

Actual versus ‘Natural’ Rates of Suicide: Evidence from the USA

Abstract

This paper employs a high dimensional variable selection technique to select a subset of suicide determinants from 167 potential factors, which are then used to estimate ‘natural’ suicide rates for US states by least squares dummy variables. Over the period 2005-2017, all states are found to have a non-zero and positive natural suicide rate, below their respective actual average rate. Higher actual rates suggest deterioration in socioeconomic conditions; inaccessible and unaffordable mental health care for certain sections of the population and inadequate implementation of measures to identify and reduce suicidal mortality. Evidence-informed policies aiming for zero-suicide target could draw inspiration from exemplar states, to direct resources towards states with greater relative differences between actual average and natural suicide rates.

Keywords: Zero suicide policy; High dimensional panel data; Natural rate of suicide; USA

JEL codes: C23, I12

1. Introduction

Suicide is recognised as a serious mental and public health issue. It inflicts emotional burdens on surviving families and causes economic losses estimated at billions of dollars per year. Suicide cuts through all regions and socio-demographic levels across the globe. In a recent systematic analysis, (Naghavi, 2019) found that the total number of deaths from intentional self-inflicted injury rose by 6.7% globally between 1990 and 2016, the age-adjusted global mortality rate due to suicide decreased by a third (33%) over that time.

In contrast to these global trends, the age-adjusted suicide rate in the USA has been on the rise. Over the period 1999-2017, age-adjusted suicide rate rose by 33%, from 10.5 per 100,000 standard population to 14.0, with around 47,000 lives lost to suicide in the USA in 2017 alone¹. In line with the 2012 National Strategy for Suicide Prevention, the Federal government is committed to develop, implement and evaluate suicide prevention strategies to reverse rising suicide rates. As part of this strategy, the zero-suicide concept, embodying an aspirational goal of zero suicides, was initiated. The model was premised on the belief that any suicide mortality by in-patients or out-patients in health and behavioural health care systems is preventable by systematic improvement to healthcare processes via quality improvement. It is difficult to establish unambiguously that a zero suicide-like policy reduces suicide mortality, even for just those individuals in behavioural healthcare. However, pure reliance on established regional/national statistics is also problematic, since any such data will also feature suicides by individuals with no or very limited engagement with formal healthcare systems. Thus, in terms of helping to assess broad regional and national progress towards a zero suicide policy target, it is reasonable to explore what non-zero value in the suicide rate is consistent with zero suicide policy.

The contribution of our paper is to provide an estimate of such a non-zero value (or natural rate) of suicide for the USA at state level, using data from 2005 to 2017. The concept of a ‘natural rate of suicide’, i.e the societal suicide rate that would prevail even if social and economic conditions were all ideal (for example, no divorce, obesity, unemployment, etc) was first explored by Yang & Lester (1991)Click or tap here to enter text.. They asserted that the societal suicide rate can never be zero, even if socioeconomic characteristics were conducive to a suicide-free society. We re-examine Yang and Lester’s (1991) proposition that natural suicide rates are non-zero and positive. The novelty of our paper is that we employ a high dimensional variable selection technique to tease out a subset of covariates amongst 167 potential determinants of suicide obtained from the literature, as opposed to pre-selecting variables to include in the model. This method offers more accurate estimates of natural suicide rates as variables included in the model are chosen through a data-driven process. As observed by Yang and Lester (1991), author-

¹ National Center for Health Statistics. Public-use data files: Mortality multiple cause files. Available from: https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple.

determined suites of regressors may yield a zero or even negative value potentially reversing the natural rate conjecture. Moreover, multiple regression models with preselected variables often include predictors not necessarily related to the response variable, resulting in unnecessary model complexity (Hastie et al., 2009).

There is a growing literature on variable selection and inference in a high dimensional setting (see, for example, Belloni et al., 2012; Belloni et al., 2014 for theoretical explanation and Hoff et al., 2009; Panagiotidis et al., 2018; (Cui et al., 2020) for empirical work). Due to its computational efficiency, Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Tibshirani (1996) has been useful in high dimensional least squares estimations (Belloni et al., 2016). LASSO is superior to other variable selection methods such as Subset selection and Ridge regression as it produces models that are more interpretable and stable (Tibshirani, 1996).

We follow the method discussed in Belloni et al. (2016) and use Cluster-LASSO, a variant of LASSO that accommodates a clustered covariance structure. Belloni et al. (2016) was the first paper that addressed within group correlation and heteroskedasticity in a fixed effect panel data setting. Once the suite of regressors is selected through Cluster-LASSO, we use the Least Squares Dummy Variable (LSDV) estimator to obtain the natural rate of suicide for each state. This enables comparison of actual and natural suicide rates and offers a means of evaluating progress in counter-suicide initiatives and policies.

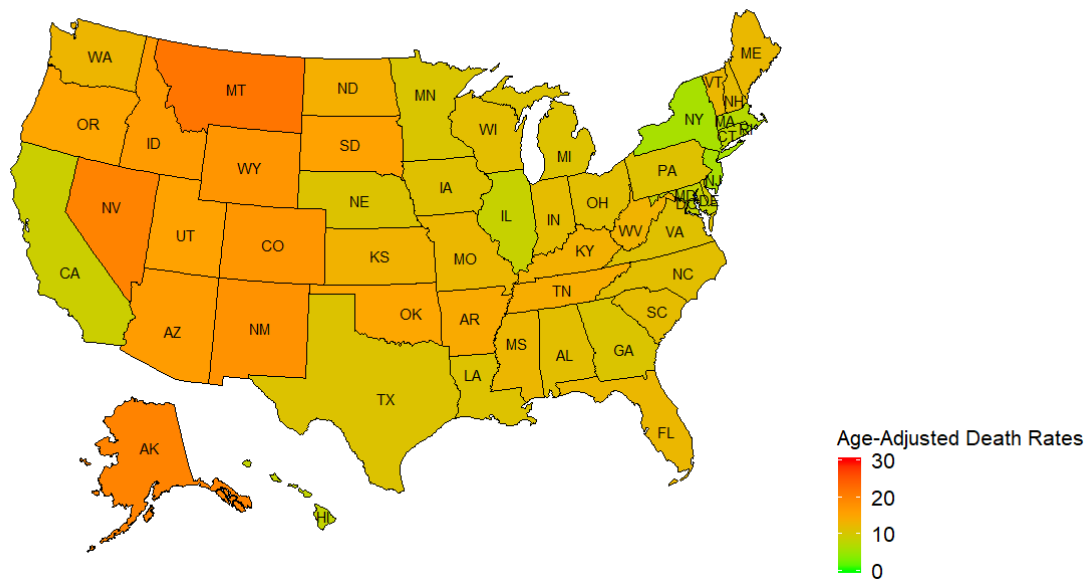
The rest of the paper is set out as follows: Section 2 presents a brief background on suicide rates across the USA. Section 3 briefly reviews the relevant literature. Data and methodology are discussed in Section 4. Estimation results and discussion are presented in Section 5. Section 6 sets out robustness checks and Section 7 concludes.

2. Background on Suicide Rates across the USA

From the mid-1980s, suicide rates in the USA dropped from 12.4 per 100,000 population (in 1985) to 10.4 per 100,000 population in 2000, with majority of states registering declines. Since then, there has been a noticeable reversal in this trend. As shown in Figures 1 and 2, between 2005 and 2017, nearly every state registered an increase in suicide rate, ranging from 2.53% in Nevada to 114.81% in Delaware. Alarming, thirty-two states (or 64%) experienced a rise of over 30%.

Figure 1: Age-adjusted suicide rates per 100,000 population for the USA, 2005

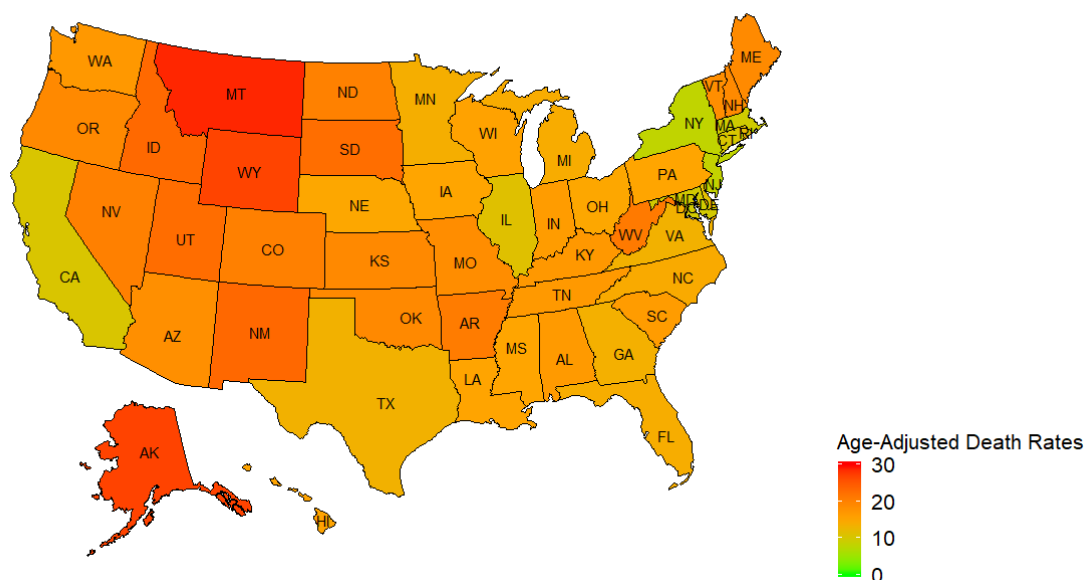
Suicide Mortality by State: 2005



Authors' computation. Data source: WISQARS, Injury Mortality Reports, Centers for Disease Control. www.cdc.gov

Figure 2: Age-adjusted suicide rates per 100,000 population for the USA, 2017

Suicide Mortality by State: 2017



Authors' computation. Data source: WISQARS, Injury Mortality Reports, Centers for Disease Control. www.cdc.gov

States in the west and mid-west regions exhibited the most significant increase in suicide rates over 2005-2017 period. Concurrently, states in the east such as Virginia and North Carolina, have seen the lowest rises over time. Mirroring the large geographic variation in rates of change of suicide, there was an equally sizeable geographic difference in the rates of death from self-harm in the USA. Montana had the highest number of suicides in 2005 at 21.7 per 100,000 which increased to 28.9 in 2017. The lowest recorded figure was for District of Columbia. The mountainous mid-western region of America is home to a collection of states, sometimes referred to as the 'Suicide Belt'. This area consists of 9 states, Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah and Wyoming. These states are geographically close and have consistently higher rates of suicide than most other states.

3. Suicide and the Natural Rate of Suicide: A Brief Retrospect

There is a rich body of theoretical and empirical work examining the motivations of suicides (see, for example, Platt, 1984; Chen et al., 2012; Vondra et al., 2019). Hamermesh & Soss (1974) were the first to propose an economic explanation to suicide through the utility maximisation framework. In particular, when discounted lifetime utility dropped below a certain threshold, suicide became an attractive option. The real-option perspective on suicide was later considered by Dixit and Pindyck (1994). Suicide decisions have also been explained through the risk-taking implications of utility maximisation approach by Becker & Posner, 2004; real option approach and Knightian uncertainty by Miao & Wang (2011).

Empirical studies have typically employed regression analysis to find the determinants of suicide. Different measures of suicide rates have been used as the dependent variable and a range of estimation methods applied to different datasets at country, regional or group level. The set of socio-economic determinants varied across studies. Chen et al. (2012) grouped these factors into three: major economic factors (income, education, income inequality, economic growth and unemployment), demography and household (female labour force participation, divorce, birth rate, migration and population, household size, age, gender and cohorts effect), and others (religion, homicide, geographical and climatic conditions, civil liberty and quality of governance and health care and alcohol consumption). The authors employed a meta-regression method to investigate why existing empirical results vary. They found evidence of publication bias with respect to several socio-economic factors. For detailed review of studies on drivers of suicide, see Chen et al. (2012).

Another strand of the literature tested the natural rate of suicide hypothesis. Yang and Lester (1991) were the first to estimate the natural rate for 48 contiguous American states in 1980. They estimated several regression equations using socio-economic variables such as divorce rates, interstate migration rates and church non-attendance. When social conditions were assumed to be 'ideal' i.e., no divorce, no interstate migration and 100% church attendance, the natural rate of suicide was 6 per 100,000 population per year (compared to 12 per 100,000 actual suicide rate). Yang & Lester (2009) further conducted a cross-sectional

analysis for 11 nations and time series estimation for the period 1950 to 1985 for 13 nations to explore the non-zero suicide natural rate hypothesis. In both instances, when the two explanatory variables in the regressions, divorce and unemployment rates were set to zero, the constant terms were positive and non-zero, indicating a positive and non-zero natural rate of suicide for these countries. (Kunze & Anderson (2002) also examined the validity of Yang and Lester's (1991) proposition for a balanced panel of all 50 states plus District of Columbia in the USA over 1985-1995 period. They used age-adjusted suicide rates and several predictor variables, including unemployment rate, median household income, divorce rate, nonmetropolitan residents, Christian church adherents, poverty rate, single occupant households and percent of non-white population. Their fixed effects estimates lent support to the positive non-zero hypothesis, albeit at a magnitude of 1.28 per 100,000 population, much lower than Yang and Lester (1991). Despite using state-level data, however, the study did not provide state-specific estimates of natural suicide rate.

A handful of studies have estimated the natural rate of suicide using non-USA data. While investigating the economic determinants of suicides in Finland over 1878-1994 period using an error correction model, Virén (1999) found support for a positive natural rate. His model included various demographic and structural variables, namely gender distribution, average age and urbanisation, GDP per capita, bankruptcies, and unemployment. The results were indicative of an association between suicide rate and economic factors. Andrés & Halicioglu (2011) provided further evidence on the natural rate of suicide hypothesis using the autoregressive-distributed lag (ARDL) approach to cointegration for 15 OECD countries, over the period 1970–2004. Their results were consistent with the presence of positive natural rates, with Turkey registering the lowest and Japan the highest rate.

It is not surprising that the natural rate, gauged by the constant term of regression models, remains highly sensitive to the set of factors included in the model (Yang and Lester, 1991). While most existing studies capture economic, demographic and social factors in one way or another, the choice of variables remains arbitrary. We propose a novel variable selection approach, the clustered variant of Least Absolute Shrinkage and Selection Operation (LASSO), to test the natural rate hypothesis. This framework allows us to select variables systematically from a wide array of suicide determinants found in the literature, which are then used to estimate the natural rate. The methodology is described in more detail in the next section.

4. Data and Methodology

4.1 Data

We consider a balanced panel of time-series and cross-section data for 50 states plus District of Columbia (DC) from 2005 to 2017. Table 1 provides the descriptive statistics and expected sign of variables (main effects) used in our selection procedure. Variable definitions and data sources are summarised in Table

A1 in the Appendix. Choice of regressors and their expected signs are drawn from the consensus established in the literature (see for example Chen et al., 2012; Collins et al., 2021). As discussed in Belloni et. al. (2016), we include interactions of initial (year 2005) values of all covariates with a linear, quadratic, and cubic term in time, to account for the fact that some features of states may not only be correlated to the level of demographic/economic and health/social factors but also to their evolution. We also extend the list of covariates by creating interaction terms among variables that control for demographic/economic and health/social factors. The inclusion of these interaction terms along with the main effects brings the total number of predictors to 167.

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	Expected Sign
AGESUICIDE	663	14.056	4.208	4.400	29.600	
<i>Demographic and Economic measures</i>						
POPDEN	663	390.102	1406.047	1.166	11302.480	–
PERCENTAGRI	663	2.645	2.315	0.035	13.060	–
DR	663	6.755	3.908	0.800	14.600	+
CUR	663	6.052	2.174	2.400	13.800	+
GDPPC	663	52725.970	20079.490	32770	183971	–
NONWHITE	663	0.229	0.136	0.034	0.754	–
CRISIS	663	0.462	0.499	0.000	1.000	+
CSI	663	80.923	10.558	63.700	96.800	–
EDUCLESSHS	663	11.447	3.645	4.100	20.500	+
EDUCCOL	663	29.410	6.995	16.100	62.200	–
<i>Health and Social measures</i>						
ALCOHOL	663	39.754	18.294	10.600	69.100	+
HHSIZE	663	2.570	0.167	2.080	3.190	–
OBESITY	663	27.705	3.789	17.000	38.100	+/-
SMOKE	663	19.089	3.585	8.800	29.000	+/-
POVLINE	663	5518261	6672103	18479	38800000	+
FERTILITY	663	34.536	6.836	14.100	55.300	–
FLABOUR	663	59.894	3.858	48.093	68.651	–

4.2 Methodology

Belloni et. al. (2016) provided a framework that applied LASSO to panel data models with additive unobserved individual specific heterogeneity in a high-dimensional setting, whilst also accounting for some form of dependence within the same cross-sectional unit of observation. Individual specific heterogeneity was eliminated in the variable selection process and overall contribution of time-varying variables could be captured by a relatively small number of available variables. The Cluster-LASSO approach accommodated within-state dependence and partialled out fixed effects.

Hence, following Belloni et al. (2016), we consider the model:

$$y_{it} = x'_{it}\beta + \alpha_i + \delta_t + \varepsilon_{it} \quad i = 1, \dots, n \quad t = 1, \dots, T \quad (1)$$

where the dependent variable, y_{it} is measured as the rate per 100,000 population of state i 's age-adjusted suicide rate over annual time t , x_{it} are covariates, α_i and δ_t are state and time fixed-effects respectively, and ε_{it} is an idiosyncratic disturbance term with mean zero conditional on covariates, but which may have dependence within an individual state.

We eliminate the fixed effects parameters from our estimation before LASSO selection, where we define:

$$\tilde{y}_{it} = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it}$$

Control variables $\tilde{x}_{it}$ and disturbance term have similar notations. Hence our ‘within’ model can be defined as

$$\tilde{y}_{it} = \beta \tilde{x}'_{it} + \tilde{\varepsilon}_{it}$$

The Cluster-LASSO estimator with data-driven penalty loadings that estimates $\hat{\beta}$ is defined by solving the following penalised minimisation problem on the within model:

$$\hat{\beta} = \arg \min_b \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\tilde{y}_{it} - b \tilde{x}'_{it})^2 + \frac{\lambda}{nT} \sum_{j=1}^p \hat{\phi}_j |b_j|$$

where λ is the main tuning parameter that dictates the amount of regularisation in the LASSO procedure and serves to balance overfitting and bias concerns. $\{\hat{\phi}_j\}_{j=1}^p$ is the covariate specific penalty loading which allows us to handle data which may be dependent within state, heteroscedastic and non-Gaussian (see Belloni et al. (2016) for a detailed discussion). Overall, this approach minimises the sum of squared deviations but also imposes an l_1 regularisation penalty to set some coefficients to zero, thus determining a set of predictors that best explains the response variable.

To find the natural rate of suicide, we use the Least Squares Dummy Variable (LSDV) estimator:

$$y_{it} = \alpha + \beta x'_{it} + \mu_i + \varepsilon_{it} ; i = 1, 2, \dots, N ; t = 1, 2, \dots, T \quad (2)$$

where α is the constant term of interest i.e the societal suicide rate when all variables in the regression are set to zero, y_{it} are observations of the response variable, x_{it} are covariates which are selected by the Cluster-LASSO approach, μ_i is the latent state effect and ε_{it} is the disturbance.

Model selection is executed through *lassopack*, a suite of programs for regularised regressions in STATA (see (Ahrens et al., 2020a, 2020b)). This package offers three approaches for selecting the penalisation (‘tuning’) parameter: information criteria, the classical K-fold cross validation and theory-driven (or ‘rigorous’) penalisation. Information criteria can be justified theoretically and are easy to calculate but do

not perform well when assumptions of independence and homoskedasticity are violated. Cross validations are computationally expensive but robust to heteroskedasticity and work well for prediction tasks. However, as noted by Ahrens et. al. (2020b), cross validation may not be the best choice for model identification, given that the training sample is set at n/K in a K -fold cross validation, whereas consistent model selection requires the training sample to be small and the validation sample close to n . We, therefore, use rigorous penalisation (*rlasso* in STATA) to select the tuning parameter λ . The objective functions to calculate λ follow Belloni et al. (2012) and are described in Ahrens et. al. (2020).

5. Results and Discussion

The predictors selected by Cluster-LASSO procedure are summarised in Table 2. These are Obesity, GDP per capita interacted with Obesity, Alcohol interacted with Percentage of female population in labour force, and interactions of initial values of Obesity and Smoke with linear term in time. The overall penalty level λ value is also reported.

Table 2: Variables Selected (Cluster-LASSO)

OBESITY	Obesity
GDPPC*OBESITY	GDP per capita x Obesity
ALCOHOL*FLABOUR	Crude prevalence of alcohol consumption x Percentage of female population in labour force
$t \times OBESITY_0$	$t \times (Obesity)_0$
$t \times SMOKE_0$	$t \times (Smoking\ prevalence\ among\ adults)_0$
Overall penalty level $\lambda = 213.062$	

After the variable selection procedure, we use Least squares Dummy Variable (LSDV) estimation procedure to obtain coefficient μ_i for each state. The natural suicide rate for each state is then calculated as $\alpha + \mu_i^2$. These are shown in Table 3 below, along with the actual average suicide rates for 2005-2017. We rank the states according their estimated natural suicide rate.

Our results suggest that the ten states having the highest natural rates are Montana (21.54), Wyoming (20.55), Alaska (20.02), New Mexico (19.27), Nevada (17.78), Utah (17.52), Idaho (17.47), Colorado (16.84), Arizona (15.90) and Oklahoma (15.54), all measured per 100,000 population. The estimated natural rates are all lower than the actual average rates across all states. Possible factors explaining the higher rates of suicidal mortality compared to an ‘acceptable’ level include worsening of economic and social well-being among some segments of the population, lack of affordable mental health care, compounded by disparities in access due to race and social class and inadequate interventions and follow-ups by health care providers to reduce deaths by suicide (Weir, 2019).

² The quality of fixed effects estimates depends on the number of time periods (T). As a result of restricted data availability, time coverage is limited to 13 years (T=13). Compared to the number of fixed effects parameters (N=51), T is relatively small, which may introduce some noise in the estimation of the natural rates.

Table 3: Actual average and Natural rate of suicide by State, 2005-2017 (Based on Cluster - LASSO covariates).

Ranking¹	State	Natural Rate (per 100,000)	Actual Average² Rate (per 100,000)	% Difference³
1	Montana	21.539	22.862	6.142
2	Wyoming	20.547	22.977	11.827
3	Alaska	20.021	22.877	14.265
4	New Mexico	19.265	20.585	6.852
5	Nevada	17.782	19.246	8.233
6	Utah	17.517	19.115	9.123
7	Idaho	17.467	18.938	8.422
8	Colorado	16.843	18.208	8.104
9	Arizona	15.895	17.162	7.971
10	Oklahoma	15.544	17.292	11.245
11	South Dakota	15.294	17.238	12.711
12	Oregon	15.079	16.746	11.055
13	Arkansas	14.767	16.331	10.591
14	Vermont	13.918	15.231	9.434
15	Kentucky	13.612	15.269	12.173
16	West Virginia	13.476	15.815	17.357
17	Maine	13.453	16.085	19.564
18	North Dakota	13.437	14.762	9.861
19	Missouri	13.344	15.138	13.444
20	Kansas	13.270	15.031	13.271
21	Tennessee	13.164	14.900	13.187
22	Washington	12.806	14.138	10.401
23	Indiana	12.452	13.692	9.958
24	New Hampshire	12.441	14.192	14.074
25	Florida	12.345	13.738	11.284
26	Alabama	12.214	13.892	13.738
27	South Carolina	11.876	13.615	14.643
28	Mississippi	11.609	13.231	13.972
29	Wisconsin	11.526	13.262	15.062
30	Iowa	10.915	12.854	17.765
31	Pennsylvania	10.830	12.685	17.128
32	Ohio	10.787	12.462	15.528
33	North Carolina	10.775	12.554	16.510
34	Michigan	10.716	12.392	15.640
35	Hawaii	10.448	12.015	14.998
36	Louisiana	10.345	12.762	23.364
37	Virginia	10.157	12.192	20.035
38	Georgia	9.942	11.800	18.688
39	Minnesota	9.915	11.854	19.556
40	Texas	9.447	11.662	23.447
41	Nebraska	9.420	11.538	22.484
42	Delaware	9.247	11.623	25.695

43	Rhode Island	8.591	10.092	17.472
44	California	8.079	10.115	25.201
45	Illinois	7.450	9.554	28.242
46	Connecticut	7.046	9.115	29.364
47	Maryland	6.995	9.077	29.764
48	Massachusetts	6.189	8.185	32.251
49	New York	5.438	7.562	39.058
50	New Jersey	5.339	7.300	36.730
51	DC	0.699	5.892	742.918

¹Ranked by Natural rate ²Averaged over 2005-2017 ³Calculated as (Actual-Natural)/Natural*100

6. Robustness Checks

As a first robustness check, we apply Cluster-LASSO to reduced samples, with the first and last 25% of observations removed in turn, to ascertain that variables selected are not sensitive to sample size. As shown in Table 4, there is considerable overlap in the variables selected in the samples.

Table 4: Variables Selected: Full and Reduced Samples (Cluster-LASSO)

Full Sample: 2005-2017	Reduced Sample 1: 2008-2017 (First 25% observations removed)	Reduced Sample 2: 2005-2014 (Last 25% observations removed)
OBESITY	FLABOUR	GDPPC*OBESITY
GDPPC*OBESITY	ALCOHOL*FLABOUR	$t \times \text{ALCOHOL}_0$
ALCOHOL*FLABOUR	$t \times \text{OBESITY}_0$	$t \times \text{SMOKE}_0$
$t \times \text{OBESITY}_0$	DR*OBESITY	$t \times \text{FLABOUR}_0$
$t \times \text{SMOKE}_0$		$t^2 \times \text{PERCENTAGRI}_0$
$\lambda = 213.062$	$\lambda = 181.69$	$\lambda = 165.90$

While Cluster-LASSO successfully deals with heteroskedasticity and autocorrelation, it is susceptible to covariate dependence. To mitigate the undesired effect stemming from such correlation, we use Factor-adjusted regularised model selection (FarmSelect) proposed by (Fan et al., 2020) as a second robustness check. This procedure transforms highly correlated covariates into weakly correlated ones by separating latent factors from idiosyncratic components. This method is implemented using FarmSelect package in R. As FarmSelect is not readily applicable to panel data, we ignore individual state-specific effects and use pooled data instead. Table 5 presents the covariates selected and corresponding Lambda values. The first 25% and last 25% of our data are also removed to ensure robustness.

Table 5: Variables Selected: Full and Reduced Samples (FarmSelect)

Full Sample: 2005-2017	Reduced Sample 1: 2008-2017 (First 25% observations removed)	Reduced Sample 2: 2005-2014 (Last 25% observations removed)
FERTILITY	FERTILITY	HHSIZE
HHSIZE	HHSIZE	$t \times \text{POPDEN}_0$
$t \times \text{POPDEN}_0$	CSI	CUR*NONWHITE
$t^3 \times \text{NONWHITE}_0$	$t \times \text{POPDEN}_0$	CUR*POVLIN
CUR*POVLIN	$t^3 \times \text{NONWHITE}_0$	POPDEN*NONWHITE
POPDEN*NONWHITE	CUR*FERTILITY	POPDEN*HHSIZE
POPDEN*HHSIZE	POPDEN*NONWHITE	POPDEN*FLABOUR
DR*EDUCCOL	POPDEN*HHSIZE	DR*EDUCCOL

DR*FLABOUR	DR*FLABOUR	EDUCLESSHS*FLABOUR
NONWHITE*FERTILITY	NONWHITE*GDPPC	OBESITY*POVLINE
PERCENTAGRI*FERTILITY	PERCENTAGRI*FERTILITY	FERTILITY*POVLINE
PERCENTAGRI*HHSIZE	ALCOHOL*OBESITY	HHSIZE*POVLINE
GDPPC*SMOKE	OBESITY*POVLINE	
OBESITY*POVLINE		
$\lambda = 0.022$	$\lambda = 0.022$	$\lambda = 0.021$

Given the nature of FarmSelect procedure, variables selected are distinct from Cluster-LASSO. Nevertheless, both methods share common covariates: Obesity, GDP per capita, Female labour population and Smoking are selected, regardless of the method used. To obtain the natural rate for each state, we apply LSDV estimation once more. Most of the estimated natural suicide rates are again lower than the actual average rates – with the exception of Montana, Wyoming, South Dakota, North Dakota, Wisconsin and DC (Table 6). Since the actual suicide rate for each state is calculated as an average over the sample period, variations in the data are concealed. Most states with natural rates higher than actual averages have larger dispersions in the data, and actual rates in later years are a lot higher than earlier years. Actual average suicide rates are in fact higher than the estimated natural rate for these states in more recent years. For example, Wyoming’s actual average of 25.3 post 2011 is higher than the rate of 21.0 in the earlier period as well as the estimated natural rate of 24.2 (all per 100,000 population). A similar pattern emerges for Montana, South Dakota, North Dakota, and Wisconsin.

Table 6: Actual average and Natural rate of suicide by State, 2005-2017 (Based on FarmSelect covariates).

Ranking ¹	State	Natural Rate (per 100,000)	Actual Average ² Rate (per 100,000)	% Difference ³
1	Wyoming	24.172	22.977	-4.944
2	Montana	23.114	22.862	-1.090
3	Alaska	20.413	22.877	12.071
4	New Mexico	18.801	20.585	9.489
5	Idaho	18.721	18.938	1.159
6	Utah	18.486	19.115	3.403
7	South Dakota	17.313	17.238	-0.433
8	North Dakota	16.476	15.815	-4.012
9	Colorado	16.463	18.208	10.600
10	Nevada	15.547	19.246	23.792
11	DC	15.526	5.892	-62.051
12	Oklahoma	15.490	17.292	11.633
13	Oregon	15.453	16.746	8.367
14	Wisconsin	15.073	13.262	-12.015
15	Arizona	14.622	17.162	17.371
16	Vermont	14.157	15.231	7.586
17	Arkansas	14.030	16.331	16.401
18	Kansas	13.759	15.031	9.245
19	Maine	13.305	14.762	10.951

20	Kentucky	13.178	15.269	15.867
21	Washington	12.558	14.138	12.582
22	Missouri	12.543	15.138	20.689
23	New Hampshire	12.338	14.192	15.027
24	Iowa	11.963	12.854	7.448
25	Tennessee	11.561	14.900	28.882
26	West Virginia	11.270	16.085	42.724
27	Indiana	10.887	13.692	25.765
28	Nebraska	10.617	11.538	8.675
29	Alabama	10.609	13.892	30.945
30	Florida	10.453	13.738	31.426
31	Pennsylvania	10.222	12.685	24.095
32	Minnesota	10.128	11.854	17.042
33	Louisiana	10.094	12.762	26.432
34	South Carolina	9.991	13.615	36.273
35	Mississippi	9.704	13.231	36.346
36	Virginia	9.573	12.192	27.358
37	Texas	9.570	11.662	21.860
38	Michigan	9.410	12.392	31.690
39	North Carolina	9.346	12.554	34.325
40	Ohio	9.279	12.462	34.303
41	Delaware	8.640	11.623	34.525
42	Georgia	7.934	11.800	48.727
43	Hawaii	7.914	12.015	51.820
44	Rhode Island	7.337	10.092	37.549
45	Connecticut	6.568	9.115	38.779
46	Illinois	6.496	9.554	47.075
47	California	6.274	10.115	61.221
48	Maryland	6.001	9.077	51.258
49	Massachusetts	5.890	8.185	38.964
50	New Jersey	5.031	7.300	45.100
51	New York	4.252	7.562	77.846

¹Ranked by Natural rate

²Averaged over 2005-2017

³Calculated as (Actual-Natural)/Natural*100

Although the reported natural rates for many states are slightly higher compared to those listed in Table 3, the rankings are similar. Our results from FarmSelect suggest that the ten states having the highest natural rates are Wyoming (24.17), Montana (23.11), Alaska (20.41), New Mexico (18.80), Idaho (18.80), Utah (18.49), South Dakota (17.31), North Dakota (16.48) and Colorado (16.46), Nevada (15.55) all per 100,000 population.

Overall, our results are consistent under different variable selection approaches, with natural rates being lower than the actual average rates. This provides a potentially useful set of policy implications and source of guidance in informing resource allocation and policy design to attain ‘zero suicide’ targets. In essence,

the results generated offer, in practical terms, a transparent means to improve outcomes in such activity. This may arise from choosing to engage with a more concentrated focus on the reduction of the relative disparity between natural and actual rates. The generated results also provide a means to selectively identify geographically proximate ‘benchmark’ states (with significantly lower relative disparities). This would facilitate much higher resolution pairwise or groupwise comparison of healthcare, mental health and other domains of public policy (employment, education etc). Assuming a genuine desire to reduce relative disparities in suicide rates, there are potential opportunities to be exploited from highlighting nearby and perhaps more familiar exemplar guides to help prompt or suggest consideration of potential policy shifts, experiments and interventions.

7. Summary and Concluding Remarks

A robust means of estimating the natural rate of suicide in the USA at state level has been unfolded. An empirical application shows how state level societal suicide rate is determined by a range of factors, identified using a technique designed for high dimensional panel data (Cluster- LASSO). Estimates of the natural suicide rate are non-zero and positive in all states, hence supporting Yang and Lester’s (1991) proposition. These estimates range from 0.70 to 21.54 per 100,000 population and are well below the actual average rates. Higher actual average suicide rates are likely due to a lack of accessibility to mental health care; an inability of many to pay for such care and inadequate mental health care practices to screen and follow up individuals with a high risk of suicidal behaviour.

Our results suggest the ten states having the highest natural rates are Montana (21.54), Wyoming (20.55), Alaska (20.02), New Mexico (19.27), Nevada (17.78), Utah (17.52), Idaho (17.47), Colorado (16.84), Arizona (15.90) and Oklahoma (15.54), all measured per 100,000 population. Most of these states are part of the ‘Suicide Belt’.

Further evidence on robust estimation of natural rates is timely, given increasing focus on suicides emerging from the economic aftermath of the Covid-19 pandemic and increasing profile to ‘zero suicide’ initiatives in behavioural healthcare systems across the world. In terms of mental health and suicide prevention resource allocation, the results of this study provide an evidence base suggesting that policy focus and greater resources can rationally be directed to those states with the greatest relative difference between actual and natural suicide rates. Overall, the approach suggested in this study offers one pragmatic macro-level pathway towards identifying and achieving a zero-suicide target beyond the confines of behavioural healthcare systems and in the context of more accessible public statistical data. In principle, it also offers a means to support improvements in suicide reduction by focused consideration on the relative disparities in actual and natural suicide rates across states and helping highlight appropriate exemplar states which others might look to for inspiration and guidance on adjusting or changing policy in order to reduce suicide mortality.

Acknowledgement

We would like to thank the Editor and anonymous reviewers for their helpful comments and suggestions. We also express our appreciation to Lerato Dixon for her invaluable assistance in data collection and Prof Robert Ackrill for his useful feedback on the paper. We remain grateful to Dr Gareth Liu-Evans from Liverpool University for his guidance and suggestions. The usual disclaimer applies.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Ahrens, A., Hansen, C.B., Schaffer, M.E., 2020a. LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation. Statistical Software Components.
- Ahrens, A., Hansen, C.B., Schaffer, M.E., 2020b. lassopack: Model selection and prediction with regularized regression in Stata. The Stata Journal: Promoting communications on statistics and Stata 20. <https://doi.org/10.1177/1536867X20909697>
- Andrés, A.R., Halicioglu, F., 2011. Testing the hypothesis of the natural suicide rates: Further evidence from OECD data. Economic Modelling 28. <https://doi.org/10.1016/j.econmod.2010.10.004>
- Becker, G.S., Posner, R.A., 2004. Suicide: An economic approach.
- Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. Econometrica 80. <https://doi.org/10.3982/ECTA9626>
- Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on Treatment Effects after Selection among High-Dimensional Controls. The Review of Economic Studies 81. <https://doi.org/10.1093/restud/rdt044>
- Belloni, A., Chernozhukov, V., Hansen, C., Kozbur, D., 2016. Inference in High-Dimensional Panel Models With an Application to Gun Control. Journal of Business & Economic Statistics 34. <https://doi.org/10.1080/07350015.2015.1102733>
- Chen, J., Choi, Y.J., Mori, K., Sawada, Y., Sugano, S., 2012. SOCIO-ECONOMIC STUDIES ON SUICIDE: A SURVEY. Journal of Economic Surveys 26. <https://doi.org/10.1111/j.1467-6419.2010.00645.x>
- Collins, A., Cox, A., Kizys, R., Haynes, F., Machin, S., Sampson, B., 2021. Suicide, sentiment and crisis. The Social Science Journal 58. <https://doi.org/10.1016/j.soscij.2019.04.001>
- Cui, H., Rajagopalan, S., Ward, A.R., 2020. Predicting product return volume using machine learning methods. European Journal of Operational Research 281. <https://doi.org/10.1016/j.ejor.2019.05.046>
- Dixit, A.K. and Pindyck, R.S., 1994. *Investment under Uncertainty*. Princeton University Press.
- Fan, J., Ke, Y., Wang, K., 2020. Factor-adjusted regularized model selection. Journal of Econometrics 216. <https://doi.org/10.1016/j.jeconom.2020.01.006>
- Hamermesh, D.S., Soss, N.M., 1974. An Economic Theory of Suicide. Journal of Political Economy 82. <https://doi.org/10.1086/260171>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction, Springer.
- Hoff, A., Gribkovskaia, I., Laporte, G., Løkketangen, A., 2009. Lasso solution strategies for the vehicle routing problem with pickups and deliveries. European Journal of Operational Research 192. <https://doi.org/10.1016/j.ejor.2007.10.021>
- Kunce, M., Anderson, A.L., 2002. A Natural Rate of Suicide for the U.S., Revisited. OMEGA - Journal of Death and Dying 44. <https://doi.org/10.2190/B5TP-BAXF-T0TJ-FFW7>
- Miao, J., Wang, N., 2011. Risk, uncertainty, and option exercise. Journal of Economic Dynamics and Control 35. <https://doi.org/10.1016/j.jedc.2010.11.005>

- Naghavi, M., 2019. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*. <https://doi.org/10.1136/bmj.l94>
- Panagiotidis, T., Stengos, T., Vravosinos, O., 2018. On the determinants of bitcoin returns: A LASSO approach. *Finance Research Letters* 27. <https://doi.org/10.1016/j.frl.2018.03.016>
- Platt, S., 1984. Unemployment and suicidal behaviour: A review of the literature. *Social Science & Medicine* 19. [https://doi.org/10.1016/0277-9536\(84\)90276-4](https://doi.org/10.1016/0277-9536(84)90276-4)
- Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Vandoros, S., Avendano, M., Kawachi, I., 2019. The association between economic uncertainty and suicide in the short-run. *Social Science & Medicine* 220. <https://doi.org/10.1016/j.socscimed.2018.11.035>
- Virén, M., 1999. Testing the “natural rate of suicide” hypothesis. *International Journal of Social Economics* 26. <https://doi.org/10.1108/03068299910370535>
- Weir, K., 2019. Worrying trends in U.S. suicide rates [WWW Document]. *Monitor on Psychology*, 50 (3), 24. URL <https://www.apa.org/monitor/2019/03/trends-suicide> (accessed 11.2.21).
- Yang, B., Lester, D., 2009. Is there a natural suicide rate? *Applied Economics Letters* 16. <https://doi.org/10.1080/13504850601018189>
- Yang, B., Lester, D., 1991. Is There a Natural Suicide Rate for a Society? *Psychological Reports* 68. <https://doi.org/10.2466/pr0.1991.68.1.322>

OLD References (to delete)

Ahrens, A., Hansen, C.B., Schaffer, M.E. 2018 (updated 2020). Lassopack: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation. <http://ideas.repec.org/c/boc/bocode/s458458.html>

Ahrens, A., Hansen, C.B. and M.E. Schaffer. 2020. Lassopack: model selection and prediction with regularized regression in Stata. *The Stata Journal*, 20 (1), 176-235. <https://doi.org/10.1177/1536867X20909697>

Andrés, A.R. and Halicioglu, F., 2011. Testing the hypothesis of the natural suicide rates: Further evidence from OECD data. *Economic Modelling*, 28 (1-2), 22-26. <https://doi.org/10.1016/j.econmod.2010.10.004>

Becker, G.S. and Posner, R.A., 2004. *Suicide: An economic approach*. University of Chicago.

Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80 (6), 2369-2429. <https://doi.org/10.3982/ECTA9626>

Belloni, A., Chernozhukov, V. and Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81 (2), 608-650. <https://doi.org/10.1093/restud/rdt044>

Belloni, A., Chernozhukov, V., Hansen, C. and Kozbur, D., 2016. Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, 34 (4), 590-605. <https://doi.org/10.1080/07350015.2015.1102733>

Chen, J., Choi, Y.J., Mori, K., Sawada, Y. and Sugano, S., 2012. Socio-economic studies on suicide: A survey. *Journal of Economic Surveys*, 26 (2), 271-306. <https://doi.org/10.1111/j.1467-6419.2010.00645.x>

Collins, A., Cox, A., Kizys, R., Haynes, F., Machin, S. and Sampson, B., 2020. Suicide, Sentiment and Crisis. *The Social Science Journal*, 1-18. <https://doi.org/10.1016/j.sosci.2019.04.001>

Cui, H., Rajagopalan, S. and Ward, A.R., 2019. Predicting Product Return Volume Using Machine Learning Methods. *European Journal of Operational Research*, 281 (3), 612-627. <https://doi.org/10.1016/j.ejor.2019.05.046>

Dixit, A.K. and Pindyck, R.S., 1994. *Investment under Uncertainty*. Princeton University Press.

Fan, J., Ke, Y. and Wang, K., 2020. Factor-adjusted regularized model selection. *Journal of Econometrics*, 216 (1), 71-85. <https://doi.org/10.1016/j.jeconom.2020.01.006>

- Hamermesh, D.S. and Soss, N.M., 1974. An economic theory of suicide. *Journal of Political Economy*, 82 (1), 83-98. <http://www.jstor.org/stable/1830901>
- Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics.
- Hoff, A., Gribkovskaia, I., Laporte, G. and Løkketangen, A., 2009. Lasso solution strategies for the vehicle routing problem with pickups and deliveries. *European Journal of Operational Research*, 192 (3), 755-766. <https://doi.org/10.1016/j.ejor.2007.10.021>
- Kunce, M. and Anderson, A.L., 2002. A natural rate of suicide for the US, revisited. *OMEGA-Journal of Death and Dying*, 44 (3), 215-222. <https://doi.org/10.2190/B5TP-BAXF-T0TJ-FFW7>
- Miao, J. and Wang, N., 2011. Risk, uncertainty, and option exercise. *Journal of Economic Dynamics and Control*, 35 (4), 442-461. <https://doi.org/10.1016/j.jedc.2010.11.005>
- Naghavi, M., 2019. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*, 364, 194. <https://doi.org/10.1136/bmj.194>
- Panagiotidis, T., Stengos, T. and Vravosinos, O., 2018. On the determinants of bitcoin returns: A LASSO approach. *Finance Research Letters*, 27, 235-240. <https://doi.org/10.1016/j.frl.2018.03.016>
- Platt, S., 1984. Unemployment and suicidal behaviour: a review of the literature. *Social Science & Medicine*, 19 (2), 93-115. [https://doi.org/10.1016/0277-9536\(84\)90276-4](https://doi.org/10.1016/0277-9536(84)90276-4)
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58 (1), 267-288. <http://www.jstor.org/stable/2346178>.
- Vandoros, S., Avendano, M. and Kawachi, I., 2019. The association between economic uncertainty and suicide in the short-run. *Social Science & Medicine*, 220, 403-410. <https://doi.org/10.1016/j.socscimed.2018.11.035>
- Virén, M., 1999. Testing the “natural rate of suicide” hypothesis. *International Journal of Social Economics*, 26 (12), 1428-1440. <https://doi.org/10.1108/03068299910370535>
- Weir, K., 2019. Worrying trends in U. S. Suicide rates. *Monitor on Psychology*, 50 (3), 24. <http://www.apa.org/monitor/2019/03/trends-suicide>

Yang, B. and Lester, D., 1991. Is there a natural suicide rate for a society? *Psychological Reports*, 68 (1), 322-322. <https://doi.org/10.2466/pr0.1991.68.1.322>

Yang, B., & Lester, D., 2009. Is there a natural suicide rate? *Applied economics letters*, 16 (2), 137-140. <https://doi.org/10.1080/13504850601018189>

Appendix

Table A1. Variable definitions and sources

Variable	Definition	Source
<i>Suicide Rates</i>	Suicide rates (deaths per 100,000 people)	
AGESUICIDE	Age-adjusted suicide rate (all mechanisms)	Centers for Disease Control WISQARS Injury Mortality Reports
<i>Demographic and Economic measures</i>		
CUR	Civilian unemployment rate	Bureau of Labor Statistics
POPDEN	Population density	U.S. Census Bureau
DR	Divorce rate	U.S. Census Bureau
NONWHITE	Share of the population that is non-white (Black, American Indian, Alaska Native, Asian, Hawaiian, and other)	U.S. Census Bureau
CRISIS	Post-2007 financial crisis dummy	Authors' Created
CSI (Consumer Sentiment Index, 1996 = 100)	Quantifies consumers' perceptions of their own financial situation and of the general economy in near and long term	Thomson Reuters Datastream; University of Michigan, Surveys of Consumers
PERCENTAGRI	Total population in agriculture, forestry and fishing and hunting, and mining of population of working age (16 and over)	U.S. Census Bureau
GDPPC	Gross domestic product per capita	Bureau of Economic Analysis
EDUCLESSHS	Highest grade of school completed is less than high school	Centers for Disease Control and Prevention
EDUCCOL	Highest grade of school completed is college and above	Centers for Disease Control and Prevention
<i>Health and social measures</i>		
ALCOHOL	Crude prevalence of alcohol consumption	Centers for Disease Control and Prevention
OBESITY	BMI greater than or equal to 30.0	Centers for Disease Control and Prevention
SMOKE	Smoking prevalence among adults	Centers for Disease Control and Prevention
FERTILITY	Percentage of women between 15 and 50 years old who gave birth within the last 12 months	U.S. Census Bureau
HHSIZE	Average household size	U.S. Census Bureau
POVLIN	Percentage below 100 percent of the poverty line	U.S. Census Bureau
FLABOUR	Percentage of female population in labour force	U.S. Census Bureau