Estimates of effects of income, public health and medical interventions on mortality decline.

March 8, 2018

1 The problem and classic solutions

1.1 McKeown:

Formulates the problem as follows: the secular mortality decline in England begins roughly in 1830-40 and proceeded without interruption past 1980. There are four main candidate explanations for the initial (1840-1920) decline: (a) change of balance between hosts and infectious disease agents; (b) improvements in nutrition; (c) public health (water, sewage, food storage and transportation) and (d) medical interventions. McKeown's inferential strategy is to use some historical evidence and dismiss (a), (c) and (d) leaving only (b) standing. But in his work one finds not one shred of empirical evidence to support the idea. The post-McKeown research is designed to find that evidence. Below we review the most important work done. Unless otherwise noted we follow the following notation throughout:

$E(t)$ =life expectancy at birth at year t
$Y(t)$ =income per capita at year t
$Z(t)$ =other covariates at year t

1.2 Preston I:

Starts from the idea that individual income causally influences health and mortality. From this he jumps to an aggregate model where life expectancy is a function of aggregate income. The model is logistic and has a ceiling parameter that helps to deal with the concavity of the relation. There are no other variables in the model, not even a measure of income inequality which should be included given the non-linear relation between income and mortality at the individual level. As in McKeown, the quantity of interest is the contribution to mortality decline associated with changes in income. This is obtained as follows (the parameterization below is not exactly what Preston used but is more general than his)

$$E(t_1) = \alpha_1/(1 + exp(\beta_1 - \gamma_1 Y(t_1))$$

$$E(t_2) = \alpha_2/(1 + exp(\beta_2 - \gamma Y(t_2))$$

Presumably unmeasured factors, including public health (water, sewage, infrastructure) and medical technology are reflected in parameters $\alpha's$.Their influence is retrieved by computing counterfactual values of $E(t_1)$ and $E(t_2)$: the first counterfactual uses the independent variables in $t_1(1930)$ in the second equation and the second uses values of the independent variables in t_2 $(1940 - 60)$ in the first equation. These counterfactuals estimates are measures of the magnitude of the income independent shift in life expectancy. And the shift is supposedly induced by exogenous interventions that operate between t_1 and t_2. They have limitations: (i) they are not properly weighted (Preston uses the arithmetic average of both counterfactual measures rather that employing better justified weights); (ii) they reflect changes in the other two parameters but particularly in the γ's. These are not interpretable as income independent changes; (iii) they include effects of both public health and medical

interventions as neither is measured and at least one of them is definitely related to national income (public health in the form of infrastructure). Note that, in theory at least, one could have as many cross sections for $E(t)$ as desired and multiple counterfactuals could be computed. This has been done with additional data up until 2010 (see Hussain below as an example).

1.3 Preston II:

In a second paper Preston changes strategy and estimates the following models

$$E(t_1) = \alpha_1 + \beta_1 lnY(t_2) + \gamma_1 Z(t_1)$$

$$E(t_1) = \alpha_2 + \beta_2 lnY(t_2) + \gamma_2 Z(t_2)$$

In addition, a change model is also estimated

$$\Delta E(t_1, t_2) = \alpha_3 + \beta_3 ln\Delta Y(t_1, t_2) + \gamma_3 \Delta Z(t_1, t_2)$$

where the symbol $\Delta G\ (t_1, t_2)$ denotes change in the function G between (t_1, t_2). Using the first two models one can compute counterfactuals (1940 and 1970). In the third model the counterfactual is immediate and results from predicting the dependent variable (a change) using observed income change and then subtracting from the total change (observed or predicted).

1.4 Rodgers:

Uses a number of model specifications the most important of which are:

$$E(t) = \alpha + \beta(1/lnY)^2$$

$$E(t) = \alpha + \beta(1/Y)^2$$

and in all cases controlling for Gini Index and setting a ceiling value for $E(t)$ to deal with non-linearities. This ceiling parameter is not emprically estimated but guessed. Since Rodgers is only interested in the returns to income, his analysis does not clarify the role of other determinants. However, one could use parameter estimates of the model to replicate shift analysis as before.

1.5 Palloni-Wyrick:

Use Box Cox but assigning 3 extreme values (0, -1 and 1) to the Box_Cox transforming parameters resulting in the following models

$$E(t) = a + blnY(t)$$

$$lnE(t) = a + blnY(t)$$

$$E(t)^{-1} = a + bln(Y(t)$$

Palloni-Wyrick used standardized death rates rather than life expectancy but interpretations are similar. Shift analysis is also possible within each of the three specifications. In addition, they estimate separately models by causes of death (causes specific standardized mortality rates). This makes analysis of shifts more interpretable: for example, a shift analysis of mortality due to infectious diseases should show a larger contribution of "medical" interventions than, say, the shift of death due to cancers. By the same token, the added values of shifts associated with each of cause of deaths must (approximately) be equal to the shift in the total death rate.

1.6 Hussain:

Computes multiple "Preston" curves for different periods (as suggested above). But his preferred model is a time series that enables testing "Granger causality". The model consists of two equations, one for E and one for Y as follows:

$$E(t) = \alpha + \sum_l \lambda_l E(t-l) + \sum_l \delta_l Y(t-l)$$

$$Y(t) = \alpha + \sum_l \mu_l Y(t-l) + \sum_l \theta_l E(t-l)$$

Although this model is not designed to answer the key question, namely, how much of the total change in life expectancy within a single period is due to changes in income, one could construct a statistic to answer it. However, I think the statistic will not be comparable to the others we computed above using "shift" analysis due to presence of the lagged term for life expectacny (a value added model useful to remove the influence of omitted covariates)

2 Non classic solutions

There are a number of ways to estimate parameters that depart from the specifications above. These are all based on the idea that one should fully use the panel nature of the data and, in particular: (a) introduce country fixed and random effects; (b) lagged variables; (c) dynamic panel estimation schemes (as, for example, Arellano-Bond) that permit easy constructions of IV to remove endogeneity.

2.1 Birchenall:

As Palloni-Wyrick, he uses death rates (not standardized) instead of life expectancy. He also uses mortality rates by causes of death and his model can be written as follows:

$$D_{ij}(t) = \alpha_j + \beta_j lnY_i(t) + \gamma_j A_i(t) + \delta_j Z_i(t)$$

$D_{ij}(t)$ is death rate in country i due to cause j and $A_i(t)$ refers to (unmeasured) technology that influences mortality at time t. This variable eventually becomes $A_i(t) = t$, namely, a secular trend that picks up changes in $D_{ij}(t)$ not explained by income, $Y_i(t)$. $Z_i(t)$ refers to additional controls. He then suggests a Oaxaca-Ramson decomposition but excluding the term $\gamma_j A_i(t)$. The effect of time (γ_j) is simply a measure of the "shift" in the death rate due to cause j due to factors other than income. Two key features of this approach are (i) the inclusion of $A_i(t) = t$ and (ii) the strategies to estimate the model including country fixed effects, lagged values of the death rates, first differences of income, and lagged values of other variables (Arellano-Bond approach)

2.2 Pritchett-Summers

The model is similar to that of Birchenall and so is the estimation approach:

$$lnE_i(t) = \alpha + \beta lnY(t) + \delta Z(t) + \mu_t + c_i$$

with i again denoting country. The model includes country and time fixed effects. The time fixed effects pick up the shifts due to exogenous improvements and diffusion. They estimate this model using instrumental variables and go on to use multiple instruments, separately and jointly, and with the conventional technique: first predict $Y(t)$ using an instrument and then use predicted values of $Y(t)$ to regress with $E(t)$. In this paper, the shift analysis is replaced by estimation of coefficient of time fixed effects.

2.3 Oeppen multilevel model

Oeppen formulates a model that recognizes the hierarchical nature of the data and estimate country-specific parameters. The general model is as follows:

$$E_i * (t) = (\beta_{oo} + \beta_{oi}) + (\beta_{1o} + \beta_{1i}) * t + (\beta_{2o} + \beta_{2i}) * Y_i(t)$$

where i denotes country and t denotes a measure of time since, say, 1900. One can add other controls to the model. The key is to estimate country specific parameters affecting the shifts over time, the effects of income, and the unmeasured or unexplained effects (constant). The value $E_i * (t)$ refers to a transform of life expectancy that removes non-linearities: $E_i * (t) = ln([E_i(t)/(100 - E_i(t))]$, a sort of log odds of life expectancy (which complicates interpretation of the coefficients!). The model is estimated using HLM software. Note that there many parameters to estimate: a set of at least three for total sample and 3 per country.

3 Our approach

In the chapter on determinants I would like to carry out analysis that considers the following issues: (a) functional form of model; (b) uncertainty of model and (c) uncertainty of measure of life expectancy.

3.1 Functional form of the model

The existing approaches can be classified into those that do not exploit the panel nature of the data (all models in section 1 above) and those that do (all models in section 2 above). I think we should only pose models that fully use the cross sectional time series nature of the data. Of all the approaches in section 2, Oeppen's is the most complete because it considers country-time-specific effects in a full-blown manner. THis surely leads to overfitting but we will deal with that in due time. The parameters β_{oi}'s are tantamount to country-fixed effects; the parameters β_{10} are equivalent to time fixed effects but are country specific as β_{1i} modifies the "baseline" time fixed effects. Each country is allowed to diverge from the average in terms of effects of income on the measure of mortality. The original specification does not exclude the possibility of using lagged values of mortality measures to remove endogeneity. It does also allow the introduction of other controls and doing so should modify the estimates of time-fixed effects and, therefore, the shifts. The only modification I would introduce a change in the definition of the dependent variable to make interpretation of coefficients more palatable; use log form for both life expectancy and income. This settles the issue of functional form.

3.2 Model specification and uncertainty: I

The specification of the model can be done in a number of ways. The description below is from the simplest to the most complex:

1. *The role of time*: We use a continuous variable for time but with modifications. The modifications are theoretically motivated as follows: we know that along the entire period under scrutiny there were some exogenous shocks that affected all countries, some more, some less. These shocks are: (i) malaria, hookworm and yellow fever eradication campaigns during 1920-1950 in countries of the Caribbean and Central America. We can get very precise dates of onset of these campaigns as well as indicators of their duration and effectiveness; (ii) diffusion of antibiotics and sulfas: these may not have taken place before 1940 and were fully introduced in different countries at different times, but almost surely after 1940 in all of them. Their effects may have operated with different lags; (iii) massive vaccination campaigns starting only with EPC WHO programs (after 1975); (iv) PHC and other interventions after 1985 or so. One would expect that at each of these times (or within narrow windows of time following the key dates) there should be mortality shifts reflected in the data. To assess these we can modify the basic model as follows: let T_j be the year that exogenous intervention event j takes place for sure and define a dummy variable $\tau_j = 0$ for all $t \leq T_j$ and $\tau_j = 1$ for all $t > T_j$. The model then becomes:

$$lnE_i * (t) = (\beta_{oo} + \beta_{oi}) + (\beta_{1o} + \beta_{1i}) * t + (\beta_{2o} + \beta_{2i}) * lnY_i(t) + (\varphi_{o3} + \varphi_{oi}) * \tau_j$$

and the parameter φ reflects the shift in mortality levels after time T_j attributable to the exogenous shock. In theory, one could introduce representation of more than one exogenous event. In practice, there will be identification problems. A number of remarks are in place. First, the size

of the shift is allowed to be country specific. Second, we did not allow for lagged effects of the interventions but we know that interventions such as diffusion of antibiotics is country-specific and may even depend on levels of income. Interactions with income can be handled easily but intercountry lag heterogeneity is more difficult to accommodate without creating identification problems. To allow lags we should make τ_j country-specific in the following sense: suppose that for country i the lag for exogenous intervention j is assumed to be l_{ij} ; then $\tau_{ij} = 1$ if $t > T_j + l_{ij}$. Third, one could also let the response of $lnE_i * (t)$ to the passage of be variable and depend on the exogenous event. In this case we are assuming that $\beta_{10} = \beta_{100} + \xi_{0j} * \tau_j$ with a similar specification for β_{1i}. This is, however, different that allowing for the event to produce a one-step shift of levels of life expectancy

2. *The role of the past*: the magnitude of life expectancy gains in an interval of time may depend on the nature and/or levels of mortality in the past. A country where malaria is pervasive for example may stand to gain a lot more from eradication campaigns than what would be expected from the size of deaths rates due to malaria alone. This is because there are synergies involving the immune system and exposure/contraction of other diseases. To pick some of these effects, which otherwise must be reflected in estimates of "shifts", one could control for levels of malaria death rates at a time before t or, equivalently, using a control for the lagged value of mortality due to malaria. A similar reasoning applies to other illnesses which cause immunodepression and whose reductions will lead to synergies or multiplying effects;

3. *The role of public health*: controls for variables that reflect expenditures in infrastructure such as access to potable water, sewage and electricity will depress the effects of income and the magnitude of the "time shifts" defined above. The model could be estimated first without controls and then with controls. One can then compare the estimated magnitude of the shifts (coefficients of the variables for time) in both cases to evaluate how much of the original shift is associated with public health.

4. *Key statistics*: for each model we define the key statistic to be the magnitude of change in life expectancy associated with non-income changes. In the simplest model where there is a single variable for time (the secular trend) the statistic can be computed by first choosing a time interval, say t_1 and $t_1 + k$ over which we want to assess the shift. We then compute

$$\Delta E_i(t, t+k) \simeq E_i(T_{*j}) * (\beta_{1o} + \beta_{1i}) * k$$

where T_* is set to be the midpoint of the interval. This is equivalent to the expected change in life expectancy in the absence of income changes during the period. Because there are country specific effects, the statistic will differ by countries and can be compared with the average value of the statistic in the sample.

If we want to capture effects of shifts due to exogenous events we need to employ the model that adds the τ_j and ensure the interval of time over which we compute the counterfactual changes include T_j as the middle of the period. Thus, suppose that one examines the period $(t, t+k)$ and that we center this interval of time on T_j, the timing of the jth intervention. Then, we can compute the change in $E(t)$ for each country i over that period of time associated with shifts in medical technology/public health as follows (need to verify this):

$$\Delta E_i(t, t+k) \simeq E_i(T_j)\{(\beta_{1o} + \beta_{1i}) * k + k/2 * (\varphi_{o3} + \varphi_{oi})\}$$

and this can be compared with the total change in $E_i(t)$ during the period to obtain an estimate of the "shift" in country i. If there are multiple exogenous events over the entire period, we just compute sums of the above terms over all events.

5. *Controls*: in addition to the above, there are other uncertainties regarding model specification. These include nature of control variables, and the inclusion (or not) of lagged values of variables, including the lagged value of $E(t)$ and the functional form for time (linear? quadratic?). Control variables include indicators of (a) public health (water, sewage, electricity); (b) female education; (c) historical indicators that will enter as lagged variables (state strength, Rockefeller investments in eradication/sanitation;vaccination coverage; fertility decline)

To account for uncertainty of models we proceed within a Bayesian framework as described below.

3.3 Model specification and uncertainty:II

We need to keep in mind the following:

"Statistical analysis should begin with the careful development of a model based on theory and previous research (Gelman and Rubin 1995)."

"BMA is best used as a subsequent robustness check to show that our inferences are not overly sensitive to plausible variations in model specification."

"If there is uncertainty about functional form of model, BMA is not there to help (Montgomery and Nyhan, 2010)"

Suppose that consideration of theories implicit above about the influence of income, technology and public health leads to select H competing models, M_1,...M_H. These models differ only in terms of the variables included and not in the functional form. Suppose also that in each model we compute the statistic of interest for each country under observation, that is an estimate of the magnitude of the shift associated to medical technology. Let this estimate for model M be ϕ_M. The marginal posterior distribution of ϕ across all models is

$$\pi(\phi|data) = \sum_{i=1}^{H} \pi(\phi|data, M_i)\pi(M_i|data)$$

where (a) $\pi(\phi|data, M_i)$ is the posterior of ϕ given the data (and under M_i); this quantity depends on the likelihood of the data under model M_i and the prior distribution of the parameters of M_i; (b)$\pi(M_i|data)$ is the posterior model probability. In practice, what we want to do is to generate a meta estimate of ϕ using a weighted average of the values of ϕ associated with each model. The main problem with the above is the computation of $\pi(M_i|data)$ which requires to know (compute) the marginal likelihood $\pi(Data|M_i)$. To avoid this computation, one can directly compute $\pi(M_i|data)$.There are a number of ways to do this.

1. *BMA*: Bayesian model averaging uses a measure of probability for a model that depends on the (normalized) BIC (or AIC or other Bayesian form of the log likelohood with penalties) of each model and a prior for the model.

2. *tBMA*: proposed by Burgette many years ago uses the expression $\pi(M_i|data) = \pi(M_i|\lambda) = \lambda^{|M_i|}$(1-$\lambda$)$^{p-|M_i|}$where p is the total number of variables considered by the analyst and |M_i|is the # of parameters considered in model M_i.The value of λ is chosen through a "leave-one-out" cross validation process (that is, model is estimated excluding one observation at a time and predicting the value of the left out observation).

3. *Stacking*: proceeds as suggested by Burgette, but is more general. Essentially one generates a vector of weights, each element of it associated with a model. The selection of the vector is done through a leave one out process after model parameters have been estimated using Bayesian procedures.

3.4 Model specification and uncertainty:III

I suggest we proceed by choosing some form of the very general model proposed by Oeppen. The default would be the general model that contains country specific constants and regression coefficients.

We then consider inclusion of following covariates: income, indicators of water, sewage, electricity, female education. In addition, we add covariates as needed to pick up (a) effect of exogenous interventions and (b) possible interactions effects.

Due to data limitations we will need to divide the estimation into two stages.

- In the first stage we ONLY consider one measured covariate, income, and use all the data available from 1890-1900 on. We can also introduce covariates associated with the timing of exogenous interventions.
- In the second stage we use information from 1950 on, a period for which we have data on other indicators.

In both stages we focus on the statistic of interest, namely, the magnitude of the shift, however defined. And these we estimate with uncertainty and compare inferences with those we would have generated had we used a traditional frequentist approach.

3.5 Model specification and uncertainty:IV

There is one more wrinkle in our approach. The foregoing only refers to life expectancy (at birth or at any other age). However, for the period 1950-2010 we also have information on causes of deaths and we can compute for the entire period the contributions of age-group specific causes of death to the gains in life expectancy. For each country we will have the variables $\Delta E(x, c)$ or number of years of life gained between 1950 and 2010 attributable to changes in causes of death c in age group x (these age groups will be very coarse and capture childhood, adolescence, young adults, adult and older people).

In theory we can use these measures as dependent variables and model the effects of time and other covariates, much as we did with life expectancy. In particular, we can compute the shift statistics for each one of them and assess how they compare with those estimated for total life expectancy and how they behave relative to expected patterns (from theories).

The key problem is that we will add too many models: if we have four causes of death and four age groups we will need to replicate the model estimation and the the model uncertainty drill 16 times.

3.6 Measurement uncertainty

The above strategies do not consider the fact that for each country-year we have multiple estimates of life expectancy, each associated with a probability of being within 5% of the true value. To include this into estimation and properly account for the uncertainty this may cause, we can employ two methods.

- Choose the weighted median or average of the estimates available for each country year and proceed with the resulting values as the "observed" data.
- Apply the BMA or stacking strategy to each one of multiple bootstraps samples of the values of life expectancy for each country year. The bootstrap samples are created using sampling with replacement but with weights proportional to the probabilities that each estimate is within 5% of the correct value