THE HAZARDS of TRANSFORMING INFANT MORTALITY with LOGS: THE GENERAL PATTERN of 20TH CENTURY INFANT MORTALITY DECLINE in 22 COUNTRIES

David Bishai
Marjorie Opuni-Akuamoa
Margaret Weden

September 21, 2006

150 Word Abstract
Statisticians since Preston have approached the task of exploring factors associated with Infant Mortality Rate (IMR) decline by first applying a logarithmic transformation to IMR. This paper assesses how correct the log transform is by nesting the log transform within the more general Box Cox $[(y^{\lambda}-1)/\lambda]$ transformation. If $\lambda=1$ the Box Cox transform amounts to a simple linear model. As λ approaches 0 Box Cox converges to a logarithmic specification. Studying 22 country's IMR time series for periods that typically stretch from late 1870 to late 1988 revealed the log transform is the best fit to IMR in only 4 of 22 countries. The assumption of linearity is best for 2 of 22. Further analysis shows that the logarithmic assumption can lead to substantial biases in estimating the coefficient of IMR on covariates like GDP per capita. Log-IMR is not generally the best transformation for the analysis of IMR.

Introduction: Time trends in infant mortality for the 20th century show a curvilinear pattern that most demographers have assumed to be approximately exponential. Virtually all cross country analyses of GDP vs IMR and time series analysis of infant mortality have studied the logarithm of infant mortality to account for the curvilinear time trend. However, there is no evidence that the log transform is the best fit for infant mortality time trends.

Objective: To use maximum likelihood methods to determine the best transformation to fit time trends in infant mortality reduction in the 20th century and to assess the importance of the proper transformation in identifying the relationship between infant mortality and GDP per capita.

Methods: The Box Cox transform is of the form $(y^{\lambda}-1)/\lambda$. If $\lambda=1$ the Box Cox transform amounts to a simple linear model. As λ approaches 0 Box Cox converges to a logarithmic specification.

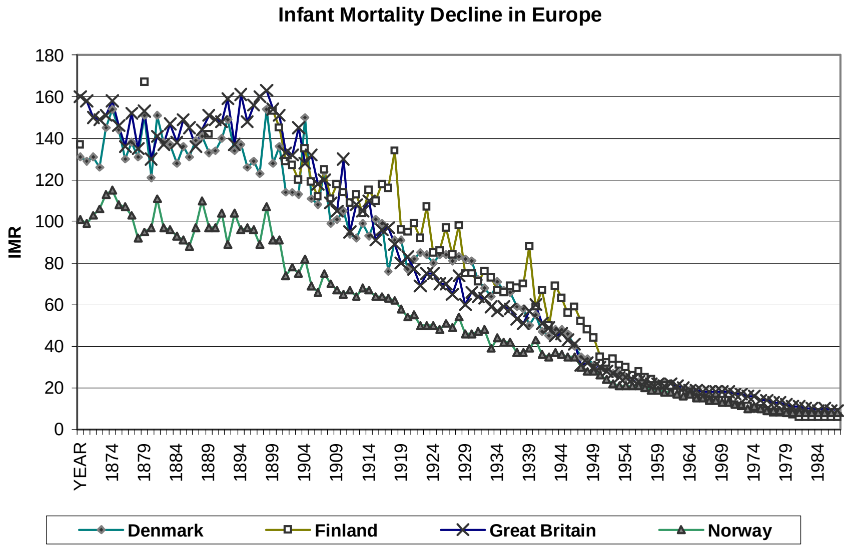
We applied the Box Cox transform to IMR and then used maximum likelihood methods to identify the best fitting value of λ for each of 22 nations' IMR data for 1870-1988 and for the pooled sample. We tested each the value of lambda in each country against the null that $\lambda=1$ and against the null that $\lambda=0$ for each country. To demonstrate the importance of selecting the proper transformation we compare regressions of log(IMR) on same year GDP per Capita against Box Cox transformed models.

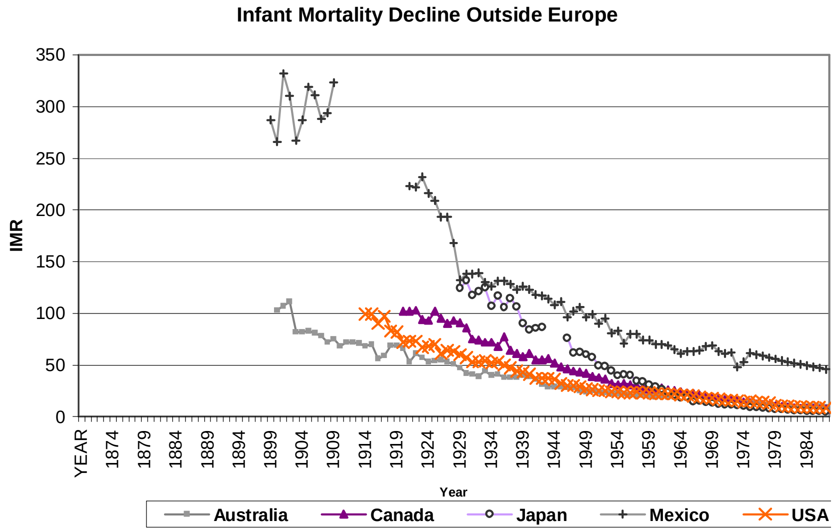
Results:
Figure 1 shows the patterns of infant mortality seen in the 22 countries, demonstrating that visual inspection of the curves cannot reliably establish the linearity or log-linearity of the IMR declines.

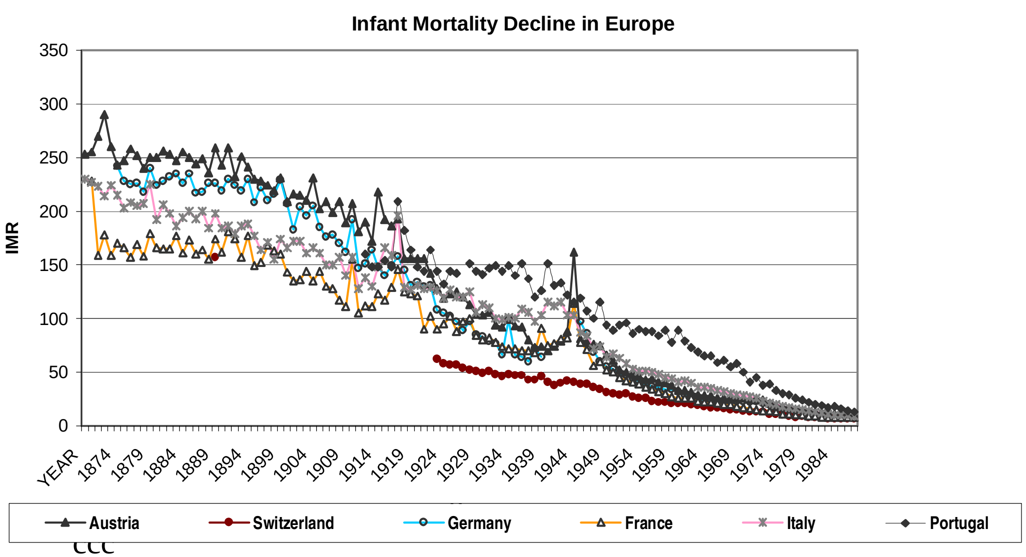
Table 1 shows the results of the Box-Cox models where IMR was regressed against year for each country. Based on Chi-Squared test statistics infant mortality decline was best described as an exponential decline in Australia, Netherlands, Uruguay, and USA. IMR decline was best described as linear in Italy, and Portugal. For the remaining 16 countries IMR decline was neither best modeled as a linear process nor as logarithmic. Table 2 compares the coefficients on GDP per capita from regressions of a transformation of IMR against GDP per capita, and time in years.

Discussion:
The assumption that IMR declines are logarithmic is enshrined in the Preston curve and in nearly all cross-country as well as time series analyses of IMR data since Preston's 1975 paper. The logarithmic assumption is seldom correct. We found that only 4 of 22 IMR time series were best described as logarithmic. Furthermore we found that the assumption that IMR declines are logarithmic can induce large biases in estimating the association between IMR and GDP per capita. Statistical analyses of IMR time series should assess the robustness of their findings to transformations other than the log transform.

Figure 1. Twentieth Century Infant Mortality Decline







ccc

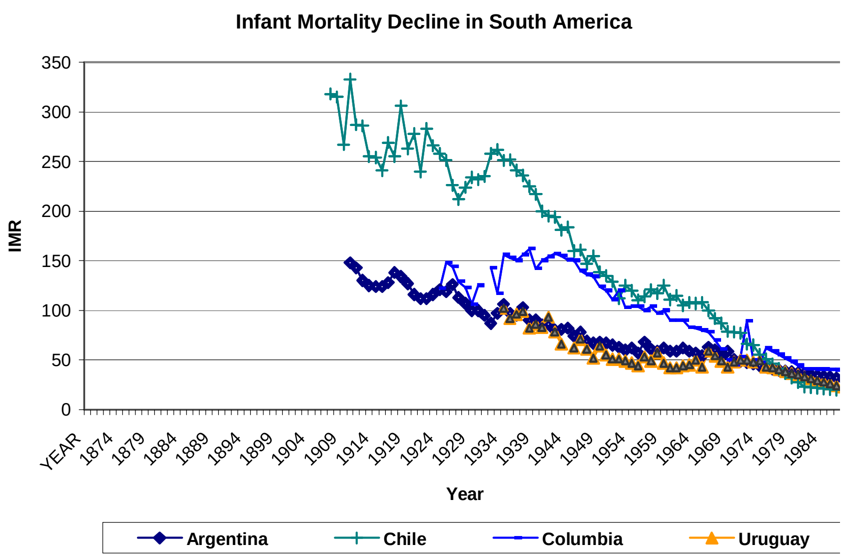


Table 1. Results from Box Cox Models in Which $(IMR^{\lambda}-1)/\lambda$ is regressed against time in years separately for each country.

Country	Years of Data	Best Fit Lambda	Chi-Squared Statistic for Null that λ=0 Results with *** Reject Log Transformation	Chi-Squared Statistic for Null that λ=1 Results with *'s Reject Linearity
Whole Sample		0.251	(184.14)***	(1377.59)***
Argentina	1911-1988	0.288	(11.471)***	(51.766)***
Australia	1890-1988	0.054	(2.163)	(207.887)***
Austria	1870-1988	0.628	(125.949)***	(41.217)***
Belgium	1870-1988	0.839	(105.406)***	(4.053)**
Canada	1921-1988	0.269	(79.775)***	(196.649)***
Switzerland	1924-1988	0.306	(66.435)***	(207.223)***
Chile	1908-1988	0.879	(119.864)***	(3.016)*
Colombia	1925-1988, excl. 1932	0.537	(5.928)**	(4.133)**
Germany	1870-1988	0.482	(105.738)***	(92.143)***
Denmark	1870-1988	0.677	(108.069)***	(26.512)***
Finland	1900-1988	0.625	(78.650)***	(32.958)***
France	1871-1988	0.776	(144.699)***	(15.209)***
Great Britain	1870-1988	0.429	(55.753)***	(75.640)***
Italy	1870-1988	0.969	(231.167)***	(.445)
Japan	1930-1988 exc.1944-1946	0.139	(15.503)***	(164.559)***
Mexico	1900-1910, 1922-1988	-0.331	(22.468)***	(174.534)***
Netherlands	1900-1988	-0.023	(.352)	(226.199)***
Norway	1870-1988	0.560	(127.990)***	(82.470)***
Portugal	1913-1988, excl. 1928	0.971	(90.991)***	(.104)
Sweden	1870-1988	0.502	(141.359)***	(135.513)***
Uruguay	1935-1988, excl. 1945	0.014	(.005)	(21.209)***
USA	1915-1988	0.010	(.079)	(192.968)***

***$p<0.01$ **$p<0.05$ *$p<0.10$

Table 2. Comparing coefficient results based on alternative assumptions about the curvature of IMR time series. In each model the independent variables are a constant, time in years, and GDP per capita in $1000 per person. Coefficients on GDP per capita are shown below.

	Box Cox Model λ="Best Fit"		Log Transform Model Assume λ=0		
	Coefficient on GDP per Capita		Coefficient on GDP per Capita		Coefficient Gap
Whole Sample	-0.142	***	-0.172	***	**-0.030**
Argentina	0.307	***	0.065	**	**-0.242**
Australia	0.018		-0.005		-0.023
Austria	-0.219	***	-0.155	***	**0.064**
Belgium	-0.153	***	-0.21	***	**-0.057**
Canada	-0.009		-0.068	***	-0.059
Switzerland	-0.055	**	-0.073	***	**-0.018**
Chile	-7.136	**	-0.13		7.006
Colombia	-0.526	***	-0.513	***	**0.013**
Germany	-0.152	***	-0.136	***	**0.016**
Denmark	-0.282	***	-0.182	***	**0.100**
Finland	-0.073	***	-0.19	***	**-0.117**
France	-0.130	***	-0.19	***	**-0.060**
Great Britain	-0.163	***	-0.146	***	**0.017**
Italy	-0.271	***	-0.192	***	**0.079**
Japan	-0.055	***	-0.053	***	**0.002**
Mexico	0.030	**	0.086	***	**0.056**
Netherlands	-0.030	***	-0.028	***	**0.002**
Norway	-0.132	***	-0.094	***	**0.038**
Portugal	-1.415	***	-0.399	***	**1.016**
Sweden	-0.157	**	-0.138	***	**0.019**
Uruguay	-0.148	**	-0.088		0.060
USA	0.004		0.001		-0.003

Italicized countries have IMR time series that are best fit as logarithmic. Bold coefficient gaps are for countries where coefficients are significant in both models.
***p<0.01 **p<0.05 *p<0.10 for null that coefficient=0