

**THE JOURNALS *ARE* FULL OF GREAT  
STUDIES BUT CAN WE BELIEVE THE  
STATISTICS?**

**Revisiting the Mass Privatisation – Mortality  
Debate.**

## **Abstract**

Cross-national statistical analyses studies based on country-level panel data are increasingly popular in social epidemiology. To provide reliable results on the societal determinants of health, analysts must give very careful consideration to conceptual and methodological issues: aggregate (historical) data are typically compatible with multiple alternative stories of the data-generating process. Studies in this field which fail to relate their empirical approach to the true underlying data-generating process are likely to produce misleading results if, for example, they misspecify their models by failing to explore the statistical properties of the longitudinal aspect of their data or by ignoring endogeneity issues. We illustrate the importance of this extra need for care with reference to a recent debate on whether discussing the role of rapid mass privatisation can explain post-communist mortality fluctuations. We demonstrate that the finding that rapid mass privatisation was a “crucial determinant” of male mortality fluctuations in the post-communist world is rejected once better consideration is given to the way in which the data are generated.

## 1. Introduction

As a result of the much enhanced availability of health-related data sets, together with an expanding body of subjective institutional indicators, increasingly sophisticated econometric software, an increased demand for interdisciplinary approaches to global health challenges and pressures stimulating widely-read medical journals to publish headline-attracting studies, many research questions traditionally falling within the domain of one discipline have been rendered ripe for appropriation by others. This is a broadly positive and welcome development. Nonetheless, especially where rather atheoretical empirical investigation takes centre-stage, it is not without risk. As Jonathan Wolff (2010), British scholar and philosopher of health, wryly observed writing in the *Guardian*, “The journals are full of great studies, but can we believe the statistics?”.

The source of these risks includes: the inherent complications (and shortfalls) in some of the data itself; the complexity of the econometric techniques (often disguised by the sophistication of the theoretical exposition and the functionality of the software); the country specificities; the myriad complications in combining methodologies; and the different norms and standards characterising and shaping expectations and know-how across different disciplinary boundaries. Of particular interest among this emerging empirical strand of ‘cross-national / time-series’ health literature are the results of increasing numbers of studies, in applied social sciences and social epidemiology, which lay claim to enhancing our understanding of the mechanisms which underpin global health trends and trajectories.

From a cursory review of recent and widely publicised work one can point to a diversity of policy relevant findings: over 50% of the 8.2 million reduction in child deaths between 1970 and 2009 are due to increased female education (Gakidou et al., 2010); social division is positively correlated with child and maternal mortality (Powell-Jackson, 2011); development assistance for health crowds out domestic government spending on health (Lu et al., 2010); indicators linked to economic and social modernization lower infant mortality rates (Shandra et al., 2004); a \$100 increase in social welfare

spending in industrialised countries reduces mortality by 1.19% (Stuckler et al., 2010); and the economic reform programmes of the IMF are associated with increased tuberculosis incidence (Stuckler et al., 2008).

So to Wolff's rhetorical question: what indeed are we to make of this strand of literature? Should we take the results at face value and assume the referees are *au fait* with the techniques applied, that the data and programming files used are reliable and have been sufficiently interrogated? That the underlying assumptions, robustness checks and declaration of caveats are as clear as they ought to be? That the aggregated proxies for political and socio-economic phenomena are reliable? And what to do when the research results and/or methods are questioned by those from across the disciplinary boundaries? These are not trivial questions since the initial publication of health findings are, quite rightly, of interest to the media and the general public, become rapidly disseminated and repeated, and have discernible impact upon public behaviour and policy making.

In this paper, we take stock of this cross-national, time-series strand of empirical global health research through the lens of another widely publicised cross-national study, namely, the 2009 *Lancet* publication which concluded that mass privatisation was the major determinant of increased male mortality in the aftermath of the collapse of communism (Stuckler et al., 2009). This provides a useful case study for a methodological discussion, not only because it was the research to which Wolff explicitly referred but also because it gave rise to an extended exchange in both the mainstream media and through various outlets of the *Lancet*, and has parallels with a not dissimilar discussion concerning the conclusion that IMF operations in the post-communist world have been associated with increased tuberculosis incidence (Stuckler et al., 2008).<sup>1</sup> In January 2009, following the publication of the mass privatisation claims, the University of Oxford (2009) issued a media release stating unequivocally that "one million working-age men died due to the economic shock of mass privatisation policies" and linked responsibility for this to Jeffrey Sachs, Lawrence Summers and Stanley Fischer. The *Economist* (2009) ran a headline, "Mass murder and the market" (though the

article itself was inclined to argue against the claims). There were multiple exchanges in the *Financial Times*, the *New York Times* and across the worldwide web. A year later, in January 2010, the *Lancet* published two pieces (Gerry et al., 2010; Earle and Gehlbach, 2010) that presented formal tests of the original findings of Stuckler and colleagues and found them wanting.<sup>2</sup>

We do not to call into question the value of this strand of research,<sup>3</sup> but rather highlight a number of key methodological and conceptual issues that, if ignored, may render some results less reliable than others. This matters because, first, the debate stems from the use of a type of data that is very widely available and increasingly called upon for investigating important questions in social epidemiology and applied social sciences. Second, whether and through which mechanisms policy decisions (by governments or international organisations) impact upon health is of paramount importance to policy makers and other interested parties. Third, the ‘mass privatisation – mortality’ debate about those ‘million lives’ awaits resolution and is relevant to policy in the affected countries particularly. Fourth, it may serve to stimulate and inform the collection of individual level data and/or promote greater understanding in domains which are becoming increasingly multidisciplinary.

Studies which fail to explore the statistical properties of the longitudinal aspect of their data, which ignore issues of endogeneity, and which fail to think about the underlying data generating process may, through misspecifying their models, produce misleading results. This is precisely what happened in the case of the debate about whether mass privatisation was a crucial determinant of male mortality in the post-communist world. By overlooking the underlying process of health production, mass privatisation has been unjustifiably associated with increasing male mortality. From this case study, we cull a number of important lessons for empirical health research.

We proceed as follows: in the next section we briefly review the key conceptual and practical challenges facing cross-national, time-series health studies; in section 3, we catalogue some of the key

empirical challenges as illustrated by the mass privatisation-mortality debate; section 4 discusses some lessons for related research; and section 5 presents our conclusions.

## **2. Cross-national panel data**

Since Auster et al. (1969) first examined an aggregate health production function, a great deal of cross-national health research has been, and continues to be, based on the estimation of such functions. This literature addresses an ever-growing need to understand the socioeconomic determinants of health, in the face of booming health-care and pharmaceutical industries, spiralling public sector medical costs, ageing populations and the growing scrutiny of the role of government in delivering health and well-being. While the empirical results are mixed, a consensus has emerged showing that health is a product of lifestyles and behaviours (e.g. tobacco, alcohol and diet), health care consumption and provision (e.g. private and public medical expenditures and treatments), socioeconomic states (e.g. inequality, wealth, work and education), the environment (e.g. physical infrastructure, climate and water) and institutional frameworks (e.g. regulatory structures). However, there is little by way of consensus on what the appropriate methodology is for explaining health production or for disentangling the relative roles of the constituent effects.

The methodological controversies are both conceptual and empirical. One main group of concerns relates to the extent to which the empirical model being estimated reflects the data generating process captured by theory (i.e. the health production function). Where there is no formal (structural) model capturing the production of health and from which testable hypotheses stem, the likelihood that we are left with a series of singular, self-contained hypotheses increases unavoidably. This likelihood is exacerbated with aggregate level data, where there is rarely reference to the problem of aggregation itself or its relationship with the model being estimated. The actual relationship that these studies are exploring takes place at the individual level but is being estimated at the aggregate level. Gravelle et al. (2002) demonstrate the profound conceptual difficulties with this kind of approach and conclude

that to test the determinants of health requires individual level data. Of course, such data are not always available and so we rightly seek to take what insights we can from the data that are available.

On the empirical side, beyond the ubiquitous worries regarding data quality and comparability across time and space, concerns relate to sample size and coverage, the use of indicator variables to proxy for social, economic and political phenomena, and the choice of estimation technique given the true data generating process. The most obvious constraint is that the number of countries and years for which there are data provides only very limited coverage, particularly for less-developed parts of the world or for newly collected indicators, such as the recently emerging proxies for institutional quality. Faced with this constraint, the most widely adopted solution is to perform the empirical analysis on ‘pooled’ ‘panel’ data. This sidesteps the small ‘N’ problem and is typically justified through reference to arguments around greater degrees of freedom, the possibility of statistically ‘netting out’ the shocks common across all countries (by including time effects) and of reducing any bias stemming from the omission of time-invariant variables (by including country controls).

In pooling the data there are of course well-known violations of the Gauss-Markov assumptions motivating ordinary least squares regression estimates. The error terms are likely to be both heteroskedastic (systematically different variances across countries) and serially correlated (not independent across time periods) as well as contemporaneously correlated in the presence of shocks common to all countries. Various practical solutions to these problems have emerged and become commonplace in the health literature, including: panel correction of the standard errors; heteroskedastic robust error corrections and the incorporation of time and unit dummies. While the identification of the problems stemming from the non-independence of observations is correct, the statistical ‘adjustments’ available should be understood and justified in the context of the data generating process.

An additional empirical challenge pertains to how we conceive of the data themselves; the causal relationships which drive them and the mode through which we can best garner them to address the research questions of interest. Even when not testing a formal behavioural model we cannot escape important questions about the real underlying model. Indeed, whatever we are doing, the *actual* model matters and in the case of health, this is particularly prescient, since we know that individual health outcomes evolve slowly in response to changes in tastes, preferences, technology and policy and that they both respond to and influence socioeconomic outcomes. In other words, health outcomes are necessarily *dynamic* (time persistent) and their relationships with other variables *endogenous*. These facts cannot be ignored.

To summarise, not all empirical work has to be (or indeed can be) grounded in the testing of parameters stemming from theoretical models. Nevertheless, thought should always be given to the underlying data generation process, since the practice of combining an atheoretical approach with a loose empirical framework, can result in incorrect conclusions. Even without a formal model, carefully thinking about *how the data are generated* (the true underlying model) can limit the risks and help to inform conceptual and empirical choices with regard to specification and estimation. We now pursue this discussion through the lens of the recent controversial debate asserting that mass privatisation in post-communist Eastern Europe was a “crucial determinant” of fluctuations in male mortality rates (Stuckler et al., 2009 – hereafter SKM).

### **3. The Mass Privatisation – Mortality argument**

As is clear from Figure 1 and explained in Gerry et al. (2010), following a rise, associated with the anti-alcohol campaign in the mid-80s, the countries of the post-communist world suffered a rapid decline in life expectancy during the early 1990s. There are rich and illuminating seams of demographic (Zakharov and Ivanova, 1996; Shkolnikov and Cornia, 2000; Vishnevsky, 2003; World Bank, 2005), epidemiological (Leon et al., 1997; McKee et al., 1998; Marmot and Bobak, 2000; McKee and Shkolnikov, 2001; Leon et al., 2007; Zaridze et al., (2009)), and social science (Shapiro,



1995; Shkolnikov et al., 1998; Cockerham, 2000; Cornia and Panicià, 2000; Davis, 2000; Brainerd and Cutler, 2005) literature identifying the nature of the trends and fluctuations captured by figure 1. This research concurs that there was indeed sharp and unprecedented fluctuations in life expectancy during the 1980s and 1990s and that these effects were felt more acutely in the western countries of the FSU (the so-called ‘mortality belt’), where the longer-run deteriorating trends were also more clearly defined. While much remains unexplained, it is now agreed that the increased dangerous use of alcohol (and surrogate alcohol) and the stress associated with socio-economic upheavals across the region was closely associated with these fluctuations.

It is in this context that SKM embark on their ambitious task of identifying the role that mass privatisation played in driving the early 1990s rise in mortality. Through analysis of an unbalanced cross-national dataset, covering the period 1989-2002, they conclude that “Rapid mass privatisation as an economic transition strategy was a crucial determinant of differences in adult mortality trends” and that therefore the social and health costs borne by countries adopting rapid mass privatisation have been far greater.

### *The Data*

In revisiting these claims, we use the exact same data as that used by SKM. The data provide an unbalanced panel of 289 observations, derived from 24 countries, for up to 14 years between 1989 and 2002. An important first step in presenting transparent cross-national panel estimates involves carefully detailing the countries and time periods included in the actual estimates. In doing this for the SKM estimates, table 1 makes clear why this is an important step. Russia is included from 1989 while the other FSU countries are included only from 1991; the Czech Republic is included from 1990 but Slovakia only from 1993 (the year when the two separate countries came into existence); Slovenia, one of the prime examples of rapid mass privatisation, is missing from the sample entirely. The data used in empirical estimates, particularly when the number of countries is just 24, should make conceptual and contextual sense and cataloguing them clearly is essential (particularly if you are

claiming that the start of the data series represents a structural break). As Maddala (1999) notes: “The countries used in the panel study are as crucial (if not more so) as the choice of the estimation method”.

### *SKM's static analysis*

The core specification on which the claims of SKM are based is a static estimate in which the log of male mortality is regressed on a selection of aggregate socio-economic, demographic and political proxies plus a dummy variable capturing mass privatisation (see appendix 1), as follows:

$$(1) \text{MR}_{it} = \alpha + \beta_1 \text{PRIV}_{it} + \beta_2 \text{GDP}_{it} + \beta_3 \text{LIB}_{it} + \beta_4 \text{TRADE}_{it} + \beta_5 \text{DEM}_{it} + \beta_6 \text{DEP}_{it} + \beta_7 \text{WAR}_{it} + \beta_8 \text{URBAN}_{it} + \beta_9 \text{EDUC}_{it} + \mu_i + \varepsilon_{it}$$

In (1)  $\mu_i$  captures the country fixed effects and  $\varepsilon_{it}$  is the error term. The main results discussed in their original text are the estimates of a Prais-Winsten pooled OLS model with panel-corrected standard errors based on an unexplained AR(1) process, supplemented with a country fixed effects estimate with clustered standard errors. In table 2 we explore a necessary second step in the analysis of pooled cross-national data, namely whether and how country and time effects should be accounted for in the estimates. The main fixed effects and Prais-Winsten results of SKM are reproduced as estimates (2) and (5) respectively.

Comparing the simple pooled OLS regression estimates (1) with a range of alternative effects models confirms that, as expected, country and time effects are important. The inclusion of simple country effects (2) has a dramatic impact upon the results: the coefficient on the mass privatisation dummy is halved; the sign *and* significance of GDP per capita, price liberalisation and foreign exchange liberalisation all change, while tertiary education switches from positive and significant to negative and significant. Substituting country effects with time effects (3) sees the results revert to those implied by pooled OLS, since the effects of mass privatisation, GDP and higher education are all overestimated in the absence of netting out cross-sectional variation that is otherwise captured in the

set of explanatory variables.<sup>4</sup> As expected, incorporating time and country effects (4) reverts closer to (2), though tertiary education is no longer significant.<sup>5</sup>

In this context, it comes as no surprise that when we test formally for group-wise heteroskedasticity, cross-sectional correlation and serial correlation we find strong evidence of all three. This being so, the significance tests associated with estimates (1) through (4) are invalid and the fixed effects results non-robust. There are corrections available for some of the reported statistical violations but the distinctive pooled OLS assumption, arising from the fact that these are panel data, is that the errors are uncorrelated over time. This leads us to estimate (5), the main SKM estimate, and one that recognises serial correlation in the error terms while also including country fixed effects. Now, let us just be clear about what the country fixed effects and the adjustments for serial correlation actually achieve.

First, fixed effects: in arguing that “our regression models also use a set of country dummy variables to hold constant fixed aspects of national surveillance infrastructure, initial country conditions and pre-existing societal characteristics, and predispositions to higher mortality”, SKM are quite right that netting out the effects of time invariant observed and unobserved variables is necessary in order to capture the determinants of mortality fluctuations. However, country dummies do not remove the effects of how those same variables interact with time-varying variables, nor do they remove the impact of any unobserved time variant societal or behavioural characteristics. In this dynamic model these time variant effects are correlated with other regressors and therefore the determinants of mortality fluctuations simply cannot be isolated through effects models alone.

Now turning to serial correlation: in their review of this paper, they claim to address autocorrelation by clustering standard errors as per the “best recommendation of Wooldridge”, and that “when you try to model the autocorrelation, you ‘pull’ things out of the data (based on your estimate of the autocorrelation coefficient)”. In actual fact, it is SKM’s Prais-Winsten results themselves (5) which are the ones explicitly based on an (unexplained) AR(1) estimate of the autocorrelation coefficient.

We simply argue that the correct model specification requires the inclusion of a dynamic term. This is quite different.

In any event, Wooldridge (2002) actually suggests that if tests show the presence of serial correlation in fixed effects regressions then “at a minimum” the standard errors indeed require adjustment. This is the axis around which the methodological and conceptual disagreement between SKM and their critics revolves. Specifically, when tests of the autocorrelation assumptions indicate serial correlation, as they do in these data (see table 2), there are two conceptually distinct ways of approaching it. One approach – that of SKM – assumes that the specification and coefficients of the model are correct but, that the standard errors are biased and therefore need adjustment. A second approach posits that the serial correlation stems from an incorrectly specified model and in particular, the omission of a variable capturing persistence in the data generating process. In this case, the coefficients themselves will not be correct and the model itself should be recast in a dynamic form.

#### *More on the dynamic approach*

Health outcomes are a product of choices made over diet and lifestyle; access to health care; innovations in health technologies; the structure of the population; government policy; social norms and mores; and exposure to the environment. These are all factors which evolve over time, continually influencing mortality patterns and distinctively conditioning the health impacts of ‘shocks’ such as economic collapse, epidemics, natural disasters and mass privatisation. This being so, then country-level mortality rates are ‘state dependent’. To illustrate this more clearly still, consider a flu epidemic or an economic meltdown occurring in a particular year: the logic of a static (SKM) estimate assumes that the epidemic and the economic collapse have no impact on health beyond the year in which they occur, other than through the cross-sectional effect of the variables incorporated in the regression. That is, the static approach maintains that these shocks do not have lasting harmful effects on the health of vulnerable populations over the longer-term, through nutrition, schooling effects and so on. This assertion contradicts much of what we have learnt from the rich demographic, epidemiological

and social science literature discussed at the start of this section, which recognises both the immediate and longer-run relationships between health outcomes and health inputs. SKM are therefore wrong: there are lasting and evolving – time variant and unobserved – effects in the annual mortality data and these effects (just like the time invariant ones) have to be taken into account. The static analysis of SKM does not do this and is therefore misspecified. In fact, their dummy variable for mass privatisation (essentially a slope shift for 11 countries) likely picks up many of these unobserved persistent effects in the data. In short, the presence of serial correlation in the pooled error terms reflects a property of the underlying data generating process and is a crucial part of the explanation for mortality fluctuations.

One possible solution is to make the OLS and effects estimates dynamic through the incorporation of a lag of the dependent variable. However, in the case of OLS, serial correlation in the error terms ensures the ensuing estimates are biased and inconsistent (Greene, 2007), while for fixed effects, based as it is on mean-differencing, an additional bias emerges through the correlation between the lag term and the individual effects. Specifically, since the individual effect is fixed over time, both the dependent variable and its lag are correlated with the country effect (Nickell, 1981).

In fact, the way that the data generating process should be modelled very much depends upon the stationarity of the dependent variable. The graphical evidence from figures 1 and 2 is suggestive of non-stationarity. Most countries in this sample experienced a short-run spike in male mortality at the start of the 1990s which maps onto a longer-term rise dating back to the late 1960s and which itself had been interrupted by the 1980s anti-alcohol campaign (Shkolnikov & Cornia, 2000). In other words, the data series being examined seems to exhibit both cycles and trends, at least for some panels, with the potential to cause further bias in the estimates (Baltagi, 2008). The available tests for stationarity with unbalanced panel data, covering a short time-period and with gaps in some panels are rather limited. Applying the Im-Pesaran-Shin test (Im, Peseran & Shin, 2003) and the Fisher-type tests (Choi, 2001) provides some evidence that the data are in fact stationary, but this evidence really

amounts only to stating that for at least one of the panels we cannot reject the null that it does not have unit root. For such a short panel, the power of these tests is simply very low.

If we do suspect non-stationarity, a possible solution would be to take first differences and estimate a dynamic model. First difference models have the distinctive advantage of only requiring weak (sequential) exogeneity, in the sense that the differenced error term should be uncorrelated with the differenced explanatory variable terms. This is a much weaker requirement than the fixed effects model has because it allows future values of the regressors to be correlated with the error term, which is precisely the case if a lagged dependent variable is incorporated to capture the dynamics of a persistent series (Cameron and Trivedi, 2009). However, although differenced estimates would represent an improvement on simple static estimates, they do not solve the persistence conundrum because the lagged dependent variable is a function itself of the remaining lagged explanatory variables and therefore measures the weighted average of those variables (Baltagi, 2008). In other words, the lagged effect is not actually a measure of mortality persistence in this case since it measures the dynamics of all of the explanatory variables based on the (wrong) assumption that those dynamics are identical. We need a way of circumventing this problem by instrumenting the differenced variables.

A Generalised Method of Moments technique, due to Arellano and Bond (1991), Arellano and Bover, (1995), Blundell and Bond (1998), Bond (2002) and discussed by Roodman (2009), allows us just such an escape through instrumenting the differenced variables with their available lags in levels (difference GMM) and lags in first differences (system GMM). Instrumental variable estimation in this first differencing model draws on lags from two or more periods ago on the (testable) assumption that the associated errors are serially uncorrelated and thus that the differenced errors are uncorrelated with the differenced dependent variable. We adopt the ‘System GMM’ approach in estimating the following system of equations.

$$(2) \text{ MR}_{it} = \alpha + \beta_1 \text{PRIV}_{it} + \beta_2 \text{X}_{it} + \epsilon_{it}$$

$$(3) \Delta MR_{it} = \beta_1(\Delta MR_{it-1}) + \beta_2(\Delta PRIV_{it}) + \beta_3(\Delta X_{it}) + (\Delta \epsilon_{it})$$

Lagged differences are used as instruments in the level equation (2) and correspondingly lagged levels are used as instruments in the difference equation (3).

System GMM also allows us to distinguish between endogenous, exogenous and pre-determined variables. This allows us to address another important weakness with the analysis of SKM, picked up by Gerry et al. (2010). Aside from serial correlation in the error term, there are nearly always sources of endogeneity in the specification of the health production function in that the explanatory variables themselves are correlated with past and possibly current realisations of the error term. For example, if government spending is an explanatory variable and government spending is higher in countries with higher mortality rates, then the beneficial impact of spending will be systematically underestimated. In the mass privatisation – mortality example, a case can be made arguing that several of the explanatory variables are endogenous; most obviously, GDP per capita is both a contributor to improved health and a product of improved health.

To test the claims of SKM within this framework, we first replicate their estimates, using one-step system GMM estimators robust to both heteroskedasticity and arbitrary patterns of autocorrelation within countries, treating the independent variables as exogenous and including the lagged dependent variable as predetermined but not strictly exogenous (Roodman, 2009). We then treat GDP, mass privatisation and war as endogenous variables, before additionally incorporating the EBRD reform indicators as endogenous. We report estimates, in table 3, based on using lags two through four (thus restricting the instrument count (Roodman, 2009)) to instrument for the endogenous variables but the results are not qualitatively affected by increasing the number of lags. We also report the AR(2) test for autocorrelation and the Hansen test for the validity of instruments. Autocorrelation indicates that lags of the dependent variable are endogenous; AR(1) is expected, by construction, in first differences so we check for AR(1) in levels by looking for evidence of AR(2) in differences. We find no such evidence. The Hansen test, unlike the Sargan test, is robust to both heteroskedasticity and

autocorrelation and confirms, in all cases, that our instruments are valid. In constructing our instrument matrix we collapse the number of instruments by limiting them to one for each variable and lag distance, rather than the default of one for each period, variable and lag distance. For small samples, this has limited impact on statistical efficiency and allows us to avoid the bias that arises from over-fitting the instruments.

While it is important to interpret GMM results with caution, we consider that, of all estimators requiring controls for persistence, unobserved country effects and instruments for endogenous socioeconomic variables, it is the system GMM which produces the most realistic and accurate estimates, in the context of the real data generating process. Our results are robust to different lag structures and instrument matrices.

It is clear from table 3 that adult male mortality is persistent such that the deaths occurring in any given year during the transition period are highly dependent on deaths in the previous year and the unobserved, time variant factors associated with those. The value of the coefficient should lie between upper and lower bounds provided by dynamic specifications of OLS and fixed effects estimates. This condition holds as the upward biased OLS coefficient is 0.96, while the downward biased fixed effects estimate is 0.37.

In column (1) we re-estimate the SKM specification dynamically. We find that the only significant result is that countries liberalising their foreign exchange and trade regimes the most, experienced lower rises in male mortality. We don't, however, believe that this specification correctly captures the data generating process, so: in column (2), we specify mass privatisation, war and GDP as endogenous variables; in column (3) we add price liberalisation and foreign exchange and trade reforms to that list; and in column (4) we explore the assertion of SKM that "even the anticipation of bad news can increase mortality"<sup>6</sup>, by including a lag of their mass privatisation variable.



The results are striking.<sup>7</sup> We find no evidence, in any plausible dynamic specification, that rapid mass privatisation was associated with changes in male mortality during this period. Instead we find that rapid progress with economic reforms was actually associated with lower male mortality, whether captured through price or trade liberalisation. We find some evidence that GDP per capita is associated with lower mortality increases and we find evidence at the margins that higher education enrolment rates and higher freedom house scores are associated with greater mortality. However, the former result stems from SKM's use of a GDP measure based on current prices rather than Purchasing Power Parity prices. Using the latter – less arbitrary – variable, GDP is no longer significant. The education result may seem surprising but it is actually consistent with the raw data. It was those countries with higher levels of 'education' (e.g. Russia, Ukraine, Belarus, Latvia, Estonia) that experienced higher male mortality in the early 1990s, while countries, such as Albania, Macedonia, Tajikistan and Uzbekistan, with lower enrolment rates, also reported lower mortality.

#### **4. Discussion and lessons**

In applied social sciences and social epidemiology, there has been a growing body of research based on cross-national panel data. Such panels are attractive because they allow researchers to circumvent the problems of 'small N' and they contain more information than equivalent cross-sections. However, these data also exhibit myriad cross-sectional and temporal dependencies which can easily result in over-interpreting or mis-interpreting the information in the data: as Cameron and Trivedi (2009) note, "NT correlated observations have less information than NT independent observations". Failure to address these correlations and their causes can result in incorrect statistical inference and/or incorrect model specification and can lead commentators to wonder whether 'we can believe the statistics'.

Using a recent well-publicised controversy concerning mass privatisation and mortality we have explored the challenges and limitations of using this type of aggregate data by paying particular attention to its temporal dependencies. Our findings therefore fall into two related categories: those

directly concerning our re-examination of the controversy; and a more general set of lessons relating to the use of cross-national panel data in the health field.

### *Mass privatisation-mortality controversy*

We find that the claims of SKM, associating mass privatisation with mortality fluctuations in the post-communist world, do not withstand scrutiny of the temporal dependencies within their data. In particular, their results depend on the assumption that health outcomes do *not* evolve over time as a product of tastes, preferences, individual choices, technological changes, demographics, ethnic composition, migration, government policy, social norms and mores, and exposure to the environment. This assumption, which amounts to the claim that time varying observed and unobserved factors are incidental cannot be correct. Each of the elements listed above is continually influencing mortality patterns and conditioning the health impacts of ‘shocks’ such as economic collapse, epidemics, natural disasters and mass privatisation. The argument of SKM that they aren’t is difficult to follow: aside from its arrant counter-intuitiveness, it is at odds both with the findings from an abundant literature spanning demography, epidemiology and social sciences and with their own recognition of the need to control for time *invariant* effects.

In short, their ‘model’ is incorrectly specified: their static aggregate analysis, remote from the underlying health production function, does not capture the real time varying relationships between health inputs and outputs. Any plausible specification that allows for dynamics in the true underlying model finds no evidence that rapid mass privatisation was associated causally, or otherwise, with short-run changes in male mortality in the early 1990s, post-communist world.

### *Lessons for cross-national panel research*

Casting our eye to the bigger picture there are a number of procedural, conceptual and methodological lessons for research in this field and with this type of data that we would like to see adopted:

- (i) The data matter and, in particular, the specifics of which countries and years are incorporated in the actual estimations needs detailing in published work.
- (ii) The panel structure of the data requires careful discussion and formal testing to establish the statistical importance of groupwise heteroskedasticity, cross-sectional correlation and serial correlation. Remedial action needs carefully explaining.
- (iii) The temporal dimension of the data merits distinct attention. Whether to ‘adjust’ the standard errors and think of serial correlation as some sort of ‘nuisance’ or to think of temporal dependence as capturing something about the real relationships in the data is a key conceptual decision and can’t simply be overlooked – misspecification matters for the coefficients, not just the standard errors.
- (iv) The underlying data generating process is important with aggregate analysis, particularly when the real outcomes are generated at the individual level. The data will surely be compatible with various data-generating processes and so, while not all empirical work should, or could, test parameters gleaned from formal mathematical modelling, empirical remoteness from an underlying model should always be a concern.

When applied to cross-national social epidemiology and health studies, these points translate into a clear message. Health outcomes are dynamic, generated at the individual level and are endogenous to a number of key input variables. As a minimum, the research analysis should situate the data in relation to the underlying data generating process (the true model) and should present results in both a static and a dynamic framework, alongside a careful discussion of endogeneity.

## **5. Conclusion**

It is obviously impossible to control perfectly for, or to understand all of, the complications associated with cross-national health studies. Cross-national panel analysis should and will be carried out (and GMM is certainly not a panacea in this regard), particularly if findings can help foster the collection and exploration of individual level longitudinal data. However, much greater care is needed in the execution of these studies, and much greater caution is needed before prescribing headline -grabbing

policy recommendations. A further reflective conclusion might be that more careful interdisciplinary cooperation in refereeing is required by journals publishing interdisciplinary health studies.

In commenting on macroeconomics, Leamer (2010) is controversially and provocatively categorical that “our understanding of causal effects in macroeconomics is virtually nil, and will remain so”.

While arguable for macroeconomics, the same assertion is more appropriate for describing our ability to capture causality at the aggregate health level, where the time-series and coverage of relevant data is even more constrained and unreliable than that facing macroeconomists. With this in mind, we do agree with Leamer that “we need to change the culture and regard the finding of “no persuasive evidence in these data” on the same footing as a “statistically significant and sturdy estimate.””.

Indeed, maybe there is a lesson here too for the publishers of academic research. Reporting on a recent scandal involving clinical trials, the *Times Higher Education Supplement* (Ince, 2011) captured it neatly: “Journals must acknowledge that falsifiability lies at the heart of the scientific endeavour.”

To this end, surely it is time to accept that there really is no persuasive evidence that rapid mass privatisation caused the mortality fluctuations seen in Eastern Europe.

## References

- Arellano, M., & Bond, S. (1991) Some tests for specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*. 58(2), 277-97.
- Arellano, M., & Bover, O. (1995) Another look at the instrumental variables estimation of error components models. *Journal of Econometrics*. 68, 29-51.
- Auster, RD., Leveson, I., & Sarachek, D. (1969) The production of health: An exploratory study. *Journal of Human Resources*. 4, 411-36.
- Baltagi, BH. (2008). *Econometric Analysis of Panel Data*. (4th ed.). New York: Wiley.
- Blundell, R., & Bond, S. (1998) Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*. 87, 115-143.
- Bond, S. (2002) Dynamic panel data models: A guide to micro data methods and practice. Working Paper 09/02. *Institute for Fiscal Studies*. London.
- Brainerd, E., & Cutler, D. (2005) Autopsy of an Empire: understanding mortality in Russia and the former Soviet Union. *Journal of Economic Perspectives*. 19, 107-130.
- Cameron, AC., & Trivedi, PK. (2009) *Microeconometrics Using Stata*. A Stata Press Publication, StataCorp LP, Texas.
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*. 20, 249-272.
- Cockerham, W. (2000) Health lifestyles in Russia. *Social Science & Medicine*. 51: 1313-1324.
- Cornia, GA., & Panicià, R. (eds.) (2000) *The Mortality Crisis in Transitional Economies*. Oxford, Oxford University Press.
- Davis, C. (2000) Transition, health production and medical system effectiveness. In Cornia,

- GA, & Paniccià, R. (eds.) *The Mortality Crisis in Transitional Economies*. Oxford, Oxford University Press
- Economist, The. (2009) Did economic reform kill millions of Russians? *The Economist*. January 22.
- Earle, J., & Gehlbach, S. (2010) Mass Privatisation and the Post-Communist Mortality Crisis: Is there really a relationship? *The Lancet*. 375(9712), 372.
- Gakidou, E., Cowling, K., Lozano, R., & Murray CJL. (2010) Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *The Lancet*. 376:959–974.
- Gerry, CJ., Mickiewicz, TM., & Nikoloski, Z. (2010). Did Mass privatisation really increase Post-Communist Male Mortality? *The Lancet*. 375(9712), 371.
- Gravelle, H., Wildman, J., & Sutton, M. (2002) Income, income inequality and health: what can we learn from aggregate data? *Social Science and Medicine*. 54, 577-589.
- Greene, WH. (2007) *Econometric Analysis* (6<sup>th</sup> Edition). Prentice Hall.
- Im, KS., Pesaran, MH., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*. 115, 53-74.
- Ince, D. (2011) Systems Failure. *Times Higher Education*. 5 May 2011: 32-37.
- Leamer, E. (2010). Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives*. 24(2), 31–46. DOI:10.1257/jep.24.2.31.
- Leon, D., Chenet, L., Shkolnikov, V., Zakharov, S., Shapiro, J., Rakhmanova, S., Vassin, S., & McKee, M. (1997) Huge variation in Russian mortality rates 1984-1994: Artefact, alcohol or what? *Lancet*. 350: 383-88.
- Leon, D., Saburovo, L., Tomkins, S., Andreev, E., McKee, M. & Shkolnikov, V. (2007) Hazardous alcohol drinking and premature mortality in Russia: a population-based control study. *The Lancet*. 369(9578): 2001-2009.

- Lu, C., Schneider, MT., Gubbins, P., Leach-Kemon, K., Jamison, D., & Murray, CJL. (2010) Public financing of health in developing countries: a cross-national systemic analysis. *The Lancet*. 375 (9723), 1375-1387.
- Maddala, GS. (1999) On the Use of Panel Data Methods with Cross-Country Data. *Annales d'Economie et de Statistique*. 55/56, 429-49.
- Marmot, M., & Bobak, M. (2000) Psychosocial and biological mechanisms behind the recent mortality crisis in Central and Eastern Europe. In Cornia, GA., & Panicià, R. (eds.). *The Mortality Crisis in Transitional Economies*. Oxford, Oxford University Press.
- McKee, M., Sanderson, C., Chenet, L., Vassin, S., & Shkolnikov, V. (1998) Seasonal variation in mortality in Moscow. *Journal of Public Health*. 20(3): 268-274.
- McKee, M. & Shkolnikov, V. (2001) Understanding the toll of premature death among men in Eastern Europe. *British Medical Journal*. 323: 1051-1055.
- Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*. 49, 1417-26.
- Powell-Jackson, T., Basu, S., Balabanova, D., McKee, M., & Stuckler, D. (2011). Democracy and growth in divided societies: A health-inequality trap? *Social Science and Medicine*. 73, 33-41.
- Roodman, D. (2009) How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*. Vol. 9, (1): 86-136.
- Shandra, JM., Nobles, J., London, B., & Williamson, JB. (2004) Dependency, democracy, and infant mortality: a quantitative, cross-national analysis of less developed countries. *Social Science and Medicine*. Vol. 59, No.2: 321-333.
- Shapiro, J. (1995) The Russian Mortality Crisis and its Causes. In *Russian Economic Reform in Jeopardy?* Aslund Anders, ed. London and New York: Pinter Publishers, 149-178.
- Shkolnikov, V., Cornia, GA., Leon, DA., & Meslé, F. (1998) Causes of the Russian mortality crisis: Evidence and interpretations. *World Development*. 26 (11): 1995-2011.

- Shkolnikov, VM., Cornia, GA. (2000) The population crisis and rising mortality in transitional Russia. In: Cornia, GA. & Panicià, R. (Eds) *The Mortality Crisis in Transition Economies*. New York: Oxford University Press; 2000 pp. 253-179.
- Stuckler, D., King, LP. & Basu, S. (2008) International Monetary Fund Programs and Tuberculosis Outcomes in Post-Communist Countries. *PLoS Med* 5(7): e143. doi:10.1371/journal.pmed.0050143.
- Stuckler, D., King, L., & McKee, M. (2009) Mass privatisation and the post-communist mortality crisis: a cross-national analysis. *The Lancet*. 373(9661), 399-407.
- Stuckler, D., Basu, S., & McKee, M. (2010) Budget crises, health, and social welfare programmes. *British Medical Journal*. 340, 3311.
- Stuckler, D., King, L., & McKee, M. (2010) How to make a mortality crisis disappear: statistical manipulation. *The Lancet*. 375(9712), 373.
- University of Oxford (2009)  
[http://www.ox.ac.uk/media/news\\_releases\\_for\\_journalists/090115.html](http://www.ox.ac.uk/media/news_releases_for_journalists/090115.html)
- Vishnevsky, A. (2003) The Depopulated Superpower. *Russia in Global Affairs*. No. 3.  
<http://eng.globalaffairs.ru/numbers/4/488.html>.
- Wolff, J. (2010). The journals are full of great studies, but can we believe the statistics? *The Guardian*, Tuesday 4 May.
- World Bank. (2005) Dying Too Young: Addressing Premature Mortality and Ill Health Due to Non-Communicable Diseases and Injuries in the Russian Federation. Washington DC.
- Zakharov, SV., & Ivanova, EI. (1996) Fertility Decline and Recent Changes in Russia: On the Threshold of the Second Demographic Transition. In *Russia's Demographic "Crisis"*, ed. Julie DaVanzo and Gwen Farnsworth. RAND.
- Zaridze, D., Maximovitch, D., Lazarev, A., Igitov, V., Boroda, A., Boreham, J., Boyle, P.,



Peto, R., & Boffetta, P. (2009) Alcohol poisoning is a main determinant of recent mortality trends in Russia: evidence from a detailed analysis of mortality statistics and autopsies. *International Journal of Epidemiology*. 38 (1):143-53

## Appendix 1: Variable definitions

Variable abbreviation	Variable definition
MR <sub>it</sub>	Log of adult male standardises mortality rates (age 15-59)
PRIV <sub>it</sub>	0 until a country has transferred the ownership of 25% of large state-owned enterprises to the private sector through vouchers and insider give-aways; 1 thereafter
GDP <sub>it</sub>	Log GDP per head in current \$US (World Bank)
LIB <sub>it</sub>	EBRD price liberalisation index
TRADE <sub>it</sub>	EBRD foreign exchange and trade liberalisation index
DEM <sub>it</sub>	Democratisation index of Freedom House
DEP <sub>it</sub>	Population dependency ratio (World Bank)
WAR <sub>it</sub>	Dummy for military conflict (World Bank)
URBAN <sub>it</sub>	Percentage of the population in urban settings (World Bank)
EDUC <sub>it</sub>	Percentage of the population with tertiary education (World Bank)

\* Descriptive statistics and correlations are detailed in Stuckler et al (2009).

## Tables

Table 1: Estimation sample

Year coverage	Countries
14 years (1989-2002)	Bulgaria, Hungary, Poland, Romania, Russia
13 years (1990-2002)	Czech Republic (1990 – 2002)
12 years (1991-2002)	Armenia, Belarus, Croatia, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Turkmenistan, Ukraine, Uzbekistan
11 years (1992-2002)	Azerbaijan
11 years (1991-2001)	Macedonia
10 years (1993-2002)	Slovakia
10 years (1989-1996; 2000-2001)	Albania
8 years (1991-1997; 1999)	Tajikistan

Table 2: Pooled mortality estimates

Variable	(1) POOL	(2) FE_C	(3) FE_T	(4) FE_CT	(5) PW_S
Mass privatisation	.28*** (.04) 6.35	.14*** (.02) 5.74	.28*** (.04) 6.66	.14*** (.02) 5.70	.13*** (.03) 5.11
Log GDP per capita	.03 (.03) 0.81	-.14*** (.02) -5.67	.06* (.03) 1.91	-.13*** (.02) -5.65	-.12*** (.03) -4.18
Price liberalisation	-.08*** (.03) -2.80	.01 (.01) 0.66	-.02 (.04) -.038	.02 (.01) 1.54	.01 (.01) 0.47
Foreign exchange and trade	-.07** (.03) -2.23	.01 (.01) 1.11	-.12*** (.03) -3.65	-.00 (.01) -0.38	.01 (.01) 0.72
Democracy	-.01 (.01) -1.06	-.01*** (.00) -2.87	-.02** (.01) -2.09	-.01*** (.00) -2.90	-.01* (.01) -1.75
War	.03 (.07) 0.41	.27*** (.03) 10.29	-.12* (.07) 1.75	.23*** (.03) 8.13	.22*** (.05) 4.54
Dependency ratio	.01*** (.00) 3.19	.02*** (.00) 6.38	.00 (.00) 1.14	.01*** (.00) 2.71	.02*** (.00) 4.20
Higher education	.01*** (.00) 5.63	-.00*** (.00) -3.00	.02 (.00) 7.46	-.00 (.00) -1.37	-.00** (.00) -2.19
Urban percentage	.01*** (.00) 3.93	.03*** (.01) 5.17	.01 (.01) 1.51	.03*** (.01) 4.42	.03*** (.01) 3.55
Constant	5.21*** (.38) 13.54	4.17*** (.30) 14.11	5.37*** (.38) 14.08	4.50*** (.31) 14.37	4.21*** (.54) 7.85
<i>N</i>	289	289	289	289	289
<i>r</i> <sup>2</sup>	.37	.95	.46	.96	.98
<i>F</i> ( <i>C</i> )	-	.00***	-	.00***	.00***
<i>F</i> ( <i>T</i> )	-	-	.00***	.00***	-
<i>F</i> ( <i>CT</i> )	-	-	-	.00***	-
<i>Serial correlation</i>	.03**	.03**	.03**	.03**	-
<i>C-S correlation</i>	-	.00***	.00***	.00***	-
<i>Group Heterogeneity</i>	.00***	.00***	.00***	.00***	-

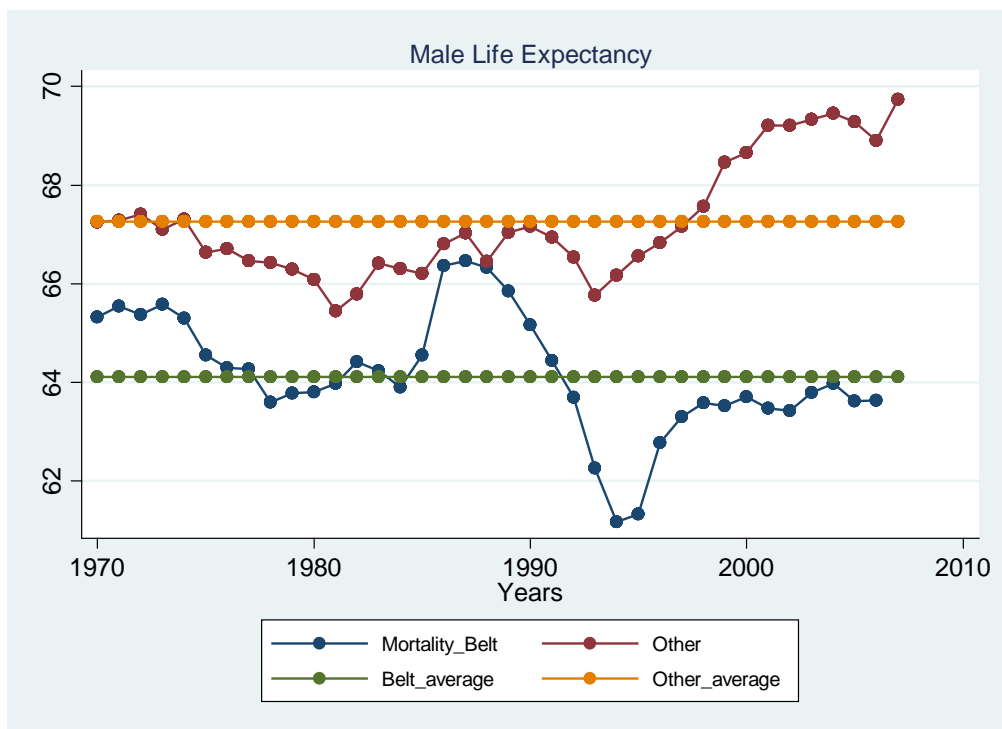
Notes: Columns 1-5 represent respectively pooled OLS, Country Fixed Effects (or technically Least Square Dummy Variables), Time Fixed effects, Time/Country Fixed Effects, Prais-Winsten pooled OLS with AR(1). Coefficient, standard errors (in parentheses) and t-statistics reported to 2 decimal places. \*\*\*/\*\*/\* indicates significance at the 1%/5%/10% level respectively.

Table 3: System-GMM mortality estimates

<b>Variable</b>	<b>GMM_1</b>	<b>GMM_2</b>	<b>GMM_3</b>	<b>GMM_4</b>
Lagged mortality	.96*** (.09) 10.09	.81*** (.08) 10.20	.88*** (.07) 13.22	.89*** (.07) 12.18
Mass privatisation	.02 (.03) 0.77	-.00 (.05) -0.10	-.00 (.03) -0.12	-.03 (.14) -0.21
Lag of mass privatisation	-	-	-	.02 (.15) 0.15
Log GDP per capita	.01 (.01) 0.39	-.08 (.05) -1.59	-.07* (.04) -1.92	-.07** (.03) -2.16
Price liberalisation	-.01 (.02) -0.71	-.05* (.03) -1.80	-.02 (.02) -0.87	-.01 (.02) -0.64
Foreign exchange and trade	-.03*** (.01) -3.13	-.00 (.02) -0.20	-.04** (.02) -2.11	-.05** (.02) -2.45
Democracy	.00 (.00) 1.12	.00 (.01) 0.43	.01* (.01) 1.80	.01* (.01) 1.87
War	-.01 (.03) -0.50	.04 (.07) 0.62	-.03 (.05) -0.62	-.05 (.05) -1.07
Dependency ratio	.00 (.00) 0.02	.00 (.00) 0.35	-.00 (.00) -0.29	-.00 (.00) -0.48
Higher education	.00 (.00) 0.46	.00** (.00) 2.35	.00** (.00) 2.11	.00 (.00) 1.62
Urban percentage	.00 (.00) 0.21	.01 (.00) 1.58	.00 (.00) 1.01	.00 (.00) 0.96
Constant	.30 (.47) 0.64	1.53*** (.46) 3.32	1.21*** (.40) 3.05	1.16*** (.49) 2.40
<i>N</i>	263	263	263	263
<i>Number of groups</i>	24	24	24	24
<i>Number of instruments</i>	24	24	30	32
<i>AR(2): p-value</i>	.55	.43	.55	.46
<i>Hansen: p-value</i>	.13	.33	.60	.67

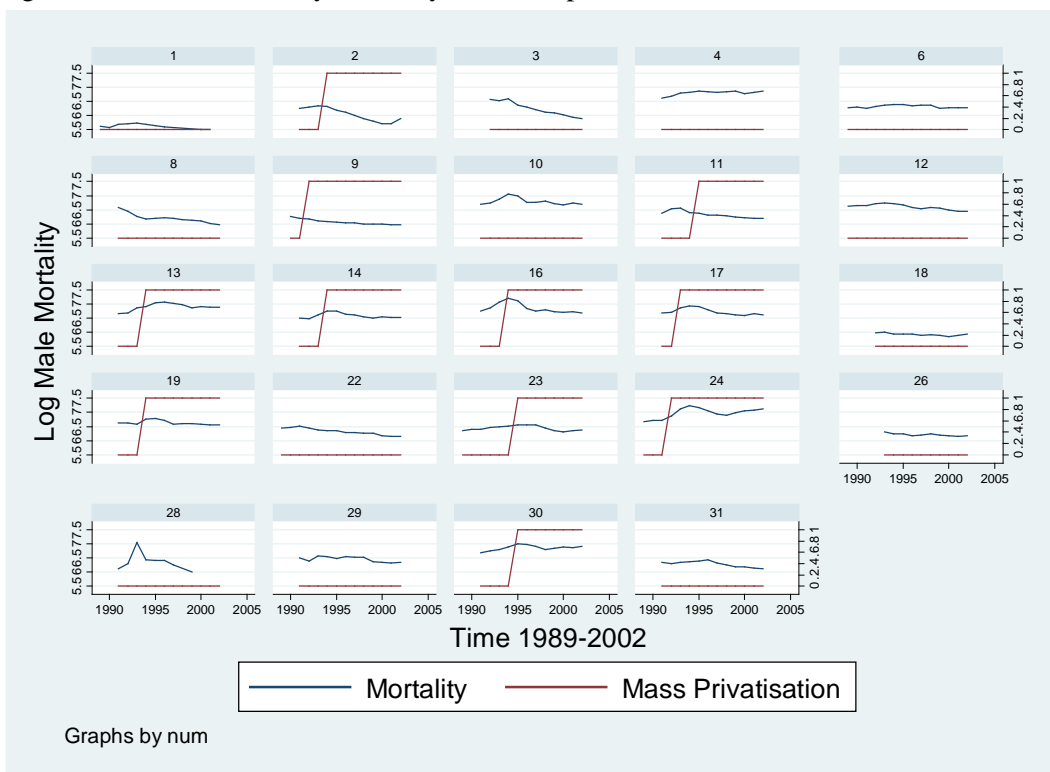
Notes: Columns 1-5 represent three differently specified System-GMM estimates described in the text. Coefficient, standard errors (in parentheses) and t-statistics reported to 2 decimal places. \*\*\*/\*\*/\* indicates significance at the 1%/5%/10% level respectively.

Figure 1: Male Life Expectancy



Source: WHO Health for All Database, January 2009

Figure 2: Individual country mortality and mass privatisation time-series



Notes (mass privatisers in bold): 1 = Albania; 2 = **Armenia**; 3 = Azerbaijan; 4 = Belarus; 6 = Bulgaria; 8 = Croatia; 9 = **Czech Republic**; 10 = Estonia; 11 = **Georgia**; 12 = Hungary; 13 = **Kazakhstan**; 14 = **Kyrgyz Republic**; 16 = Latvia; 17 = Lithuania; 18 = Macedonia; 19 = **Moldova**; 22 = Poland; 23 = **Romania**; 24 = **Russia**; 26 = Slovakia; 28 = Tajikistan; 29 = Turkmenistan; 30 = **Ukraine**; 31 = Uzbekistan.

---

<sup>1</sup> See <http://www.imf.org/external/np/vc/2008/072308.htm> for the IMF's response to this study and <http://tinyurl.com/5r2kywa> for subsequent methodological discussion.

<sup>2</sup> The response of Stuckler, King and McKee is published (unrefereed) on the *Lancet* website and can be enjoyed here: (<http://download.thelancet.com/mmc/journals/lancet/PIIS0140673610601602/mmc1.pdf>)

<sup>3</sup> Research in this spirit often presents important new data (Gakidou et al, 2010), embraces highly topical questions (Stuckler et al, 2010) or tackles important questions of political economy (Lu et al, 2010).

<sup>4</sup> Specifically, the estimates in (3) ignore both the inverse U-shaped evolution of mortality during this period (figure 2) and the upward trends in mass privatisation, GDP per capita and higher education.

<sup>5</sup> This approach provides a weighted average of the previous two effects models, in which the weight is a function of the relative variation in the two dimensions (Baltagi, 2008). Thus the countries and years included become doubly important since this determines the weighting (Maddala, 1999).

<sup>6</sup> This assertion was made in their review of this paper.

<sup>7</sup> SKM suggest that the smaller number of significant results is reason enough to doubt our findings. However, we are comfortable with these findings. The theory and the evidence regarding the short-run relationship between health on the one hand, and GDP, democracy, education, war, urbanization and the dependency ratio on the other, is inconclusive.