PAA 2011 Extended Abstract.

Clique analysis of mortality co-movements: A new life expectancy time series analysis

Andrew Noymer[*,†,‡]
noymer@uci.edu

Tanya Jukkala[§]
tanya.jukkala@sh.se

Christopher S. Marcum[*]
cmarcum@uci.edu

17 September 2010

Abstract

In this paper we combine two separate types of analysis to form a new way of looking at mortality declines throughout the world. First, we use the Goodman-Grunfeld test to determine co-movements of life expectancy values. Goodman-Grunfeld time series analysis gives a nonparametric test of co-movement between two time series; it tests whether co-movements or counter-movements (i.e., aligned signs of first differences) are greater than would be expected, accounting for the trends of each time series, and correcting for the serial correlation of each series. Using this technique, we performed an all-pairwise-comparisons analysis of countries with life expectancy time series data in the Human Mortality Database (N=40). We are interested in co-movememts not counter-movements so we restrict the analysis thusly and use one-sided tests. From this analysis, we get a matrix of countries whose life expectancies are expanding in lock-step with each other,

---

[*] Department of Sociology, University of California, Irvine

[†] Department of Population Health and Disease Prevention, University of California, Irvine

[‡] Health and Global Change (HGC) project, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

[§] Stockholm Centre on Health of Societies in Transition (SCOHOST), Södertörns University, Stockholm, Sweden

statistically speaking. We then subject this connectivity matrix to clique analysis from the toolbox of social network analysis. We find that there are many groups of countries whose life expectancy is moving in lockstep. The largest is seven countries: Belgium, Denmark, Finland, Italy, Netherlands, England and Wales, Scotland; and there are 8 cliques of size 6, 13 of size 5, and 14 of size 4. We conclude that, historically at least, mortality decline is in remarkable synchrony worldwide, pointing to shared fortunes of environmental factors such as crop failures and epidemics.

1 Overview

1.1 Analytic strategy

Our analytic strategy has two parts. In the first part we use time series analysis, using the Goodman-Grunfeld test to establish pairs of countries where life expectancy is co-moving. In the second part, we use social network analysis to identify “cliques” — that is, groups of countries where every pairwise comparison within the group is significant. The cliques represent groups of countries that are all moving together in mortality — when $e(0)$ in one goes up, the others go up, and when $e(0)$ in one declines, the others decline, within a given level of statistical certainty (we use 5%).

We analyze all countries (N=40) in the Human Mortality Database (HMD). Two countries are co-moving if they posses a significant ($p < 0.05$) Goodman-Grunfeld test statistic for co-movement. We are interested in co-movements, only. There is no theoretical reason to anticipate that two countries may be counter-moving in life expectancy; therefore, we use a one-sided tests throughout.

We use a Neyman-Pearson not Fisherian framework so we do not consider very small p-values to be any “stronger” a connection than ones just below the threshold of significance. We do not use Bonferroni (or similar) adjustments, because our final outcome is clique membership, and this imposes a de facto correction for multiple comparisons.

We use social network analysis to find the cliques within the output matrix of the GG analysis. In § 2 below, we give more details of the techniques.

1.2 Interpretation of results

We find 17.5% of the HMD countries are in one large mortality-decline clique. We find this to be a relatively large proportion. It is not surprising that all the HMD countries have declining mortality; this is a basic fact of population studies. But we consider it to be quite surprising for so many countries to be moving together at such a fine (12-month) time scale.

We consider these results to be an elaboration, and amplification, of the finding that mortality has declined following a 'universal pattern' (Tuljapurkar et al., 2000). The presence of cliques of mortality decline shows that groups of countries not only have similar patterns of morality decline, but show that, *even on a year-to-year basis*, groups of countries follow each other in ups and downs of mortality.

As we refine our results, we will better present the geographic dimensions of the analysis. We find that cliques are strongest among geographically-close countries. For example, the largest clique (Belgium, Denmark, Finland, Italy, Netherlands, England and Wales, Scotland) is all northern European countries (except one southern European country, Italy). This points toward shared environmental effects such as crop failures and geographically-spreading epidemics. Shared social effects would produce more dispersed cliques, since interlocked economies can occur across the globe (for example, Great Britain and Australia) — however, we should stress that this dimension of the analysis is still preliminary.

2 Data analysis

2.1 Goodman-Grunfeld time series analysis

We use the Goodman-Grunfeld nonparametric test of comovement between two time series Goodman and Grunfeld (1961). This test is based on a chi-squared analysis of signs of differences of each time series, cross-classified in a 2×2 table. Thus, for two countries, we analyze if life expectancy, $e(0)$, both increases, or both decreases, or move in either permutation of opposite directions, on a year-to-year basis. Under the null hypothesis of no comovement between the two data series, the expected frequencies of the four cells of the table would be equal. Moreover, the Goodman-Grunfeld test corrects for serial correlation.

A sketch of the test follows; the particulars are given in the cited works Goodman and Grunfeld (1961); Goodman (1963). Take two time series $X = \{X_0, X_1, \ldots, X_n\}$ and $Y = \{Y_0, Y_1, \ldots, Y_n\}$, with subscripts for time periods. Create dummy (indicator) variables, $U = \{U_0, U_1, \ldots, U_{n-1}\}$ and $V = \{V_0, V_1, \ldots, V_{n-1}\}$ respectively, coded $\{0, 1\}$, to indicate period-to-period increase. That is, $U_i = 1$ if $X_{i+1} - X_i > 0$, and similarly for V_i and the Y series.

Then cross-classify U and V in a 2×2 table, the cell counts of which are labeled (left to right) a, b across the top row and c, d across the bottom row; the counts of comovements are a and d and the countermovements are b and c The Goodman-Grunfeld test statistic is:

$$\frac{a - A}{\sqrt{n[(a+b)(a+c)(b+d)(c+d)/n^4 + 2ef]}} \sim N(0,1), \text{where:}$$

$$e = \sum_{i=0}^{n'-1} U_i U_{i+1}/n' - [(a+b)/n]^2, \text{ and } f = \sum_{i=0}^{n'-1} V_i V_{i+1}/n' - [(a+c)/n]^2,$$

and where $n = a + b + c + d$, $A = (a+b)(a+c)/n$, and $n' = n - 1$ is the number of sequential pairwise comparisons of U and V. The term in ef is the correction for serial correlation. The test statistic is normally distributed because of the equivalence between a normal and a χ^2 distribution with one degree of freedom. Although not shown in the above formula, we use the 'continuity correction' (Yates, 1984), as Goodman and Grunfeld advise. The Goodman-Grunfeld test is more conservative (i.e., harder to reject the null hypothesis of no comovement) than a naïve χ^2 analysis of the comovements. We use one-sided tests throughout.

2.2 GG results

The connectivity matrix is given in the symmetric square matrix immediately below. Countries from the HMD are organized in rows and columns. A "1" indicates co-movement ($p < 0.05$, one sided) in the Goodman-Grunfeld test, and a "0" indicates lack of significance ($p \geq 0.05$). The diagonal has been coded as "1" since a country is co-moving with itself, but this is of course arbitrary — the diagonal has no meaningful interpretation. A country index will be added. These results form the input of

40 × 40 connectivity matrix:

$$\begin{pmatrix}
1000100000000010000000000000000000000000 \\
0100000100110000000000000000110100000000 \\
0010000001000000000011000010110000000010 \\
0001000010110100001000010100000010011000 \\
1000100000000000000000000011000000000000 \\
0000010000000000001000010100000000001101 \\
0000001000000000000000000000000000000000 \\
0100000110111010001000000010110100000000 \\
0001000110101000001000010100000010011000 \\
0010000001000000000011010000100000000000 \\
0101000110111010001000010100000111011100 \\
0101000100110110001000010000000011010000 \\
0000000110101000000001000110010011000001 \\
0001000000010100001000010100000001011001 \\
1000000100110010001010000010010001000000 \\
0000000000000001000000011000000000011100 \\
0000000000000000100000000101000000001100 \\
0000000000000000010000000000000000000000 \\
0001010110110110001000010000010001011000 \\
0000000000000000000100001000000000100000 \\
0010000001000010000011010000100000000000 \\
0010000001001000000011000000100000000010 \\
0000000000000000000000100010000000000000 \\
0001010011110101001010010100000001011100 \\
0000000000000001000100001000010000010101 \\
0001010010101100100000010100000010001101 \\
0010100100001010000000100011110100000000 \\
0000100000000000100000000011000100000000 \\
0110000101000000000011000010100000000010 \\
0110000100001010001000001010010000000000 \\
0000000000000000000000000000001000000000 \\
0100000100100000000000000011000100000100 \\
0001000010111000000000000100000011010001 \\
0000000000111110001000010000000011011001 \\
0000000000000000000100000000000000100000 \\
0001000010110101001000011000000011011100 \\
0001010010100101101000010100000001011100 \\
0000010000100001100000011100000100011100 \\
0010000000000000000001000000100000000010 \\
0000010000001100000000001100000011000001
\end{pmatrix}$$

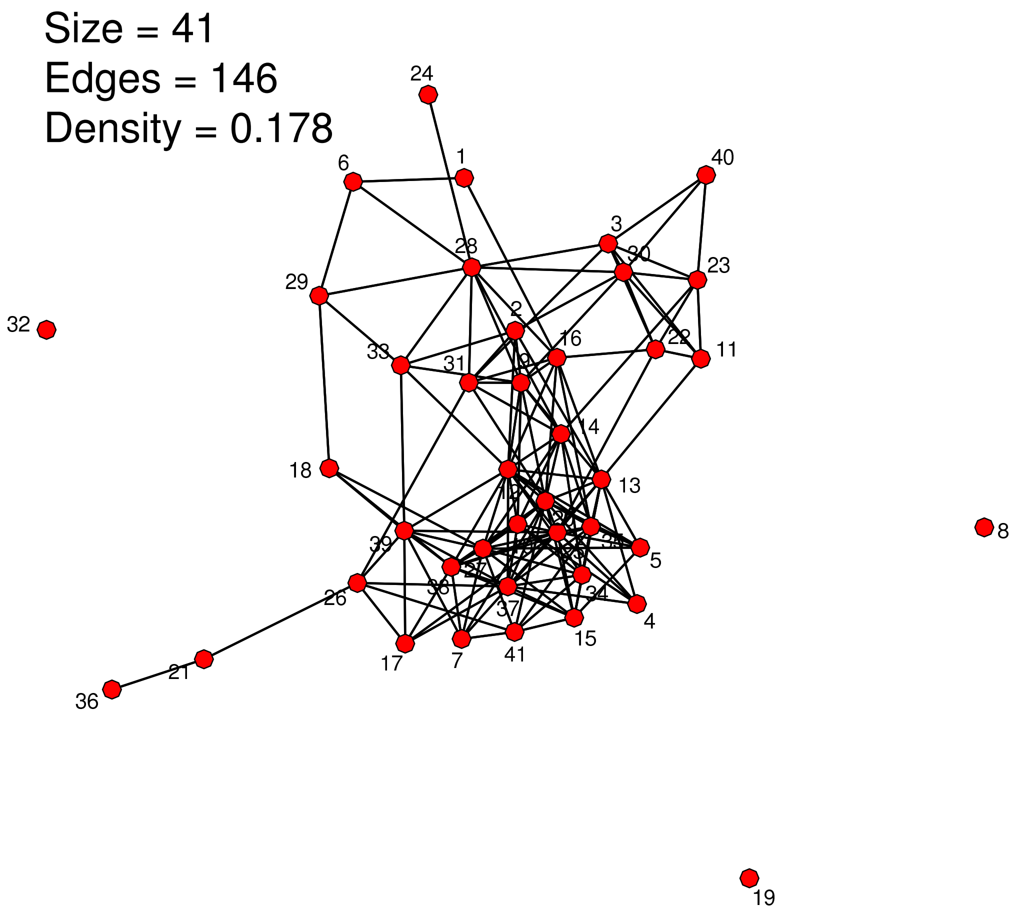

Figure 1: Social network of countries with co-moving mortality declines.

2.3 Social network analysis

Cliques are important subfeatures of connected networks. Cliques help to identify cohesive subgroups in a network, and to highlight local redundancy in information channels. In social network analysis, a clique is defined as a maximal complete subgraph (Wasserman and Faust, 1994). That is, it is a grouping of vertices in a network such that each vertice in the group is connected to all other members of that group. Clique size is determined by the maximal number of vertices present in the complete subgraph that constitutes the clique.

A clique census was conducted on the network featured in Figure 1. This a relatively sparse network with three isolates and a moderate network density of only 0.19. The algorithm used for the clique census was a variation of Makino and Uno's (2004) work as implemented by Butts (2009), which is efficient on networks of this size and density. The results of the clique census are given in Table 1. In this table, the columns index the size of the clique and the rows index the vertex labels of the clique members. Excluding the three isolates (i.e. a cohesive subgroup of size 1), there are a total of 60 cliques in this network. The largest clique size is order 7, while the modal clique size is order 4, with 14 cliques of that size and 27 unique members. The membership into a clique with the modal size is non-trivial; 68% of the actors in this network belong to a clique of this size. However, the distribution of clique co-membership is sparse, as depicted in Figure 2. The off-diagonal cells of this sociomatrix are shaded in proportion to the number of times actor i share a clique with actor j, where i and j index rows and columns. The shading of the cells on the diagonal of this sociomatrix are proportional to the i^{th} actor's total number of clique memberships.

Table 1: Vertex Clique Membership By Clique Size

		Clique Size					
	Isolates	**2**	**3**	**4**	**5**	**6**	**7**
Clique Members	31	23,27	27,28,32	17,26,37,38	16,24,36,37,38	4,9,11,24,26,37	4,9,11,19,24,36,37
	18	20,35	5,27,28	16,25,36,38	6,24,26,37,38	11,19,24,34,36,37	
	7	20,25	6,26,40	6,19,24,37	11,24,36,37,38	11,12,19,24,34,36	
		17,28	14,34,40	3,22,29,39	11,24,26,37,38	4,14,19,24,36,37	
		25,40	14,26,40	2,8,11,32	4,9,11,33,36	14,19,24,34,36,37	
		25,30	3,27,30	8,15,19,30	4,9,11,26,33	12,14,19,24,34,36	
		15,21	3,27,29	8,15,27,30	11,12,33,34,36	4,12,14,19,24,36	
		1,15	10,21,24	8,13,27,30	4,11,12,33,36	4,11,12,19,24,36	
		13,22	11,32,38	13,26,33,40	4,14,24,26,37		
		1,5	2,8,30	13,33,34,40	3,10,21,22,29		
			2,8,29	11,13,33,34	11,12,15,19,34		
			8,27,32	8,9,11,19	8,11,12,15,19		
			8,27,29	8,9,11,13	9,11,13,26,33		
				2,8,11,12			
N	3	10	13	14	13	8	1

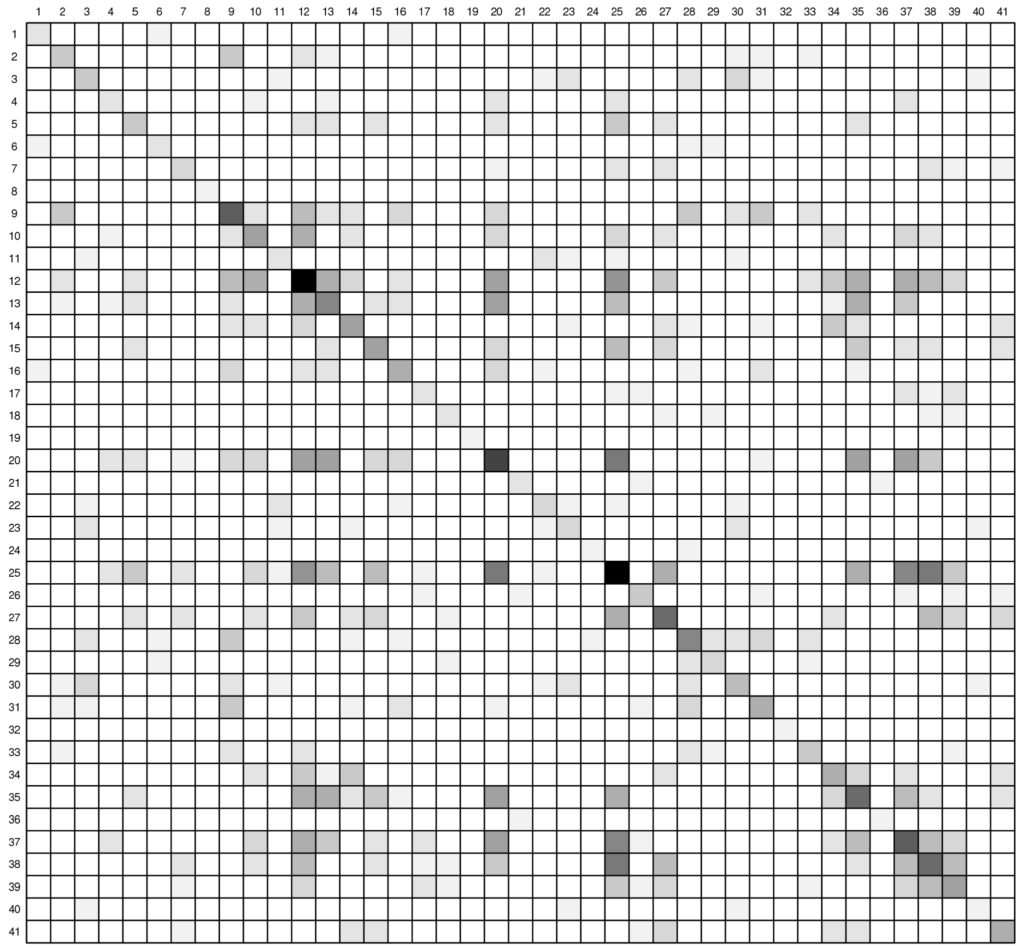

Figure 2: Sociomatrix; see text for discussion.

Works Cited

Butts, Carter T. 2009. “`sna`: Tools for Social Network Analysis. version 2.0.” R package.

Goodman, Leo A. 1963. “Tests based on the movements in and the comovements between m-dependent time series.” In Carl F. Christ, Milton Friedman, Leo A. Goodman, Zvi Griliches, Arnold C. Harberger, Nissan Liviatan, Jacob Mincer, Yair Mundlak, Marc Nerlove, Don Patinkin, Lester G. Telser, and Henri Theil (eds.), *Measurement in economics: Studies in mathematical economics and econometrics in memory of Yehuda Grunfeld*, chap. 10, pp. 253–269. Stanford University Press, Stanford, California.

Goodman, Leo A. and Yehuda Grunfeld. 1961. “Some nonparametric tests for comovements between time series.” *Journal of the American Statistical Association* 56(293):11–26.

Makino, Kazuhisa and Takeaki Uno. 2004. “New algorithms for enumerating all maximal cliques.” In Torgen Hagerup and Jyrki Katajainen (eds.), *Algorithm theory — SWAT 2004: 9th Scandinavian workshop on algorithm theory, Humlebæk, Denmark, July 8–10, 2004, Proceedings*, pp. 260–272. Springer, Berlin.

Tuljapurkar, Shripad, Nan Li, and Carl Boe. 2000. “A universal pattern of mortality decline in the G7 countries.” *Nature* 405:789–792.

Wasserman, Stanley and Katherine Faust. 1994. *Social network analysis: Methods and applications*. Cambridge University Press, Cambridge.

Yates, F. 1984. “Test of significance for 2×2 contingency tables [with discussion].” *Journal of the Royal Statistical Society, Series A* 147(3):426–463.