Methods for "A Time Series Approach to Estimating Excess Mortality Rates in Puerto Rico, Post-Maria"[1]

Menzie Chinn[2]

August 10, 2018

Procedure:

Estimate relationship between mortality as recorded and population in a log-log relationship (with monthly dummies) over the 2010M01-2016.M12 period (Maria landfall is 20 September 2017), using OLS and quantile regression. Use the regression to obtain fitted values over the entire 2010M01-2018M03 period (truncating 2018M04-05 period assuming those figures are provisional). Use the 2017M09-2018M03 fitted values as a counterfactual, and (1) compare against recorded figures, and (2) compare against the 95% CI upper bound of the counterfactual.

Data

Mortality data as reported by Puerto Rico vital statistics system, and Santos-Lozada and Howard (2017). Population estimates from *World Economic Outlook* database (April 2018), cubic interpolation to monthly frequency.

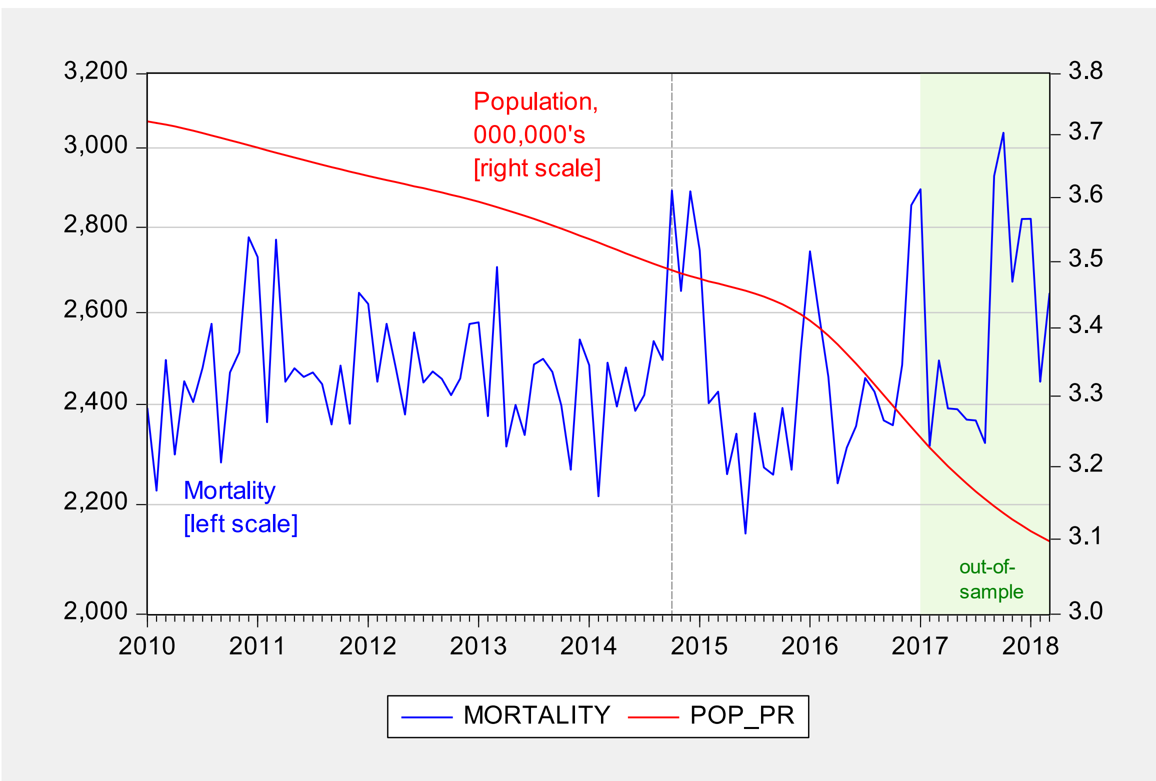

Figure 1: Mortality per month (blue, left log scale), and population (red, right log scale), cubic interpolation from IMF *World Economic Outlook* database data. Out of sample period shaded light green.

[1] http://econbrowser.com/archives/2018/08/a-time-series-approach-to-estimating-excess-mortality-rates-in-puerto-rico-post-maria

[2] Email: mchinn@lafollette.wisc.edu

Regression Specifications

Baseline regression: Estimate on monthly data, 2010M01-2016M12, using OLS regression:

$$m_t = \beta_0 + \delta(OCT14) + monthly\ dummies + u_t$$

Where *m* is log mortality, and OCT14 is a dummy variable taking on a value of 1 in October 2014.

Population Augmented Regression: Estimate on monthly data, 2010M01-2016M12, using OLS and quantile regression (where *pop* is log population).

$$m_t = \beta_0 + \beta_1 pop_t + \delta(OCT14) + monthly\ dummies + u_t$$

Empirical Results

Dependent Variable: LOG(MORTALITY)
Method: Least Squares
Date: 08/10/18 Time: 17:20
Sample: 2010M01 2016M12
Included observations: 84
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.866824	0.020328	386.9860	0.0000
M02	-0.095336	0.018418	-5.176269	0.0000
M03	-0.020215	0.027164	-0.744187	0.4592
M04	-0.106299	0.025572	-4.156797	0.0001
M05	-0.081843	0.022870	-3.578639	0.0006
M06	-0.094278	0.028871	-3.265510	0.0017
M07	-0.063950	0.021104	-3.030301	0.0034
M08	-0.059599	0.025492	-2.337892	0.0222
M09	-0.091181	0.025184	-3.620534	0.0005
M10	-0.075533	0.021566	-3.502368	0.0008
M11	-0.073395	0.027287	-2.689747	0.0089
M12	0.027161	0.026340	1.031202	0.3059
DUMMY_OCT14	0.178067	0.008007	22.23795	0.0000

R-squared	0.529973	Mean dependent var	7.807738
Adjusted R-squared	0.450532	S.D. dependent var	0.060449
S.E. of regression	0.044809	Akaike info criterion	-3.231443
Sum squared resid	0.142555	Schwarz criterion	-2.855246
Log likelihood	148.7206	Hannan-Quinn criter.	-3.080215
F-statistic	6.671273	Durbin-Watson stat	1.084493
Prob(F-statistic)	0.000000	Wald F-statistic	193.5423
Prob(Wald F-statistic)	0.000000		

Date: 08/10/18   Time: 17:43
Sample: 1974Q1 2015Q4
Included observations: 164

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|******|	.|******|	1	0.873	0.873	127.34	0.000
.|***** |	*|. |	2	0.730	-0.136	216.91	0.000
.|**** |	*|. |	3	0.587	-0.077	275.25	0.000
.|*** |	.|. |	4	0.469	0.013	312.62	0.000
.|*** |	.|* |	5	0.391	0.085	338.81	0.000
.|** |	.|. |	6	0.323	-0.043	356.77	0.000
.|** |	*|. |	7	0.236	-0.137	366.45	0.000
.|* |	.|. |	8	0.147	-0.053	370.22	0.000
.|. |	*|. |	9	0.035	-0.153	370.44	0.000
*|. |	.|. |	10	-0.066	-0.045	371.21	0.000
*|. |	.|. |	11	-0.126	0.056	374.05	0.000
*|. |	.|. |	12	-0.171	-0.048	379.30	0.000
*|. |	.|. |	13	-0.195	-0.008	386.16	0.000
**|. |	.|. |	14	-0.210	-0.010	394.20	0.000
**|. |	.|* |	15	-0.208	0.077	402.09	0.000
*|. |	.|. |	16	-0.202	-0.019	409.58	0.000
**|. |	*|. |	17	-0.212	-0.104	417.86	0.000
**|. |	*|. |	18	-0.242	-0.125	428.80	0.000
**|. |	*|. |	19	-0.276	-0.079	443.12	0.000
**|. |	*|. |	20	-0.315	-0.097	461.88	0.000
**|. |	.|. |	21	-0.315	0.069	480.72	0.000
**|. |	.|. |	22	-0.296	-0.011	497.57	0.000
**|. |	*|. |	23	-0.281	-0.071	512.78	0.000
**|. |	.|. |	24	-0.262	0.001	526.16	0.000
**|. |	.|. |	25	-0.245	0.052	537.88	0.000
**|. |	.|. |	26	-0.226	0.019	547.96	0.000
*|. |	.|. |	27	-0.185	0.046	554.74	0.000
*|. |	.|. |	28	-0.143	-0.035	558.86	0.000
*|. |	.|. |	29	-0.102	-0.040	560.97	0.000
*|. |	.|. |	30	-0.066	-0.042	561.85	0.000
.|. |	.|. |	31	-0.031	0.034	562.05	0.000
.|. |	*|. |	32	-0.019	-0.116	562.13	0.000
.|. |	.|. |	33	-0.001	-0.008	562.13	0.000
.|. |	.|. |	34	0.008	-0.035	562.14	0.000
.|. |	*|. |	35	-0.009	-0.107	562.15	0.000
.|. |	.|. |	36	-0.026	-0.002	562.30	0.000

There is evidence of serial correlation. Hence, inference is conducted using HAC robust standard errors.

In order to adjust for the smaller population which implies a smaller “normal” rate of mortality/month, I run a log-log regression:

```
Dependent Variable: LOG(MORTALITY)
Method: Least Squares
Date: 08/10/18   Time: 18:34
Sample: 2010M01 2016M12
Included observations: 84
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
        bandwidth = 4.0000)
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(POP_PR)	0.147249	0.174183	0.845370	0.4008
C	7.679160	0.217920	35.23850	0.0000
M02	-0.095134	0.019103	-4.979991	0.0000
M03	-0.019802	0.027377	-0.723319	0.4719
M04	-0.105667	0.025845	-4.088446	0.0001
M05	-0.080984	0.023136	-3.500380	0.0008
M06	-0.093184	0.028757	-3.240362	0.0018
M07	-0.062615	0.021694	-2.886295	0.0052
M08	-0.058015	0.025345	-2.288967	0.0251
M09	-0.089342	0.025658	-3.482093	0.0009
M10	-0.073707	0.021701	-3.396426	0.0011
M11	-0.071030	0.027649	-2.568938	0.0123
M12	0.029797	0.027024	1.102635	0.2740
DUMMY_OCT14	0.179984	0.006084	29.58270	0.0000

R-squared	0.536770	Mean dependent var	7.807738
Adjusted R-squared	0.450741	S.D. dependent var	0.060449
S.E. of regression	0.044800	Akaike info criterion	-3.222199
Sum squared resid	0.140494	Schwarz criterion	-2.817063
Log likelihood	149.3323	Hannan-Quinn criter.	-3.059337
F-statistic	6.239439	Durbin-Watson stat	1.111064
Prob(F-statistic)	0.000000	Wald F-statistic	280.7819
Prob(Wald F-statistic)	0.000000		

The interpretation of the regression is that a one percentage point increase in the population induces a 0.15 percent increase in the per month mortality rate.

```
Date: 08/10/18   Time: 18:36
Sample: 2010M01 2016M12
Included observations: 84
==========================================================================
                          Partial
 Autocorrelation        Correlation            AC       PAC   Q-Stat   Prob
==========================================================================
      . |***    |          . |***    |     1  0.394  0.394  13.545  0.000
      . |*.     |          . | .     |     2  0.196  0.048  16.925  0.000
      . |*.     |          . | .     |     3  0.115  0.027  18.107  0.000
      . | .     |          . | .     |     4  0.038 -0.027  18.241  0.001
      . | .     |          . | .     |     5  0.018  0.002  18.269  0.003
      .*| .     |          .*| .     |     6 -0.121 -0.153  19.631  0.003
      .*| .     |          .*| .     |     7 -0.168 -0.089  22.264  0.002
      .*| .     |          . | .     |     8 -0.164 -0.059  24.835  0.002
      **| .     |          .*| .     |     9 -0.205 -0.108  28.887  0.001
      .*| .     |          .*| .     |    10 -0.191 -0.067  32.446  0.000
      .*| .     |          . | .     |    11 -0.165 -0.045  35.149  0.000
      **| .     |          .*| .     |    12 -0.210 -0.140  39.575  0.000
      . | .     |          . |*.     |    13 -0.026  0.114  39.641  0.000
      . | .     |          . | .     |    14  0.014  0.001  39.661  0.000
      . | .     |          . | .     |    15  0.016 -0.029  39.687  0.001
      . | .     |          .*| .     |    16 -0.048 -0.131  39.936  0.001
      . | .     |          . |*.     |    17  0.051  0.081  40.218  0.001
      . | .     |          .*| .     |    18  0.023 -0.111  40.275  0.002
      . | .     |          . | .     |    19 -0.011 -0.060  40.287  0.003
      . | .     |          . | .     |    20  0.025  0.017  40.360  0.004
      . | .     |          . | .     |    21  0.013 -0.027  40.380  0.007
      . |*.     |          . |*.     |    22  0.129  0.115  42.306  0.006
      . | .     |          .*| .     |    23  0.006 -0.108  42.310  0.008
      . | .     |          . | .     |    24 -0.006 -0.017  42.315  0.012
      . | .     |          . | .     |    25 -0.027 -0.045  42.403  0.016
      . | .     |          . | .     |    26  0.003  0.036  42.404  0.022
      . | .     |          . | .     |    27  0.043  0.010  42.636  0.028
      . | .     |          . | .     |    28  0.068  0.035  43.238  0.033
      . | .     |          . | .     |    29  0.044  0.045  43.489  0.041
      . | .     |          . | .     |    30  0.047 -0.009  43.781  0.050
      . | .     |          . | .     |    31  0.043  0.017  44.027  0.061
      . | .     |          .*| .     |    32 -0.057 -0.117  44.485  0.070
      .*| .     |          . | .     |    33 -0.100 -0.063  45.901  0.067
      . | .     |          . | .     |    34 -0.057  0.057  46.362  0.077
      . | .     |          . | .     |    35 -0.027 -0.039  46.471  0.093
      .*| .     |          **| .     |    36 -0.191 -0.211  51.966  0.041
==========================================================================
```

The residuals appear Normal, as shown in the histogram below.

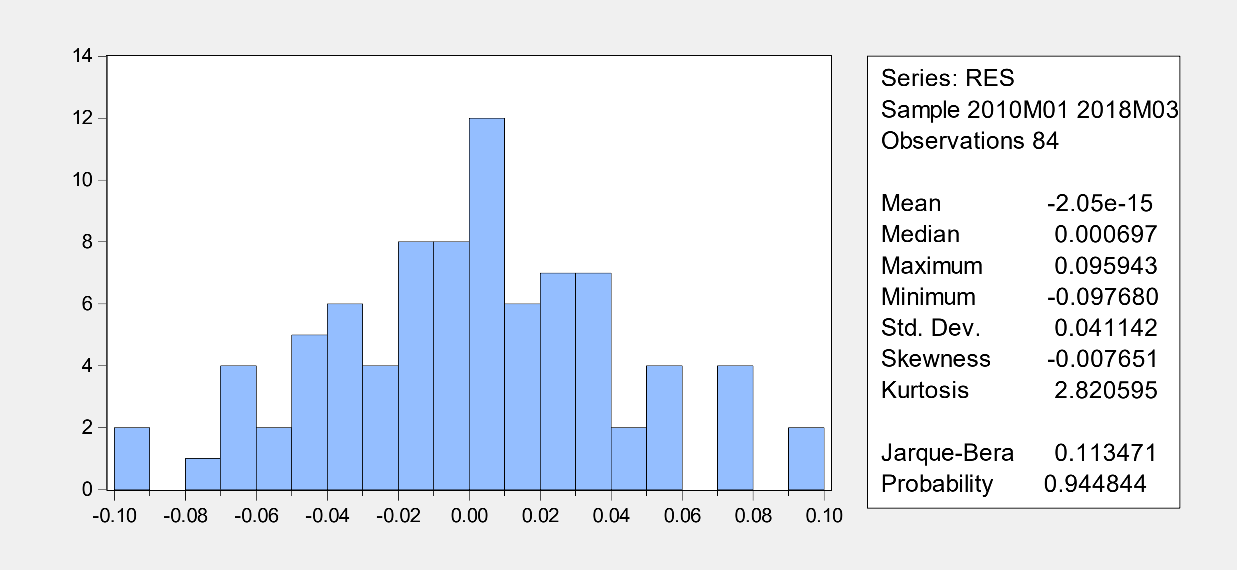

Figure 2: Residuals from log-log regression.

Hence, I use this specification as the equation to generate the key counterfactual.

Note a quantile regression would yield a larger coefficient on *population*, but is not justified on the basis of the diagnostics.

For completeness sake, I report the estimates:

Dependent Variable: LOG(MORTALITY)
Method: Quantile Regression (Median)
Date: 08/10/18 Time: 17:24
Sample: 2010M01 2016M12
Included observations: 84
Huber Sandwich Standard Errors & Covariance
Sparsity method: Kernel (Epanechnikov) using residuals
Bandwidth method: Hall-Sheather, bw=0.22184
Estimation successful but solution may not be unique

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(POP_PR)	0.319807	0.177654	1.800166	0.0761
C	7.457890	0.234453	31.80969	0.0000
M02	-0.093383	0.041294	-2.261409	0.0268
M03	-0.041078	0.039115	-1.050178	0.2972
M04	-0.119054	0.040657	-2.928251	0.0046
M05	-0.082573	0.037348	-2.210924	0.0303
M06	-0.081421	0.038707	-2.103537	0.0390
M07	-0.061344	0.036667	-1.673003	0.0988

M08	-0.046976	0.038134	-1.231857	0.2221
M09	-0.070510	0.042069	-1.676056	0.0982
M10	-0.073234	0.037114	-1.973228	0.0524
M11	-0.061715	0.052257	-1.180982	0.2416
M12	0.009579	0.044844	0.213597	0.8315
DUMMY_OCT14	0.185248	0.030040	6.166683	0.0000

Pseudo R-squared	0.309758	Mean dependent var	7.807738
Adjusted R-squared	0.181570	S.D. dependent var	0.060449
S.E. of regression	0.047213	Objective	1.280008
Quantile dependent var	7.805475	Restr. objective	1.854433
Sparsity	0.095375	Quasi-LR statistic	48.18267
Prob(Quasi-LR stat)	0.000006		

I plot the excess mortality series below:

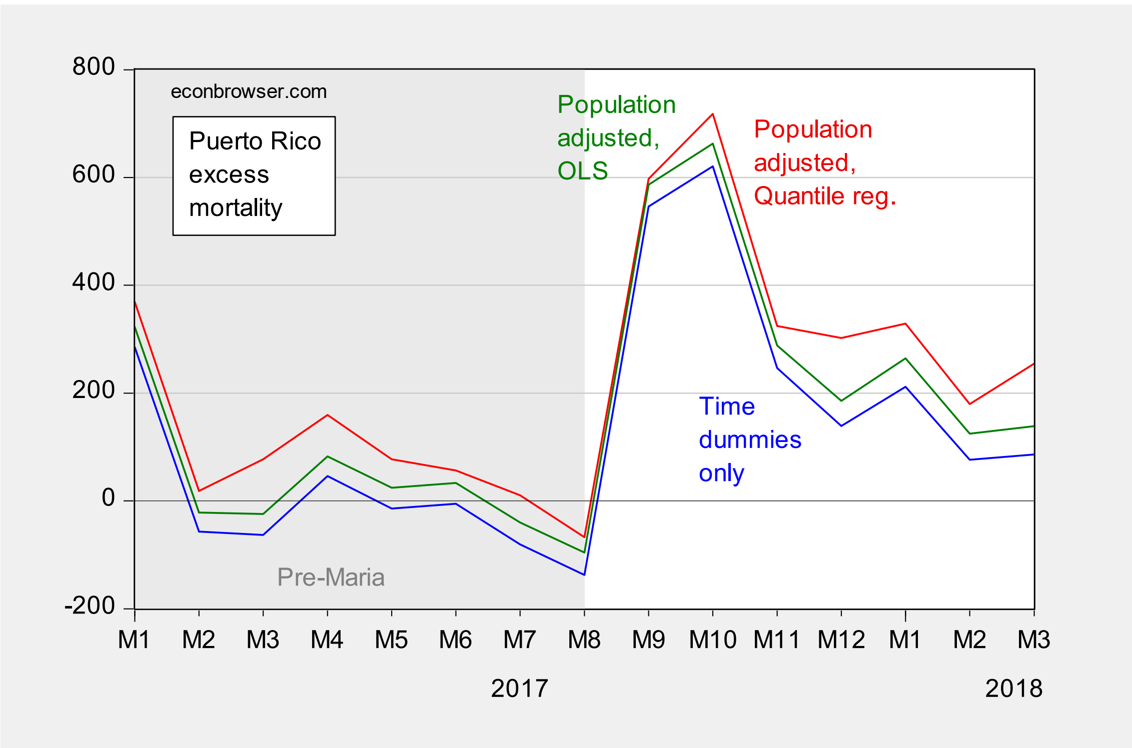

Figure 3: Deviations from predicted values, for simple time dummies OLS model (blue), OLS model adjusting for population (green), and Quantile Regression model adjusting for population (red). Gray shading denotes pre-Maria sample. Source: Author's calculations.

Estimates for deviations from upper 95% confidence interval, and for deviations from conditional means, below:

Table 1: Estimates of Excess Deaths, 2017M09-2018M03

	above 95% upper bound				deviation from mean		
	OLS, time dummies	QREG, pop	LS, pop		OLS, time dummies	QREG, pop	LS, pop
2017M01	33.10	58.81	65.86		284.74	369.13	322.34
2017M02	-286.37	-234.24	-262.28		-57.00	18.44	-21.79
2017M03	-307.32	-171.68	-271.57		-63.05	77.36	-24.07
2017M04	-171.51	-75.16	-143.11		45.86	159.50	82.49
2017M05	-231.07	-162.91	-208.30		-14.22	77.40	24.11
2017M06	-234.84	-187.40	-214.17		-5.51	56.45	33.06
2017M07	-297.30	-234.76	-274.26		-80.63	10.19	-40.21
2017M08	-365.57	-318.78	-345.85		-137.30	-67.38	-96.12
2017M09	*325.73*	*347.93*	*344.11*		546.13	597.52	586.51
2017M10	*404.71*	*477.25*	*431.94*		620.56	718.08	662.66
2017M11	*9.82*	-3.02	*28.03*		246.38	324.29	288.27
2017M12	-122.73	-2.69	-105.91		138.89	302.00	185.52
2018M01	-39.90	*7.93*	-0.92		211.74	328.97	264.79
2018M02	-153.37	-82.37	-125.97		76.00	179.82	124.55
2018M03	-158.32	-2.29	-118.07		85.95	254.66	138.55
TOTAL	**740.25**	**833.11**	**804.09**		**1925.65**	**2705.33**	**2250.86**

<u>Findings:</u>

Using the conservative approach of taking only entries above the 95% upper bound yields a baseline estimates of 740 excess deaths. This s below the Santos-Lozada and Howard (2018) estimate because (1) I have taken a different approach to estimating the conditional mean, and (2) accounted for serial correlation in calculating the standard errors. Accounting for population change, the excess rises to 804, still using HAC robust standard errors.

Using the deviation from conditional mean, and summing using the baseline model yields excess deaths of 1926; using the population adjusted OLS approach, I obtain an estimate of 2251 excess deaths (2705 using quantile regressions).

Note that these estimates extend only through March 2018. Presumably, there are still excess deaths given the destruction of infrastructure, and associated illnesses (e.g., the leptospirosis epidemic).

References

Santos-Lozada, Alexis R., and Jeffrey T. Howard. Estimates of excess deaths in Puerto Rico following Hurricane Maria. *SocArxiv*. (2017):1-8.

Santos-Lozada, Alexis R., and Jeffrey T. Howard. "Use of death counts from vital statistics to calculate excess deaths in Puerto Rico following Hurricane Maria." *JAMA* (2018).