

The deterministic component D_{it} can be empty, include cross-section intercepts and/or linear time trends. The F_t component denotes a vector of common factors, π_i is a matrix of factor loadings and u_{it} is the idiosyncratic disturbance. The dependent variable y_{it} on stochastic regressors x_{it} are assumed relationally cross-section dependent with the dependence determined by the common factors F_t. Interestingly, Banerjee and Carrion-i-Silvestre (2015) (BCIS) show that the set of common factors affecting the dependent variable can be different from those impacting the regressors by defining π_i as block-diagonal.

The BCIS framework is very flexible and implies a broader definition of cointegration. Generally, cointegration among $Y_{it} = (y_{it}, x'_{it})'$ requires F_t to be integrated of order zero, I(0), so that the observables capture all the common trends. However, BCIS allows F_t to be I(0), I(1) or a combination of both. Specifically, the general model that relates the dependent variable and the regressors becomes,

$$y_{it} = D_{it} + x'_{it}\beta_{it} + \pi_i \, F_t + u_{it} \,, \tag{6}$$

where β_{it} denotes the vector of parameters.[13] To this point, cross-section dependence has been presumed to be the result of observable common factors which is not feasible, generally, in practice where unobservable factors must be estimated. BCIS follows Bai and Ng (2002, 2004) using principal components to estimate the common factors and panel information criteria to determine the number of factors. In order to derive idiosyncratic disturbance test statistics, BCIS proposes six model specifications by varying the deterministic components and the cointegrating vector. Using any specification, they suggest testing the null hypothesis of no cointegration using ADF type regression results. The performance-preferred, pooled, panel statistic (between dimension) then becomes,

$$N^{-1/2} Z_{\hat{t}}(\hat{\lambda}) = N^{-1/2} \sum_{i=1}^{N} t_{\hat{\rho}_i}(\hat{\lambda}_i) \,, \tag{7}$$

where $t_{\hat{\rho}_i}(\hat{\lambda}_i)$ are the t-ratios from the model specific ADF-type regressions on the recovered estimated residuals from the model specific version of equation (6). Note that Bai and Ng (2004) prefer to combine p-values (see equation (4)). The notation $(\hat{\lambda}_i)$ refers to the break fraction parameters. Unknown break points are determined as the dates that minimize the sequence of the ADF t-ratio test statistics. This panel test statistic allows for a large degree of heterogeneity and is shown to converge to standard normal post standardization (see Pedroni 1999, 2004 and Banerjee and Carrion-i-Silvestre 2015 for more).

Herein, our application centers on the estimation of unemployment rate impacts on age-adjusted state suicide rates. This data allows a balanced panel of 49 contiguous jurisdictions over the time dimension of 81 years (see Table 1). Recall the CD test statistic rejects the null hypothesis of weak cross-sectional dependence at the < 5% level. Table 4 reports the results of the BCIS factor-based cointegration testing.

---

[13] Banerjee and Carrion-i-Silvestre (2015) denote the deterministic component as $f_i(t)$.

Table 4: BCIS panel cointegration test

	Not Transformed	**Transformed**
Common Factors	6	1
MQc	25.02	
ADF t-stat		-1.548
Separate State Rejections	15	16
Pooled Z_t	-3.17**	-4.23**
-Schwert Criteria Max Common Factors: 10		
-Significant at the < 5% level (**) for all D_{it} . See BCIS (2015) for tables containing critical values.		
-Results for one unknown break point shown.		
-Allowance for structural change in level and trend.		

Two sets of results are presented depending on whether or not the variables are transformed by dividing them by their standard deviations when using principal components. Without transformation, Bai and Ng (2002) criteria select 6 common factors. The *MQ* statistic fails to reject the null of retaining all of the common factors (interestingly for any value of λ) therefore *all* common factors are non-stationary. With transformation, Bai and Ng criteria selects one common factor that tests out as non-stationary. The calculated structural break modestly impacts the test results. In both sets, however, the pooled idiosyncratic tests reject the null for all D_{it} specifications with one unknown break modeled at the < 5% level. To conclude, the variables (in the vector $Y_{it} = (y_{it}, x'_{it})'$) natural log suicide rates and natural log unemployment rates are *not* cointegrated because at minimum one common factor is non-stationary.

In order to highlight a contrast, we test the panel for cointegration following Pedroni (1999, 2004).[14] Pedroni's popular panel methodology is Engle and Granger (1987) based examining the residuals of a spurious regression using non-stationary variables. If the suite of variables are cointegrated the residuals should be integrated of order zero. Ensuring broad applicability, Pedroni proposes several tests that allow for heterogeneous intercepts and coefficients across cross-sections. Under the null hypothesis of no cointegration, the residuals will be non-stationary. There are two alternative hypotheses: homogenous (common autoregressive coefficients) or within-dimension panel; and heterogeneous (individual autoregressive coefficients) or between-dimension group mean. Pedroni (2004) defines five specific test statistics, three within-dimension panel and two between-dimension group mean. The first three statistics: 'panel rho', 'panel t' and 'panel variance ratio (v)' are analogous to the semiparametric treatments examined by Phillips and Perron (1988) and Phillips and Ouliaris (1990) for conventional time-series data. The two grouped mean statistics, 'group rho' and 'group t' were also adapted from Phillips and Ouliaris (1990). Additionally, we will include parametric ADF versions of the panel and group mean statistics for comparison (see Pedroni 1999). Table 5 presents the cointegration tests for the natural log transformed variable suite.

[14] Pedroni (1999, 2004) relies on cross-sectional independence.

Table 5: Pedroni residual cointegration tests

Observations: 3,969		
Cross-sections included: 49		
Null Hypothesis: No cointegration		
Trend assumption: Deterministic intercept and trend		
Automatic lag length selection based on Akaike information criterion with a max lag of 11		
Newey-West automatic bandwidth selection and Bartlett kernel		
LN SUICIDE on LN UNEMPLOYMENT		
Within-dimension common AR	**Statistic**	**p-value**
Panel: v	2.294	0.011
Panel: rho	-1.671	0.048
Panel: t	-2.246	0.012
Panel: Augmented Dickey - Fuller	-1.387	0.084
Between-dimension individual AR	**Statistic**	**p-value**
Group: rho	-1.581	0.058
Group: t	-2.369	0.009
Group: Augmented Dickey - Fuller	-1.282	0.099

In a Monte Carlo experiment of the small sample properties of the statistics, Pedroni (2004, pp. 609-617) finds, for T in the range of 80, that the 'panel v' and 'group rho' statistics suffer size distortions that lead to persistent failure to reject the null in simulation.[15] This was not the case herein. Pedroni makes the point, for shorter panels, "if the 'group rho' statistic rejects the null, one can be relatively confident of the conclusion". Conversely, the t-statistics were size distorted such as to over reject the null of no-cointegration in simulation.[16] In our case, the 'panel v', 'panel rho' and 'panel t' statistics are all sufficient to reject the null at the < 5% level while the 'group t' statistic rejects at the < 1% level. In light of the Pedroni Monte Carlo results, we conclude that Pedroni's methodology, that relies on cross-section independence, indicates cointegration of suicide rates and unemployment rates.

4 Panel Estimation Scenarios

In our first of four cases, the presumption is that the researcher ignores cross-section error dependence. First-generation panel unit root tests (see Table 3) appear to reject the null of non-stationarity and the researcher proceeds with conventional panel data estimation. The estimated model becomes,

$$S_{it} = \alpha + U_{it}\beta + \varepsilon_{it} \,, \tag{8}$$

where S_{it} is the dependent variable (natural log of suicide rates), α is a scalar constant, U_{it} is the natural log of unemployment rates that vary across states i and over years t, β is the homogeneous estimated long-run coefficient and ε_{it} denotes the overall error term. The error term is comprised of three components,

[15] The panel 'rho' statistic is shown to be somewhat neutral regarding size distortion for $T = 80$.
[16] Pedroni shows that when the T dimension reaches 150, the size distortion of the statistics wane.

$$\varepsilon_{it} = \mu_i + \gamma_t + \nu_{it}, \tag{9}$$

where μ_i denotes the unobservable state specific effects, γ_t represents the latent year specific effects and ν_{it} is the remainder stochastic disturbance.[17] The component μ_i is time-invariant and will account for state specific effects not included in the right-hand-side (RHS). The year effects, γ_t , are state-invariant and capture any unobserved, time-varying effects common to all states. The remainder disturbance ν_{it} varies with states and time and is assumed orthogonal to U_{it}, μ_i and γ_t with a mean of zero and a constant variance σ_ν^2.

Generally, two specifications of equation (8) are considered and differ based on their treatment of μ_i and γ_t. First, 'fixed effects' (FE) treats μ_i and γ_t as fixed but unknown constants differing across states and years. This specification is easily estimated by including state and year dummy variables in the RHS of equation (8) (Least Squares Dummy Variable (LSDV) estimator). However, when *N* and/or *T* are exceedingly large, LSDV suffers from the loss of precious degrees of freedom. Alternatively, estimates can be obtained by transforming the data into deviations from their respective means ('within' estimator). The two fixed effects estimation methods described reveal two crucial defects: (i) any year-invariant and/or state-invariant observable variables cannot be estimated, and (ii) the estimator is not fully efficient because, in certain cases, it ignores variation across states and/or time. Second, 'random effects' (RE) assumes that the μ_i and γ_t are random variables, distributed independently across states and time with variances σ_μ^2 and σ_γ^2. Estimates of this specification are based on transformations of the data into deviations from weighted respective means where the weights are based on, generally, the estimated variances of the components in equation (9), *N* and *T* (Feasible General Least Squares (FGLS) estimator, see Baltagi 2021). Unbiased robust estimates of the variance components are best obtained from pooled ordinary least squares (OLS) and LSDV estimators. The potential correlation of μ_i and γ_t with the RHS variables is a defect of the random effects construct. If these correlations are present, random effects estimation yields biased and inconsistent estimates of β. Conversely, the fixed effects estimator is not impacted by this lack of orthogonality.

Hausman (1978) outlines a specification test of the null hypothesis of orthogonality between the latent effects and regressors where $H_0 : E(\mu_i|U_{it}) = E(\gamma_t|U_{it}) = 0$. By failing to reject the null, both fixed effects and random effects are unbiased and consistent, but fixed effects is less efficient. When the null is rejected, fixed effects is unbiased and consistent but random effects is not. Accordingly, if the null is not rejected the two estimates should not differ systematically. The test of the null considers the difference between the two estimators, $\hat{g} = \hat{\beta}_{FE} - \hat{\beta}_{RE}$, within the sampling error. Hausman (1978) formally derives the chi-squared test statistic based on the Wald criterion,

$$\chi_K^2 = \hat{g}'[Var(\hat{g})]^{-1}\hat{g}, \tag{10}$$

where *K* degrees of freedom equals the number of estimated slope coefficients. The center positive definite matrix should be based on robust covariance estimates. In summary, the random effects specification requires exogeneity of all regressors and the components in equation (9). Conversely, the fixed effects model allows for endogeneity of all the regressors and μ_i, γ_t − but ignores observable state- and/or year-invariant regressors. Table 6 presents the first case estimation results indicating a significant, positive long-run association between U.S. unemployment rates and suicide rates.

---

[17] Potential bias correction of the ν_{it} follows Newey and West (1987).

Table 6: Conventional two-way model

Variable/Test	FE Coefficients/Statistic
Constant	2.443***
LN Unemployment	0.070***
R Squared	0.84
Hausman (df) FE vs RE	11.901***(1)
F-test State Effects (df)	359.466*** (48, 3839)
F-test Period Effects (df)	23.169*** (80, 3839)
Estimated with Newey and West (1987) robust covariance matrix for unspecified autocorrelation. Observations: 3,969. *** significance at the < 1% level.	

In our second case, the researcher continues to ignore cross-section dependence but recognizes that panels with T being sufficiently large require vigilant consideration of time-series properties.[18] What appears to be a strong regression relationship in case one above could be entirely spurious due to underlying characteristics of the time-series processes. Along with variable series integration, the panel application is further complicated by possible heterogeneity of the parameters. Recall, conventional panel estimation specifications presume homogeneity of long-run parameters. Regarding variable integration, a closer examination of conventional unit root test results, provided in Table 3, show that the conclusion of stationarity for the natural log of suicide rate series is not statistically imperious. Paired with the conclusion of cointegration detected by the Pedroni testing shown in Table 5, our practitioner decides to estimate a dynamic heterogeneous panel model, specifically a panel autoregressive distributed lag (ARDL) model (Pesaran et al 1999). The model estimated is shown to be,

$$S_{it} = \sum_{j=1}^{p} \lambda_{ij} S_{i,t-j} + \sum_{j=0}^{q} \delta_{ij}^{/} U_{i,t-j} + \mu_i + \varepsilon_{it}, \tag{11}$$

where S_{it} is the natural log of suicide rates, λ_{ij} are the coefficients of the lagged dependent variables, p is the lag selection for the dependent variable, δ_{ij} are the coefficients of the regressor (U, natural log of unemployment rates) and the respective lags, q is the lag selection for the regressors, μ_i are the state specific effects and ε_{it} denotes the remainder disturbance. Overall, T must be large enough so the model can be estimated for each of the 49 jurisdictions. Additionally, time trends or other types of fixed regressors can be included in the RHS of equation (11).

In certain cases it may be best to work with a re-parameterization of equation (11),

$$\Delta S_{it} = \phi_i S_{i,t-1} + \beta_i^{/} U_{it} + \sum_{j=1}^{p-1} \lambda_{ij}^{*} \Delta S_{i,t-j} + \sum_{j=0}^{q-1} \delta_{ij}^{*/} \Delta U_{i,t-j} + \mu_i + \varepsilon_{it}, \tag{12}$$

[18] This second scenario appears most often in the studies, using panels, reviewed by Milner et al (2014), Chang and Chen (2017), Lin and Chen (2018) and Vandoros and Kawachi (2021).

where the first two terms on the RHS reflect the long run estimates, Δ indicates first differencing of the variable series and ϕ_i denotes the state specific error correction coefficients denoting the speed of adjustment to long-run equilibrium. Based on the results of a Hausman (1978) type specification test, our researcher estimates a pooled mean group (PMG/ARDL) model.[19] This estimator allows the intercepts, short-run coefficients and error variances to differ across states, though the long-run estimates are homogenous. Estimation results are depicted in Table 7 and show support of both long- run and short-run associations between suicide and unemployment rates though the speed of adjustment coefficient is narrowly significant at the < 10% level. Generally, cointegration is ascertained from the significance of the long run coefficients, specifically $\phi_i \neq 0$ statistically. In this case, evidence of a long-run relationship is weak and appears suspect.

Table 7: Panel PMG/ARDL results

	Panel PMG/ARDL	
Variable	**Coefficient**	**P-value**
Long-Run		
LN Unemployment	0.081	0.052
Short-Run		
Constant	0.261	0.000
ΔLN Suicide(-1)	-0.323	0.000
ΔLN Suicide(-2)	-0.102	0.000
ΔLN Unemployment	0.034	0.000
ΔLN Unemployment(-1)	0.067	0.000
ΔLN Unemployment(-2)	-0.001	0.888
Cointegration variable ϕ	-0.108	0.097
Wald, F-test[a]	93.65	0.000

[a]The null hypothesis of the Wald test: Unemployment does not Granger (1969) cause Suicide (short run); H_0: ΔlnUnemp = ΔlnUnemp(-1) = ΔlnUnemp(-2) = 0.
Akaike information criterion (AIC) lag selection: 3 lags (*p* & *q*).

To this point our hypothetical researcher has ignored the potential presence of error cross-section dependence. Wrongly assuming cross-sectional independence has lead to incorrect testing inference as described above. Pesaran CD tests clearly indicate cross-section dependence in the variables and regression suite. PANIC unit root testing confirms that suicide rates and unemployment rates are indeed integrated series of order one. BCIS panel cointegration testing validates that the two series are not cointegrated due to at least one non-stationary common factor, suggesting no long-run relationship of the variables. In order to examine short-run associations, our case number three follows, again, Pesaran et al (1999) and estimates a panel vector autoregressive (VAR) model. This specification does not rely on the cointegration of the non-stationary variables. The model is defined,

$$S_{it} = \sum_{j=1}^{p} \lambda_{ij} S_{i,t-j} + \sum_{j=0}^{q} \delta'_{ij} U_{i,t-j} + \mu_i + \varepsilon_{it}\,. \tag{13}$$

[19] The Hausman test statistic does not reject the null of equality between the mean group estimator (Pesaran and Smith 1995) and PMG. In this case, PMG is more efficient.

It is important to note that the long run and state specific error correction coefficients do not appear in this parameterization of the model due to the lack of cointegration of the variable series. Moreover, the practitioner is not interested in short-run reverse causality. The first two columns of Table 8 present the fixed-effects results of case three.[20] For the panel VAR results, the Granger (1969) causality null hypothesis becomes: unemployment does not Granger cause suicide. A Wald test of log unemployment and its respective lags, jointly equaling zero, is rejected at the < 1% level. In this case we can conclude that rising unemployment rates are linked, in the short run, to increases in suicide rates in the U.S.

Table 8: Panel VAR comparisons

	Panel VAR		**CS Augmented VAR**	
Variable	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Constant	0.253	0.000	0.325	0.000
LN Suicide(-1)	0.540	0.000	0.459	0.000
LN Suicide(-2)	0.355	0.000	0.418	0.000
LN Unemployment	0.051	0.000	0.003	0.802
LN Unemployment(-1)	0.019	0.024	0.011	0.504
LN Unemployment(-2)	-0.031	0.000	-0.016	0.193
Wald, F-test	108.79	0.000	0.75	0.525
AIC lag selection	2 lags (*p* & *q*)		2 lags (*p* & *q*)	

In case three, the researcher was diligent in accounting for cross-section error dependence in the unit root and cointegration testing, yet ignored any potential impact on the panel VAR estimation. In our fourth and final case we correct this misspecification. When cross-sectional dependence is present in the panel, regression estimates can be inconsistent. For example, when the dependence is linked to unobserved common factors, parameter estimates are inconsistent if the factors and regressors are correlated. Chudik and Pesaran (2015) developed an estimator that approximates the common factors by adding cross-sectional averages to the RHS. Cross-sectionally augmenting equation (13) leads to,

$$S_{it} = \sum_{j=1}^{p} \lambda_{ij} S_{i,t-j} + \sum_{j=0}^{q} \delta'_{ij} U_{i,t-j} + \sum_{j=0}^{z} \psi'_{ij} \bar{v}_{t-j} + \mu_i + \varepsilon_{it}\,, \tag{14}$$

where $\bar{v}_{t-j} = (\bar{S}_{i,t-j}, \bar{U}_{i,t-j})$ defines the cross-section averages and $z \approx \sqrt[3]{T}$ denotes the lag floor. Herein, for the cross-section averages, only the base of the variables are added (no lags) in order to avoid multicollinearity issues. Short-run estimates of the augmented model are shown in the last two columns of Table 8.[21] The Wald F-test fails to reject joint insignificance of the log unemployment variables. When properly accounting for cross-section dependence in *all* modeling aspects, we find no statistically imperious association, short-run or long-run, between suicide rates in the U.S. and cyclical unemployment rates. Results highlight the important role cross-sectional dependence, non-stationarity and cointegration play when searching for short-run and long-run series links. Findings in this examination bolster those critical of the unemployment-suicide association found in many ecological treatments (see Laanani et al 2015 for a thoughtful critique). The inconsistencies highlighted herein temper any meaningful causal interpretation.

---

[20] A large Hausman statistic of 113.32 with 5 degrees of freedom soundly rejects the random effects specification.
[21] For comparison, a reviewer wanted to include a fully estimated CS-ARDL model with the long-run coefficients calculated. Results, reaching the same conclusion of insignificant short and long-run associations, are included in the Appendix.

5 Concluding remarks

In 2021, close to 48,000 died from suicide in the U.S., up roughly 4% from 2020 (Curtin et al 2022). Since 2000, the U.S. has seen more than a 34% increase in age-adjusted suicide rates. Recent reviews of the sociological and economic-based ecological studies of suicide find cyclical unemployment to be a key suicide risk factor, though the evidence presented is mixed at best. A new vein of this literature is emerging and links the ambiguity of ecological associations to faulty statistical methodologies. Conventional time-series methods applied suffer from inadequate period observations, the inability to sufficiently control for latent confounders and limiting distributions that rely on complex functionals of Weiner processes. In contrast, for those persistent on using ecological designs, the use of non-stationary panel data methods may provide advantages when the limited time dimension can be augmented by including cross-section variation and when normal limiting distributions, leading to better statistical inference, are favored.

Including cross-section variation in the specification potentially adds a new layer of complexity. If the cross-section units are cointegrating (dependent) and independence is presumed, incorrect statistical inference and inconsistent coefficient estimation can result. Cross-sectional dependence has to be tackled in all modeling aspects − unit root testing, cointegration testing and model estimation. Herein, we fully address the import of cross-sectional dependence on the ecological relationship between U.S. unemployment rates and suicide rates using an 81-year panel of the 48 contiguous states and the District of Columbia. When proper allowances are made for cross-section dependence at each step of the examination, we find no significant statistical association, short-run or long-run, between suicide rates in the U.S. and cyclical unemployment rates. Results of this sequential analysis highlight potential sources of the ambiguity found in the literature attempting to link unemployment with suicide. This last point gives rise to a broader limitation in ecological suicidology crucial to the direction of future research. Studying jurisdictional suicide rates rather than understanding individual behavior is problematic. Identifying the actual causes of self-harm during economic crises appears paramount when designing costly public health policies aimed at combating suicide occurrence.

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica 70(1), pp. 191-221.

Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. Econometrica 72(4), pp. 1127-1177.

Bai, J. and Ng, S. (2010). Panel unit root tests with cross-section dependence: A further investigation. Econometric Theory 26, pp. 1088–1114.

Bailey, N., Kapetanios, G. and Pesaran, M. (2016). Exponent of cross-sectional dependence: estimation and inference. Journal of Applied Econometrics 31(6), pp. 929-960.

Baltagi, B. (2021). Econometric Analysis of Panel Data. Sixth Edition. Chapter 12. Springer: Switzerland.

Banerjee, A., and Carrion-i-Silvestre, J.L. (2015). Cointegration in panel data with breaks and cross-section dependence. Journal of Applied Econometrics 30, pp. 1-23.

Beck, N. and Katz, J. (1995). What to do (and not to do) with time-series cross-section data. American Political Science Review 89(3), pp. 634-647.

Beenstock, M. and Felsenstein, D. (2019). The Econometric Analysis of Non-Stationary Panel Data. Springer: Switzerland.

Chang, T. and Chen, W. (2017). Revisiting the relationship between suicide and unemployment: evidence from linear and non-linear cointegration. Economic Systems 41(2), pp. 266-278.

Cheng, K., Durmaz, N., Kim, H. and Stern, M. (2012). Hysteresis vs. the natural rate of US unemployment. Economic Modeling 29(2), pp. 428-434.

Chudik, A. and Pesaran, M. (2015). Common correlated effects estimation of hetergeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, pp. 393-420.

Curtin, S., Garnett, M. and Ahmad, F. (2022). Provisional numbers and rates of suicide by month and demographic characteristics: United States, 2021. Vital Statistics Rapid Release Number 24, Report 120830.

Curtin, S. and Hedegaard, H. (2019). Suicide rates for females and males by race and ethnicity: United States, 1999 and 2017. National Center for Health Statistics, Health E-Stat, June edition.

Douglas, J. (1967). The Social Meaning of Suicide. Princeton, NJ: Princeton University Press.

Durkheim, E. (1897/1951). Suicide. Translated by Spaulding, J. and Simpson, G. Glencoe, IL: Free Press.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time-series with a unit root. Journal of the American Statistical Association 74, pp. 427-431.

Dickey, D. and Fuller, W. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, pp. 1057-1072.

Engle, R. and Granger, C. (1987). Co-integration and error correction: representation, estimation and testing. Econometrica 55, pp. 251-276.

Fedden, H. (1938). Suicide: A Social and Historical Study. London: Peter Davies.

Feng, S. and Hu, Y. (2013). Misclassification errors and the underestimation of the U.S. unemployment rate. American Economic Review 103(2), pp. 1054-1070.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, pp. 424-438.

Hamermesh, D. and Soss, N. (1974). An economic theory of suicide. Journal of Political Economy 82(1), pp. 83-98.

Hatzius, J. and Stehn, S. (2012). Comment on: The ins and outs of forecasting unemployment: using labor force flows to forecast the labor market, by Barnichon, R. and Nekarda, C. Brookings Papers on Economic Activity 45(2), pp. 118-123.

Hausman, J. (1978). Specification tests in econometrics. Econometrica 46(6), pp. 1251-1271.

Henry, A. and Short, J. (1954). Suicide and Homicide. Glencoe, IL: Free Press.

Holderness, C. (2016). Problems using aggregate data to infer individual behavior: Evidence from law, finance, and ownership concentration. Critical Finance Review 5(1), pp. 1-40.

Im, K., Pesaran, M. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, pp. 53-74.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. Journal of Econometrics 90, pp. 1–44.

Laanani, M., Ghosn, W., Jougla, E. and Rey, G. (2015). Impact of unemployment variations on suicide mortality in Western European countries (2000-2010). Journal of Epidemiology and Community Health 69, pp. 103-109.

Levin, A., Lin, CF., and Chu, CS. (2002). Unit root tests in panel data: asymptotic and finite - sample properties. Journal of Econometrics 108(1), pp. 1-24.

Lin, Y.H. and Chen, W.Y. (2018). Does unemployment have asymmetric effects on suicide rates? Evidence from the United States: 1928-2013. Economic Research - Ekonomska Istraživanja 31(1). pp. 1404-1417.

Milner, A., Page, A. and LaMontague, A. (2014). Cause and effect in studies on unemployment, mental health and suicide: a meta-analytic and conceptual review. Psychological Medicine 44, pp. 909-917.

Newey, W. and West, K. (1987). A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, pp. 150-161.

O'Connell, P. (1998). The overvaluation of purchasing power parity. Journal of International Economics 44, pp. 1-19.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. Oxford Bulletin of Economics and Statistics 61, pp. 653–670.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. Econometric Theory 20, pp. 597–625.

Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. University of Cambridge Working Papers in Economics No. 0435.

Pesaran, M. (2007). A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics 22, pp. 265-312.

Pesaran, M. (2015). Testing weak cross-sectional dependence in large panels. Econometric Reviews 34, pp. 1089-1117.

Pesaran, M. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, pp. 79-113.

Pesaran, M., Shin, Y. and Smith, R. (1999). Pooled mean group estimation of dynamic heterogeneous panels. Journal of the American Statistical Association 94(446), pp. 621-634.

Phillips, P. (1987). Time series regression with a unit root. Econometrica 55(2), pp 277-301.

Phillips, P. and Moon, H. (1999). Linear regression limit theory for nonstationary panel data. Econometrica 67, pp. 1057-1111.

Phillips, P. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. Econometrica 58, pp. 165–193.

Phillips, P. and Perron, P. (1988). Testing for a unit root in time series regressions. Biometrika 75, pp. 335–346.

Stock, J. and Watson, M. (1988). Testing for common trends. Journal of the American Statistical Association 83, pp. 1097-1107.

Vandoros, S. and Kawachi, I. (2021). Economic uncertainty and suicide in the United States. European Journal of Epidemiology 36, pp. 641-647.

Warshauer, M. and Monk, M. (1978). Problems with suicide statistics for whites and blacks. American Journal of Public Health 68(4), pp. 383-388.

Westerlund, J. and Breitung, J. (2013). Lessons from a decade of IPS and LLC. Econometric Review 32, pp. 547-591.

Appendix

Table A1: CS-ARDL results

	CS-ARDL	
Variable	**Coefficient**	**P-value**
Long-Run		
LN Unemployment	0.013	0.452
Short-Run		
Constant	0.502	0.000
ΔLN Suicide(-1)	-0.282	0.000
ΔLN Suicide(-2)	-0.091	0.000
ΔLN Unemployment	0.035	0.068
ΔLN Unemployment(-1)	0.071	0.187
ΔLN Unemployment(-2)	-0.002	0.760
Cointegration variable ϕ	-0.207	0.101
Wald, F-test[a]	1.98	0.114

[a]The null hypothesis of the Wald test: Unemployment does not Granger (1969) cause Suicide (short-run);
H_0: ΔlnUnemp = ΔlnUnemp(-1) = ΔlnUnemp(-2) = 0.
Akaike information criterion (AIC) lag selection: 3 lags (*p* & *q*).

Data issues

The lack of accuracy of suicide statistics is well documented (Douglas 1967; Hamermesh and Soss 1974; Warshauer and Monk 1978; Curtin and Hedegaard 2019). Underreporting or misclassified death certificates appear to be the main source of the deficiencies. Some argue that the undercounting could be as much as one-third overall (Warshauer and Monk 1978). Attitudes, held by informants and examiners, toward suicide seem to be the main influence of the errors in the variable. These errors may imply that any cross-country analyses of suicide rates are likely "worthless" (Hamermesh and Soss 1974). Within the U.S., suicide rates by jurisdiction also suffer from these errors in the left-hand-side (LHS) which could easily be correlated with regressors and various unobserved factors. Moreover, state-level unemployment rates are not derived without their share of measurement error. Official unemployment rates, by jurisdiction, have been shown to substantially underestimate the true rate due to misclassification errors in the labor force participation denominator (Feng and Hu 2013). Errors in the RHS produce inconsistent estimates denoted as 'least squares attenuation'. If all of this wasn't enough, there is a vast general literature critical of the use of aggregate data to explain heterogeneous individual occurrence (see Holderness 2016 for an excellent review). Statistical properties and the biases introduced by using aggregated per capita or averaged data have yet to be adequately explained. Aggregation defects likely plague the ecological literature cited in the introduction section above and unfortunately the examination herein. Whether there is a solution to these deleterious data issues is perhaps a blurred question, is the cat not already out of the bag?