Journal of Statistical and Econometric Methods, Vol.11, No.3, 2022, 13-29
ISSN: 2241-0384 (print), 2241-0376 (online)
https://doi.org/10.47260/jsem/1132
Scientific Press International Limited

The Tenuous Ecological Divorce and Unemployment Link with Suicide: A U.S. Panel Analysis 1968-2020

Mitch Kunce[1]

Abstract

In 2020, close to 46,000 people died from suicide in the U.S. Globally, rates of suicide have declined some in the last 20 years − not so for the U.S. In the last two decades, the U.S. has seen more than a 25% increase in age-adjusted suicide rates. Recent reviews of the sociologically grounded ecological studies of suicide find jurisdictional divorce and unemployment rates to be key suicide risk factors. However, a new vein of this literature is beginning to scrutinize long-established ecological links based on faulty statistical methodologies that previously ignored variable non-stationarity and the lack of series cointegration. The purpose herein is to fully dissect the tenuous ecological relationships between U.S. annual divorce rates, unemployment rates and suicide rates using a 53 year non-stationary panel of all 50 states and the District of Columbia. Results suggest no statistically imperious association, short-run or long-run, between suicide rates in the U.S. and the long-established risk factors divorce and unemployment rates. Implications of this dissection advocate for a shift in research focus to the individual − controlling for idiosyncratic specific-effects and key social processes.

JEL classification numbers: B55, C18, C49.
Keywords: Suicide rates, Cointegration, Ecological analysis, Non-stationary panel.

[1] DouglasMitchell Econometric Consulting Laramie, WY USA.

Article Info: *Received:* August 12, 2022. *Revised:* August 26, 2022.
Published online: September 1, 2022.

1. Introduction

A recent review of the suicide literature, examining suicide from the sociological perspective, identified divorce and unemployment as abiding key factors associated with increased suicide risk (Kołodziej-Sarzyńska et al. 2019).[2] The authors examined 469 peer-reviewed papers from the period 2000-2017 – eventually filtering the review's focus on 63 publications. The risk factors, divorce and unemployment, appeared most often as covariates in the papers reviewed.[3] In the past, Kunce and Anderson (2002) were critical of the documented ecological divorce-suicide link on methodological grounds. They argued that conventional single cross-section (time periods $T = 1$) or time-series (grouped units $N = 1$) specifications could neither identify nor control for unobserved group- and/or time-specific effects possibly correlated with divorce. Concurrently, Steven Stack (2002) reviewed 134 older studies including 795 findings on divorce and suicide characterizing the aggregate divorce-suicide association as potentially spurious and suggested a shift in focus away from the ecological to the individual. Regarding unemployment, Stack (2000) reviewed 130 articles and reports on economic factors and suicide – again critical of the purported aggregate unemployment-suicide association. Stack dismissed any cause and effect interpretation by past authors and concluded that the supposed ecological unemployment-suicide link was corrupted by multiple confounding factors, many simply unobservable.

As forwarded by Kołodziej-Sarzyńska et al. (2019), the new millennium started off with innovative ecological treatments attempting to explain the variation in aggregate suicide rates. The focus of this new vein of literature centered on the non-stationarity of the suicide rate time-series and, accordingly, potential right-hand-side (RHS) risk variables (divorce and unemployment included). The thought being that non-stationarity and lacking variable cointegration was the latent, underlying reason Steven Stack and others were so critical of the older ecological associations. Altinanahtar and Halicioglu (2009) claimed to be the first to empirically examine suicide time-series associations using a cointegration approach − ARDL autoregressive-distributed lag (Pesaran et al. 1998, 2001). Table 1 lists examples of this new vein of ecological suicide literature noting the type of data analyzed, T periods, method(s), and results for the divorce and/or unemployment covariates.[4] The table is representative of the overall state of this new vein, generally more focus on time-series specifications rather than pooled panel treatments. Perhaps this is a result of data availability or, more candidly, the novelty and inherent complexities involved in non-stationary panel data analysis, which we will explore below. In any case, newer results shown for divorce and unemployment significance in explaining suicide rate variation continue to be mixed at best.

[2] Milner et al. (2013) and Yip et al. (2015) also provide extensive reviews.

[3] Roughly 75% of the studies included both divorce and unemployment rates on the right-hand-side.

[4] Not intended as an exhaustive list.

Table 1: Cointegration ecological suicide studies

Paper	Data; *T* Periods; Method	Divorce	Unemployment
Altinanahtar and Halicioglu (2009)	Time series, Turkey; 34; ARDL	Insignificant	Not included
Inagaki (2010)	Time series, Japan; 57; FMOLS, DOLS	Insignificant	+ * Total sample
Andrés and Halicioglu (2010)	Time series, Denmark; 37; ARDL	Insignificant	+ * Total sample + * Males + * Females
Ceccherini-Nelli and Priebe (2011)	Time series, UK, US[5], France, Italy; 106, 98, 35, 32; Cointegration correlation	Not included	+ * Total sample UK + * Total sample US
Andrés et al. (2011)	Time series, Japan; 53; ARDL	+ * Total sample + * Males	+ * Females
Walsh and Walsh (2011)	Time series, Ireland; 42; Cointegrated AR(1) error	Not included	+ * Males
Okada and Samreth (2013)	Time series, 13 O.E.C.D.; 48 longest, 30 shortest; Country specific ARDL	+ * in 9 countries	Not included
Qiang Sun and Zhang (2016)	Time series, UK; 31; ARDL	+ * Total sample	+ * Total sample
Chang and Chen (2017)	Time series, US; 86; ARDL	Insignificant	+ * Total sample
Phiri and Mukuku (2020)	Time series, South Africa; 20; ARDL	+ * Total sample	Insignificant
================	==================	===========	===========
Matsubayashi and Ueda (2011)	21 O.E.C.D. panel; 25; Cointegrated two-way fixed effects	+ * Total sample + * Males	Insignificant
Kerr et al. (2011)	48 U.S. panel; 53; Differenced GLS	Insignificant	+ * Total sample + * Males + * Females
Barth et al. (2011)	18 country panel; 25; VEC one-way fixed effects	+ * Total sample + * Males + * Females	+ * Females
Jalles and Andresen (2015)	10 Canadian province panel; 9; Panel VAR, two-way fixed effects	Insignificant	+ * Total sample + * Females
(+ *) a positive significant relationship at conventional levels.			
(FMOLS) fully modified ordinary least squares; (DOLS) dynamic ordinary least squares; (GLS) generalized least squares; (VEC) vector error correction			

[5] Prior to 1946, data collection on deaths in the U.S. was the responsibility of the Bureau of the Census. These early data suffer from quality, conformity and completeness issues. For a historical treatment of the evolution of death data collection in the U.S., see Kunce (2022).

Within conventional panel analysis, the time dimension T is generally short. In this case, the time-series asymptotics of the data are generally a side-issue that is usually of little interest. Of late, so called 'macro' panels (large N and T) have risen to the forefront. So much so that limiting distributions of double indexed integrated processes had to be developed (Phillips and Moon 1999). The prospect of N and T being large has split the thinking on model estimation into two factions. The first is critical of pooling the data and promotes heterogeneous specifications for each grouped unit (see Baltagi 2021 for a review). Trend variables included in the RHS of these single time-series models can help control for latent factors that are time-varying within the grouped units. One drawback of these single unit specifications is that estimates critically rely on T being sufficiently large. Many of the time-series examples depicted in Table 1 appear to be deficient in this regard. The second faction champions applying time-series procedures to panels. Those in this camp are not resistant to pooling spatial data[6], extend as much heterogeneity as possible, pay close attention to non-stationarity and examine cointegration of all variable combinations (see Beenstock and Felsenstein 2019 for a review). The aim of 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 purpose of the examination herein is to fully dissect the tenuous ecological relationships between U.S. annual divorce rates, unemployment rates and suicide rates using a 53 year panel of all 50 states and the District of Columbia. The risk factors divorce and unemployment are emphasized due to their historical, well documented link with suicide. The focus, however, is rooted in statistical method and should not be considered a test of any specific suicide theory. In section 2 we will examine the underlying stochastic process generating each variable's series via unit root testing. Section 3 focuses on cointegration structures and tests for them. Section 4 presents results from three panel estimators for comparison and lastly section 5 concludes.

[6] Of course paying particular attention to cross-section correlation.

2. Panel Unit Root Test

Testing for unit roots in single equation, long time-series ($t = 1, \ldots, T$) specifications is now common practice among applied researchers. Single equation tests, however, suffer from a low power defect when examining shorter time spans. A popular remedy for this problem is to use panel unit root tests that augment power by exploiting cross-sectional information. Conventional panel unit root tests have been criticized, of late, for assuming that cross-section cointegrating relationships are not present (Westerlund and Breitung 2013). Assuming cross-section independence, when perhaps dependence is in play, tends to distort the size of the estimated test statistics that reject the null of non-stationarity too often. Pesaran (2007) proposes a test statistic for individual variable cross-section dependence,

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

where $\hat{\rho}_{ij}$ denotes the pair-wise correlation coefficient from the residuals of cross-sectioned ($i = 1, \ldots, N$) Augmented Dickey and Fuller (1979, 1981) regressions.[7] The *CD* statistic, testing the null of independence, is distributed asymptotically normal and possesses adequate small sample properties. Table 2 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, divorce rate and unemployment rate. Natural logarithmic transformations are used as a means to remove growth in variance, of all rates, over time. Test statistics shown are sufficient to reject the null hypothesis of cross-section independence at the < 10% level. Each separate variable, particularly divorce rates, could be characterized as mildly cross-sectional dependent.[8]

[7] For convenience, we may duplicate the use of some equation variable symbols as the paper proceeds. Consider similar symbols and definitions equation specific.

[8] Following Pesaran (2007) regarding the strict null hypothesis. Results are similar when the focus shifts to the 48 contiguous states plus DC.

Table 2: Data descriptions, sources, descriptive statistics and CD tests

Suicide Rate. Age Standardized per 100,000 total state population. National Center for Health Statistics, National Vital Statistics System, 1968-2020. Year 2000 age standardization applies. Mean 13.64, STD 3.89. Natural Log Suicide Rate, Mean 2.57, STD 0.28. Pesaran 2007 CD, 1.88 (p-value 0.0601).
Divorce Rate. Based on counts of divorce decrees granted by a state (of occurance) per 1,000 total state population. National Center for Health Statistics (NCHS), National Vital Statistics System, 1968-2020. California, Georgia, Hawaii, Indiana, Louisiana, Minnesota and New Mexico, have stopped or do not consistently report divorce occurrence to the NCHS. This missing data can be recovered from the U.S. Bureau of the Census, American Community Survey (ACS), Public Use Microdata Sample (PUMS). See Mayol-Garcia et al. (2021) for a similar application of ACS PUMS data. Mean 4.37, STD 1.82. Natural Log Divorce Rate, Mean 1.41, STD 0.35. Pesaran 2007 CD, 1.65 (p-value 0.0989).
Unemployment Rate. Seasonally adjusted average annual rates by state, in percent of labor force. U.S. Bureau of Labor Statistics, 1968-2020.[9] Mean 5.75, STD 2.07. Natural Log Unemployment Rate, Mean 1.69, STD 0.35. Pesaran 2007 CD, 1.94 (p-value 0.0524).

Faced with, though mild, cross-section 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.[10] 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},

$$S_{it} = D_{it} + \lambda_i' F_t + \eta_{it}, \qquad (2)$$

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 $\beta_i t$.

[9] 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.

[10] 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'.

Herein, factor selection follows the information criteria proposed by Bai and Ng (2002). Relative to the number of cross-sections (*N*) and time periods (*T*), the number of common factors are usually small. Multivariate common factors from equation (2) are tested using the modified version of the, more general, *Qc* test developed by Stock and Watson (1988). 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,

$$P_{\hat{\eta}} = \frac{-2\sum_{i=1}^{N} \ln\left(p_{\hat{\eta}_i}\right) - 2N}{2\sqrt{N}}, \tag{3}$$

where $p_{\hat{\eta}_i}$ denotes the probability values from the cross-sectioned Augmented Dickey-Fuller tests. It is important to note that tests on the common factors are asymptotically independent of tests on the idiosyncratic components. 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 3 provides the PANIC results for the natural log series of each variable separately.

Table 3: PANIC results

	LN Suicide		**LN Divorce**		**LN Unemployment**	
	Lags	**ADF Stat**	**Lags**	**ADF Stat**	**Lags**	**ADF Stat**
Alabama	1	-2.11	0	-1.10	0	-1.55
Alaska	2	-2.21	0	0.56	8	-0.56
Arizona	4	-0.89	0	-1.23	0	-1.14
Arkansas	3	-3.31***	7	-2.67***	0	-1.48
California	0	-1.31	2	0.58	1	-2.09**
Colorado	1	-1.41	10	1.24	0	-1.55
Connecticut	0	-2.71**	6	0.96	1	-3.06***
DC	4	-1.79	3	-1.72	0	-0.89
Delaware	1	-1.66	2	-0.34	0	-1.43
Florida	5	-0.20	9	1.49	1	-3.16***
Georgia	7	-1.50	2	0.12	0	-1.47
Hawaii	7	-0.77	3	-0.73	2	-2.08**
Idaho	1	-3.84***	3	-1.02	0	-1.43
Illinois	8	-1.79	0	1.19	0	-1.85*
Indiana	3	-1.07	5	0.19	0	-2.28**
Iowa	0	-1.31	0	-1.50	5	-1.53
Kansas	2	-1.42	2	-1.09	3	-0.07
Kentucky	1	-1.95	4	-1.20	9	-0.99
Louisiana	3	-1.14	8	5.12	2	-1.87*

Maine	1	-2.23	3	0.06	0	-1.01
Maryland	0	-4.36***	4	0.44	1	-1.85*
Massachusetts	1	-2.03	0	0.23	0	-2.29**
Michigan	1	-1.27	3	0.74	1	-2.21**
Minnesota	2	-2.30	2	0.67	3	-3.54***
Mississippi	2	-0.81	0	-1.10	0	-0.61
Missouri	1	-1.15	2	-0.60	2	-1.26
Montana	4	-0.99	0	0.30	1	-2.36**
Nebraska	0	-3.22***	1	1.33	3	-1.24
Nevada	4	-0.78	2	-1.99**	1	-0.43
New Hampshire	2	-1.44	5	0.30	1	-1.90*
New Jersey	7	-1.40	6	1.39	0	-2.74***
New Mexico	4	-0.90	5	-3.13***	0	-1.54
New York	7	-2.30	10	0.93	0	-3.39***
North Carolina	9	-2.28	7	-0.73	2	-0.98
North Dakota	4	-0.85	0	0.30	8	-1.10
Ohio	1	-1.30	0	-1.76	0	-2.09**
Oklahoma	3	-1.68	2	0.50	5	-1.30
Oregon	4	-0.72	0	-0.41	2	-1.52
Pennsylvania	1	-1.88	1	-0.99	1	-2.44**
Rhode Island	1	-1.05	3	-0.24	1	-1.60
South Carolina	1	-2.12	0	-0.14	0	-1.39
South Dakota	4	-1.29	1	-0.69	1	-1.59
Tennessee	3	-1.03	0	-1.51	0	-2.26**
Texas	9	-2.30	1	2.26	0	-0.88
Utah	7	-0.85	0	-1.23	0	-0.37
Vermont	2	-1.23	1	-1.11	0	-0.68
Virginia	10	-1.03	7	-0.91	1	-1.10
Washington	1	-2.16	0	1.25	0	-2.38**
West Virginia	0	-3.77***	7	0.08	1	-0.75
Wisconsin	3	-1.26	10	1.01	0	-1.59
Wyoming	2	-1.04	0	-1.92*	0	-2.09**
Null Rejections		6		4		20
Common Factors	7	68.58	7	12.07	6	14.63
Idiosyncratic		3.14***		0.02		11.74***
Trend assumption: Deterministic intercept and trend.						
Significance at 1% (***), 5% (**), 10% (*) levels. ADF - augmented Dickey-Fuller.						

Regarding the idiosyncratic components, individual cross-section unit root tests show null hypothesis (unit root) rejections in 6 states for suicide rates, in 4 states for divorce rates and in 20 states for unemployment rates. The last row of Table 3 shows the pooled idiosyncratic component test (equation (3)) for each panel variable. 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 cross-sections exists. As such, the pooled test mirrors a panel test for no cointegration among all cross-sections for each variable separately. The pooled statistic rejects no cointegration among cross-sections for suicide and unemployment rates separately. Divorce rates are consistent with pooled idiosyncratic non-stationarity.[11]

As depicted near the bottom of Table 3, multiple common factors are determined for each panel variable. In the multivariate common factor case ($r > 1$) testing proceeds using an iterative procedure. Herein, we apply the more general *MQc* test which corrects for serial correlation, of arbitrary form, through non-parametric estimation. *MQc* parallels the multivariate procedure suggested by Phillips (1987). The null hypothesis states that r common factors have at most r common stochastic trends. As in our case for each variable, failure to reject the null of retaining the common factors indicates that *all* common factors are non-stationary. Overall, results are consistent with non-stationarity in U.S. suicide rates, divorce rates and unemployment rates, all pervasive in the common factors and finite in the idiosyncratic components.[12]

3. Cointegration

Generally, non-stationary variables should not be used in conventional regression models. Most of the criticism of the earlier ecological suicide work where T is sufficiently large, as reviewed by Steven Stack and others, may stem from the non-stationarity of these three key variables. However, what if there is a linear combination of suicide rates on divorce rates and unemployment rates that is stationary? In other words, the errors of this combination are integrated of order zero, I(0). Series that satisfy this requirement are desired and defined as cointegrated (Engle and Granger 1987). A natural first step in the analysis of cointegration is to establish that it is indeed a characteristic of the data. As in the unit root analysis above, cross-section dependence plays an important role in cointegration testing. Pesaran (2004, 2015) suggests a cross-section dependence test similar to equation (1) but based on the average of pair-wise correlation coefficients of the residuals from a pooled panel regression. In order to obtain these residuals, we regress log suicide rates on log divorce rates and log unemployment rates in a simple panel least squares specification.[13] The Pesaran CD test statistic is 1.772 with a p-value of 0.076 which is consistent with weak cross-sectional dependence (Pesaran 2015).[14]

---

[11] This does not imply that all idiosyncratic components are non-stationary. The failure to reject the null implies a finite number of non-stationary components.

[12] All first differenced tests confirm I(0) when differenced, stationarity.

[13] 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.

[14] This test statistic ($N = 51$) fails to reject the null of weak cross-sectional correlation in the residuals at the < 5% level and of course rejects at the < 10% level. The test statistic and (p-value) for the 48 contiguous states and DC are 1.891 (0.059).

In light of what we consider weak cross-sectional error correlation, we proceed to test the panel for cointegration following Pedroni (1999, 2004).[15] Pedroni's 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 4 presents the relevant cointegration tests for the natural log transformed variable suite.[16]

Table 4: Pedroni residual cointegration tests

Observations: 2,703		
Cross-sections included: 51		
Null Hypothesis: No cointegration		
Trend assumption: Deterministic intercept and trend		
Automatic lag length selection based on Akaike information criterion with a max lag of 10		
Newey-West automatic bandwidth selection and Bartlett kernel		
LNSUICIDE on LNDIVORCE, LNUNEMPLOYMENT		
Within-dimension common AR	**Statistic**	**p-value**
Panel: rho	-1.459	0.072
Panel: t	-1.639	0.051
Panel: Augmented Dickey - Fuller	-0.574	0.283
Between-dimension individual AR	**Statistic**	**p-value**
Group: t	-1.229	0.109
Group: Augmented Dickey - Fuller	-0.373	0.355

Note that the 'group rho' along with the 'panel v' (variance ratio) test statistics are not included in the table. 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 50, that the two statistics excluded suffer size distortions that lead to persistent

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

[16] A reviewer, on an earlier version of this paper, requested Pedroni tests on the log linear form and comparative Kao (1999) ADF test statistics. These analogous results are found in the Appendix.

failure to reject the null in simulation.[17] This was the case for all specifications we tested. Pedroni makes the point, for short 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.[18] This latter finding from Pedroni's experiment is the reason we focus on the t-statistic results in Table 4. In our case, the 'panel t' statistic is sufficient to reject the null at the < 10% level while the 'group t' statistic fails to reject at conventional levels. In light of the Pedroni Monte Carlo results, we are on the inference fence − leaning toward a conclusion of no cointegration. In simulation, the t-statistics soundly reject due to the size distortion when T = 50, herein they struggle to do so.

4. Panel Estimation

With unit root presence in each series, evidence of weak cross-section dependence and a conclusion of no cointegration of the variable suite, we follow Pesaran et al. (1999) and estimate a panel vector autoregressive (VAR) model. This specification does not rely on the cointegration of the non-stationary variables. Given data on states, $i = 1, 2, ..., N$ and years $t = 1, 2, ..., T$, the model becomes,

$$S_{it} = \alpha + \sum_{j=1}^{p} \lambda_{ij} S_{i,t-j} + \sum_{j=0}^{q} \delta'_{ij} X_{i,t-j} + \mu_i + \gamma_t + \varepsilon_{it}, \quad (4)$$

where S_{it} is the natural log of suicide rates, α is a scalar constant, 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 regressors (divorce and unemployment) and their respective lags, q is the lag selection for the regressors, μ_i are the state specific effects, γ_t are the year specific effects and ε_{it} denotes the remainder disturbance. Generally, two specifications of equation (4) are considered. Fixed effects treats μ_i and γ_t as fixed yet unknown constants differing across states and over time. Alternatively, random effects assumes that μ_i and γ_t are random, distributed independently across states and over time. The potential correlation of μ_i and γ_t with the dependent variable lags and variables in $X_{i,t-j}$ is a primary consideration. If these correlations are present, random effects estimation yields biased and inconsistent estimates of λ and δ and the variances of μ_i , γ_t and ε_{it} . In contrast, the fixed effects estimator is not impacted by this lack of orthogonality but is not fully efficient (in certain cases) since it ignores variation across states and over time. The choice of estimator generally rests on statistical considerations and hypothesis testing. Hausman (1978) outlines a specification test of the null hypothesis of orthogonality between the latent effects and regressors. The large

[17] The panel 'rho' statistic is shown to be somewhat neutral regarding size distortion for small T.
[18] Pedroni shows that when the T dimension reaches 150, the size distortion of the statistics wane.

Hausman statistics shown in Table 5 below favor the fixed-effects VAR specification.

Results for the panel VAR estimation are depicted in Table 5. Given the evidence of unit root and no cointegration for this particular variable combination, we include spurious panel least squares and conventional two-way fixed-effects estimates for comparison only.

Table 5: Regression comparisons

	Two-way Panel VAR		**Panel Least Squares**		**Two-way FE**	
Variable	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
Constant	0.595	0.000	2.139	0.000	2.322	0.000
LNDivorce	0.123	0.072	0.361	0.000	0.134	0.057
LNUnemployment	0.002	0.921	0.103	0.069	0.037	0.159
LNSuicide(-1)	0.381	0.000				
LNSuicide(-2)	0.366	0.000				
LNDivorce(-1)	-0.081	0.091				
LNDivorce(-2)	0.001	0.976				
LNUnemployment(-1)	0.026	0.460				
LNUnemployment(-2)	-0.028	0.285				
Observations	2601		2703		2703	
Akaike info criterion	2 lags					
Hausman (df) FE vs RE	171.06***(8)				106.63***(2)	
F-test State Effects (df)	4.035*** (50,2492)				117.485*** (50,2598)	
F-test Period Effects (df)	7.467*** (50,2492)				25.516*** (52,2598)	
Note: All models estimated with robust panel corrected standard errors (PCSE) and covariance, Beck and Katz (1995). *** significance at the < 1% level.						

For the panel VAR results, Granger (1969) causality null hypotheses become: divorce does not Granger cause suicide; unemployment does not Granger cause suicide. A Wald test of each regressor and its respective lags, jointly equaling zero, fail to reject each null at the < 5% level.[19] Post rigorous dissection, we find no statistically imperious association, short-run or long-run, between suicide rates in the U.S. and the long-established risk factors divorce and unemployment rates.[20] In contrast, the spurious panel least squares specification finds positive statistically significant associations while the spurious two-way fixed effects estimator links divorce and suicide only (at conventional levels). Results highlight the important

---

[19] H_0: LNDiv = LNDiv(-1) = LNDiv(-2) = 0, Chi-square 6.496, p-value 0.089. H_0: LNUnemp = LNUnemp(-1) = LNUnemp(-2) = 0, Chi-square 1.348, p-value 0.717.

[20] We are not ruling out that some other, what John Hood-Williams (1996) would call 'jumbled', suite of regressors, including divorce and unemployment, would yield differing results.

role that series non-stationarity and cointegration play when searching for short-run and long-run variable links.

5. Concluding remarks

According to recent data from the World Health Organization, roughly 700,000 people take their own lives each year (W.H.O. 2021). In 2020, close to 46,000 died from suicide in the U.S. alone. Globally, rates of suicide have declined some in the last 20 years − not so for the U.S. In the last two decades, the U.S. has seen more than a 25% increase in age-adjusted suicide rates. Recent reviews of the sociologically grounded ecological studies of suicide find jurisdictional divorce and unemployment rates to be key suicide risk factors − perhaps even influencing the recent rise in suicide rates in the U.S. However, a new vein of this literature is beginning to scrutinize long-established ecological links arguing faulty statistical methodologies. It is clear, going forward, that series non-stationarity and cointegration must not be ignored by those persistent on using ecological designs. Moreover, the use of non-stationary panel data methods may provide advantages when N with T are relatively large and normal limiting distributions are desired.

Notwithstanding, there is a vast general literature critical of the use of aggregate data to explain heterogeneous individual occurrence (see Holderness 2016 for a 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 the examination herein. Sociological approaches to the study of suicide appear preoccupied with the potential association or link of aggregate risk factors to suicide rates. Finding these tenuous ecological links, arguably, does little to advance the understanding of the individual act of suicide. Practical implications of this paper (again echoing Steven Stack) call for a shift in focus to a smaller unit of analysis. Preference is given to individual-level specifications controlling for socioeconomic factors and individual specific effects (e.g. childhood circumstances and mental health). Incorporating sociological impacts into individual specific analysis may further the understanding of the complexities of suicide and perhaps assist in how to more effectively intervene.

References

[1] Kołodziej-Sarzyńska, M., Majewska, M., Juchnowicz, D. and Karakuła-Juchnowicz, H. (2019). Risk factors of suicide with reference to the theory of social integration by Émile Durkheim. Psychiatria Polska 53(4), pp. 865–881.

[2] Milner, A., Page, A. and LaMontague, A. (2013). Long-term unemployment and suicide: a systematic review and meta-analysis. PLoS ONE, 8(1), paper 51333.

[3] Yip, P., Yousuf, S., Chan, CH., Yung, T. and Wu, K. (2015). The roles of culture and gender in the relationship between divorce and suicide risk: a meta analysis. Social Science and Medicine 128, pp. 87-94.

[4] Kunce, M. and Anderson, A. (2002). The impact of socioeconomic factors on state suicide rates: a methodological note. Urban Studies 39(1), pp. 155-162.

[5] Stack, S. (2002). Quantitative suicidology: individual and aggregate level approaches. Archives of Suicide Research 6(1), pp. 61-67.

[6] Stack, S. (2000). Suicide: a 15-year review of the sociological literature part I. Suicide and Life-Threatening Behavior 30(2), pp. 145-162.

[7] Altinanahtar, A. and Halicioglu, F. (2009). A dynamic econometric model of suicides in Turkey. The Journal of Socio-economics 38(6), pp. 903-907.

[8] Pesaran, M. and Shin, Y. (1998). An autoregressive distributed-lag modeling approach to cointegration analysis. Econometric Society Monographs 31, pp. 371- 414.

[9] Pesaran, M., Shin, Y. and Smith, R. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics 16(3), pp. 289- 326.

[10] Inagaki, K. (2010). Income inequality and the suicide rate in Japan: evidence from cointegration and La-var. Journal of Applied Economics 13(1), pp. 113-133.

[11] Andrés, A. and Halicioglu, F. (2010). Determinants of suicide in Denmark: evidence from time series data. Health Policy 98(2-3), pp. 263-269.

[12] Ceccherini-Nelli, A. and Priebe, S. (2011). Economic factors and suicide rates: associations over time in four countries. Social Psychiatry and Psychiatric Epidemiology 46, pp. 975-982.

[13] Kunce, M. (2022). Non-stationarity in U.S. all-cause mortality rates: probing the usefulness of the idiom 'excess death'. European Journal of Statistics and Probability 10(2), pp. 1-13.

[14] Andrés, A., Halicioglu, F. and Yamamura, E.(2011). Socio-economic determinants of suicide in Japan. The Journal of Socio-economics 40, pp. 723-731.

[15] Walsh, B. and Walsh, D. (2011). Suicide in Ireland: the influence of alcohol and unemployment. The Economic and Social Review 42(1), pp. 27-47.

[16] Okada, K. and Samreth, S. (2013). A study on the socio-economic determinants of suicide: evidence from 13 European O.E.C.D. countries. The Journal of Socio-economics 45, pp. 78-85.

[17] Qiang Sun, B. and Zhang, J. (2016). Economic and sociological correlates of suicides: multilevel analysis of the time series data in the U.K. Journal of Forensic Sciences 61(2), pp. 345-351.
[18] 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.
[19] Phiri, A. and Mukuku, D. (2020). Does unemployment aggravate suicide rates in South Africa? Some empirical evidence. Review of Social Economy 78(4), pp. 532-560.
[20] Matsubayashi, T. and Ueda, M. (2011). The effect of national suicide prevention programs on suicide rates in 21 O.E.C.D. nations. Social Science and Medicine 73, pp. 1395-1400.
[21] Kerr, W., Subbaraman, M. and Ye, Y. (2011). Per capita alcohol consumption and suicide mortality in a panel of U.S. states from 1950-2002. Drug and Alcohol Review 30(5), pp. 473-480.
[22] Barth, A., Sögner, L., Gnambs, T., Kundi, M., Reiner, A. and Winker, R. (2011). Socioeconomic factors and suicide: an analysis of 18 industrialized countries for the years 1983-2007.Journal of Occupational and Environmental Medicine 53(3), pp. 313-317.
[23] Jalles, J. and Andresen, M. (2015). The social and economic determinants of suicide in Canadian provinces. Health Economics Review 5, pp. 1-12.
[24] Phillips, P. and Moon, H. (1999). Linear regression limit theory for nonstationary panel data. Econometrica 67, pp. 1057-1111.
[25] Baltagi, B. (2021). Econometric Analysis of Panel Data. Sixth Edition. Chapter 12. Springer: Switzerland.
[26] Beenstock, M. and Felsenstein, D. (2019). The Econometric Analysis of Non-Stationary Panel Data. Springer: Switzerland.
[27] Westerlund, J. and Breitung, J. (2013). Lessons from a decade of IPS and LLC. Econometric Review 32, pp. 547-591.
[28] Pesaran, M. (2007). A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics 22, pp. 265-312.
[29] 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.
[30] Dickey, D. and Fuller, W. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, pp. 1057-1072.
[31] Mayol-Garcia, Y., Gurrentz, B. and Kreider, R. (2021). Number, timing, and duration of marriages and divorces. Current Population Report. U.S. Census Bureau.
[32] Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. Econometrica 72(4), pp. 1127-1177.
[33] Bai, J. and Ng, S. (2010). Panel unit root tests with cross-section dependence: A further investigation. Econometric Theory 26, pp. 1088–1114.

[34] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica 70(1), pp. 191-221.
[35] Stock, J. and Watson, M. (1988). Testing for common trends. Journal of the American Statistical Association 83, pp. 1097-1107.
[36] Phillips, P. (1987). Time series regression with a unit root. Econometrica 55(2), pp 277-301.
[37] Engle, R. and Granger, C. (1987). Co-integration and error correction: representation, estimation and testing. Econometrica 55, pp. 251-276.
[38] Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. University of Cambridge Working Papers in Economics No. 0435.
[39] Pesaran, M. (2015). Testing weak cross-sectional dependence in large panels. Econometric Reviews 34, pp. 1089-1117.
[40] Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. Oxford Bulletin of Economics and Statistics 61, pp. 653–670.
[41] 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.
[42] Phillips, P. and Perron, P. (1988). Testing for a unit root in time series regressions. Biometrika 75, pp. 335–346.
[43] Phillips, P. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. Econometrica 58, pp. 165–193.
[44] Kao, C. (1999) Spurious regression and residual-based tests for cointegration in panel data. Journal of Econometrics 90, pp. 1–44.
[45] 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.
[46] Hausman, J. (1978) Specification tests in econometrics. Econometrica 46(6), pp. 1251-1271.
[47] 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.
[48] Granger, C. (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, pp. 424-438.
[49] Hood-Williams, J. (1996). Studying suicide. Health & Place 2(3), pp. 167-177.
[50] World Heath Organization. (2021) Suicide worldwide in 2019: global health estimates. Geneva: World Health Organization.
[51] 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.

Appendix

Table A1: Kao residual cointegration tests

Observations: 2,703		
Cross-sections included: 51		
Null Hypothesis: No cointegration		
Trend assumption: Deterministic intercept		
Automatic lag length selection based on Akaike information criterion with a max lag of 10		
Newey-West automatic bandwidth selection and Bartlett kernel		
	ADF Statistic	**p-value**
Log Linear Specification	1.374	0.085
Log Log Specification	1.398	0.081

Table A2: Pedroni residual cointegration tests

Observations: 2,703		
Cross-sections included: 51		
Null Hypothesis: No cointegration		
Trend assumption: Deterministic intercept and trend		
Automatic lag length selection based on Akaike information criterion with a max lag of 10		
Newey-West automatic bandwidth selection and Bartlett kernel		
LNSUICIDE on DIVORCE, UNEMPLOYMENT		
Within-dimension common AR	**Statistic**	**p-value**
Panel: rho	-1.387	0.083
Panel: t	-1.575	0.058
Panel: Augmented Dickey - Fuller	-0.776	0.219
Between-dimension individual AR	**Statistic**	**p-value**
Group: t	-1.118	0.132
Group: Augmented Dickey - Fuller	-0.292	0.385