EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

Econ Journal Watch

Scholarly Comments on Academic Economics
Volume 18, Issue 1, March 2021

In Memoriam

In Memoriam

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

Econ Journal Watch 18(1)
March 2021: 3–20

Immigration's Effect on Institutional Quality: The Place of Simpler Evidence

Garett Jones[1] and Ryan Fraser[2]

Link to Abstract

The idea that the inclusion of a multitude of control variables will necessarily improve (and will not worsen) causal inference is a methodological urban legend at best.
—Julia Rohrer (2018, 28)

An important question is whether migration affects institutional quality over decades.[3] Since productivity and experience with reasonably competent governance differ enormously across countries, the most important papers (including Dimant, Krieger, and Redlin 2015; Clark et al. 2015; and Bologna Pavlik, Lujan Padilla, and Powell 2019) have focused on whether large migration flows from relatively poor or corrupt countries predict lower institutional quality. These papers have reported regression results that use the stock or flow of a nation's immigrants from such origin countries as a control variable, and then use those migration variables to predict subsequent institutional quality. Unfortunately, when reporting the relationship between immigration from relatively poor or corrupt countries and

1. George Mason University, Fairfax, VA 22030.
2. U.S. Bureau of Labor Statistics, Washington, DC 20212. All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.
3. For suggestive evidence that the *threat* of emigration can improve a nation's institutions, see Hall (2016), who reports that institutional quality has tended to improve more in nations where residents can exit the nation more easily. Ease of exit is proxied solely by the ratio of length of national borders to coastline—hence it is implicitly a measure of border crenulation—and so further investigation into proxies for low relative cost of emigration is certainly warranted.

subsequent changes in institutional quality, none of the important papers just cited reports simple correlations or scatterplots to give readers a sense of the underlying data. We rectify that omission here. We draw attention to the potential of overcontrol bias—in particular, of controlling for proxies of the dependent variable—to obscure strong, important statistical relationships in data.

What feeling does your statistical intuition prompt inside of you when you consider the following investigations?

- A regression of restaurant success on chef expertise that also controls for that restaurant's online rating
- A regression of long-run inflation on central bank independence that also controls for long-run money growth
- A regression of firm output growth on that firm's access to financial markets that also controls for the firm's capital and labor growth

In each of these cases, it's plausible that the conventional, expected correlations between

- restaurant success and chef expertise,
- inflation and central bank independence, and
- firm growth and access to finance

could well be rendered statistically insignificant if the researcher controlled for these extra variables. Overcontrol bias can occur when a regression of an outcome Y on a proposed cause X also controls for other proxies of Y or for other proxies of X. Psychologist W. Joel Schneider (2007) illustrates one version of overcontrol bias with a quip: "I am as tall as the Rocky Mountains! (After controlling for barometric pressure)."

When regressions include as controls a combination of proxies of the proposed causal independent variable, proxies of the dependent variable, or factors that are proposed as intermediate causal channels between proposed cause and effect, such regressions can fall prey to overcontrol bias. Controlling for intermediate channels may seem intuitive, but it can eclipse obvious relationships. Obvious bivariate relationships can be reduced, eliminated, or even see the sign change when controlling for mediating channels. How should economists change their empirical approaches to account for overcontrol bias? Psychologist Julia Rohrer offers simple, practical advice in an influential recent essay:

> Mediators are causally affected by the independent variable… A solid rule of thumb is that researchers should not control for such posttreatment variables. (Rohrer 2018, 34)

Rohrer's essay on overcontrol bias is the best recent overview of subject and is strongly recommended. As she notes:

> Often, the analysis follows the rationale that "more control" is always better than less. Models resulting from such an approach have been labeled "garbage-can regressions." (Rohrer 2018, 28)

An alternative to the mediator-heavy regression often exists: plainness and simplicity in evidence. Economists test their models by checking to see whether the models replicate durable correlations, sometimes called 'stylized facts,' such as:

- the positive relationship between the savings rate and GDP per capita;
- the roughly 2/3 share of national income that goes to workers;
- the positive relationship between central bank independence and inflation;
- the relatively stable return to capital across the decades in rich countries, combined with a rising return to labor; and
- the nearly-perfect, 1-for-1 relationship between long-run money growth and inflation between countries.

Such stylized facts are simple—one might even say simplistic. They are rarely derived from multivariate regressions; instead they are simple patterns, scatterplots, time trends, correlations. Economists check to see whether a particular model can replicate them—if the model fits them, that's not a complete test of the model, but it's a good sign.

We report, to our knowledge for the first time, simpler evidence about the relationship between ten- to twenty-year changes in immigration from relatively poor or corrupt countries and two-decade changes in institutional quality. To reduce concerns about data mining, we solely use relevant data from a *Southern Economic Journal* article on the topic by Jamie Bologna Pavlik, Estefania Lujan Padilla, and Benjamin Powell (2019—henceforth "BLP").

The indices used in the institutional quality literature lack transparent units, and the functional form—linear or non-linear—between immigrant experience and institutional quality is unclear. For those reasons, we emphasize the Spearman and Kendall *rank correlation* results. These two methods test only for the rank-order relationship between pairs of numbers, and hence dramatically reduce the risk that outliers can drive results. Economists are overwhelmingly more likely to use the Pearson correlation coefficient (the familiar *r*), but Pearson's *r* is formally designed for linear relationships where the variables are normally distributed. For cases where there is no particular reason to believe in a linear relationship between the two variables, or where the variables aren't normally distributed, a rank

correlation statistic like Spearman's or Kendall's is often recommended. As Wassily Hoeffding noted in 1957 in the pages of *Econometrica*:

> Ranking methods are generally based on fewer assumptions than the usual numerical methods; [and] they are invariant under arbitrary changes of the scale of the measurements… (Hoeffding 1957)

Psychologists are particularly likely to use Spearman correlations. In part they do so because their units of measurement of psychological behavior are often arbitrary, and closer to rankings than true unit measurements; in part they do so because there is little reason to think the relationship between any two measurements will be linear, and little reason to think the measures are themselves normally distributed. Anthony Bishara and James Hittner (2012) review the advice given in widely used psychometrics textbooks:

> By far, the most frequent recommendation was to use Spearman's rank-order correlation—the argument being that Spearman's nonparametric test would be more valid than Pearson's test when parametric assumptions are violated. (Bishara and Hittner 2012)

As noted, we also report the Kendall rank correlation, often denoted by the Greek letter τ. The difference between Spearman and Kendall is somewhat like the difference between horseshoes and basketball: With the Spearman correlation, being off from a perfect rank-order match by a little bit is better than being off by a lot. With Kendall, by contrast, 'an inch is as good as a mile,' since Kendall only tracks whether, for two observations, the two variables have the same (concordant) or different (discordant) rank orders. So Spearman is more like horseshoes (close counts for something), while Kendall is more like basketball (a miss is a miss).[4]

Since we have no idea whether the true relationship between immigration and institutional quality—if one even exists—will be linear or nonlinear, and since the rates of migration are obviously non-normal—both highly skewed and massively kurtotic—we place the most weight on the Spearman and Kendall rank correlation results, not on the Pearson correlations. We do, however, also report conventional Pearson correlations (which presume linearity and normality) and

4. Spearman's ϱ is calculated like Pearson's r, but using rank orders of each variable instead of the actual value of the variable. Thus, Spearman's ϱ is a simple correlation coefficient, but for rankings. Kendall's rank correlation, by contrast, counts up the percentage of all possible pairs of observations where the rank order is concordant (same order) rather than discordant (different order), and then uses that fact to calculate Kendall's τ. With Kendall unlike with Spearman (or Pearson) the magnitude of the concordance or discordance is irrelevant. The magnitude of a Kendall correlation is not directly comparable to that the of the Spearman (or Pearson) correlation.

tentatively discuss magnitudes as well. In addition, we offer, to our knowledge for the first time, scatterplots of such data. This short paper, in other words, is designed to rectify omissions, to fill gaps that in other economics literatures are routinely filled in.

We report the rank correlations between the 10- to 20-year change in the percentage of a nation's population from relatively poor or corrupt countries (in a precise sense defined below) and the 20-year change in the Fraser Institute's Economic Freedom of the World (EFW) index. The Spearman rank correlation between these migration measures and EFW is always in the range of −0.41 to −0.49, while the Kendall rank correlation is always between −0.27 and −0.32. Therefore, in this sample, more immigration from relatively poor or corrupt nations is associated with relative declines in institutional quality. These rank correlation relationships are always significant at the 0.5-percent level. We appear to be the first to report such a negative rank correlation relationship.

In their recent book *Wretched Refuse? The Political Economy of Immigration and Institutions*, Alex Nowrasteh and Benjamin Powell report:

> Regardless of the immigration measure used or the precise regression specification, we have not found a single instance in which immigration is associated with less economic freedom. (Nowrasteh and Powell 2020, 132)

As we show below, a resort to simpler methods rather than heavily controlled regressions might have led to a different statement.

The Pearson correlation (which again, presumes linearity and normality) between the same migration measures and a nation's change in Control of Corruption—one of the components of the World Bank's *Worldwide Governance Indicators*—is always negative and in the range of −0.11 and −0.21, and statistically significant at the 10-percent level in three out of four specifications. We offer a suggestive structural model that could frame future discussions of immigration and institutional quality, then report the stylized facts with correlations and scatterplots to illustrate this heretofore unreported pattern of relationships. As a suggestive extension, we also report regression results that simultaneously control for immigration from relatively poor countries as well as immigration from all other countries. While F-tests cannot reject the equality of the two coefficients, the coefficient on immigration from relatively poor countries is much larger in absolute value and always negative.

Our statistical model

We offer a simple model to frame the discussion of overcontrol bias and to motivate the reporting of change-on-change correlations. The idea is more general than the model below and can be summed up as follows:

> If *levels* of traits of people within a nation contribute to the nation's *level* of institutional quality, then *changes* in those national traits will cause *changes* in a nation's institutional quality, perhaps with a lag.

We do not have data on the traits of people. Instead, we have data on average measures of institutional quality (and income per capita) in an immigrant's nation of origin. We thus test the hypothesis that some traits that shape home-country institutional quality and productivity tend to migrate, to some degree, to an immigrant's destination country, perhaps with a lag. And we do not dispute that causality also works in the opposite direction, that is, that institutions affect the traits of people.

Nation *i*'s level of institutional quality in year t, I_{it}, is generated according to the following linear model. Suppressing the constant, we have:

$$I_{it} = \gamma_i + \theta_t + \alpha L_{it} + \varepsilon_{it} \tag{1}$$

Here, γ_i is a country-specific factor that might represent long-standing cultural and even geographic drivers of institutional quality—factors that won't materially change over a few decades. The θ_t factor captures global average shifts in institutional quality, such as the move toward markets in the 1990s, and also captures worldwide period-level biases in the measurement of the index itself. L_{it} is the critical variable: it represents the percentage of nation *i*'s population who are migrants from countries with 'low' levels of either productivity or transparent governance. In BLP (2019), the standard for 'low' is either one standard deviation below country *i*'s level of productivity or one standard deviation above country *i*'s level of corruption. Of course, that implies that for countries with the very lowest levels of productivity or transparent governance, those values in BLP are mathematically equal to zero. The parameter a, which in principle may be positive, negative, or zero, captures how much migrants from weaker-performing countries shape institutional quality, if at all.

While we do not formally model the social process by which this might occur, one possible channel is especially worth noting. One of us has written recently in the American context about "one particular way immigration is likely already affecting institutions: through a populist backlash channel" (Jones 2018, 342). Hostility to new arrivals is, alas, part of the human experience, and the

backlash of a majority to newcomers from relatively poor and corrupt countries may push nations away from economic freedom and neutral competence and t0ward policies and practices that favor insiders. The degree (if any) to which this backlash channel, a shifting median voter channel (e.g., Caplan 2011), or a cultural change channel (e.g., Alesina and Giuliano 2015) might be important we leave to future research.

Finally, ε_{it} captures all other omitted, time-varying country-level drivers of national institutional quality. We do *not* assume ε_{it} is i.i.d., and in particular we do not assume the epsilon term is uncorrelated with L_{it}. However, we do strongly suggest that omitted, country-specific, time-varying institution-improving variables are likely to be *positively* correlated with L_{it}, since national economic success tends to attract migrants of all skill levels. Thus, the modest negative correlations we report below may well be understatements: migrants surely try to go to places that are good, and they plausibly try to go to places that are getting better. It is plausible, therefore, that our omitted variable bias leads to statistical results that understate the relationship we report.

Taking first differences of (1) and collecting nuisance terms into $\tilde{\varepsilon}_{it}$ yields:

$$\Delta I_{it} = \alpha \Delta L_{it} + \tilde{\varepsilon}_{it} \tag{2}$$

Here $\tilde{\varepsilon}_{it} = \Delta \varepsilon_{it} + \Delta \theta_t$. The first term of $\tilde{\varepsilon}_{it}$ captures changes in country-specific time-varying factors (again, plausibly *positively* correlated with L_{it}, since migrants seek prosperity), and the second term captures global trends in measured institutional quality.

Specification (2), which we use below, certainly has its limitations, but take a moment to compare it to specifications used in BLP. They have the following general form, where the circumflex, inverted-circumflex, and bar versions of I_{it} indicate alternative institutional quality measures, and the betas indicate their coefficients:

$$I_{it} = \beta_1 \hat{I}_{it} + \beta_2 \check{I}_{it} + \beta_3 \bar{I}_{it} + \beta_4 I_{i(t-1)} + \alpha L_{i(t-1)} + \varepsilon_{it} \tag{3}$$

For example, in their Tables 6a, 6b, 8, and 10 (BLP, 1253, 1254, 1256, 1258), paired with their discussion in the associated subsection (ibid., 1252–1254), the authors *always* control for all of the following: the Fraser Institute's Economic Freedom of the World index, the Center for Systemic Peace's Polity Score, Freedom House's Freedom of the Press score, the size of the shadow economy, and average GDP per capita. The dependent variable is included as a lagged dependent variable, while

the other institutional controls in their regressions are averaged over a 20-year period or even a five-year period depending on the specification, and hence include contemporaneous data in those averages. Thus, there is substantial risk—even a presumption, to our mind—of overcontrol bias in BLP's results, since recent immigration levels, $L_{i(t-1)}$, could influence all of these potential mediators, and since institutional change is usually slow. In the aforementioned language of Rohrer, these institutional and income controls are potentially "posttreatment variables;" and so these specifications in BLP appear quite likely to run afoul of her (2018) aforementioned "solid rule of thumb" that one should not control for such variables.

BLP do report some simpler results, but only for a different set of immigration measures. In their Table 2, they provide what they call "Baseline Results" of the relationship between various "Immigrant Stocks" and corruption—but only when the "Immigrant Stocks" are broken down into immigrants from OECD countries versus those from non-OECD countries (BLP, 1249). They report:

> The only statistically significant association between immigration and corruption that we found, in these baseline regressions, was that a higher immigrant stock, in 1995, that originated from OECD origin countries was associated with a lower level of corruption in 2015. (BLP, 1249)

It is perhaps notable that they find that greater immigration from OECD countries predicts lower future corruption relative to immigration from non-OECD countries. These are stocks rather than flows—levels rather than medium-run changes—but nevertheless they are examples of simpler statistical evidence, and Tables 6a and 6b (BLP, 1253, 1254) *never* include the extra controls for near-contemporaneous institutional quality and productivity. Such statistical clarity is welcome. However, the authors do not report such simpler baseline results for the immigration measures we discuss below.

Data

All data are from Bologna Pavlik, Lujan Padilla, and Powell (2019), and were generously provided by the authors. The data in levels were all drawn from their "Data" worksheet and transformed into changes accordingly. BLP report complete data for 110 countries. Summary statistics for the change variables are reported in Table 1.

TABLE 1. Summary statistics: changes in migrant population and changes in institutional quality

	Change in percentage of population who have immigrated from poorer countries, 1995–2005	Change in percentage of population who have immigrated from poorer countries, 1995–2015	Change in percentage of population who have immigrated from more corrupt countries, 1995–2015	Change in Economic Freedom of the World, 1995–2015	Change in Control of Corruption, 1995–2015	Change in Control of Corruption, 1995–2005 annual mean to 2005–2015 annual mean
Mean	0.33	1.07	1.67	0.665	−0.036	−0.028
Median	0.01	0.06	0.04	0.505	−0.018	−0.038
Standard deviation	1.34	3.24	3.70	0.93	0.37	0.22
Skewness	2.55	3.30	2.74	0.68	0.26	0.52
Kurtosis	15.22	15.21	11.13	4.81	5.29	5.24

Note: All data are derived from Bologna Pavlik, Lujan Padilla, and Powell (2019).

The three migrant change variables all report changes in the percentage of nation *i*'s population that came from countries where either the level of income is one standard deviation *below* that of country *i* or where the score on the corruption index (described below) is one standard deviation *worse* than that of country *i*. Migrant population data are from the World Bank's *International Migrant Stock by Destination and Origin* data set.

Notice that these migrant change variables are all focused on lower-performing countries, around the *bottom* 15 percent of the global sample—and BLP did not report simple "Baseline Results" looking at immigration from this group of countries. By contrast, the above-mentioned OECD results in BLP, for which simple "Baseline Results" *were* reported, focus on immigrants from the world's *best performing* economies. BLP then compared immigration from OECD countries (of which today there are 37) to immigration from all non-OECD countries, arguably a version of 'the best versus the rest.' While BLP's OECD results checked to see whether large numbers of immigrants from countries with institutional quality well *above* the global mean predicted better institutional quality in the destination country[5]—finding that, on average, the answer was yes—the results below focus

5. Furthermore, the OECD includes some countries that both have relatively weak institutional quality, and that have a substantial amount of emigration—so percentage of immigrants from OECD countries may be a poor proxy for percentage of immigrants from countries with high institutional quality. For example, Mexico, an OECD member, was in 2015 the country of origin for 6.7 million immigrants according to the UN, the most of any OECD country and double the number of the next-highest OECD source country, Germany. Mexico's 2015 Control of Corruption score was −0.77, compared to Germany's +1.84. Likewise, Turkey and Italy are also OECD countries with substantial amounts of emigration (2.6 million and 3.3 million, respectively), and have Control of Corruption scores of −0.03 and +0.02, respectively. Thus, additional proxies for the institutional quality of emigrant source countries beyond OECD membership appear warranted.

mostly on immigrants from countries with institutional quality well *below* the global mean. BLP reported simple results on whether immigration from *elite* countries has a statistical relationship with corruption. We do the same for immigration from the world's *poorest and most corrupt* countries.

There is high skewness in the migration level data: for instance, in 2015 the median country in the sample has 1.0 percent of its population from such poorer countries, while the mean value is 4.4 percent, and the standard deviation is 11.5 percent. Of course, for the very poorest countries, the number of migrants from countries substantially poorer is mathematically equal to zero (a fact true for both our estimates and for the original BLP estimates). Our correlations use changes not levels, but similarly display that same high level of positive skewness—a sign the data are far from normally distributed. Two of the migrant population measures are the change in the percentage of nation *i*'s population between 1995 to 2005, and between 1995 to 2015, who are immigrants from substantially poorer countries. The third measure is the same change in the percentage of the nation's population between 1995 to 2015 who are from substantially more corrupt countries. These timeframes are used because they are reported in the BLP data set, because none include changes in migrant population that occur *after* the reported changes in institutional quality, and because the 1995 to 2005 measure in particular allows us to take a first look at the possible lag structure between changes in a nation's population traits and changes in its institutional quality.

We also use BLP's data on the Fraser Institute's Economic Freedom of the World (EFW, values ranged from 0 to 10) and the corruption component of the World Bank's *Worldwide Governance Indicators* (with values ranged from −2.5 to +2.5). In both indices, the larger, more positive number is better. Thus a *higher* Control of Corruption (COFC) score implies *greater* control of corruption. Again, we use changes in three relevant measures for which the levels are already calculated in the BLP data set. Two of the three are changes in the EFW and COFC indices between 1995 and 2015. Also, for the COFC, we use their data on the average level of COFC between 1995 and 2005, and again between 2005 and 2015, to calculate the average change between the two periods; this will reduce the effect of year-to-year noise in the estimates, and again allows us to offer a suggestive test of the lag structure.

To be clear about the direction of the relationship, we present a few typical observations. Singapore's EFW score is 8.81, while Venezuela's is 4.92; Switzerland's Control of Corruption score is 2.14, while Pakistan's is −0.81. Higher values accord with conventional expectations of higher institutional quality.

What do simpler methods show?

Scatterplots are reported in Figures 1, 2, and 3, using World Bank codes as data labels; Spearman rank correlations are reported in Table 2, Kendall rank correlations in Table 3, and Pearson correlations in Table 4.

As the first columns of Table 2 and Table 3 make clear, the rank-order relationship between change in migrants from poor countries as a percentage of population and change in economic freedom is reasonably strong, with rank correlations between −0.41 and −0.49 for Spearman and −0.27 to −0.32 for Kendall. Since these two types of correlation coefficients are calculated in quite different ways, the magnitudes of the coefficients are not comparable to each other, though we should note that a Kendall correlation of −0.3 implies that 65 percent of all possible pairs of observations are ranked in the same (negative) direction.[6] All of these relationships are statistically significant at conventional levels, and hence quite unlikely to be pure coincidence. Recall that rank correlations give no weight to the magnitude of a change, and hence give no particular weight to outliers. Ours is the first paper to report this stylized fact, a finding that certainly deserves further theoretical and empirical inquiry. All of the reported rank correlations are strictly negative.

TABLE 2. Spearman rank correlations: change in migrant population vs. change in institutional quality

	Economic Freedom of the World, 1995–2015	Control of Corruption, 1995–2015	Control of Corruption, 1995–2005 annual mean to 2005–2015 annual mean
Poor, ’95–’05	−0.41* [0.00]	−0.17 [0.07]	−0.21* [0.03]
Poor, ’95–’15	−0.49* [0.00]	−0.16 [0.09]	
Corrupt, ’95–’15	−0.42* [0.00]	−0.11 [0.23]	

Note: ‘Poor’ and ‘Corrupt’ denote percentage of the destination country’s total population who are immigrants from countries one standard deviation poorer or more corrupt than the destination country. Asterisk indicates significance at 5-percent level, p-values in brackets.

6. Professor of Health Research Methods Ronán Michael Conroy in 2015 helpfully pointed out the relevant formula on a ResearchGate discussion of the Kendall rank correlation (link); it’s just a reorganization of the definition of Kendall’s τ.

TABLE 3. Kendall rank correlations: change in migrant population vs. change in institutional quality

	Economic Freedom of the World, 1995–2015	Control of Corruption, 1995–2015	Control of Corruption, 1995–2005 annual mean to 2005–2015 annual mean
Poor, '95–'05	−0.27* [0.00]	−0.12 [0.07]	−0.14* [0.03]
Poor, '95–'15	−0.32* [0.00]	−0.11 [0.08]	
Corrupt, '95–'15	−0.28* [0.00]	−0.08 [0.21]	

Note: 'Poor' and 'Corrupt' denote percentage of the destination country's total population who are immigrants from countries one standard deviation poorer or more corrupt than the destination country. Asterisk indicates significance at 5-percent level, p-values in brackets.

The first two scatterplots—where each of the 110 observations is a country—illustrate the negative relationship—obviously non-linear, and closer to an L-shaped hyperbolic—between the change in the percentage of each nation's population who are immigrants from relatively poor countries and the change in that nation's EFW score. The dozen or so outliers that are essentially irrelevant to Table 2 and 3's rank correlation estimates are obvious in the scatterplots, but of course that is a reason to place more weight on the rank correlations, and less on the visible outliers. The third scatterplot illustrates one of the weakest correlations—between the change in percentage of a nation's population from poor countries and the change in Control of Corruption.

Figure 1. 10-year change in migrant population from poor countries and 20-year change in Economic Freedom of the World

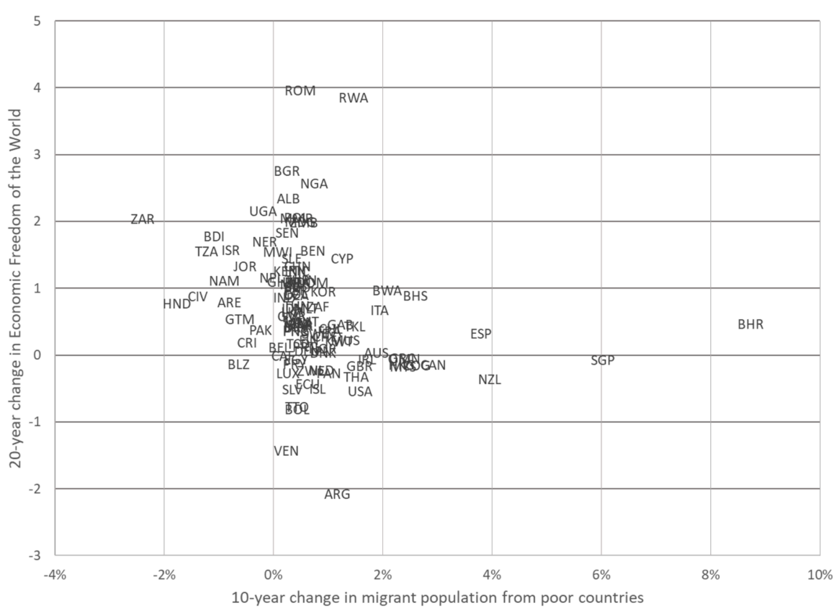


Figure 2. 20-year change in migrant population from poor countries and 20-year change in Economic Freedom of the World

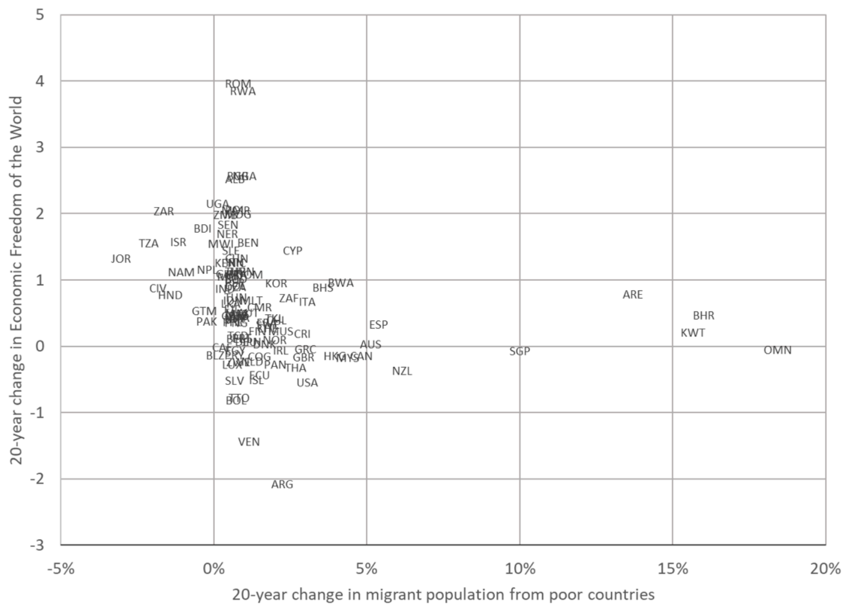


Figure 3. 10-year change in migrant population from poor countries and 20-year change in Control of Corruption

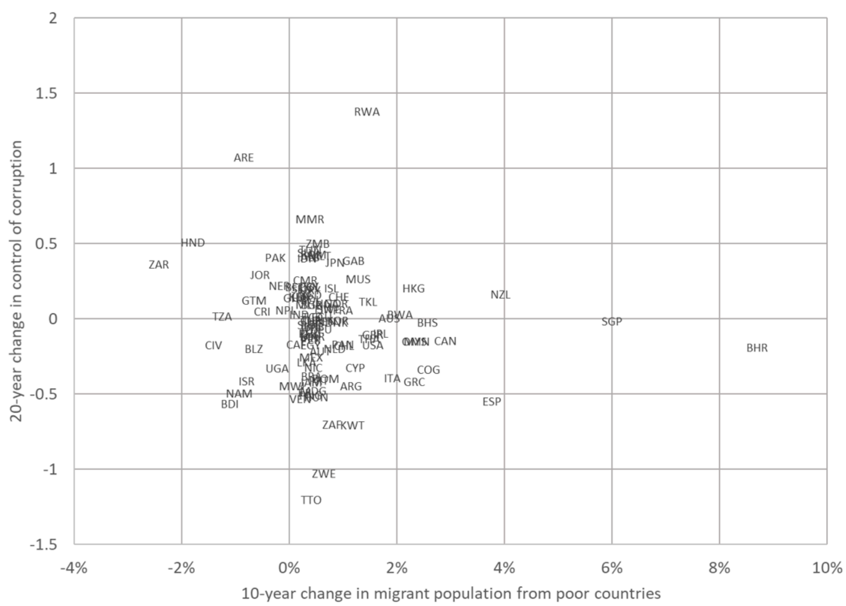


TABLE 4. Pearson correlations: change in migrant population vs. change in institutional quality

	Economic Freedom of the World, 1995–2015	Control of Corruption, 1995–2015	Control of Corruption, 1995–2005 annual mean to 2005–2015 annual mean
Poor, '95–'05	−0.27* [0.00]	−0.11 [0.23]	−0.17 [0.08]
Poor, '95–'15	−0.24* [0.01]	−0.05 [0.64]	
Corrupt, '95–'15	−0.28* [0.00]	−0.05 [0.60]	

Note: 'Poor' and 'Corrupt' denote percentage of the destination country's total population who are immigrants from countries one standard deviation poorer or more corrupt than the destination country. Asterisk indicates significance at 5-percent level, p-values in brackets.

The Pearson correlations, the *r* familiar to economists, are reported to ease discussion of magnitudes, even though the proper functional form is unlikely to have the linear form implied by a Pearson correlation. Consider the −0.27 correlation between the change in migrant share from poor countries between 1995 to 2005 and the change in economic freedom from 1995 to 2015. As a mere matter of prediction—not causation—a linear model would predict that a one standard deviation rise in the change in nation *i*'s population who are from relatively poor countries (1.34 percent of country *i*'s population) would be associated with a decline equal to −0.25 in nation *i*'s EFW score. For 2015 this is approximately the gap between higher-ranked Israel and lower-ranked Iceland, or between higher-ranked Uganda and lower-ranked Indonesia. The United States' value for the change in the U.S. population who are migrants from poor countries between 1995 and 2005 was +1.2 percent, so one would predict a decline in the U.S. EFW score of −0.24.

These results may help explain the tendency toward global convergence in institutional quality reported by Joshua Hall (2016), who notes that there has been substantial convergence in economic freedom around the world since 1980, and hence a decline in the global dispersion of economic freedom scores. Migration from nations with lower EFW scores to those with higher EFW scores would tend to generate both the stylized facts we report here and the stylized fact of a fall in the global dispersion of economic freedom. This is just one suggestion of paths for future work that can build upon the results presented here.

Immigration from relatively poor versus other nations

As a robustness check and as an elementary horserace, we report basic re-

gressions that compare whether immigration from the poorest nations has a weaker or stronger relationship with changes in institutional quality than immigration from other nations. We look at the relationship between the percentage change in migrant population from relatively non-poor countries over the period 1995 to 2005 and changes in our institutional variables, EFW and COFC, over the period 1995 to 2015. Our measure for relatively non-poor countries is drawn directly from the BLP data set: it is a residual, a measure of total migrant population minus migrant population from relatively poor countries, as a percent of the destination country's population. Relatively poor countries are defined as before: countries one standard deviation poorer than the destination country. The two immigration measures will simply be called 'poor' and 'non-poor.'

TABLE 5. Regression results: migrant population from relatively poor versus non-poor countries

	Economic Freedom of the World, 1995–2015 [poor alone]	Economic Freedom of the World, 1995–2015 [poor + non-poor]	Control of Corruption, 1995–2015 [poor alone]	Control of Corruption, 1995–2015 [poor + non-poor]
Poor, '95–'05	−18.65* (6.47)	−17.23* (7.14)	−3.13 (2.61)	−3.40 (2.89)
Non-poor, '95–'05		−4.19 (8.83)		0.80 (3.57)
F-Test: Poor = Rich		0.94		0.59
R^2	0.072	0.074	0.013	0.014
N	110	110	110	110

Note: 'Poor' denotes percentage of the destination country's total population who are immigrants from countries one standard deviation poorer than the destination country. 'Non-poor' denotes percentage of the destination country's total population who are immigrants from any non-poor country, using the same one standard deviation cutoff: 'Non-poor' is thus a residual, total immigrants minus immigrants from poor countries as a percentage of the destination country's population. Asterisk indicates significance at 5-percent level, standard errors in parentheses; constant not reported. The 5-percent critical value for the relevant test of F(1,107) is 3.9.

We report OLS regression results in Table 5. Across the board, the migration variables are once again better predictors of economic freedom than of control of corruption—this may well be driven by the fact that EFW is an aggregate institutional quality variable, a combination of many plausibly important institutional quality factors, while control of corruption is just one among many plausibly important institutional quality factors. The EFW measure may thus contain a stronger overall signal of institutional quality relative to control of corruption. When compared to a simpler regression that only controls for migration from poor countries, including a control for immigrants from non-poor countries offers little improvement in fit. The association between migration from poor countries and EFW is statistically significant, but the non-poor-country immigration statistic is

not statistically significant. When control of corruption is the dependent variable, no variable is significant. Perhaps more important than the lack of statistical significance for non-poor-country immigration is the fact that in the EFW regression the non-poor-country coefficient is less than one-fourth the size of the poor-country coefficient. When control of corruption is a dependent variable—where again, neither control is statistically significant at conventional levels—the non-poor country coefficient has a flipped sign—consistent with the theory that non-poor-country migration improves institutional quality—but even in absolute value the non-poor-country coefficient is still less than one-fourth of the value of the poor country coefficient.

F-statistics for the test of equality of poor- and non-poor-country coefficients are never rejected at conventional levels—so a skeptic could hold the view that both coefficients had the same value, somewhere between the poor-country and non-poor-country values. However, the best unrestricted estimates indicate that immigration from poor countries predicts a substantially greater change in relative institutional quality than immigration from non-poor countries.

Conclusion

In their famous paper "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence," Alberto Alesina and Lawrence Summers (1993) never report a correlation, never report statistical significance. Indeed, they only reported data for sixteen countries. That paper helped create a revolution in how economists think about the relationship between institutions and economic outcomes. As they note:

> The results here are not conclusive in the sense that we have looked at the data only in a very straightforward way; more detailed analysis…is warranted… Our results here do, however, create some presumption that the inflation benefits of central bank independence are likely to outweigh any output costs. (Alesina and Summers 1993, 159)

The present paper adds slightly more statistical detail compared to Alesina and Summers (1993), and has a much larger sample, but the point is similar: that looking "at the data…in a very straightforward way" is the beginning of wisdom. The stylized facts presented here will, we hope, lead to further scientific inquiry into the structural causes of cross-country differences in institutional quality.

Appendix

Data and code related to this research is available from the journal website (**link**).

References

Alesina, Alberto, and Paola Giuliano. 2015. Culture and Institutions. *Journal of Economic Literature* 53(4): 898–944. **Link**

Alesina, Alberto, and Lawrence H. Summers. 1993. Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence. *Journal of Money, Credit and Banking* 25(2): 151–162.

Bishara, Anthony J., and James B. Hittner. 2012. Testing the Significance of a Correlation with Nonnormal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychological Methods* 17(3): 399–417.

Bologna Pavlik, Jamie, Estefania Lujan Padilla, and Benjamin Powell. 2019 (BLP). Cultural Baggage: Do Immigrants Import Corruption? *Southern Economic Journal* 85(4): 1243–1261.

Caplan, Bryan. 2011. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*, new ed. Princeton, N.J.: Princeton University Press.

Clark, J. R., Robert Lawson, Alex Nowrasteh, Benjamin Powell, and Ryan Murphy. 2015. Does Immigration Impact Institutions? *Public Choice* 163(3–4): 321–335.

Dimant, Eugen, Tim Krieger, and Margarete Redlin. 2015. A Crook Is a Crook…But Is He Still a Crook Abroad? On the Effect of Immigration on Destination-Country Corruption. *German Economic Review* 16(4): 464–489.

Hall, Joshua C. 2016. Institutional Convergence: Exit or Voice? *Journal of Economics and Finance* 40: 829–840.

Hoeffding, Wassily. 1957. Review of *Rank Correlation Methods* by M. G. Kendall. *Econometrica* 25(1): 181–183.

Jones, Garett. 2018. The Populist Revolts of 2016: A Hidden Cost of Immigration? *Independent Review* 22(3): 341–347.

Nowrasteh, Alex, and Benjamin Powell. 2020. *Wretched Refuse?: The Political Economy of Immigration and Institutions*. Cambridge, UK: Cambridge University Press.

Rohrer, Julia M. 2018. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* 1(1): 27–42. **Link**

Schneider, William Joel. 2007. IQ and Wealth: The Smart Rich and the Dumb Poor. IQsCorner.com, May 2. **Link**

About the Authors

Garett Jones is a macroeconomist at George Mason University. He is the author of *Hive Mind* and *10% Less Democracy*, both with Stanford University Press. In the past, he has held editorial positions with *The New Palgrave Dictionary of Economics*, *Journal of Neuroscience, Psychology, and Economics*, and *Econ Journal Watch*, and has worked in the United States Senate. His email is jonesgarett@gmail.com.

Ryan Fraser is an entry level Economist at the Office of Employment and Unemployment Statistics in the U.S. Bureau of Labor Statistics. He recently graduated with a Bachelor of Science in Economics from George Mason University with honors in the major in spring of 2020. His research has focused on the economics of immigration and applications of price theory. He lives and works in Fairfax, Virginia.

Bologna Pavlik, Lujan Padilla, and Powell's reply to this article
Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6023/

EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 21–34

Simpler Evidence on Immigration and Institutions: An Assessment

Jamie Bologna Pavlik[1], Estefania Lujan Padilla[2], and Benjamin Powell[3]

LINK TO ABSTRACT

The "New Economic Case for Immigration Restrictions" has challenged the consensus view that moving from existing restrictive immigration policies to unrestricted, or free, immigration would generate massive global gains in output—the so-called "trillion-dollar bills on the sidewalk" (Clemens 2011). Those making the new case for immigration restrictions generally posit that factors responsible for the low productivity in immigrants' origin countries could migrate with immigrants and undermine the high productivity in destination countries, thus wiping out the forecasted trillion-dollar gains. The new economic case for immigration restrictions is an empirical conjecture. However, it was presented (by, e.g., Borjas 2014; 2015) as a theoretical possibility without supporting empirical evidence that immigration has, in fact, carried such a negative externality.

The new economic case for immigration restrictions does not specify the channel through which the factors responsible for low productivity in origin countries are transmitted to destination countries, but the potential impact immigrants have on formal or informal institutions (norms) are one plausible channel. As George Borjas (2015, 961) asks, "What would happen to the institutions and social norms that govern economic exchanges in specific countries after the entry/exit of perhaps hundreds of millions of people?" One of us (Powell), along with coauthors J. R. Clark, Robert Lawson, Alex Nowrasteh, and Ryan Murphy, published the first paper examining how immigration impacts one formal institution related to

1. Texas Tech University, Lubbock, TX 79409. We thank Robert Lawson, Ryan Murphy, J. R. Clark, and Alex Nowrasteh for helpful comments on an earlier draft of this paper.
2. Texas Tech University, Lubbock, TX 79409.
3. Texas Tech University, Lubbock, TX 79409.

productivity—economic freedom (Clark et al. 2015). Numerous papers have since investigated how immigration impacts other formal and informal institutions related to productivity. These papers have used cross-country and cross-state econometric analyses, synthetic control analyses, and analytical narrative case studies.[4] Our paper (Bologna Pavlik et al. 2019) was one such study and investigated the impact that immigration could have on corruption in a cross-country analysis.

The paper by Garett Jones and Ryan Fraser (2021) in this issue of *Econ Journal Watch* is framed explicitly as a critique of our paper (Bologna Pavlik et al. 2019) and as a critique of Clark et al. (2015) and Eugen Dimant, Tim Krieger, and Margarete Redlin (2015). Specifically, Jones and Fraser write "Unfortunately, when reporting the relationship between immigration from relatively poor or corrupt countries and subsequent changes in institutional quality, none of the important papers just cited reports simple correlations or scatterplots to give readers a sense of the underlying data. We rectify that omission here. We draw the attention to the potential of overcontrol bias—in particular, of controlling for proxies of the dependent variable—to obscure strong, important statistical relationships in data" (2021, 3–4). Unfortunately for Jones and Fraser, their critique seriously misses its mark. If Jones and Fraser (or the editor and referees of this journal) compared their own results with regards to corruption to the findings in our paper, they would find that they reconfirm our result: there is essentially no relationship between immigration and corruption in destination countries. Similarly, if Jones and Fraser had read Clark et al. (2015) more carefully they would have realized that that paper never examines the flows of immigrants from relatively poorer or more corrupt countries as Jones and Fraser claim. Thus, we are unsure of how that paper could suffer from the overcontrol bias that Jones and Fraser claim that they correct for. In short, Jones and Fraser's central claim that these studies suffer from overcontrol bias is simply false.

Two of the three papers Jones and Fraser (2021) claim suffer from overcontrol bias examine the impact of immigration on corruption—Bologna Pavlik et al. (2019) and Dimant, Krieger, and Redlin (2015)—while Clark et al. (2015) is the only paper that examines the impact of immigration on economic freedom that they claim suffers from overcontrol bias. Jones and Fraser (2021) use the data set from our paper (Bologna Pavlik et al. 2019) to attempt to show how overcontrol bias masks important relationships between immigration and both economic freedom and corruption. Although they use our data set, our paper never

4. See, e.g., DeBacker et al. (2015); Dimant, Krieger, and Redlin (2015); Clemens and Pritchett (2016); Powell, Clark, and Nowrasteh (2017); Padilla and Cachanosky (2018); Forrester et al. (2019); Nowrasteh, Forrester, and Blondin (2019); Arif et al. (2020); Nowrasteh and Powell (2020); Yao, Bolen, and Williamson (2020; 2021).

examined the impact immigration had on economic freedom; Clark et al. (2015) is the only paper cited by Jones and Fraser that examined that relationship. Economic freedom, not corruption, is the dependent variable in most of the statistically significant results reported by Jones and Fraser (2021). In fact, the phrase *economic freedom* appears 20 times in their paper, and the abbreviation for its measure, "EFW," appears 15 times, while the word *corruption* appears 33 times and the abbreviation for its measure, "COFC," appears only 5 times. It seems reasonable to conclude that Jones and Fraser (2021) is as much a critique of Clark et al. (2015) as it is of our paper (Bologna Pavlik et al. 2019) and to evaluate it as such.[5]

We proceed briefly as follows. The next section illustrates how Jones and Fraser (2021) reconfirm the findings of Bologna Pavlik et al. (2019). The section after illustrates how their critique does not apply to Clark et al. (2015). We then critically examine the potential value in Jones and Fraser's analysis as a standalone empirical contribution to the literature on the relationship between immigration and economic freedom.

Immigration and corruption

In Bologna Pavlik et al. (2019) we studied the relationship between immigration and corruption between 1995 and 2015. Our baseline results examined the relationship between the initial immigrant stock, the subsequent immigrant inflow, and the interaction of the stock and inflow with no contemporaneous controls and only a control for the initial 1995 level of corruption (Bologna Pavlik et al. 2019, 1249 Table 2). Jones and Fraser (2021, 10) note that this is precisely the "simpler statistical evidence" they would like to see. Bologna Pavlik et al. (2019) find no general relationship between immigration and changes in corruption. The paper then goes on to look at the same relationship with contemporaneous controls and the results remained largely unchanged (2019, 1250 Table 3). That still remained true after we used another measure of corruption (International Country Risk Guide (ICRG) in place of the World Bank measure) that allowed us to control for prior trends in corruption for the decade preceding our analysis (2019, 1257 Table 9). The paper goes on to look at the effect of immigration at different levels of corruption and economic freedom in destination countries and the effect of immigrants from origin countries with income that is a standard deviation lower or corruption that is a standard deviation higher than in their destination countries.

5. We will not directly evaluate Jones and Fraser (2021)'s merits as a critique of Dimant, Krieger, and Redlin (2015) though, to some extent, we do so indirectly by examining the relationship between immigration and corruption.

These later regressions based on immigrant origin are the ones that Jones and Fraser take issue with.

Having already shown that the "simpler" evidence in our baseline results did not change when we add controls, we opted to use the fully specified model when further breaking down our data—a fairly standard practice. However, Jones and Fraser are correct that we did not report baseline results without controls for the impact of immigrants from only lower income or more corrupt countries. But what did we find in fully controlled regressions? Our results examining the impact of these two groups over a 20-year period report no statistically significant relationship (Bologna Pavlik et al. 2019, 1253 Table 6a). When we switch to an alternative measure of corruption (ICRG) and control for the trend in corruption for prior periods our results are a little more mixed. Those results find no relationship between immigrants from lower income countries and changes in corruption but do find that immigrants from more corrupt origins are associated with decreased corruption in destination countries at conventional levels of statistical significance. Overall, that finding is an outlier in our paper. How do we write up our results? In our results section we state, "Overall, our results indicate that there is no general long-run association between immigration and corruption. … Finally, we do not find support for the idea that immigrants from poorer or more corrupt countries will import their origin country's corruption to their destination country" (ibid., 1254). We reiterate that in the conclusion though we do acknowledge our one outlier finding: "We find limited evidence that increased migration from countries with more corruption may actually reduce corruption in the destination country" (ibid., 1259). Overall, we find a null result.

What do Jones and Fraser find in their simpler and not 'overcontrolled' analysis of the relationship between immigration from poorer or more corrupt countries and subsequent changes in corruption? Essentially, a null result. In their Tables 2, 3, and 4 they look at simple correlations (Spearman rank, Kendall rank, Pearson) of the 20-year change in corruption score with two measures ('95–'05 and '95–'15) of the change in the immigrant population from poorer countries and one measure ('95–'15) of the change in the migrant population from more corrupt countries (Jones and Fraser 2021, 13, 14, 16). None of their nine correlations are statistically significant. Jones and Fraser also run two regressions on the impact of immigrants from poorer countries on corruption in destination countries in Table 5 and similarly report no statistically significant relationship (ibid., 17). Tables 2–4 do report something labeled "Control of Corruption, 1995–2005 annual mean to 2005–2015 annual mean" for the measure of immigrants from poor countries from '95–'05 but not for their other two measures. It is not clear what these results are measuring as the labels are unclear and they are not described in text of their results, but two of the three measures are statistically significant at the 10 percent level. It

would be safe to characterize their overall result as finding no relationship between immigrants from poorer or more corrupt countries and changes in corruption in destination countries.

Since Jones and Fraser (2021, 9–10) take issue with the possibility of "over-control bias" in our Tables 6a, 6b, 8, and 10, it would have been more straightforward to test for such bias by replicating our results and then removing the variables that they suspect are responsible for overcontrol bias. We do that here for our baseline (Bologna Pavlik et al. 2019, 1253 Table 6a) in this article's Table 1.

TABLE 1. The effect of immigration from relatively 'worse' origin countries on corruption, with and without basic controls

	Dependent variable: Corruption in 2015							
	Lower-income migrants				More-corrupt migrants			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
'Worse' immigrant stock	0.527 (0.760)		0.097 (1.033)		0.828 (0.587)		0.800 (0.640)	
'Worse' immigrant net inflow		0.543 (1.766)		−2.053 (2.109)		0.538 (1.215)		−0.939 (1.525)
Corruption 1995	0.927*** (0.028)	0.929*** (0.029)	0.807*** (0.074)	0.810*** (0.078)	0.908*** (0.032)	0.924*** (0.034)	0.805*** (0.078)	0.819*** (0.082)
GDP per capita			−0.050 (0.066)	−0.008 (0.063)			−0.079 (0.065)	−0.028 (0.068)
EFW			0.214** (0.084)	0.246*** (0.091)			0.188** (0.085)	0.230** (0.090)
Polity			−0.008 (0.015)	−0.014 (0.015)			0.001 (0.014)	−0.010 (0.015)
Shadow economy			−0.002 (0.001)	−0.002 (0.001)			−0.002 (0.002)	−0.002 (0.002)
Human capital			0.097 (0.078)	0.039 (0.092)			0.113 (0.069)	0.063 (0.084)
Freedom of the press			0.002 (0.004)	0.002 (0.004)			0.003 (0.004)	0.002 (0.004)
Constant	−0.037 (0.036)	−0.030 (0.037)	−1.214** (0.550)	−1.633** (0.628)	−0.058 (0.038)	−0.032 (0.037)	−0.904* (0.543)	−1.435** (0.672)
Observations	110	110	100	100	110	110	100	100
R^2	0.888	0.887	0.902	0.903	0.892	0.887	0.905	0.902
Adjusted R^2	0.886	0.885	0.893	0.895	0.890	0.885	0.896	0.894

Notes: Robust standard errors given in parentheses. *, **, *** denote statistical significance at the 10-, 5-, and 1-percent levels respectively. All 'basic' controls are averaged using all available values from 1995–2015.

All contemporaneous controls have been removed and we control only for initial levels of corruption at the start of the period such that we are focusing on *changes* in corruption over a 20-year period. One might be able to construct an argument that controlling for initial levels of corruption leaves a collider bias when examining how accumulated stocks of immigrants impact the subsequent 20-year change in

corruption but is implausible to argue that subsequent 20-year flows of immigrants somehow impacted initial levels of corruption. Table 1 compares this relatively uncontrolled regression with the results reported in Table 6a in our original paper.[6] When we remove the additional controls there is still no statistically significant relationship between immigrants from lower-income countries and corruption in destination countries though one measure does change signs to become positive (less corrupt). Similarly, the results when jettisoning the controls in the regressions examining the impact of immigrants from relatively more corrupt origins remain statistically insignificant and one measure also changes signs to become positive (less corrupt). Thus "overcontrol" is not biasing our previously reported results.

So did our paper overcontrol and thus hide a harmful impact of immigration on corruption, as implied by Jones and Fraser? Most of our results find no relationship between immigration and corruption. That's the conclusion we emphasized. When Jones and Fraser look at the relationship between immigrants from only lower-income or more corrupt countries and subsequent changes in corruption with no controls, most of their results find no relationship. When we modify our main table that they took issue with and report uncontrolled results, we again do not find the bias claimed by Jones and Fraser. Jones and Fraser's simple data and our update to Table 6a are one more robustness test on our null results. These results remain largely the same. Thus, there was no overcontrol bias masking an otherwise significant harmful relationship between immigration and corruption. Thanks.

Immigration and economic freedom

Economic freedom is the dependent variable in most of the statistically significant correlations found in Jones and Fraser's paper. The majority of their textual discussion of their findings is dedicated to discussing economic freedom rather than corruption. Perhaps the bulk of their contribution then is to fix the "overcontrol bias" in the Clark et al. (2015) paper that investigated the impact of immigration on economic freedom, rather than the Bologna Pavlik et al. (2019) paper, which investigated the impact of immigration on corruption. After all, in their opening paragraph Jones and Fraser cite both of these papers and then explicitly state "Unfortunately, when reporting the relationship between immigration

6. Please note that we inverted the Control of Corruption measure in Bologna Pavlik et al. (2019) while Jones and Fraser (2021) did not. Here we do *not* invert it so that Table 1 is more easily comparable to Jones and Fraser's results. Thus in Table 1 positive coefficients signal greater control of corruption, i.e., less corruption.

from relatively poor or corrupt countries and subsequent changes in institutional quality, *none* of the important papers just cited reports simple correlations or scatterplots to give readers a sense of the underlying data" (Jones and Fraser 2021, 3–4, our emphasis). We challenge Jones and Fraser, the editor of EJW, or readers to find any regressions in Clark et al. (2015) that break out the relationship between immigrants from origin countries one standard deviation poorer or more corrupt than destination countries as was done in Bologna Pavlik et al. (2019) and Jones and Fraser's critique. Clark et al. (2015) report 32 different regression results but none break out the data in the way that Jones and Fraser erroneously assert.

The closest that Clark et al. (2015) come to looking at the impact of immigrants from poorer countries is when they separate the stock of immigrants in a destination country by whether they came from OECD or non-OECD origin countries (regressions 2, 7, and 12 in Clark et al. 2015, 327, 328, 329). But note that even in this case "overcontrol bias" is not a problem. Regression 2 in Table 2 of the original article contains no other contemporaneous controls, controlling only for the initial level of economic freedom (ibid., 327). This exact same regression appears in Bologna Pavlik et al. (2019, 1249) with corruption replacing economic freedom and in that case Jones and Fraser (2021, 10) note that these are "examples of simpler statistical evidence" that they would like to see. Thus, this cannot be the result that they take issue with, yet Clark et al. (2015) never attempt to look at any other measure of immigration from relatively poorer origins. Clark et al. do add additional controls as they continue analyzing the impact of immigrants from OECD and non-OECD origins and find that in each case, non-OECD immigrants are associated with larger subsequent improvements in economic freedom. The magnitude of the coefficient and its statistical significance increases as additional controls are added, but it remains that the results are significant in the baseline results with no contemporaneous controls. Scholars can legitimately debate which model most accurately captures the relationship but there is no "overcontrol bias" masking a hidden relationship.

Clark et al. (2015) was the first paper to empirically examine the new economic case for immigration restrictions. As such, that paper focused on the most general relationship between total immigrant stocks and flows and a measure of institutional quality (economic freedom) associated with higher productivity. It is unsurprising that papers building on that contribution have continued to refine how they look at data to examine how different origins, destinations, regions, times, or outcomes contribute to our understanding of the merits or demerits of the new economic case for immigration restrictions. To our knowledge, no one has examined the impact of immigration on economic freedom in a cross-country setting while specifically looking at the impact of immigrant flows from relatively poorer or more corrupt origins as we did in our paper examining the impact of

those immigrant flows on corruption. Rather than issuing misguided claims about overcontrol bias in existing studies, Jones and Fraser could have addressed this gap in the literature. Unfortunately, since they did not make a full analysis of the impact of these immigrant flows on economic freedom the sole focus of their study, their contribution to filling this gap is inadequate at best.

What do Jones and Fraser contribute?

Jones and Fraser (2021) is the first study to look at how immigrants from relatively poorer countries impact economic freedom in destination countries in a cross-country analysis. This is an important topic that is worthy of inquiry. Unfortunately, Jones and Fraser contribute only a scatterplot, a few simple correlations, and two OLS regressions to this inquiry. While scatterplots and correlations are a fine starting point, that is all that they are, a starting point. In this case they are also a misleading starting point, as we will show by replicating and then building on Jones and Fraser's Table 5 OLS regressions (2021, 17).

There are hundreds of papers that have used economic freedom as an explanatory variable over the last 25 years, but more recently a sizable literature has developed examining economic freedom as the dependent variable.[7] The most consistent finding across studies that examine changes in economic freedom is that higher initial levels of economic freedom are negatively associated with subsequent increases in economic freedom. In other words, the higher the initial economic freedom level, the harder it is to improve. This is dictated, at least in part, by the construction of the index itself. The earlier literature studying economic freedom as an independent variable has found it strongly correlated with higher income levels, and it is well known that immigrants tend to immigrate from poorer countries to richer countries.[8] This all implies that Jones and Fraser (2021) have built-in omitted variable bias in their simple correlations. We rectify this by replicating their OLS regression and then controlling for the initial level of economic freedom in Table 2. Since their OLS analysis analyzed the impact of subsequent flows of immigration, it is not plausible that initial levels of economic freedom were somehow caused by these subsequent flows. Thus, overcontrol via collider bias is not possible.

Note that Jones and Fraser (2021) claim in their text that they are following the definition of relatively poor as used in Bologna Pavlik et al. (2019): "'Relatively

7. *See* Hall and Lawson (2014) for a survey of the literature using economic freedom as an explanatory variable and see Lawson et al. (2020) for a survey of the literature examining economic freedom as a dependent variable.

8. Additionally, Ashby (2010) finds that even once per capita GDP is controlled for higher economic freedom is still associated with higher immigration.

poor countries' are defined as before: countries one standard deviation poorer than the destination country" (Jones and Fraser 2021, 17). However, in examining their data and code it turns out that they mistakenly coded relatively poor as simply immigrants from the poorest 50 percent of countries. Our Table 2 uses their stated definition—one standard deviation—but in our appendix we include a table conducting an identical exercise with their mistaken coding of relatively poor as bottom 50 percent and show essentially the same results.

Regressions 9 and 10 in Table 2 replicate Jones and Fraser's result and, like their Table 5 (Jones and Fraser 2021, 17), find that greater immigrant flows are negatively associated with subsequent changes in economic freedom. Regressions 11 and 12 in Table 2 repeat these same regressions while adding only a control for the initial level of economic freedom. As can be seen, in both regressions the impact of immigrant inflows from relatively poorer countries loses its statistical significance, the coefficients decrease substantially in magnitude, and in one case changes sign to become positively associated with subsequent increases in economic freedom. Regressions 13 and 14 in Table 2 repeat this same exercise while changing our immigrant inflow variable to include the entire 20-year time period. The net inflow of immigrants from poorer countries remains statistically insignificant in both regressions. However, now both measures have positive coefficients indicating a greater immigrant inflow would be positively associated with larger improvements in economic freedom.

TABLE 2. The effect of immigration from relatively 'worse' origin countries on economic freedom

	Dependent Variable: Change EFW 1995–2015					
	Inflow 1995–2005				Inflow 1995–2015	
	(9)	(10)	(11)	(12)	(13)	(14)
'Poor' immigrant net inflow	−19.682** (7.590)	−15.690* (8.510)	2.714 (6.252)	−0.011 (6.658)	0.595 (2.265)	0.540 (2.271)
'Non-poor' immigrant net inflow		−8.101 (6.483)		7.224 (6.148)		2.369 (3.406)
Economic Freedom 1995			−0.489*** (0.054)	−0.507*** (0.056)	−0.483*** (0.052)	−0.501*** (0.058)
Constant	0.755*** (0.094)	0.748*** (0.094)	3.671*** (0.330)	3.786*** (0.344)	3.645*** (0.322)	3.742*** (0.352)
Observations	110	110	110	110	110	110
R^2	0.059	0.068	0.466	0.473	0.466	0.468
Adjusted R^2	0.0499	0.0505	0.456	0.458	0.456	0.453

Notes: Standard errors given in parentheses. *, **, *** denote statistical significance at the 10-, 5-, and 1-percent levels respectively.

We do not view these results as definitive in the least. The impact of immigrants from poorer (or less free, more corrupt, etc.) countries on destination

country institutions is a subject worthy of a much more thorough investigation than is appropriate as a response to Jones and Fraser. Our results here only indicate that their simple analysis sheds little light on the relationship.

Conclusion

Jones and Fraser (2021) fails massively as a critique of Bologna Pavlik et al. (2019) and Clark et al. (2015). Jones and Fraser claim that both studies suffer from overcontrol bias and thus the results of these studies cannot be trusted. When Jones and Fraser attempt to show the simple relationship between immigration from poorer or more corrupt countries and corruption without any controls they find essentially no relationship. That is generally the same conclusion Bologna Pavlik et al. (2019) found and emphasized. Thus, overcontrol bias is not a problem in that study. Jones and Fraser are even more misguided in their claim that they are correcting for overcontrol in Clark et al.'s (2015) examination of the impact of immigration on economic freedom because Clark et al. never examine the relationship between flows of immigrants from poorer or more corrupt countries and economic freedom as Jones and Fraser claim! We wonder how such a misguided framing even made it through a peer review process.[9]

If there is any merit in Jones and Fraser's paper it is that they are the first paper to *start to examine* the relationship between immigrants from poorer or more corrupt countries and the impact they have on economic freedom on destination countries. Unfortunately, we must emphasize 'start to examine' because, as written, that investigation is seriously underdeveloped. As we illustrated in the previous section, once the most standard control from the literature studying changes in economic freedom as a dependent variable is included (the initial level), Jones and Fraser's claimed negative relationship between immigration flows and changes in economic freedom loses statistical significance and reverses sign.

Where does this leave the state of the debate about the empirical relevance of the new case for immigration restrictions? Jones and Fraser have given us no reason to discount the general findings of the two papers they criticize. They do find new results that were previously uninvestigated showing simple scatterplots and correlations between immigration from poorer or more corrupt origins are

9. Our wonder has subsequently decreased after correspondence with EJW editor Dan Klein where he encouraged us to consider removing this sentence because in his view editors and referees wouldn't normally check the legitimacy of citations such as Clark et al. (2015) in Jones and Fraser's paper (2021). In fact, the editor emailed one of us (Powell) asking for a copy of Clark et al. (2015) after we submitted our reply to the accepted Jones and Fraser (2021) critique. Readers can judge for themselves whether an editor or a journal's referees should be familiar with one of the main papers a submission purports to critique.

correlated with subsequent decreases in economic freedom, but that result falls apart with a single standard control. In contrast, a multitude of studies have found either no statistically or economically significant relationship between immigration and economic freedom or a positive relationship.[10] Nowrasteh and Powell (2020) comprehensively document all of the studies to date that empirically assessing the new case for immigration restrictions and conclude that "there is no Q.E.D. here. We cannot rule out that, in some cases, in some places, from some particular immigrant flows, a negative externality that undermines formal and informal institutions or norms related to productivity does exist. However, in general, our findings should make scholars skeptical of how widely relevant the new case for immigration restrictions is" (p. 284). Jones and Fraser have given us no cause to revise that conclusion.

Appendix

Data and code related to this research is available from the journal website (**link**).

TABLE A1. The effect of immigration from relatively 'worse' origin countries on economic freedom, using Jones and Fraser's actual measures of immigration inflows (bottom 50 percent of countries defined as poor)

	Dependent Variable: Change EFW 1995–2015					
	Inflow 1995–2005				Inflow 1995–2015	
	(15)	(16)	(17)	(18)	(19)	(20)
'Poor' immigrant net inflow	−18.648*** (6.465)	−17.229** (7.143)	0.836 (5.394)	−2.214 (5.620)	0.595 (2.265)	0.540 (2.271)
'Non-poor' immigrant net inflow		−4.193 (8.827)		12.025* (6.875)		2.369 (3.406)
Economic Freedom 1995			−0.483*** (0.054)	−0.508*** (0.056)	−0.483*** (0.052)	−0.501*** (0.058)
Constant	0.727*** (0.089)	0.733*** (0.090)	3.645*** (0.335)	3.778*** (0.341)	3.645*** (0.322)	3.742*** (0.352)
Observations	110	110	110	110	110	110
R^2	0.072	0.073	0.466	0.481	0.466	0.468
Adjusted R^2	0.0629	0.0562	0.456	0.466	0.456	0.453

Notes: Standard errors given in parentheses. *, **, *** denote statistical significance at the 10-, 5-, and 1-percent levels respectively.

10. Powell, Clark, and Nowrasteh (2017); Padilla and Cachanosky (2018); Nowrasteh, Forrester and Blondin (2019); Arif et al. (2020); Nowrasteh and Powell (2020); Yao, Bolen, and Williamson (2020; 2021).

References

Arif, Imran, Adam Hoffer, Dean Stansel, and Donald Lacombe. 2020. Economic Freedom and Migration: A Metro Area-Level Analysis. *Southern Economics Journal* 87(1): 170–190.

Ashby, Nathan. 2010. Freedom and International Migration. *Southern Economic Journal* 77(1): 49–62.

Bologna Pavlik, Jamie, Estefania Lujan Padilla, and Benjamin Powell. 2019. Cultural Baggage: Do Immigrants Import Corruption? *Southern Economic Journal* 85(4): 1243–1261.

Borjas, George J. 2014. *Immigration Economics*. Cambridge, Mass.: Harvard University Press.

Borjas, George J. 2015. Immigration and Globalization: A Review Essay. *Journal of Economic Literature* 53(4): 961–974.

Clark, J. R., Robert A. Lawson, Alex Nowrasteh, Benjamin Powell, and Ryan Murphy. 2015. Does Immigration Impact Institutions? *Public Choice* 163(3–4): 321–335.

Clemens, Michael A. 2011. Economics and Emigration: Trillion-Dollar Bills on the Sidewalk? *Journal of Economic Perspectives* 25(3): 83–106. **Link**

Clemens, Michael A., and Lant Pritchett. 2016. The New Economics Case for Migration Restrictions: An Assessment. *Journal of Economic Development* 138: 153–164.

DeBacker, Jason, Bradley T. Heim, and Ahn Tran. 2015. Importing Corruption Culture from Overseas: Evidence from Corporate Tax Evasion in the United States. *Journal of Financial Economics* 117(1): 122–138.

Dimant, Eugen, Tim Krieger, and Margarete Redlin. 2015. A Crook Is a Crook…But Is He Still a Crook Abroad? On the Effect of Immigration on Destination-Country Corruption. *German Economic Review* 16(4): 464–489.

Forrester, Andrew C., Benjamin Powell, Alex Nowrasteh, and Michelangelo Landgrave. 2019. Do Immigrants Import Terrorism? *Journal of Economic Behavior & Organization* 166: 529–543.

Jones, Garett, and Ryan Fraser. 2021. Immigration's Effect on Institutional Quality: The Place of Simpler Evidence. *Econ Journal Watch* 18(1): 3–20. **Link**

Hall, Joshua C., and Robert A. Lawson. 2014. Economic Freedom of the World: An Accounting of the Literature. *Contemporary Economic Policy* 32(1): 1–19.

Lawson, Robert A., Ryan Murphy, and Benjamin Powell. 2020. The Determinants of Economic Freedom: A Survey. *Contemporary Economic Policy* 34(4): 622–642.

Nowrasteh, Alex, Andrew C. Forrester, and Cole Blondin. 2020. How Mass Immigration Affects Countries with Weak Economic Institutions: A Natural Experiment in Jordan. *World Bank Economic Review* 34(2): 533–549.

Nowrasteh, Alex, and Benjamin Powell. 2020. *Wretched Refuse? The Political Economy of Immigration and Institutions*. Cambridge, UK: Cambridge University Press.

Padilla, Alexandre, and Nicolás Cachanosky. 2018. The Grecian Horse: Does Immigration Lead to the Deterioration of American Institutions? *Public Choice* 174(3–4): 351–405.

Powell, Benjamin, J. R. Clark, and Alex Nowrasteh. 2017. Does Mass Immigration Destroy Institutions? 1990s Israel as a Natural Experiment. *Journal of Economic Behavior & Organization* 141: 83–95.

Yao, Lili, J. Brandon Bolen, and Claudia R. Williamson. 2020. Are Economic Arguments Against Immigration Missing the Boat? The Institutional Effect of the Mariel Boatlift. Working paper.

Yao, Lili, J. Brandon Bolen, and Claudia R. Williamson. 2021 (forthcoming). The Effect of Immigration on US State-Level Institutions: Evidence from the Immigration Reform and Control Act. *Public Choice*.

About the Authors

Jamie Bologna Pavlik is a research fellow at the Free Market Institute and an assistant professor of agricultural and applied economics in the College of Agricultural Sciences and Natural Resources at Texas Tech University. Dr. Bologna Pavlik earned her B.S. in business economics from the University of Penn State at Behrend and her Ph.D. in economics from West Virginia University, where she received the department's Best Doctoral Student Award. Dr. Bologna Pavlik's research interests are in economic growth and development, with a focus on corruption within institutions. Her email address is jamie.bologna@ttu.edu.

Estefanía Luján Padilla is a research associate at the Free Market Institute. She earned her B.A. in Economics with a minor in Finance from the Universidad Francisco Marroquín (UFM) in Guatemala and a M.S. in Agricultural and Applied Economics at Texas Tech University. Prior to joining Texas Tech University, she served as Project Coordinator at the School of Economic Sciences at UFM, a staff researcher at UFM Market Trends, and provided research assistance for projects focused on issues in Guatemala. She has also served as an instructor at the School of Economic Sciences and Henry Hazlitt Center at UFM. Her email address is estefania.lujan-padilla@ttu.edu.

Benjamin Powell is the executive director of the Free Market Institute and a professor of economics in the Jerry S. Rawls College of Business Administration at Texas Tech University. He is a Senior Fellow with the **Independent Institute**, and the secretary-treasurer of the **Southern Economic Association** and the **Association of Private Enterprise Education**. He earned his B.S. in economics and finance from the University of Massachusetts at Lowell, and his M.A. and Ph.D. in economics from George Mason University. He is the author of *Out of Poverty: Sweatshops in the Global Economy* (Cambridge University Press 2014), co-author of *Socialism Sucks: Two Economists Drink Their Way Through the Unfree World* (Regnery 2019) and *Wretched Refuse? The Political Economy of Immigration and Institutions* (Cambridge 2020), and editor or co-editor of four other books including *The Economics of Immigration: Market-Based Approaches, Social Science, and Public Policy* (Oxford University Press 2015). He is author of more than 75 scholarly articles and policy studies. Prof. Powell's research findings have been reported in hundreds of popular press outlets including *The Wall Street Journal* and *The New York Times*. He also writes frequently for the popular press. His popular writing has appeared in the *Chicago Tribune*, *New York Post*, *The Dallas Morning News* and elsewhere. He has appeared on Fox News Channel, CNN, MSNBC, Showtime, CNBC, and he was a regular guest commentator on Fox Business Network's *Freedom Watch* and *Stossel*. His email address is benjamin.powell@ttu.edu.

Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6024/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 35–51

Alive and Kicking: Mortality of New Orleans Medicare Enrollees After Hurricane Katrina

Robert Kaestner[1]

LINK TO ABSTRACT

In a recent article in the *American Economic Review*, Tatyana Deryugina and David Molitor (2020) analyzed the effect of Hurricane Katrina, which occurred in New Orleans in late August 2005, on mortality of residents of New Orleans enrolled in Medicare. The article is titled, "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina."

I reproduce the abstract in full and apply boldface to parts:

> We follow Medicare cohorts to estimate Hurricane Katrina's long-run mortality effects on victims initially living in New Orleans. Including the initial shock, the hurricane improved eight-year survival by 2.07 percentage points. **Migration to lower-mortality regions explains most of this survival increase**. Those migrating to low-versus high-mortality regions look similar at baseline, but their subsequent mortality is 0.83–1.01 percentage points lower per percentage point reduction in local mortality, quantifying **causal effects** of place on mortality among this population. Migrants' mortality is also lower in destinations with healthier behaviors and higher incomes but is unrelated to local medical spending and quality. (Deryugina and Molitor 2020, 3602, my emphases)

I offer several criticisms of Deryugina and Molitor (2020—hereafter "DM"), but let me start with two illustrations that highlight the intuition of some of my criticisms.

1. University of Chicago, Chicago, IL 60637.

First, consider a time just prior to Hurricane Katrina, say July 2005, and two men, both age 70, one living in New Orleans and one living in Richmond, Virginia, which is one of DM's comparison regions. Assuming a world like ours except that Hurricane Katrina does *not* happen, do these two men have equal chances of surviving to 2013? The answer may well be no: The New Orleans man is someone who has made it to 70 *in New Orleans*, which has a relatively high mortality rate. Making it to 70 there suggests unusually good health compared to someone who made it to age 70 in Richmond, which has a relatively low mortality rate. Thus, it is likely that the 70-year-old in New Orleans will live longer than the 70-year-old in Richmond. If so, then an analysis of the effect of Hurricane Katrina will appear to lower death rates when in fact it is simply an artifact of not comparing people with the same health.

Now consider a second scenario that involves two 70-year-olds in New Orleans in July 2005. After Hurricane Katrina, one moves to Baton Rouge because a city bus was available to take them there, while the other moves to Houston because he had a car and went to see relatives. In 2013 the man who moved to Baton Rouge has passed away, while the man who moved to Houston still lives. It is possible that the difference in their surviving was in part causally related to the conditions and environments of Baton Rouge versus Houston. But it is also possible that the causal difference was a difference in the two men, a difference that figured into the one's moving to Baton Rouge and the other's moving to Houston. If entirely so, then the man who moved to Baton Rouge and died would have died even if he had moved to Houston.

These two illustrations highlight that people differ in ways that scientists do not and often cannot observe and measure. Those differences might relate to things that are observed, however, such as where one has been living, whether one chooses to move, and to where one moves. We are creative beings, each with "a principle of motion of its own" (Smith 1976/1790, 234.17). That principle of motion, that will, moves on the basis of the soul's particularistic opportunities and constraints. Economists should always figure that inscrutable hidden factors lie behind the events that result from human will.

Who did or did not die? Mortality of movers and stayers

DM conducted a difference-in-differences analysis of annual mortality that compared the mortality rate of elderly (roughly 80 percent) and non-elderly, disabled (roughly 20 percent) Medicare enrollees living in New Orleans in 2004 in

years prior to and after Hurricane Katrina to the mortality rate of similar residents in other areas.[2] Based on the results of this analysis, DM conclude that Hurricane Katrina caused a decline in Medicare-enrollee mortality. I shall cast some doubt on this basic claim.

But first, let's grant that Hurricane Katrina caused a decline in Medicare-enrollee mortality. DM say that such a decline came mainly because of a decline in mortality among movers. However, DM never actually analyzed the causal effect of Katrina on the mortality of movers or stayers pre-to-post Katrina. The difference-in-difference approach of DM cannot identify the effect of Hurricane Katrina on those who moved and those who stayed in New Orleans, but only the effect of Hurricane Katrina on the two groups combined. So, the conclusion of DM, "Migration to lower-mortality regions explains most of this survival increase" (p. 3602), is not supported by direct evidence.

The fact that the DM analysis is of *all* residents of New Orleans (and other cities) and not just movers or just stayers is important because it seems reasonable to assume that Hurricane Katrina *increased* subsequent mortality of New Orleans stayers. There was widespread damage to infrastructure, including the destruction of most healthcare facilities (e.g., Memorial Medical Center), and a large share of medical professionals were displaced leaving a severe shortage (Rudowitz et al. 2006). DM allude to possible improvements in the healthcare infrastructure in New Orleans subsequent to Hurricane Katrina (see footnote 21 of paper), but several studies suggest that such was not the case. At best, there was marginal improvement in access provided by clinics, but this did not occur for at least a few years afterward, when Federal assistance was provided (Rittenhouse et al. 2012; Cole et al. 2015; Hamel et al. 2015). The Memorial Medical Center (i.e., Charity Hospital), which had served most low-income persons in New Orleans, was replaced by a new hospital, but that did not open until 2015. Also, there was a large decline in employment (and income) in the short run aftermath of Katrina, although by 2007–08 labor income among those who filed tax returns and could be later observed in tax data recovered to pre-Katrina levels (Deryugina et al. 2018). The DM analysis, however, is focused on elderly and disabled who mainly do not work and who suffered from the destruction and displacement caused by Hurricane Katrina. The trauma, along with a devastated health care system, a destroyed infrastructure and a severely bruised economy, all suggest that the

2. In most analyses, DM used residents in 10 cities that closely match New Orleans with respect to median earnings, population growth, and racial composition as the comparison group. The 10 cities are: Baltimore, MD; Birmingham, AL; Detroit, MI; Gary, IN; Jackson, MS; Memphis, TN; Newark, NJ; Portsmouth, VA; Richmond, VA; and St. Louis, MO. Other comparison areas are used instead of these 10 cities; the estimates using these alternatives are similar. In other analyses, DM used a larger group of areas defined by county or commuting zone.

mortality rate of elderly/disabled stayers likely increased.

It is possible to obtain an estimate of the effect of Hurricane Katrina on the mortality rate of each group with the help of a few assumptions. First, note that prior to Hurricane Katrina, for example in 2004, the mortality rate of the elderly and disabled in New Orleans was algebraically equal to the following:

$$M_{pre} = \alpha M_{pre}^{stayers} + (1 - \alpha) M_{pre}^{movers} \tag{1}$$

Equation (1) simply states that the mortality rate (M) in New Orleans prior to Hurricane Katrina (pre) is equal to the share of New Orleans residents who stayed in New Orleans after Katrina (α) multiplied by their mortality rate at that time (i.e., 2004) plus the share of New Orleans residents who moved out of New Orleans after Katrina ($1-\alpha$) multiplied by their mortality rate at that time. Similarly, for the post-Katrina period (2006 to 2013), the mortality rate of the elderly and disabled in New Orleans was:

$$M_{post} = \alpha M_{post}^{stayers} + (1 - \alpha) M_{post}^{movers} \tag{2}$$

The difference in the mortality rate of New Orleans residents is:

$$M_{post} - M_{pre} = \alpha\left(M_{post}^{stayers} - M_{pre}^{stayers}\right) + (1 - \alpha)\left(M_{post}^{movers} - M_{pre}^{movers}\right) \tag{3}$$

Equation (3) is the difference-in-differences estimate obtained by DM after netting out confounding trends using the comparison areas. From the article, we know that the estimated difference (in-differences) in the mortality rate of New Orleans residents is between −0.5 and −0.2 (see DM, 3619 Table 2).

To identify the effect of Hurricane Katrina on the mortality of movers and stayers, it is necessary to know three values: α, $\left(M_{post}^{stayers} - M_{pre}^{stayers}\right)$, and $\left(M_{post}^{movers} - M_{pre}^{movers}\right)$. The article provides estimates of α, but not the two other quantities. In 2006, approximately 50 percent of elderly and disabled residents of New Orleans had moved, but by 2013 approximately half had returned. Given these figures, I assume that the share of stayers (α) was 0.63 and the share of movers ($1-\alpha$) was 0.37. Next, I calculate different values for $\left(M_{post}^{stayers} - M_{pre}^{stayers}\right)$ conditional on assumed values for $\left(M_{post}^{movers} - M_{pre}^{movers}\right)$. For example, evidence presented by DM suggests that the mortality rate of movers converged to the mortality rate of the place they moved to. The mortality rate of the areas that New Orleans residents moved to, per DM's Online Appendix (**link**), was approximately between 5.0 and

6.0 in 2006 (DM, A-45 Table A.17). Given these figures, a very large, perhaps implausibly large, estimate of $\left(M^{movers}_{post} - M^{movers}_{pre}\right)$ is −1 percentage point. Using this estimate, the assumption that α=0.63, and the estimate of $M_{post} - M_{pre}$ (equation 3) provided by DM of −0.5, I derive an estimate of $\left(M^{stayers}_{post} - M^{stayers}_{pre}\right)$ that is equal to −0.21. This calculation suggests that the mortality rate of stayers *decreased* by 0.21 percentage points.

TABLE 1. Estimates of the effect of Hurricane Katrina on mortality of movers and stayers

Effect on movers (assumed)	Total effect (estimated by DM)	Effect on stayers (derived)
$M^{movers}_{post} - M^{movers}_{pre}$	$M_{post} - M_{pre}$	$M^{stayers}_{post} - M^{stayers}_{pre}$
−1	−0.5	−0.21
−0.8	−0.5	−0.32
−0.6	−0.5	−0.44
−0.4	−0.5	−0.56
−0.2	−0.5	−0.68
0	−0.5	−0.79
−1	−0.35	0.03
−0.8	−0.35	−0.09
−0.6	−0.35	−0.20
−0.4	−0.35	−0.32
−0.2	−0.35	−0.44
0	−0.35	−0.56
−1	−0.2	0.27
−0.8	−0.2	0.15
−0.6	−0.2	0.03
−0.4	−0.2	−0.08
−0.2	−0.2	−0.20
0	−0.2	−0.32

Table 1 provides estimates of $\left(M^{stayers}_{post} - M^{stayers}_{pre}\right)$ for different values of $\left(M^{movers}_{post} - M^{movers}_{pre}\right)$ and for three estimates of $M_{post} - M_{pre}$ provided by DM that bracket other estimates. Under the assumption that Hurricane Katrina decreased mortality of movers ($M^{movers}_{post} - M^{movers}_{pre}$), as DM conclude, estimates in Table 1 strongly suggest that the mortality rate of stayers also declined in New Orleans after Hurricane Katrina, and possibly by more than the decline in mortality among movers. However, all evidence of the devastation and disruption caused by Hurricane Katrina suggest that the mortality of stayers should, if anything,

increase.[3]

What does the evidence in Table 1 suggest about the effect of place on mortality? Not much. Whether you stayed in New Orleans or moved to another place, mortality seemed to decline. The surprising finding is that mortality declined at all for either group. It is surprising given the trauma (e.g., Laditka et al. 2010; Rhodes et al. 2010; Fussell and Lowe 2014; Calvo et al. 2015), widespread destruction of the healthcare infrastructure (Rittenhouse et al. 2012; Cole et al. 2015; Hamel et al. 2015), and destruction of the physical infrastructure and displacement that took place in varying degrees for both movers and stayers.

And there is other evidence casting doubt on the claim that the decline in mortality after Hurricane Katrina was due to moving. Estimates show that movers were much more likely to be black (22 percentage points) and under-age-65 (12 percentage points) than non-movers (DM, A-35 Table A.7). If moving was the cause of the decline in mortality, then we would expect the mortality rates of black and under-age-65 residents of New Orleans to go down more than that of other groups. Such is not the case. While the mortality rate of all black residents declined more than other race/ethnic groups after Katrina, the larger decline was not statistically significant. The mortality rate of the under-age-65 residents of New Orleans was actually higher after Katrina than older groups, although this estimate was not significant. In general, there is little systematic relationship between the proportion of a demographic group that moved and the effect of Hurricane Katrina on mortality. This is further evidence suggesting that moving was not the primary cause of the decline in mortality reported.[4]

Econometric problems

An important statistical issue that noticeably affects DM's results has to do with the way the regression model was specified. The issue is whether the model allows mortality rates to differ granularly by *age-by-year* as opposed to more coarsely by *age and year*, with the year effect being *the same* for all ages that year. The subtle distinction is important because the sample consists of a cohort of persons of different ages followed over time—meaning, as they age. Older members of the

3. DM show that mortality of stayers declined in the few years after Hurricane Katrina, although estimates were not statistically significant (DM, A-27 Figure A.13). However, this evidence comes from a problematic analysis that selects on moving status, which is endogenous. Stayers in New Orleans are unlikely to be comparable to stayers in other areas and DM provide no evidence to justify such an analysis.
4. The effect of Hurricane Katrina on mortality depends on exposure (moving) and the effect of exposure, which may differ by demographic group. Thus, there does not have to be a correlation between the extent of exposure and the effect of Hurricane Katrina.

cohort will die sooner than younger members and therefore the analysis should allow this to be the case. In most analyses, however, DM do not allow mortality rates to differ by age-by-year. When they do, results differ markedly.

Model specification

The standard demographic model used to estimate time to death and applied to the DM context is:

$$DEATH_{jat} = \alpha_{at} + \sum_{t=2005}^{2013} \gamma_t \left(NO_j{*}YEAR_t\right) + \delta_j + e_{jat} \tag{4}$$
$$j = 1, \ldots, J \quad \text{(ZIP codes)}$$
$$a = 30, \ldots, 85 \quad \text{(age)}$$
$$t = 2004, \ldots, 2013$$

Equation (4) specifies that the probability of person of age a in 2004 in ZIP code j in year t dies (*DEATH*), conditional on being alive at the beginning of year t, depends on age-by-year fixed effects (α_{at}), the interaction between a New Orleans dummy variable (*NO*) and year dummy variables, and ZIP code fixed effects.[5] An important feature of equation (4) is that there are separate year effects for persons of each age (here I use ages 30 to 85 because the youngest and oldest ages in the sample are not provided). The need to include these fixed effects stems from the fact that the probability that a person age 40 in 2004 is alive in any year after 2004 is very different than the probability a person 70 in 2004 is alive in any year thereafter. The older person has a much higher probability of dying in any year thereafter. The inclusion of age-by-year fixed effects is clearly the appropriate model, and it is the workhorse model of demographic studies of mortality (see, e.g., Rodríguez 2007).

DM do not estimate this standard model for most analyses, but only in sensitivity analyses. Instead, DM estimate:

$$DEATH_{jat} = \alpha_a + \sum_{t=2005}^{2013} \gamma_t \left(NO{*}YEAR_t\right) + \delta_j + e_{jat} \tag{5}$$

The key difference between equations (4) and (5) is that equation (5) restricts the year effects to be the same for each age.

As shown in DM, the addition of age-race-sex-year fixed effects to the regression model results in substantially smaller estimates of the effect of Hurricane Katrina on mortality (DM, A-21 Figure A.7). For the 1992 and 1999

5. There is no justification for including ZIP code fixed effects. While they likely reflect differences in individuals and environment, they are arguably unnecessary to include given the inclusion of the New Orleans dummy variable.

samples, the addition of such controls yields estimates that are not statistically significant in the period from 2009 (2010) to 2013. Thus, in these samples, any evidence of a decline in mortality post-Katrina may be due to what demographers and epidemiologists refer to as harvesting—the trauma of Hurricane Katrina speeded up the death of (i.e., harvested) some frail people, and so the spike in deaths after Katrina is naturally (biologically) followed by a temporary decline in mortality, which is then followed by a return to baseline. This is exactly what is shown in DM's plots (ibid.). Notably, DM do not calculate the cumulative mortality effect for these samples, which would be much smaller than that derived from estimates from models that exclude age-race-sex-year fixed effects.

Perhaps more worrisome is that estimates from the standard and more appropriate demographic model reveal evidence of differential pre-trends—a sign of an invalid research design—that are non-trivial in magnitude when compared to the post-Katrina estimates.

As noted, estimates reported by DM from equation (4) are substantially smaller and often not statistically significant relative to estimates from equation (5). Why would that be the case? Consider an extreme case in which all persons in New Orleans are age 40 and all persons in comparison areas are age 60. For any year post-Katrina the probability of dying would be lower for the younger residents of New Orleans than the older residents of other places even conditioning on the baseline age difference. This is consistent with the different estimates from equations (4) and (5) reported by DM and the fact that residents in New Orleans are younger than residents in other areas. Differences in race and gender may have similar confounding influences. The upshot is that when the correct model is used, estimates of the effect of Hurricane Katrina on mortality are much smaller (half the size or less) and less statistically significant. Such findings suggest the presence of the harvesting effect.

Selective mortality and unmeasured heterogeneity

A notable empirical fact reported by DM is that *all* residents in New Orleans had higher rates of mortality than most of the comparison areas (DM, 3610 Figure 1, A-19 Figure A.5). Here I return to the first of my initial illustrations. Consider two individuals who are both age 70 (approximately the mean age of sample) in 2004, but one lives in New Orleans and the other lives in Richmond. The higher rate of mortality in New Orleans suggests that the 70-year-old in New Orleans is healthier than the 70-year-old in Richmond because of selective mortality—to make it to age 70 in New Orleans with its relative mortality rate a person has to be healthier than a 70-year-old in the lower-mortality area.

Evidence that selective mortality is an issue is found in DM's figure that

shows diverging pre-trends between the mortality rate in New Orleans and other areas, with New Orleans having, as people age, a consistently larger decline in mortality than the comparison areas prior to Hurricane Katrina (DM, A-21 Figure A.7). The magnitudes of the deviations between New Orleans mortality and mortality in other areas prior to Katrina are about half as large as the post-Katrina deviations. Again, we find substantial evidence to question the basic claim that Hurricane Katrina reduced mortality.

Inference

DM's approach to inference is questionable. Their approach is the so-called cluster-robust method. DM construct standard errors allowing for non-independence of observations within baseline ZIP code without providing any justification for their choice. Moreover, it is not a statistically sound choice, and there are available alternatives. For analyses that use cohorts from 1992 or 1999 and the 10-city comparison group, the approach of Stephen Donald and Kevin Lang (2007) would be feasible and preferable. That approach was not used, however, and DM do not explain why it was not used. For analyses that use counties as comparisons for New Orleans (N=152), or Commuting Zones (N=400), instead of the 10 comparison cities, the method of Timothy Conley and Christopher Taber (2011), or randomization inference (e.g., Good 2005) could be used, but was not.

It is well-known that standard errors can be substantially different. To see the potential problem, consider DM's main analyses using the 2004 cohort and residents of New Orleans and 10 comparison cities. One view of these data is that there are only six 'observations'—three observations for New Orleans for 2004, 2005 and 2006–2013 and three observations for the comparison cities in 2004, 2005, 2006–2013. But the relevant table in DM reports sample sizes of roughly 8 million and 10 million observations, which are not very relevant, and notes that standard errors are clustered by ZIP codes with no justification (DM, 3619 Table 2). One may argue that inference should be conducted assuming six observations, which is to say that no inference can be made (Donald and Lang 2007; Cameron and Miller 2015).

Instead of applying the just-noted two feasible approaches to inference to the difference-in-differences analyses that produced the estimates underlying the article's conclusions, DM chose to alter the research design, apparently just for the purpose of conducting inference. It is a curious choice. Specifically, DM use a synthetic control (SC) approach instead of a difference-in-difference approach to estimate the effect of Hurricane Katrina on mortality. While the inference from the SC approach suggests that Hurricane Katrina had a significant, negative effect on mortality, the magnitudes and pattern of estimates differs markedly from those

obtained from the difference-in-differences approach. Estimates are much larger (nearly twice as large), and estimates do not become smaller over time as do the difference-in-differences estimates (from the best practice specification).

Applying the conclusion about statistical significance based on the inference approach from this supplemental analysis, which did not undergo any sensitivity analysis (e.g., in the method to select weights) and that is only cursorily described, to estimates obtained from the difference-in-differences analysis seems a bridge too far. The authors offer no justification for applying the SC-derived conclusion to the difference-in-difference analysis. More importantly, there is no reason that randomization (permutation) inference (or the approach of Conley and Taber 2011) could not have been conducted in the context of a difference-in-differences analysis—there was no need to move to a different research design to conduct such inference (e.g., Kaestner 2016). Finally, no SC analysis was conducted for the 2004 cohort because it is infeasible, but the alternative approaches to inference could have been used for this cohort. In sum, the lack of a valid approach to inference means that none of the primary estimates reported by DM can be assessed from a statistical point of view. They do not provide reliable evidence of the effect of Hurricane Katrina on mortality.

No theory and no hypotheses? No problem—we have a research design

The DM article is almost entirely an empirical exercise with virtually no attempt to use theory to generate testable hypotheses. The only theoretical statement I could find is:

> The disruption induced by extreme weather events can be used to illuminate factors that affect the accumulation or depreciation of health capital (Grossman 1972). (DM, 3602–3603)

Without any theorizing about how purposive human beings would bring about putatively observed results, the plausibility of estimates depends solely on the credibility of the research design. The interpretation of estimates becomes ad hoc.

DM refer to the finding that Hurricane Katrina reduced mortality of New Orleans residents as counterintuitive. The finding is more than counterintuitive. It is inconsistent with the canonical model of the demand for health, that of Michael Grossman (1972), which would predict an increase in mortality for residents that remained in New Orleans because of:

- worsened health after Katrina because of increased stress and adverse environmental factors such as mold (Laditka et al. 2010; Rhodes et al. 2010; Fussell and Lowe 2014; Calvo et al. 2015);
- a decrease in the productivity of investment in health and degradation of the health care system (Rittenhouse et al. 2012; Cole et al. 2015; Hamel et al. 2015);
- higher costs to investing in health, for example, because of less access and availability to medical services and higher prices of goods and services (e.g., rents rose significantly[6]); and
- a decline in income because of loss of employment and macro-economic effects.[7]

As shown in Table 1, estimates and other information found in DM suggest that mortality declined among residents who remained in New Orleans. How should we interpret this? Is it a surprising, novel finding that should make us rethink theory? Or should we question the credibility of the estimate? The atheoretical analysis and legitimate questions about the validity and reliability (inference) of the estimates in DM leads me to conclude that we should doubt the estimate.

Theory also can be used to assess whether the mortality of those who left New Orleans is likely to decrease. The Grossman (1972) model suggests the mortality of movers will be affected by worsened health, the access and price of health care, the quality of health care, and income.

Consider worsened health. It is unlikely that the age pattern of onset of illness changes much with destination location, although pollution and other environmental factors may be present. Interestingly, DM show that moving to a low-pollution area is associated with an *increase* in mortality (DM, 3626 Figure 6). DM also report that movers are differentially selected with respect to destination pollution—healthier movers are significantly more likely to go to more polluted places. Movers are also significantly selected on upward income mobility and median house value. Several other characteristics (social capital, crime rate, income segregation, and urban population) show non-trivial though only marginally significant correlations between mover characteristics and destination characteristics. Of course, economic theory suggests that a mover's destination is not

6. *See* House Financial Services Committee's Democratic Staff (2015).

7. The analysis of the effect of Hurricane Katrina on labor income of New Orleans residents in Deryugina et al. (2018) shows that over the period from 2005–13 labor income was statistically no different from the 2004 level. The method of inference used in this study suffers from the same problems as described earlier. Labor income is not as important for elderly and disabled because many do not work. Net income likely declined for these demographic groups because of increased expenditures.

random even if the decision to move were random.

To support the argument that moving is random (and to overturn standard economic theory), DM rely on the fact that mover's predicted mortality—based on race (black), sex, age (<64, 75+) and several chronic condition indicators—is not strongly correlated with destination mortality. However, these characteristics likely explain a small portion of mover mortality (not reported by DM) and this analysis is not very strong evidence against the possible bias due to the likely non-random sorting of movers by destination, particularly because there is substantial evidence that mover's destination was non-random. Others have provided similar evidence. In work I coauthored with Kevin Callison and Jason Ward, for example, we show that among elderly Medicare enrollees, sicker movers are more likely than healthier movers to select high-spending Medicare destinations (Callison et al. 2021).

The findings with respect to pollution are of particular importance because it is one of the few destination characteristics with a clear causal link to mortality. Most other destination characteristics have no such link. Consider the health behaviors of movers now residing in new destinations. Estimates in Figure 6 of DM show that moving to a high-smoking area, or a high-obesity area, increases mortality, and moving to a high-exercise area decreases mortality (DM, 3626). The causal mechanisms underlying these associations is unclear. Few people actually quit smoking. DM provide no evidence that the level of smoking in an area causes someone (e.g., a mover) to smoke less. Indeed, the trauma and stress associated with Katrina might lead one to expect an increase in smoking among movers. Even if there were differences in the cigarette prices and tobacco-control policies in an area, which DM could have measured but did not, evidence suggests that these policies would have very minor effects on elderly smoking (e.g., Callison and Kaestner 2014). Similarly, why would area obesity level cause a mover's mortality to increase? Does a 70-year-old suddenly lose weight because there is less obesity in the area she moved to? That is doubtful, and DM provide no evidence of such behavioral channel.[8] Does an elderly or disabled person start jogging because they see some people jogging near their house? Again, this is doubtful. The lack of strong causal links between the characteristics that were considered and mortality is obvious and reveal the lack of a theoretical basis of the mover analysis and the ad-hoc, empirical approach taken. Imbuing estimates of these correlations with causal meaning is inappropriate and likely misleading.

Consider the access and price of medical care. All persons in the sample are

8. Datar and Nicosia (2018) examine associations between the county obesity rate and weight status of enlisted (military) adults (mean age 37) who are arguably randomly assigned to areas. Associations are positive, but small and usually not statistically significant. Similarly, Christakis and Fowler (2007; 2008) report that the obesity and smoking of neighbors is not related to either of these health behaviors.

covered by Medicare, and a much higher proportion of New Orleans residents are covered by Medicaid as dual eligible and therefore have more generous benefits. This fact implies that there are virtually no price differences or access differences between New Orleans and destination areas that would significantly affect mover's mortality. Does the quality of care differ markedly between New Orleans and destination areas? Perhaps, but this is unmeasured. Here too, estimates in DM's Figure 6 suggest the answer is no, as measures of the health care infrastructure (not necessarily quality) in destination places is unrelated to mover mortality (DM, 3626).

Finally, consider income. Yet again, estimates in DM's Figure 6 show that moving to an area with a higher per-capita income decreases mortality, but moving to an area with high rates of elderly poverty has no effect (DM, 3626). But most movers were not working, so, here too, the causal mechanisms between area income and mover mortality are lacking.

Overall, there is little evidence consistent with theory to support the conclusion that moving reduced mortality. So, why do DM conclude that moving reduced mortality? The main reason underlying the claim is the estimate of the association between destination mortality and the probability of dying. But why should we trust this estimate in light of the documented non-random selection of movers to destinations, the inconsistent evidence (pollution and social capital) and the fact that there is no theory underlying this particular association. Why should destination mortality affect mover mortality? What is the causal mechanism? None is provided, and when theoretical causes of mortality are assessed, the evidence is weak or inconsistent with theory. In short, the empirical analysis conducted by DM on movers is at best exploratory, and at worst an atheoretical, data-mining exercise.

Conclusion

The surprising finding reported in DM that a hurricane as devastating as Katrina reduced the mortality of residents of New Orleans merits close scrutiny because it is inconsistent with intuition, theory, and prior evidence (e.g., Laditka et al. 2010; Rhodes et al. 2010; Fussell and Lowe 2014; Calvo et al. 2015). I have provided a thorough review and critical assessment of the evidence provided by DM to support their conclusions. I find the evidence wanting.

First, I show that based on the estimates reported in DM and reasonable assumptions, DM's evidence would suggest that the mortality rate of residents of New Orleans *who remained in New Orleans* decreased after Katrina. While possible, there is no external evidence or theory to support this conclusion. Moreover, when the appropriate regression model is used (equation 4), estimates of the effect of

Hurricane Katrina on mortality are greatly reduced in magnitude and there is evidence of diverging pre-trends that raise questions of the validity of the difference-in-differences research design. Then there is the serious problem related to how DM approached inference. They simply never provided a valid approach to inference despite the availability of such approaches widely utilized in the literature.

Second, an assessment of the mover analysis and the effects of destination characteristics on mover mortality reveals non-trivial selection and several inconsistencies. Moving to a place with lower pollution is associated with an increase in mortality. Note that pollution is the only direct environmental influence assessed by DM. This is important because there are, at best, distant causal links between area characteristic such as obesity, smoking, exercise and income and a mover's mortality. Thus, the conclusion by DM that moving reduced mortality, which is something never actually assessed directly, is unjustified.

I will end by questioning what we expect to learn from an atheoretical, empirical exercise of a one-off natural disaster. There will never be another Hurricane Katrina in New Orleans. New Orleans has already changed physically and institutionally, and the composition of the city's population has changed. Thus, even if another storm of the same magnitude rolled through New Orleans, it is unlikely that findings from this study would apply. The external validity of the findings, even assuming that the findings are valid, for other cities and other storms seems extremely limited. Cities differ in their infrastructure, institutions, and populations (see DM, 3609 Table 1, which shows substantial differences even among purposely matched cities). These and many other factors determine the effects of a natural disaster, such as a hurricane. As Nancy Cartwright (2013) argues, in most cases evidence from one experiment, in this case a natural experiment, do not travel well and are unlikely to be replicated in another (natural) experiment. The near absence of any plausible external validity and the absence of any tests of theoretically derived hypotheses make studies like DM's marginal in their scientific importance.

DM ask in their title: Does when you die depend on where you live? Surely it does in some cases, and we do not need a study of hurricane-related trauma and displacement from New Orleans to know that. It is well-documented that there are environmental causes of mortality, such as pollution, unsafe drinking water, poor sanitation, disease (e.g., malaria), and crime. It is well-known that environments differ geographically. Similarly, there are longstanding geographic differences in social and economic opportunities that vary by geography (e.g., Appalachia) that are a cause of low incomes, low education, and poor health behaviors. The geographical landscape of disease and mortality in the U.S. has been thoroughly assessed by the Centers for Disease Control since at least 1975 (Mason et al. 1975; Mason et al. 1976; Mason et al. 1981; Pickle et al. 1987; Pickle et al. 1990; Pickle

et al. 1996). So the interesting question is not whether where you live affects when you die, but how to affect the causes of mortality and how to facilitate mobility to diminish the contribution of geography-related mortality. The DM study is silent on these questions.

References

Callison, Kevin, and Robert Kaestner. 2014. Do Higher Tobacco Taxes Reduce Adult Smoking? New Evidence of the Effect of Recent Cigarette Tax Increases on Adult Smoking. *Economic Inquiry* 52(1): 155–172.

Callison, Kevin, Robert Kaestner, and Jason Ward. 2021. Do Supply-Side Forces Explain Geographic Variation in Health Care Use? *Economic Inquiry* 59(1): 119–139.

Calvo, Rocío, Mariana Arcaya, Christopher F. Baum, Sarah R. Lowe, and Mary C. Waters. 2015. Happily Ever After? Pre-and-Post Disaster Determinants of Happiness Among Survivors of Hurricane Katrina. *Journal of Happiness Studies* 16(2): 427–442.

Cameron, Colin, and Doug Miller. 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources* 50(2): 317–372.

Cartwright, Nancy. 2013. Knowing What We Are Talking About: Why Evidence Doesn't Always Travel. *Evidence & Policy: A Journal of Research, Debate and Practice* 9(1): 97–112.

Christakis, Nicholas A., and James H. Fowler. 2007. The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* 357: 370–379.

Christakis, Nicholas A., and James H. Fowler. 2008. The Collective Dynamics of Smoking in a Large Social Network. *New England Journal of Medicine* 358: 2249–2258.

Cole, Evan S., Claudia Campbell, Mark L. Diana, Larry Webber, and Richard Culbertson. 2015. Patient-Centered Medical Homes in Louisiana Had Minimal Impact on Medicaid Population's Use of Acute Care and Costs. *Health Affairs* 34(1): 87–94.

Conley, Timothy G., and Christopher R. Taber. 2011. Inference with 'Difference in Differences' with a Small Number of Policy Changes. *Review of Economics and Statistics* 93(1): 113–125.

Datar, Ashlesha, and Nancy Nicosia. 2018. Assessing Social Contagion in Body Mass Index, Overweight, and Obesity Using a Natural Experiment. *JAMA Pediatrics* 172(3): 239–246.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns. *American Economic Journal: Applied Economics* 10(2): 202–233.

Deryugina, Tatyana, and David Molitor. 2020 (DM). Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina. *American Economic Review* 110(11): 3602–3633.

Donald, Stephen G., and Kevin Lang. 2007. Inference with Difference-in-Differences and Other Panel Data. *Review of Economics and Statistics* 89(2): 221–233.

Fussell, Elizabeth, and Sarah R. Lowe. 2014. The Impact of Housing Displacement on the Mental Health of Low-Income Parents After Hurricane Katrina. *Social Science & Medicine* 113: 137–144.

Good, Phillip. 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 3rd ed. New York: Springer.

Grossman, Michael. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 80(2): 223–255.

Hamel, Liz, Jamie Firth, and Mollyann Brodie. 2015. New Orleans Ten Years After the Storm: The Kaiser Family Foundation Katrina Survey Project. Henry J. Kaiser Family Foundation (Menlo Park, Calif.). **Link**

House Financial Services Committee's Democratic Staff. 2015. Ten Years After Hurricane Katrina: Status of Housing in New Orleans. U.S. House Financial Services Committee (Washington, D.C.). **Link**

Kaestner, Robert. 2016. Did Massachusetts Health Care Reform Lower Mortality? No According to Randomization Inference. *Statistics and Public Policy* 3(1): 1–6.

Laditka, Sarah B., Louise M. Murray, and James N. Laditka. 2010. In the Eye of the Storm: Resilience and Vulnerability Among African American Women in the Wake of Hurricane Katrina. *Health Care for Women International* 31(11): 1013–1027.

Mason, Thomas J., Joseph F. Fraumeni Jr., Robert Hoover, and William J. Blot. 1981. *An Atlas of Mortality from Selected Diseases*. Department of Health and Human Services Publication (NIH) 81-2397. Washington, D.C.: U.S. Government Printing Office.

Mason, Thomas J., Frank W. McKay, Robert Hoover, William J. Blot, and Joseph F. Fraumeni Jr. 1975. *Atlas of Cancer Mortality for U.S. Counties: 1950–1969*. Department of Health, Education, and Welfare Publication (NIH) 75-780. Washington, D.C.: U.S. Government Printing Office.

Mason, Thomas J., Frank W. McKay, Robert Hoover, William J. Blot, and Joseph F. Fraumeni Jr. 1976. *Atlas of Cancer Mortality Among U.S. Nonwhites: 1950–1969*. Department of Health, Education, and Welfare Publication (NIH) 76-1204. Washington, D.C.: U.S. Government Printing Office.

Pickle, Linda Williams, Thomas J. Mason, Neil Howard, Robert Hoover, and Joseph F. Fraumeni Jr. 1987. *Atlas of U.S. Cancer Mortality Among Whites: 1950–1980*. Department of Health and Human Services Publication (NIH) 87-2900. Washington, D.C.: U.S. Government Printing Office.

Pickle, Linda Williams, Thomas J. Mason, Neil Howard, Robert Hoover, and Joseph F. Fraumeni Jr. 1990. *Atlas of U.S. Cancer Mortality Among Nonwhites: 1950–1980*. Department of Health and Human Services Publication (NIH) 90-1582. Washington, D.C.: U.S. Government Printing Office.

Pickle, Linda Williams, Michael Mungiole, Gretchen K. Jones, and Andrew A. White. 1996. *Atlas of United States Mortality*. Department of Health and Human Services Publication (PHS) 97-1015. Hyattsville, Md.: National Center for Health Statistics.

Rhodes, Jean, Christian Chan, Christina Paxson, Cecilia Elena Rouse, Mary Waters, and Elizabeth Fussell. 2010. The Impact of Hurricane Katrina on the Mental and Physical Health of Low-Income Parents in New Orleans. *American Journal of*

Orthopsychiatry 80(2): 237–247.

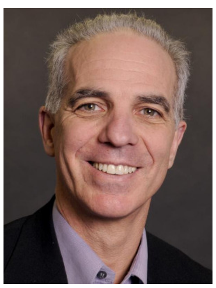
Rittenhouse, Diane R., Laura A. Schmidt, Kevin J. Wu, and James Wiley. 2012. The Post-Katrina Conversion of Clinics in New Orleans to Medical Homes Shows Change Is Possible, But Hard to Sustain. *Health Affairs* 31(8): 1729–1738.

Rodríguez, Germán. 2007. Discrete Time Models. In *Lecture Notes on Generalized Linear Models*, section 7.5. Princeton University (Princeton, N.J.). **Link**

Rudowitz, Robin, Diane Rowland, and Adele Shartzer. 2006. Health Care in New Orleans Before and After Hurricane Katrina. *Health Affairs* 25(Supp. 1): W393–W406.

Smith, Adam. 1976 [1790]. *The Theory of Moral Sentiments*, eds. D. D. Raphael and A. L. Macfie. Oxford: Oxford University Press.

About the Author



Robert Kaestner is a Research Professor at the Harris School of Public Policy of the University of Chicago. He is also a Research Associate of the National Bureau of Economic Research, an Affiliated Scholar of the Urban Institute and a Senior Fellow of the Schaeffer Center for Health Policy of USC. Prior to joining Harris, Kaestner was on the faculty of the University of Illinois, University of Illinois at Chicago, University of California, Riverside, the CUNY Graduate Center and Baruch College (CUNY). He received his Ph.D. in Economics from the City University of New York. He received his BA and MA from Binghamton University (SUNY). His research interests include health, demography, labor, and social policy evaluation. He has published over 125 articles in academic journals. Recent studies have been awarded Article of the Year by AcademyHealth in 2011 and the 2012 Frank R. Breul Memorial Prize for the best publication in *Social Services Review*. Dr. Kaestner has also been the Principal Investigator on several NIH grants focused on Medicare and Medicaid policy. Kaestner is an Associate Editor of the *Journal of Health Economics* and the *American Journal of Health Economics*, and on the Editorial Board of *Demography* and *Journal of Policy Analysis & Management*. His email is kaestner.robert@gmail.com.

Deryugina and Molitor's reply to this article
Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6025/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 52–60

Reply to "Alive and Kicking: Mortality of New Orleans Medicare Enrollees After Hurricane Katrina"

Tatyana Deryugina[1] and David Molitor[2]

LINK TO ABSTRACT

In the paper "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina," published in the November 2020 issue of the *American Economic Review* (Deryugina and Molitor 2020—hereafter "DM"), we used administrative Medicare data to estimate the short- and long-run mortality effects of Hurricane Katrina on elderly and long-term disabled victims who were initially living in New Orleans. We found that despite a substantial mortality increase in the year of the hurricane, 2005, the cumulative probability of dying in the longer run was *lower* among Hurricane Katrina victims than among several comparison groups. This surprising result is apparent in plots of raw mortality rates, difference-in-difference analyses using a variety of comparison groups, estimates from the synthetic control method, and survival model analyses.

To explain why the mortality of Hurricane Katrina victims decreased in the long run, we compared the mortality of victims who moved to higher- versus lower-mortality destination regions. We showed that these movers' ex ante predicted mortality was unrelated to the local mortality rate in the destination region, but their realized mortality was highly correlated with the destination mortality rate, demonstrating that place has a causal effect on life expectancy. Using a back-of-the-

1. University of Illinois at Urbana-Champaign, Champaign, IL 61820.
2. University of Illinois at Urbana-Champaign, Champaign, IL 61820.

envelope calculation, we showed that the estimated place effect combined with the average decline in local mortality among the victims can account for most of the mortality decline among the Hurricane Katrina victims.

Robert Kaestner (2021—hereafter "RK") questions both our finding of a mortality decrease among Hurricane Katrina victims and our analysis of mortality among survivors who left New Orleans in the aftermath of the hurricane. While many of the concerns he raises are ex ante valid, most of them have testable implications that were already addressed in the published paper. Below, we respond to RK's concerns, restating and elaborating on our original findings. Additionally, we show that the back-of-the-envelope exercise offered by RK to show that our estimates imply a decrease in mortality among New Orleans stayers is incorrect and does not fit the context.

Non-parallel trends

The first issue raised by RK is that because mortality in New Orleans was higher than other places even prior to Hurricane Katrina, residents of New Orleans in 2004—our primary sample of treated individuals—are those who survived relatively harsh initial conditions and therefore might be positively selected on health relative to control individuals with similar demographics. If so, the treated group would have experienced lower mortality rates even absent the hurricane, relative to control groups who faced more favorable initial conditions. In this case, a difference-in-differences estimate would produce what appears to be a mortality reduction due to the hurricane, even if there were no effect whatsoever. Of course, selection and estimator bias could just as well go the other way. For example, individuals who survived the harsh conditions in New Orleans prior to Katrina may have suffered health scarring, such as through earlier onset and progression of chronic conditions, leaving them *negatively* selected on health relative to others of similar demographics.

We completely agree that correctly estimating the mortality effects of Hurricane Katrina hinges on selecting a comparison group whose mortality rates parallel the counterfactual mortality among New Orleans victims. That is why our AER paper does much to address the validity of various control groups and the plausibility of the parallel-trends assumption. Our preferred specification compares mortality rates among cohorts from New Orleans versus cities that had similar baseline demographic and economic conditions and, it turns out, also had above-average mortality rates pre-Katrina. Concerns about differential mortality trends also motivate our use of pre-2004 cohorts in some specifications: if differential pre-hurricane mortality were a concern, we would expect New Orleans

mortality rates to diverge from the control group even prior to the hurricane. Yet we see no evidence of this, even in the 1992 cohort, which we observe for over a decade prior to the hurricane (DM, 3619 Figure 4).

We also, in the Online Appendix to DM (**link**), report results where the control group consists of beneficiaries from the entire United States (DM, A-23 Figure A.9) or is constructed from other high-mortality regions, which would arguably be selected similarly to the New Orleans cohort (DM, A-19 Figure A.5). We see mortality rates that are nearly indistinguishable from New Orleans in both levels and trends prior to the hurricane but diverge following the hurricane, implying mortality decreases that are very similar to what we estimate using our preferred control group. Overall, the evidence in the paper suggests that the type of positive selection posited by RK (or another type of selection that would cause the parallel trends assumption to be violated) is not present in our setting.

Non-random moves

The second issue raised by RK is that movers did not choose where to locate at random and that we "overturn standard economic theory" to "support the argument that moving is random" (RK, 46). In fact, our paper does not claim that those leaving New Orleans chose their destination at random. We explicitly acknowledge that "little systematic information is available on how victims chose where to relocate in the longer run" (DM, 3607). In Section IV.C (DM, 3620–3629), we also explicitly acknowledge the presence of some sorting along dimensions other than the local mortality rate (e.g., local pollution and social capital levels).

Moving to destinations at random is a stronger condition than needed for estimating causal effects of place on a specific outcome, like mortality. As we state in the introduction, "The relationship between local and migrant mortality describes the causal effect of place on individual mortality under the assumption that baseline mortality risk among those who move is uncorrelated with mortality rates in the destination region" (DM, 3604). In support of this assumption, we find that movers' observable risk factors, including demographics, medical spending, and chronic conditions, and predicted mortality based on these factors are largely uncorrelated with mortality rates in the destination region (DM, 3621–3622). Of course, this assumption could still be violated if there are unobservable mortality risk factors that are uncorrelated with the observed factors but are correlated with destination mortality. The scope for such a violation is higher when observables explain a small share of risk. Related to this point, RK speculates that "these characteristics likely explain a small portion of mover mortality (not reported by

DM)" (RK, 46). While reporting the R-squared for a binary outcome like mortality would not be particularly informative, DM (A-28 Figure A.14) does show that the rich set of risk factors we include in our prediction model yields out-of-sample "mortality predictions that are strongly correlated with realized mortality among the New Orleans movers" (DM, 3621).

Movers versus stayers

RK also takes issue with the fact that we never compare the mortality of New Orleans movers to the mortality of New Orleans stayers. The reason we do not do this is because the New Orleans movers are quite different on most of the observable characteristics from both the average 2004 New Orleans Medicare beneficiary (DM, 3609 Table 1) and from survivors of the hurricane who remain in New Orleans (DM, A-35 Table A.7). Thus, even with extensive controls, any direct comparison of movers and stayers is likely to be highly susceptible to the influence of unobservable differences between them.

Instead, our estimate of the share of the net mortality decline that can be explained by moving is based on a calculation that combines the average change in the local mortality among movers with the estimated effect of a change in the local mortality rate on one's own mortality. This is what we mean when we say, in the article abstract, "Migration to lower-mortality regions explains most of this survival increase" (DM, 3602). As we explicitly acknowledge and discuss in the paper, there may be other factors affecting mortality in our sample, among movers, among stayers, or both (DM, 3628).

In an attempt to infer mortality effects among stayers from our estimates, RK presents a back-of-the-envelope calculation and claims that it implies Hurricane Katrina caused a mortality decline among stayers. The calculation begins with equation (1):

$$M_{pre} = \alpha M_{pre}^{stayers} + (1 - \alpha) M_{pre}^{movers},$$

where M_{pre} is the pre-hurricane mortality rate of the New Orleans cohort, α is the share of stayers, and $1 - \alpha$ is the share of movers. The variables $M_{pre}^{stayers}$ and M_{pre}^{movers} represent the pre-hurricane mortality of stayers and movers, respectively.

This equation is ill-defined and unhelpful for evaluating our results. First, the pre-Katrina New Orleans cohort is not composed only of would-be movers and stayers—categories that by their nature require surviving the hurricane—but also includes individuals who died prior to the beginning of 2006 (the point at which we

classify victims as movers or stayers, as described on pages 3610 and 3614 of DM). Second, because it is impossible to classify someone who has died prior to or in the immediate aftermath of the hurricane as a mover or a stayer, $M_{pre}^{stayers}$ and M_{pre}^{movers} are both mechanically zero. These issues invalidate the back-of-the-envelope exercise for its intended purpose. In addition, RK's definition of movers (and thus his calculated α) does not match the definition used in our paper (DM, 3614).

Age-by-year fixed effects

To estimate the long-run mortality effects of Hurricane Katrina on its victims, in the paper we measure mortality outcomes of cohorts from New Orleans and compare these to mortality outcomes of control cohorts. Both the treatment and control cohorts age together over time. Because our primary estimating strategy does not include age-by-year fixed effects, RK argues that this approach will "generally fail to account for the fact that people of different ages, races, or sexes will have different probabilities of dying as time goes by" (RK, **abs.**). He states, "The inclusion of age-by-year fixed effects is clearly the appropriate model" (RK, 41). Yet such a model is not clearly superior to our primary specification, nor does the evidence support using age-by-year fixed effects to estimate the mortality effects of Hurricane Katrina.

Age-by-year fixed effects essentially match individuals from the treatment group to individuals of the same age in the comparison group. This may be appropriate in some settings, but if a 65-year-old in New Orleans is in worse health than a 65-year-old in other regions, someone older than 65 might be a better control. Ultimately, a valid control group is not necessarily one with demographics that are statistically identical to the control group but one whose mortality rate would have evolved in parallel with that of the New Orleans cohort absent the hurricane.

RK notes that when we control for all combinations of age, race, sex, and year in the 1992 and 1999 cohorts, mortality prior to Hurricane Katrina rises somewhat faster among the New Orleans cohort compared to the controls. RK appears to misinterpret this positive trend as demonstrating positive selection on health (i.e., that New Orleans residents are healthier than others of the same age, race, and sex). Actually, the positive mortality trend indicates *negative* selection on health: prior to Hurricane Katrina, mortality rates among the New Orleans cohort are increasing somewhat with respect to others of the same age, sex, and race. Thus, in our context, the inclusion of fixed effects for all combinations of age, race, sex, and year appears to worsen the validity of the counterfactual, suggesting that estimates from specifications that do not include such fixed effects are more reliable.

RK also points out that some of the post-Katrina estimates for the 1992 and

1999 cohorts are not statistically significant after controlling for all combinations of age, race, sex, and year. However, the presence of a small but positive pre-trend implies that post-period estimates will be upwardly biased, yielding estimated mortality decreases that are smaller than the true mortality decreases. Additionally, as we note, "The 1992 and 1999 cohorts…may only partially capture Hurricane Katrina's impact on Medicare victims, as about two-thirds (one-third) of individuals in the 1992 (1999) cohort had moved away or died before 2005" (DM, 3612–3613). Furthermore, the elderly in the 1992 (1999) Medicare cohort were at least 77 (70) by the time Hurricane Katrina struck. It should therefore not be surprising that the post-Katrina estimates are noisier for these cohorts, though as Table 2 shows, the 2006–2013 mortality declines are jointly highly significant for the 1999 cohort, even with extensive controls (DM, 3619).

Heterogeneity

Following Hurricane Katrina, residents of New Orleans who were black or under the age of 65 were more likely to move than others. Yet because these groups did not experience significantly larger mortality declines after the disaster, RK takes this as evidence "that moving was not the primary cause of the decline in mortality reported" (RK, 40).

Our paper does not claim to estimate the mortality effect of moving. Rather, we estimate how exposure to the local conditions in a place shape mortality, net of any moving effects. To do so, we compare a mover's mortality to that of *other movers* as a function of the local mortality rate in his or her destination region. Such a comparison "will control for any mortality effects that are common to all migrants" (DM, 3614). For example, it is possible for the long-run mortality of a particular group of movers to increase, on average, as a result of the hurricane even while some of the group members experience a *relative* mortality decline as a result of moving to lower-mortality regions.

How selection into moving among a demographic group maps to subsequent mortality declines also depends on whether there are heterogeneous treatment effects. A demographic group with a greater propensity to move than another demographic group may not experience larger average mortality declines if the additional movers are also those for whom place effects are smallest. Finally, the effect of place itself may vary by demographic group, which RK acknowledges (RK, 40 n.4). As we mention in the paper, "place may have a larger impact for black individuals, who made up a large share of the New Orleans victims and were also disproportionately likely to move after the hurricane, than for other races" (DM, 3629).

Local correlates of movers’ mortality

In addition to examining the correlation between movers’ mortality rates and destination region mortality, our AER paper reports how movers’ mortality correlates with other destination characteristics. RK critiques these estimated correlations as though they are meant to capture a causal relationship between a particular destination characteristic and movers’ mortality. He goes on to state “Imbuing estimates of these correlations with causal meaning is inappropriate” (RK, 46).

It is not clear to us what prompted RK to raise this point. Nowhere does the published manuscript claim that these correlations reflect the causal effect of a particular local characteristic. By contrast, as we state in the paper, “we emphasize that the estimate reflects the causal effect of the given characteristic itself only if the characteristic is uncorrelated with any other local attribute that also affects movers' mortality. Because each region is a bundle of many, often correlated, characteristics, these results should be viewed as suggestive of what actually determines place effects” (DM, 3625).

Conclusion

In his comment, RK expresses skepticism about the effects of Hurricane Katrina on the long-run mortality of its victims documented in our AER paper. He raises several potential threats to the validity of our analysis that we previously addressed and reiterates some points of caution that we emphasized in our published paper. In his criticism, RK himself makes several assumptions—such as that hurricane survivors who did not move away from New Orleans must have suffered mortality increases—that seem to be based on strong priors rather than empirical evidence.

While our finding that Hurricane Katrina reduced mortality among New Orleans residents may be surprising, we are not the first to document that Hurricane Katrina improved some outcomes for its victims. Deryugina, Kawano, and Levitt (2018) and Groen, Kutzbach, and Polivka (2020) show that Hurricane Katrina *increased* its victims’ earnings in the longer run. Sacerdote (2012) documents improved test scores among students displaced from New Orleans. Beyond Hurricane Katrina, Ruhm (2000) shows that mortality *decreases* during recessions. No study is without its limitations, but we are grateful for the opportunity to reiterate the ways in which our study adds new evidence on how Hurricane Katrina decreased mortality among its elderly and long-term disabled victims and that moving to lower-mortality places played an important role in these dynamics.

References

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns. *American Economic Journal: Applied Economics* 10(2): 202–233.

Deryugina, Tatyana, and David Molitor. 2020 (DM). Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina. *American Economic Review* 110(11): 3602–3633. **Link**

Groen, Jeffrey A., Mark J. Kutzbach, and Anne E. Polivka. 2020. Storms and Jobs: The Effect of Hurricanes on Individuals' Employment and Earnings over the Long Term. *Journal of Labor Economics* 38(3): 653–685.

Kaestner, Robert. 2021 (RK). Alive and Kicking: Mortality of New Orleans Medicare Enrollees After Hurricane Katrina. *Econ Journal Watch* 18(1): 35–51. **Link**

Ruhm, Christopher J. 2000. Are Recessions Good for Your Health? *Quarterly Journal of Economics* 115(2): 617–650.

Sacerdote, Bruce. 2012. When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics* 4(1): 109–135.

About the Authors

Tatyana Deryugina is an Associate Professor of Finance at the University of Illinois. Her research agenda focuses on environmental risk. She has studied the economic costs of both natural and man-made environmental shocks, including hurricanes, climate change, hazardous substance spills, and air pollution. She is a co-editor at the *Journal of the Association of Environmental and Resource Economists* and at *Environmental and Energy Policy and the Economy*, and she serves on the board of editors of *AEJ: Policy*. She is affiliated with the National Bureau of Economic Research, the Institute for the Study of Labor (IZA), the E2e Project, and the CESifo Research Network. Professor Deryugina holds a Ph.D. in Economics from MIT, a B.A. in Applied Mathematics from UC Berkeley, and a B.S. in Environmental Economics and Policy from UC Berkeley. Her email address is deryugin@illinois.edu.

David Molitor is an Assistant Professor of Finance at the Gies College of Business at the University of Illinois at Urbana-Champaign and a Faculty Research Fellow at the National Bureau of Economic Research (NBER). His research explores how location and the environment shape health and health care delivery in the United States. He is a Principal Investigator of the Illinois Workplace Wellness Study, a large-scale field experiment of workplace wellness conducted at the University of Illinois. His work has been supported by the National Institutes of Health, the National Science Foundation, J-PAL North America, and the Robert Wood Johnson Foundation. His research has been published in leading academic journals including the *American Economic Review*, the *Quarterly Journal of Economics*, and the *Review of Economics and Statistics*. He is a recipient of the iHEA Arrow Award (2020) and NIHCM Research Award (2020) and has also been recognized for excellence in teaching. His email address is dmolitor@illinois.edu.

Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6026/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 61–70

Critique of an Article on Machine Learning in the Detection of Accounting Fraud

Stephen Walker[1]

LINK TO ABSTRACT

This critique treats an article in *Journal of Accounting Research* entitled "Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach" by authors Yang Bao, Bin Ke, Bin Li, Y. Julia Yu, and Jie Zhang (Bao et al. 2020). In addition to the published paper, the authors provide their Matlab code with an associated dataset in a CSV file and other documents hosted at the code-sharing service Github (**link**). This paper applies their code and dataset to replicate the results and studies the key assumption driving those results.

Within the fields of accounting and finance, corporate fraud detection models have been the subject of a significant volume of work. The literature follows a long line of prediction and detection models found in the literature on capital markets. Parties with interest in these models include the investing public and regulatory bodies such as the Securities and Exchange Commission. Previous corporate frauds including Enron and Worldcom left significant damage in their wake, affecting not only their employees and investors but also the public's trust and faith in capital-market institutions. The great hope is that an early warning system can alert the Securities and Exchange Commission and investors to potential fraud and act before the fraud grows too large.

The previous standard in the accounting literature for detecting accounting fraud is known as the F-Score, which is based on a seven-variable logistic regression model published by Patricia Dechow and collaborators (2011). For modeling pur-

1. Graduate student, University of California, Berkeley, CA 94720.

poses, the best proxy for accounting fraud is the SEC-issued Accounting and Auditing Enforcement Release (AAER), an enforcement action that describes the fraud and typically orders a restatement of previously issued financial reports (e.g., 10-Ks). The observable covariates to these fraud models are financial statement ratios that might include changes in sales, accounts receivables, and inventories, in addition to indicator variables for capital-markets activity including share or debt issuances. These ratios are based on a long line of theoretical and empirical work. A novel innovation of the Bao et al. (2020) paper is that they do not use financial ratios, but rather apply raw financial variables taken directly from the financial statements.

The authors provide a dataset that includes a total of 146,045 firm-year observations from 1991–2014. The data comes from the CompuStat database. AAER data is sourced from the USC Marshall School of Business (previously the Haas School of Business). Unique AAER cases total 413 (each of which may last multiple years), and the sample's total number of fraud-case firm-years is 964. Taking the 964 AAER-affected firm-years and dividing by the total of 146,045 firm-years gives an approximation for the unconditional probability of finding fraud for any firm in any given year as 0.7 percent. Fraud is a rare event, and comparing detection rates against this unconditional expectation is important within accounting research.

Replicating the paper

Replicating the paper is relatively simple. The software Matlab is required. The Matlab code file is called "run_RUSBoost28." The dataset is a CSV file called "uscecchini28.csv." The column headers are shown in Table 1.

The dependent variable is an indicator variable equaling 1 if the AAER covered the firm-year in the data, and zero otherwise which is in the dataset's column 9, labeled *misstate*. The independent variables are 28 raw financial statement variables reported by the company in their annual report and shown in columns 10–37, which include items such as total assets and ending price per share for the period. In the Matlab code, the dataset will be divided into a training and test set. For example, the first looped-trained model was based on data covered by the period from 1991 through 2001. The model was then applied out of sample, e.g., to the year 2003, and that application generated a probabilistic score for each firm in that year. The top 1 percent of the probability scores were taken from this selection and if there is a firm in this subset with an actual AAER for that year, it is counted as a correctly identified positive hit. The fraction of correct hits is the positive predictive value. The model was run iteratively for each year in the study's test period, 2003 through 2008.

TABLE 1. CSV file (the dataset)

Position	Column	Description
1	fyear	Fiscal Year
2	gkvey	Compustat firm identifier
3	sich	4-digit Standard Industrial Classification Code (SIC)
4	insbnk	An indicator variable for financial institutions between SIC 6000–6999
5	understatement	An indicator variable if the misstate indicator involved an understatement
6	option	Not used
7	p_aaer	Identifier for AAER
8	new_p_aaer	New Identifier for AAER
9	misstate	Indicator variable for misstatement
10	act	Current Assets - Total
11	ap	Accounts Payable - Trade
12	at	Assets - Total
13	ceq	Common/Ordinary Equity - Total
14	che	Cash and Short-Term Investments
15	cogs	Cost of Goods Sold
16	csho	Common Shares Outstanding
17	dlc	Debt in Current Liabilities
18	dltis	Long-Term Debt Issuance
19	dltt	Long-Term Debt Total
20	dp	Depreciation and Amortization
21	ib	Income Before Extraordinary Items
22	invt	Inventories - Total
23	ivao	Investment and Advances Other
24	ivst	Short-Term Investments - Total
25	lct	Current Liabilities - Total
26	lt	Liabilities - Total
27	ni	Net Income (Loss)
28	ppegt	Property, Plant and Equipment - Total (Gross)
29	pstk	Preferred/Preference Stock (Capital) - Total
30	re	Retained Earnings
31	rect	Receivables Total
32	sale	Sales/Turnover (Net)
33	sstk	Sale of Common and Preferred Stock
34	txp	Income Taxes Payable
35	txt	Income Taxes - Total
36	xint	Interest and Related Expense - Total
37	prcc_f	Price Close - Annual - Fiscal

Machine learning requires measuring results using a hold-out test sample because machine learning can overfit training datasets and produce results that are too good to be true. An iterative approach is preferable because it shows results as it steps through time, which is what would be experienced in the real world, and thus adds validity to the model. A two-year (or longer) gap between the training sample and test sample is required because AAERs are not immediately known when financial reports are issued. In fact, many years can pass between the financial report and the AAER issuance. A modeler must ask (in the spirit of Senator Howard Baker): What can the model know, and when can the model know it?

One issue related to that question involves serial frauds. Some serial frauds may traverse both training and test periods since they cover more than the gap period. To address this issue, the readme file that accompanies the data and code (**link**) notes:

> The variable new_p_aaer is used for identifying serial frauds as described in Section 3.3 (see the code in "RUSBoost28.m" for more details).

Section 3.3 from their paper is reported in its entirety below, with boldface added to emphasize the action described.

> 3.3 SERIAL FRAUD
>
> Accounting fraud may span multiple consecutive reporting periods, creating a situation of so-called "serial fraud." In our sample, the mean, median, and 90th percentile of the duration of the disclosed accounting fraud cases is two years, two years, and four years, respectively, suggesting that it is **common for a case of fraud to span multiple consecutive reporting periods.** Such serial fraud may overstate the performance of the ensemble learning method if instances of fraudulent reporting span both the training and test periods. This is because ensemble learning is more flexible and powerful than the logistic regression model, and may therefore be better able to fit a fraudulent firm than a fraudulent firm-year. **Hence, enhanced performance of the ensemble learning method may result from the fact that both the training and test samples contain the same fraudulent firm**; the ensemble learning model may not perform as well when the sample contains different firms. **To deal with this concern, we break up those cases of serial fraud that span both the training and test periods. Because we have a small number of fraudulent firm-years relative to the number of non-fraudulent firm-years in any test year, we recode all the fraudulent years in the training period to zero for those cases of serial fraud that span both the training and test periods.** Although this approach helps us avoid the problems associated with serial fraud, it may also introduce measurement errors into the training data. (Bao et al. 2020, 211–212, my emphases)

In summary, serial fraud concerns AAER cases that span multiple reporting periods. However, the section does not directly address why the column *new_p_aaer* was created. Returning to the Matlab code for an explanation, Figure 1 shows the code for the model.

Figure 1. The Matlab code

```
1   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2   % run RUSBoost model with 28 raw accounting variables as features %
3   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4
5   %% set parameters
6   iters = 3000; % the number of iterations/trees of RUSBoost model
7   gap = 2; % the gap between training and testing periods, 2-year gap by default
8
9   %% train and test models
10  for year_test = 2003:2008
11      rng(0,'twister'); % fix random seed for reproducing the results
12      % read training data
13      fprintf('Running RUSBoost (training period: 1991-%d, testing period: %d, with %d-year gap)...\n',year_test-gap,year_test,gap);
14      data_train = data_reader('uscecchini28.csv','uscecchini28',1991,year_test-gap);
15      y_train = data_train.labels;
16      X_train = data_train.features;
17      newpaaer_train = data_train.newpaaers;
18      data_test = data_reader('uscecchini28.csv','uscecchini28',year_test,year_test);
19      y_test = data_test.labels;
20      X_test = data_test.features;
21      newpaaer_test = unique(data_test.newpaaers(data_test.labels~=0));
22      % handle serial frauds as described in our paper
23      num_frauds = sum(y_train==1);
24      y_train(ismember(newpaaer_train,newpaaer_test))=0;
25      num_frauds = num_frauds - sum(y_train==1);
26      fprintf('Recode %d overlapped frauds (i.e., change fraud label from 1 to 0).\n',num_frauds);
27
28      % train model
29      t1 = tic;
30      t = templateTree('MinLeafSize',5); % base model
31      % fit RUSBoost model (default parameters: learning rate: 0.1, RatioToSmallest: [1 1])
32      rusboost = fitensemble(X_train,y_train,'RUSBoost',iters,t,'LearnRate',0.1,'RatioToSmallest',[1 1]);
33      t_train = toc(t1);
34      % turn on the following line of code if you want to get feature importance
35      % [imp,ma] = predictorImportance(rusboost);
36
37      % test model
38      t2 = tic;
39      [label_predict,dec_values] = predict(rusboost,X_test); % predict frauds in the testing year
40      dec_values = dec_values(:,2); % get fraud probability
41      t_test = toc(t2);
42
43      % print evaluation results
44      fprintf('Training time: %g seconds | Testing time %g seconds \n', t_train, t_test);
45      metrics = evaluate(y_test,label_predict,dec_values,0.01); % topN=0.01
46      fprintf('Performance (top1%% as cut-off thresh): \n');
47      fprintf('AUC: %.4f \n', metrics.auc);
48      fprintf('NCDG@k=top1%%: %.4f \n', metrics.ndcg_at_k);
49      fprintf('Sensitivity: %.2f%% \n', metrics.sensitivity_topk*100);
50      fprintf('Precision: %.2f%% \n', metrics.precision_topk*100);
51      % fprintf('Importance of predictors:%d\n', output.imp);
52
53      % turn on the following lines of code if your want to save prediction results in a file
54      output_filename = ['prediction_rusboost28_',num2str(year_test),'.csv'];
55      dlmwrite(output_filename,[data_test.years, data_test.firms, y_test, dec_values],'precision','%g');
56
57  end
```

Line 10 starts the loop that runs the model iteratively stepping through each year of the test period from 2003–2008. Line 21 creates a list of unique values of AAER identifiers where the *misstate* column is not equal to zero (equal to 1) for the test set. Line 24 performs the action described in Section 3.3 and sets the *y_train* indicator values to zero where there is a match in the AAER identifiers in the training sample to the previously created list from the test sample.

The intention of Section 3.3 appears to be correctly coded in Matlab. However, what is the *new_p_aaer* field? In Table 1, the 7th position contains another

field called *p_aaer*. The *p_aaer* field is the AAER number that matches the SEC issued number, which can be searched on the SEC website (**link**). When comparing these two columns, it appears that *new_p_aaer* takes the original AAER number and adds a '1' or '2.' In fact, all but 17 AAER cases take the original AAER number and add a '1.'

I sent an email to the authors of the paper copying their editor and asked specifically about this issue. Professor Ke Bin sent the following response on behalf of the author group to all recipients of the original email (boldface added):

> As we discussed in Section 3.3 of our paper, "we recode all the fraudulent years in the training period to zero for those cases of serial fraud that span both the training and test periods." **Our serial frauds have two requirements: (1) have the same AAER id, and (2) are consecutive in our sample. "1" and "2" are suffix to distinguish serial frauds with the same AAER id but not consecutive in our sample.**

I understand the first part of the requirement. However, I do not understand the second part to the requirement—which was not described in the paper or in the online supporting documents. The serial fraud issue is a problem with the span of the fraud itself, not whether it is consecutive in their sample.

The reason that some cases are not consecutive in the sample was provided by the next explanation, given by Professor Ke when I asked why there were a few missing firm-year observations in the sample.

> We require all observations to contain non-missing values for the 28 raw accounting variables, consistent with prior studies cited in our paper. Those observation [related to the 17 AAERs] are dropped because one of the 28 raw variables are missing in WRDS COMPUSTAT database. For example, firm-years of AAER No. 2472, 2504, 2591, and 2894 are missing DLTIS (Long-term debt issuance) and firm-years of AAER No. 2754 and 3217 are missing XINT (Interest and related expense, total).

To show what Professor Ke is speaking to, Figure 2 shows the AAERs at issue. There are only 17 AAERs where *new_p_aaer* changes values because of the "not consecutive in our sample" issue out of a total of 413 unique AAERs in their sample. Additionally, a large fraction of correct cases identified by the model are related to these 17 AAERs. The number of firm-years correctly identified by the AAERs from 2003–2008 total 10 firm-years and are shown in the bolded boxes. The total correct cases identified by their model are 16 firm-years. So, 63 percent of the correct cases are associated with this issue.

Figure 2. Seventeen AAER cases with two different new AAER identifiers

AAER Number	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
857	8571		8572									
1542							15421	15421		15422		
1839								18391	18391	18391		18392
2472								24721	24721	24721	24721	
2504										25041	25041	
2591										25911	25911	
2754				27541	27541	27541		27542	27542	27542	27542	27542
2894										28941		28942
2937								29371				
2949									29491			
2957							29571	29571	29571			29572
3022									30221			30222
3045							30451	30451	30451		30452	
3156												
3217						32171	32171			32172	32172	32172
3909												
3996												

AAER Number	Out-of-Sample											
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
857												
1542												
1839												
2472	24722	24722	24722									
2504	25042	25042										
2591	25912	25912										
2754	27542	27542	27542									
2894	28942											
2937		29372	29372	29372								
2949		29492	29492									
2957												
3022												
3045												
3156	31561			31562								
3217	32172	32172	32172									
3909					39091		39092	39092	39092	39092	39092	39092
3996								39961		39962	39962	

Professor Ke's explanation is not consistent with how other variables are handled in the dataset. The statement suggests a rule that an observation is dropped if it has a missing CompuStat variable. According to the "SAS coding.pdf" file (**link**), the authors recoded *txp*, *ivao*, *ivst*, and *pstk* to 0 if they were missing. If done for these four variables, why are variables *dltis* and *xint* inconsistently handled?

However, the real issue is not these missing observations per se. Rather, it is the additional requirement that a consecutive sample be required for serial fraud identification. Section 3.3 of their paper describes the bias in machine learning related to serial fraud occurring when "both the training and test samples contain the same fraudulent firm" (Bao et al. 2020, 211–212). To illustrate, take for example AAER No. 2504. This AAER affected Delphi Corporation for the years 2000–2004 and was issued by the SEC in 2006. Summarizing Delphi in context of the Matlab code,

- If an AAER identifier from the test set matches the same identifier from the training set, the Matlab model recodes AAER's *misstate* = 1 in the training set to 0.
- As shown in Figure 2, the AAER identifier for Delphi **changes** to 25041 in the training set and to 25042 in the test set.
- Because Delphi 25042 is not in the training set, the Matlab code **will not** recode Delphi 25041's *misstate* = 1 to 0.

Because the Matlab code treats Delphi AAER No. 2504 as two different AAERs 25041 and 25042, the same fraudulent firm is contained in both the training and test samples. Therefore, the Bao et al. (2020) results are still susceptible to the problem they addressed in Section 3.3. In fact, if Delphi's AAER had not been changed, their machine learning model would not have identified the fraud for the year 2003 or 2004 contributing significantly to the published results.

Re-running the dataset: a simple change

I investigated how the authors' AAER identifier change affected the results. I return the AAER identifiers to their original values by replacing the column *new_p_aaer* with data from the *p_aaer* column in the CSV file. This avoids making any code changes within Matlab. Running their original code on this modified dataset excludes from training the additional firm-years associated exactly with these 17 unique AAER cases, but changes nothing else.

TABLE 2. Three model scenarios

Panel A. Correct cases predicted to be positive			
	(1)	(2)	(3)
Year	Published model	Re-run model	Recoded model
2003	8	7	4
2004	4	4	3
2005	2	2	1
2006	1	1	0
2007	1	1	1
2008	0	0	0
Total	16	15	9

Panel B. Positive predictive values (correct cases / # predicted positive)			
	(1)	(2)	(3)
Year	Published model	Re-run model	Recoded model
2003	13.3 percent	11.7 percent	6.7 percent
2004	6.7 percent	6.7 percent	5.0 percent
2005	3.4 percent	3.4 percent	1.7 percent
2006	1.7 percent	1.7 percent	0.0 percent
2007	1.7 percent	1.7 percent	1.7 percent
2008	0.0 percent	0.0 percent	0.0 percent
Total	4.5 percent	4.2 percent	2.5 percent

The updated results are reported in Table 2. The first column reports the results by year from output files provided by the authors in the "prediciton_rusboost28_2003-2008.zip" file (link). Correct cases total 16 for the 2003–2008 out-of-sample test, corresponding to a 4.5 percent positive predictive value, matching the reported values published. Positive predictive value, also known as precision, is calculated as the proportion of correct AAER firm-years out of the cases predicted to experience an AAER. The second column reports the results I obtain when running their original code on their original dataset, showing 15

correct cases corresponding to a 4.2 percent positive predictive value (I'm not sure why it is 15 rather than 16 as in the published paper). The third column reports the results I obtain when running their original code on the dataset with the AAER identifiers replaced by their original values, showing only 9 correct cases corresponding to a 2.5 percent positive predictive value. This value is critical because their published model compared the machine learning result with the result from a parsimonious logit model based on prior literature, which their paper reports to be 2.63 percent for positive predictive value. The updated result shows that the prior model in the literature outperforms this machine learning approach.

Conclusion

The crucial issue in the present critique is to address whether it is appropriate to give new identifiers to the AAER because there is a break in the series resulting from missing data. Since the serial fraud issue concerns the span of the AAER itself and not the sample data, there does not appear to be a logical purpose for the recoding done by the authors. Giving a new AAER identifier to these 17 unique cases out of a total of 413 disproportionately improved their reported results. Without the change, results do not improve upon the prior literature.

Appendix

Data and code related to this research is available from the journal website (**link**).

References

Bao, Yang, Bin Ke, Bin Li, Y. Julia Yu, and Jie Zhang. 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58(1): 199–235.

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28(1): 17–82.

About the Author

Stephen Walker is a Ph.D. candidate in Accounting at UC Berkeley Haas School of Business. Prior to his Ph.D. studies, Stephen worked in sellside equity research for Sanford C. Bernstein in New York City covering the transportation industry, and he trained investment professionals with Training the Street. Stephen was also a small business entrepreneur. He holds an MBA from Columbia Business School and started his career in the finance department with CSX Corporation. He can be reached via his personal website at stephenwalker.me (**link**).

Bao, Ke, Li, Yu, and Zhang's reply to this article
Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6027/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 71–78

A Response to "Critique of an Article on Machine Learning in the Detection of Accounting Fraud"

Yang Bao[1], Bin Ke[2], Bin Li[3], Y. Julia Yu[4], and Jie Zhang[5]

LINK TO ABSTRACT

This is a response to Stephen Walker's (2021) critique of our article, "Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach" (Bao, Ke, Li, Yu, and Zhang 2020). We received the final version of Walker's critique on February 24, 2021. Walker (2021) raises two empirical issues about our paper. The first one is about our treatment of missing values for the raw financial statement variables. The second one is about our treatment of serial fraud. We find no evidence that these two issues alter our paper's inferences. Walker (2021) also implies that we were not transparent in disclosure. To the contrary, we demonstrated our transparency by releasing not only our program codes but also our full data set, so that people can freely replicate and extend our findings.

Missing values

Following prior research, Bao et al. (2020) started with 28 raw financial variables for all fraud prediction models. Some observations of the 28 raw financial

1. Shanghai Jiao Tong University, Shanghai, China 200030.
2. National University of Singapore, Singapore 119245.
3. Wuhan University, Wuhan, China 430072.
4. University of Virginia, Charlottesville, VA 22903.
5. Nanyang Technological University, Singapore 639798.

variables contain missing values. Walker (2021, 67) questions why we recoded the missing values to zero for *txp*, *ivao*, *ivst*, and *pstk* but not for the other raw financial variables. Our treatment of missing values is consistent with the existing literature, and we applied the same treatment consistently for all fraud prediction models. Specifically, we follow Mark Cecchini et al. (2010) and Patricia Dechow et al. (2011) by dropping the firm-year observations with missing values. The only exceptions are the raw variables *ivao* (Investments and Advances), *ivst* (Short-term Investment), *pstk* (Preferred/Preference Stock), and *txp* (Taxes Payable), which are set to zero if missing. These four raw data items are used to calculate the financial ratios "Changes in RSST accruals" and "Changes in working capital accruals," used in the logit model based on the 14 financial ratios in our published paper. Recoding missing values of *ivao*, *ivst*, *pstk*, and *txp* to zero follows the common practice in the prior accounting literature (e.g., Richardson et al. 2005; Dechow et al. 2011; Allen, Larson, and Sloan 2013). Scott Richardson et al. (2005, 451 n.8) explain in their study that these variables "represent a balance sheet item that may not be relevant for many companies (e.g., preferred stock), so we set them to zero rather than needlessly discarding observations."

Walker's approach to dealing with serial fraud

Walker (2021) also suggests that our treatment of serial fraud observations leads to incorrect inferences. Ours is one of the first studies to directly examine the influence of serial fraud on fraud prediction. To mitigate the influence of serial fraud cases that span both the training and testing samples on our inferences, we adopted a conservative approach in the published paper by recoding the serial fraud observations to zero (i.e., non-fraud) in the training sample if the serial fraud spans both the training and test periods. Walker questioned our implementation of this conservative approach because we required serial fraud to be *consecutive in our final sample*. Walker proposes an alternative approach by recoding the serial fraud observations to zero in the training sample if the serial fraud observations spanning both the training and test periods share the same primary AAER ID (i.e., *p_AAER* in Dechow et al.'s fraud database).[6] Walker claims that our best machine learning-based fraud prediction model, RUSBoost, underperforms in terms of precision the logit model based on the 14 financial ratios when his approach is adopted.

We believe that Walker's criticism is flawed for several reasons. First, he did not recalibrate the most important parameter of RUSBoost, number of trees,

6. Note that *p_AAER* is the primary AAER number used by Dechow et al. (2011) to identify all the fraud years for the same firm that are part of the same fraud incident.

after changing the fraud training samples using his approach. As a result, he has understated the true performance of RUSBoost (see below for evidence). Second, Walker focuses only on the results in Table 3 of our published paper (using 2003–2008 as the test sample) while ignoring the results in Table 5 of our published paper (using 2003–2005, 2003–2011 and 2003–2014 as three alternative test samples), which is equally important for our inferences due to the underreporting of fraud in the latter part of our test years (i.e., 2006–2014). As explained in Sections 3.1 and 7.1 of our paper (Bao et al. 2020, 207–209, 224–225), many accounting frauds could remain undetected due to reduced regulatory enforcement of accounting fraud that approximately coincided with the 2008 financial crisis. In addition, Alexander Dyck et al. (2010) show that it takes approximately 24 months, on average, for the initial disclosure of a fraud, implying that the detected accounting fraud labels could be understated for the test years as early as 2006/2007. Therefore, the test years 2003–2005 should represent the best test period to assess the performance of different fraud prediction models. Third, there is no clear consensus on how to handle serial fraud in model building (see section 4 for more elaboration).

To address Walker's concern head on, here we replicate the results for our Tables 3 and 5 using Walker's approach. Walker does not replicate all of those results (Bao et al. 2020, 219 Table 3, 226 Table 5) using his approach. As Walker's approach would require the recoding of a significant number of fraud observations into non-fraud in the training sample (see footnote 8 below), we re-optimized an important model parameter of RUSBoost in model training (i.e., number of trees). The results are reported in the following Table 1.[7] We refer the reader to Bao et al. (2020) for the detailed definitions of the variables included in this table.

Several important findings emerge from Table 1. First, the performance of RUSBoost relative to the logit model based on the 14 financial ratios using both AUC and NDCG@k is the strongest for the test years 2003–2005, which are argued to be the best test period as noted above. Second, even though serial fraud is the most severe in 2003–2005, RUSBoost continues to outperform the logit model based on the 14 ratios after adjusting the serial fraud issue using Walker's approach. This evidence suggests that serial fraud is not driving the superior results of RUSBoost.[8] Third, contrary to Walker's claim that the precision of RUSBoost

7. The results in Table 1 are generated using Matlab R2020b on Windows 10.

8. Following Walker's approach, the percentage of recoded fraud firm years due to this serial fraud issue is close to 20 percent for each of the test years 2003–2005 but this percentage drops monotonically for the test years after 2005 (e.g., 6.12 percent for test year 2006, 3.74 percent for test year 2007, and so on). It is clear that the serial fraud issue is the most severe for the test years 2003–2005. Therefore, if serial fraud is a concern, it should negatively affect our RUSBoost model's relative performance the most for the test years 2003–2005. Our results in Table 1 do not support this prediction.

TABLE 1. Results using Walker's approach, corresponding to Tables 3 and 5 in Bao et al. (2020)

Panel A. Performance metrics averaged over the test period 2003–2005				
	Model			
Metric	RUSBoost using 28 raw financial data items	Logit using 28 raw financial data items	Logit using 14 financial ratios	SVM-FK using 28 raw financial data items
AUC	0.7428	0.6736	0.6483	0.6255
NDCG@k	0.0394	0.0084	0.0109	0.0172
Sensitivity	0.0453	0.0097	0.0137	0.0265
Precision	0.0502	0.0113	0.0128	0.0164
# of True Fraud Obs Identified	9	2	2	4
Panel B. Performance metrics averaged over the test period 2003–2008				
	Model			
Metric	RUSBoost using 28 raw financial data items	Logit using 28 raw financial data items	Logit using 14 financial ratios	SVM-FK using 28 raw financial data items
AUC	0.7228	0.6842	0.6711	0.6195
NDCG@k	0.0237	0.0042	0.0273	0.0199
Sensitivity	0.0291	0.0048	0.0399	0.0263
Precision	0.0281	0.0057	0.0262	0.0193
# of True Fraud Obs Identified	10	2	8	6
Panel C. Performance metrics averaged over the test period 2003–2011				
	Model			
Metric	RUSBoost using 28 raw financial data items	Logit using 28 raw financial data items	Logit using 14 financial ratios	SVM-FK using 28 raw financial data items
AUC	0.7171	0.6981	0.6711	0.6297
NDCG@k	0.0243	0.0105	0.0235	0.0184
Sensitivity	0.0325	0.0171	0.0349	0.0300
Precision	0.0249	0.0100	0.0222	0.0168
# of True Fraud Obs Identified	13	5	10	9
Panel D. Performance metrics averaged over the test period 2003–2014				
	Model			
Metric	RUSBoost using 28 raw financial data items	Logit using 28 raw financial data items	Logit using 14 financial ratios	SVM-FK using 28 raw financial data items
AUC	0.7196	0.7061	0.7015	0.6519
NDCG@k	0.0182	0.0101	0.0223	0.0261
Sensitivity	0.0244	0.0172	0.0345	0.0346
Precision	0.0187	0.0090	0.0185	0.0170
# of True Fraud Obs Identified	13	6	11	9

underperforms the precision of the logit model based on the 14 ratios for the test period 2003–2008, we find that the precision of RUSBoost continues to outper-

form the precision of the logit model based on the 14 ratios (2.81 percent vs. 2.62 percent). Specifically, the RUSBoost catches 10 accounting fraud firm years while the logit model catches 8 accounting fraud firm years in the test period 2003–2008. Fourth, using the AUC as a performance metric (a common performance evaluation metric in the fraud prediction literature), RUSBoost always outperforms the logit model based on the 14 ratios for all test periods in Table 1. Finally, similar to the reported results in our published paper, a counterintuitive finding in Table 1 is that the AUC of the logit model based on the 14 ratios increases over time for the test samples from 2003–2005 to 2003–2014, even though the problem of undetected fraud grows over time as noted above and therefore the reported fraud frequencies are not an accurate measure of the true fraud frequencies. Overall, we conclude that adopting Walker's approach does not alter our inferences.

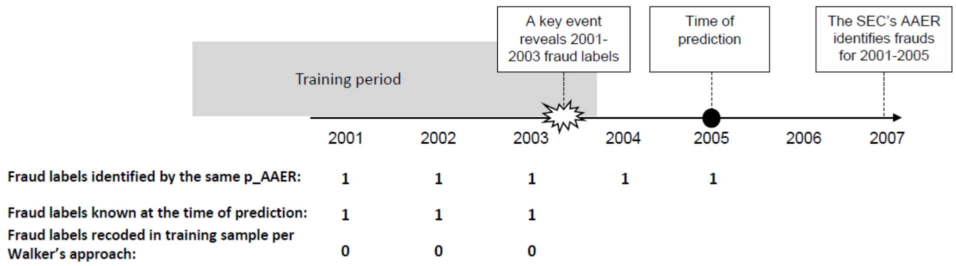
What is the optimal approach for dealing with serial fraud?

Walker's critique raises the broader question of whether there is an optimal approach for dealing with serial fraud. Ex ante, there is no clear answer to this question. As noted by Bao et al. (2020), most prior studies do not deal with the serial fraud issue at all and instead treat each fraud firm-year as an independent observation. If we follow the approach of these prior studies, as shown in Bao et al. (2020, 227 Table 6), the performance of RUSBoost would far exceed that of the logit model based on the 14 ratios.

It is also unclear whether Walker's approach of using *p_AAER* to define serial fraud is the most appropriate. This is because the AAERs issued by the SEC often come at the last stage of the fraud revelation timeline (Karpoff et al. 2017). Many other fraud revelation events (e.g., restatement, litigation, analyst report, whistleblower, etc.) can provide more timely fraud label information before the announcement of the SEC's AAERs. Hence, Walker's approach could be inappropriate and represent overkill.

To illustrate our reasoning, let's consider the following hypothetical scenario (see Figure 1 for the timeline): a firm has consecutive fraud labels from 2001 to 2005 identified by the SEC's AAER issued in 2007. These fraud labels are summarized in Dechow et al.'s database using the same *p_AAER*. The fraud labels for 2001–2003 were known to investors in 2004 due to a fraud revelation event in the middle of 2004, but the fraud labels for 2004–2005 were not known to investors until the date of AAER release at the end of 2007.

Figure 1. A serial fraud example with a key fraud revelation event during training period



Now let's assume that, right after the release of fiscal year 2005's 10-K (i.e., time of prediction in Figure 1), we wish to predict whether the firm commits fraud or not in its 2005 annual report. As the fraud labels for 2001–2003 were already known to investors in 2004, we can include the available fraud labels in 2001–2003 to train a fraud prediction model for the test year 2005. On the other hand, Walker's approach would recode the fraud labels in 2001–2003 to zero in model training for test year 2005, which seems inappropriate because the fraud labels in 2001–2003 were already known to the public in 2005 and therefore should be included in model training.[9] We reviewed a few serial fraud cases and found that cases similar to our example above existed in our sample (e.g., AAER #2591 and AAER #2819), but it is beyond the scope of our study to identify all such cases.

In conclusion, Walker's approach of relying solely on *p_AAER* ID to define serial fraud could be inappropriate, given that the reported frequency of detected fraud in our sample is very low to start with (less than 1 percent, on average, as shown in Bao et al. 2020, 211 Table 1) and a portion of the serial fraud observations, which represents the majority of the fraud observations in our sample, could have been known to the public at the time of model training.

Appendix

Code related to this research is available from the journal website (**link**).

9. Even if the fraud labels for 2001–2003 were known to investors in 2006 only (i.e., *after* the time of prediction in Figure 1), we can still use the fraud labels for the years 2001–2003 to retrain a prediction model for the test year 2005. As we did not know whether year 2005 was a fraud or not until 2007, such a retrained prediction model based on more updated fraud labels from the training years 2001–2003 should be still valuable to investors at the time of prediction 2005. Walker's approach would preclude the inclusion of these fraud labels in model training.

References

Allen, Eric J., Chad R. Larson, and Richard G. Sloan. 2013. Accrual Reversals, Earnings and Stock Returns. *Journal of Accounting and Economics* 56(1): 113–129.

Bao, Yang, Bin Ke, Bin Li, Y. Julia Yu, and Jie Zhang. 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58(1): 199–235.

Cecchini, Mark, Haldun Aytug, Gary J. Koehler, and Praveen Pathak. 2010. Detecting Management Fraud in Public Companies. *Management Science* 56(7): 1146–1160.

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28(1): 17–82.

Dyck, Alexander, Adair Morse, and Luigi Zingales. 2010. Who Blows the Whistle on Corporate Fraud? *Journal of Finance* 65(6): 2213–2253.

Karpoff, Jonathan M., Allison Koester, D. Scott Lee, and Gerald S. Martin. 2017. Proxies and Databases in Financial Misconduct Research. *Accounting Review* 92(6): 129–163.

Richardson, Scott A., Richard G. Sloan, Mark T. Soliman, and İrem Tuna. 2005. Accrual Reliability, Earnings Persistence, and Stock Prices. *Journal of Accounting and Economics* 39(3): 437–485.

Walker, Stephen. 2021. Critique of an Article on Machine Learning in the Detection of Accounting Fraud. *Econ Journal Watch* 18(1): 61–70. **Link**

About the Authors

Yang Bao is an assistant professor at Shanghai Jiao Tong University. His email address is baoyang@sjtu.edu.cn.

Bin Ke is a professor at National University of Singapore. His email address is bizk@nus.edu.sg.

Bin Li is a professor at Wuhan University. His email address is binli.whu@whu.edu.cn.

Y. Julia Yu is an assistant professor at University of Virginia. Her email address is julia.yu@virginia.edu.

Jie Zhang is an associate professor at Nanyang Technological University. His email address is zhangj@ntu.edu.sg.

Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6028/

EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 79–94

Colonial New Jersey's Paper Money: A Reply to Michener Again, and Again, and Again

Farley Grubb[1]

LINK TO ABSTRACT

In the last issue of *Econ Journal Watch*, Ronald Michener (2020) published his seventh critical comment on my research. In my replies to his previous six comments, I demonstrate that Michener is misguided (see Grubb 2005; 2006a; b; 2018b; 2019b; 2020a). I will continue that demonstration here in my reply to his seventh comment. I will demonstrate that Michener does not understand basic microeconomic theory; that Michener does not understand rational expectations or how to make it operational; that Michener does not understand my model of monetary performance; that Michener does not understand how colonial New Jersey redeemed its paper money; and that Michener does not know how to evaluate quotation evidence. In summary, Michener thinks downward sloping marginal revenue curves are demand curves, thinks that rational expectations means you must be able to expect the unexpected, does not consider risk when evaluating an asset's present value, erroneously thinks that the redemption of colonial paper money was on demand, and considers antecedent and subsequent context irrelevant for evaluating quotation evidence.

1. University of Delaware, Newark, DE 19716.

Michener does not understand basic microeconomic theory

In his Figure 1, Michener (2020, 307) draws a downward sloping marginal revenue curve and labels it a demand curve. Marginal revenue curves are not demand curves. Demand curves are average revenue curves. All economists know this. As such, Michener commits a fundamental error in microeconomic theory.

How do we know that what Michener draws in his Figure 1 and labels a demand curve is, in fact, a marginal revenue curve? In his Figure 1, Michener (2020, 307) uses an example of an asset package that yields a stream of payoffs into the future to illustrate his objection to how I calculate present value. To slightly simplify, Michener posits a package that has five $100 bills in it. You get to randomly draw one bill out today from the package, and then randomly draw out another bill one year from now from the package, then randomly draw out another bill two years from now from the package, then randomly draw out another bill three years from now from the package, and finally you get to have the last bill in the package four years from now. You do not know which bill you will get each year, but you know that each bill has a face value of $100.

What is the present value of this package—what I call the asset present value (APV)? What would you pay today for this package—your indifference price? In other words, what is its total revenue in present-value terms? It is ($\$100 + \$100e^{-r} + \$100e^{-2r} + \$100e^{-3r} + \$100e^{-4r}$), where r = the interest rate. This is basic finance. This total revenue in present-value terms is exactly the area under the curve Michener draws in his Figure 1 and erroneously labels a demand curve (Michener 2020, 307). If the area under a curve is total revenue then the curve must be a marginal revenue curve and not a demand curve, QED. This is basic calculus.

Demand curves measure the total revenue of an asset package in present-value terms by taking the present-value price of that package and multiplying that price by the relevant quantity, i.e., (P * Q). The vertical axis is the present-value price (P) which is the price of an individual dollar in the package because the horizontal axis is measured in individual dollars (see Michener 2020, 307 Figure 1). The total revenue in present-value terms is a rectangular box under a demand curve at the relevant PQ point on the demand curve. What would be the present-value price of the above package of five $100 bills? The face-value quantity supplied is $500 in Michener's Figure 1 (ibid.). What would be the present-value price (P) on the demand curve at that quantity? By definition it would have to be (P * (Q = $500)) = Total Revenue = ($\$100 + \$100e^{-r} + \$100e^{-2r} + \$100e^{-3r} + \$100e^{-4r}$). This reduces to (P * $500) = $\$100 * (1 + e^{-r} + e^{-2r} + e^{-3r} + e^{-4r})$, which further reduces to $P = 1/5 * (1 + e^{-r} + e^{-2r} + e^{-3r} + e^{-4r})$. This is exactly the present-value

price that Michener reports (2020, 307) for my method of calculating the APV, namely the price of an individual bill in present-value terms in the package. Thus, Michener's example proves that I am pricing bills along a demand curve which is what Michener insists should be done (and which the pricing Michener advocates fails to do in his own example).[2]

While all economists know that downward sloping demand curves are average revenue curves and not marginal revenue curves, such knowledge may not resonate for my non-economist history and finance colleagues.[3] For non-economists I will show Michener's error another way. If you followed Michener's advice and sold enough of the above packages of five $100 bills today for the present-value price Michener's insists they should be sold for, you would be impoverished. For illustration's sake, suppose the interest rate, *r*, that you can both borrow and lend at, is 6 percent. In his Figure 1, Michener (2020, 307) sets the present-value price of the face value of the above $500 package at e^{-4r}. At 6 percent that present value is ($500 * $e^{(-4*0.06)}$) = $393.31. This is what Michener would have you sell the above $500 face-value package for today (to be supposedly indifferent between keeping the package for yourself and selling it for $393.31 today).

Suppose you followed Michener's advice and sold the above $500 package for $393.31 today, would you come out ahead or behind at the end of four years? You would take the $393.31 today and invest it at 6 percent annual interest. After one year at 6 percent your initial $393.31 would grow to $416.91. By the next year the $416.91 would grow to $441.92. By the third year your sum would grow to $468.44. And by the fourth and final year your sum would grow to $496.55.

Now suppose you kept the package for yourself and invested the sums as

2. Michener (2020, 308) points out that as the quantity of bills from a given emission (when there is only one emission in existence) are redeemed over time their present value will rise, finally reaching face value (or 1 in his example) in the final year of redemption. He claims this fact is "utterly fatal to Grubb's argument." Michener says that Grubb (2016b, 174–178) acknowledges this utterly fatal aspect to his argument when "he touts the stability of APV as a virtue of his theory…" This is a bait and switch maneuver by Michener. While Michener's claim about the present-value path of bills is correct for an isolated and single emission when only that emission and no other emissions are in existence, such was not the case for New Jersey. New Jersey engaged in on-going and overlapping emissions throughout its history, and had legal tender laws that made the redemption of bills fungible across outstanding emissions (Grubb 2015; 2016a). Michener (2020, 309, 311–312) even acknowledges that New Jersey ran overlapping emissions. Michener fails to address how to calculate APV when emissions overlap and bills across emissions are fungible in redemption due to legal tender laws, whereas Grubb's (2016a; b) method does. The only "utterly fatal" outcome here is Michener's "utterly fatal" inability to model or even understand how New Jersey's complex emission and redemption structure worked.

3. Michener (2020, 308–309) justifies his approach by spouting "marginalist revolution." However, there is nothing in the marginalist revolution that says that downward sloping marginal revenue cures are demand curves. Michener (2020, 308–309) goes on to conflate his Figure 1 marginal revenue curve (which he erroneously labels as a demand curve) with being a marginal utility curve. Clearly Michener does not understand the marginalist revolution or basic microeconomic theory.

they came into your hands. What would you have at the end of four years? You take the package's first $100 bill and invest it at 6 percent. Thus at the end of year one you now have $206 from the package ($100 gained initially + $6 interest on that $100 + $100 from the end of year one's draw). In year two, this package has grown to $318.36. In year three, this package has grown to $437.46. By the fourth and final year this package has grown to $563.71. Thus, your loss in selling this package of five $100 bills for Michener's present-value price of $393.31 is ($563.71 − $496.55) = $67.16. If you sold enough of these packages at a loss of $67 each, you would end up impoverished.

Now supposing you sell this package of five $100 bills for 500 times the price Grubb calculates for a single dollar in the package, how will you fare? As shown above and as Michener (2020, 307) claims, Grubb's present-value price for a dollar in this package is $1/5 * (1 + e^{-r} + e^{-2r} + e^{-3r} + e^{-4r})$, which at 6 percent equals $0.8901166. What you should sell the package for (its present value) is therefore ($0.8901166 * 500) = $445.06 (as opposed to Michener's price of $393.31). If you take the $445.06 today and invest it at 6 percent annual interest you will end up at the end of four years with approximately the same sum you would have if you had not sold the package but instead invested it yourself as described above. Under Grubb's pricing advice you would be indifferent between keeping and selling the package and so break even as opposed losing $67 on every package you sold under Michener's pricing advice. This comparison shows that Michener is wrong and following his method will impoverish you, whereas following my method leaves you indifferent and so correctly measures the present value (the APV) of the package of bills.

No matter how the Michener example is set up, the same result occurs. Just to be thorough, and in anticipation of objections to the example I just used, I will give another example scenario that also matches Michener's (2020, 307) setup. Suppose you owe $100 in taxes today and the same amount each year exactly one year apart for the next four years, for a total of $500 face value in taxes owed over this period. You can pay your taxes each year in cash (dollars) or by using bills of credit designed as tax coupons and which are only usable as tax coupons. You have five $100 tax coupons, for a total of $500 face value in tax coupons. Each $100 tax coupon has a designated year it can be used in, and in no other year. Each tax coupon in your package of five tax coupons is good for a different year across the five years you owe taxes. Thus, if you keep the package of five $100 tax coupons, then you can pay all the taxes you owe today and over the next four years.

What would you sell this package of five $100 tax coupons for today in cash? What is its present value? Having sold the package for cash, will you have enough cash to pay your tax bill today and over the next four years in cash? Again, suppose the interest rate, *r*, that you can both borrow and lend at, is 6 percent. In his Figure

1, Michener (2020, 307) sets the present-value price of the face value of the above \$500 tax coupon package at e^{-4r}. At 6 percent that present value is (\$500 * $e^{(-4*0.06)}$) = \$393.31. This is what Michener would have you sell the above \$500 face-value tax coupon package for today (to be supposedly indifferent between keeping the package for yourself and selling it for \$393.31 cash today).

Suppose you followed Michener's advice and sold the above \$500 tax coupon package for \$393.31 cash today, would you come out ahead or behind at the end of four years? Could you pay all your taxes today and over the next four years with that cash? You would take the \$393.31 cash today, pay your \$100 tax bill today and invest the rest (\$293.31) at 6 percent annual interest. By next year (year one in Michener's example) at 6 percent your initial \$293.31 would grow to \$310.91. You would pay next year's \$100 tax bill in cash and be left with \$210.91 to invest for the following year. That \$210.91 would grow to \$223.96 by the start of year two. You would pay your \$100 tax bill for year two in cash and be left with \$123.56 to invest for the following year. That \$123.56 would grow to \$130.97 by the start of year three. You would pay your \$100 tax bill in cash for year three and be left with \$30.97 to invest for the following year. That \$30.97 would grow to \$32.83 by the start of year four. You owe your last \$100 tax bill in cash at the start of year four. But you do not have enough cash left. You are \$67.17 short. You are a tax delinquent. This is the same shortfall amount (\$67) generated by my prior asset-package present-value-calculation example. If you follow Michener's pricing advice the sheriff will show up and sell your property to pay back taxes or maybe put you in debtor's prison.

Now suppose you sell your five \$100 tax coupon package today for \$445.06 cash, namely the package's present value as Grubb calculates it (see above). Would you come out ahead or behind at the end of four years? Could you pay all your taxes today and over the next four years with that cash? You would take the \$445.06 cash today, pay your \$100 tax bill today and invest the rest (\$345.06) at 6 percent annual interest. By next year (year one in Michener's example) at 6 percent your initial \$345.06 would grow to \$365.76. You would pay next year's \$100 tax bill in cash and be left with \$265.76 to invest for the following year. That \$265.76 would grow to \$281.71 by the start of year two. You would pay your \$100 tax bill in cash for year two and be left with \$181.71 to invest for the following year. That \$181.71 would grow to \$192.61 by the start of year three. You would pay your \$100 tax bill in cash for year three and be left with \$92.61 to invest for the following year. That \$92.61 would grow to \$98.17 by the start of year four. You owe your last \$100 tax bill in cash at the start of year four. You are less than \$2 short, and so basically indifferent between keeping the tax coupon package and selling it today for the cash price Grubb calculates as its present value.

This second scenario yields the same outcome as the first example scenario

above. Thus, again, these comparisons show that Michener is wrong and following his method will impoverish you or turn you into a tax delinquent deadbeat, whereas following my method leaves you indifferent and so correctly measures the present value (the APV) of the package of bills or tax coupons.

Michener does not understand rational expectations

To estimate the present value of colonial bills of credit (their APVs), when they are hypothesized only to be zero-coupon bonds, requires knowing when these bills will be redeemed at face value. Some expectation or forecast of the path of future bill redemption has to be imputed to citizens. I use a rational expectations approach to impute the expected path of bill redemption. Michener (2020, 310–313) fails to grasp even the rudimentary aspects of this approach.[4]

Under this approach, citizens use all the information available to forecast the likely future. Regarding the path of future bill redemption, citizens knew the legislated plans for future redemption, they knew how, when, and why the legislature deviated from their planned redemption path in the past and so could forecast likely future deviations, citizens knew the likely chances of ongoing and renewed conflicts with native Americans and the French and how that affected bill emissions and redemptions, and so on.

Events citizens could not know in advance, by definition, would not factor into their forecast. Such events would only affect forecasts after the unpredictable event became manifest. Examples of unpredictable events would be future earthquakes, covid-19 pandemics, the American Revolution, and so on. Rational expectations separates predictable information from unpredictable information when considering how expectations are formed—using only information from the predictable or known set. Michener does not understand this.

Given the above discussion of the information citizens had, I constructed APV under the assumption that citizens correctly forecast the actual pre-1775 path of bill redemption. As a first pass, this is not a bad rendition of rational expectations when there are no unpredictable events, or when such events are minor. That said, this APV construction is still only a preliminary counterfactual measure as it abstracts from any potential risk discount (RD) that would correct and transform the APV measure into an actual risk-adjusted present-value forecast

4. Michener and I were both students of Robert Lucas Jr. at the University of Chicago at about the same time. Considering that fact leaves me dumbfounded as to Michener's lack of understanding of rational expectations.

(APV − RD)—something my model does that Michener does not consider (Grubb 2016a, 1217).

If an unpredictable event occurs, once it becomes manifest, it could then raise doubts about the legislature's ability to redeem bills in the future to the extent that the legislature actually did. That doubt would show up as a measurable positive risk discount in my model of monetary performance (Grubb 2016a, 1217). I find several episodes pre-1775 where there is a positive risk discount attached to New Jersey bills, and my model locates what the unpredictable events likely were that became manifest and so generated the risk adjustment to APV (see Grubb 2016a, 1223–1228). Michener does not understand this.

The American Revolution and the monetary havoc that ensued was an unpredictable event pre-1775. As such, you cannot use the actual redemption path of bills post-1775 as a pre-1775 forecast of the expected bill redemption path post-1775. To do so would be an error in the operational application of rational expectations. Instead, I use as my pre-1775 expectation of post-1775 bill redemption the bill-redemption path that would have been likely if there was no Revolution. Again, this accords with rational expectations methodology.

What Michener misses and what is so powerful about my modeling approach is it shows that, pre-1775, citizens actually did not expect the Revolution and the monetary havoc that ensued. If citizens had expected the Revolution and also expected that shortly after the Revolution started their bills would no longer be redeemed as they had been in the past and might even become worthless, then between 1770 and 1774 the market exchange value (MEV) of bills, what people currently paid for them in the marketplace, would have fallen well below my counterfactual asset present value (APV) measure, thus yielding a sizable risk discount, namely (MEV − APV) < 0. As Grubb (2016a, 1223, 1225) shows, there is no risk discount manifest in 1770–1774 given my APV construction, and so citizens must not have expected the Revolution and the monetary havoc that ensued.[5]

5. Michener often criticizes my research by advocating alternative approaches. He, however, never actually does what he advocates, even though he could easily do so, to see if his alternative approaches make sense. Using the actual redemption of bills post-Revolution as the expectation of what that redemption would be pre-1775 to construct APV in the early 1770s is one such example of an alternative approach advocated by Michener (2020, 310–313). What if Michener had actually applied his advocated alternative; what would he have discovered; what would he have to assume? The MEV in Figure 2 of Grubb (2016a, 1223) between 1770 and 1774 would not change as those data are observations. In that figure, the counterfactual APV as calculated using Michener's alternative approach would now be considerably lower than that shown by me using my approach. Michener's result would produce a very large transaction premium (TP), namely ($MEV - APV_{Michener}$) >> 0, between 1770 and 1774, much larger than the small TP I find. In other words, Michener would have us believe that citizens were actually paying a massive amount over and above the bills' asset present value to acquire these bills in 1770–1774 even though citizens knew (according to Michener) that the bills would be near worthless after 1775. This behavior by citizens would be patently irrational. Thus, Michener's advocated alternative approach only makes sense if you assume that colonists

That citizens did not expect the Revolution and the monetary havoc that ensued can also be gleaned from the writings of the founding fathers and the Continental Congress (Smith 1976, vol. 1; *Journals of the Continental Congress*, vols. 1 and 2). In fact, the 1787 Constitutional Convention opened with a statement by Edmund Randolph that "the havoc of paper money had not been foreseen" by the creators of the Articles of Confederation (Farrand 1966, 1:18; Grubb 2006c, 50). Virtually no responsible scholar thinks that the actual monetary havoc that occurred after 1775 was predictable from the vantage point of the early 1770s.

Michener (2020, 310–313) makes the absurd inference that if pre-1775 bill redemption could be predicted then the bill redemption during the monetary havoc that ensured during the Revolution must also have been predictable. Therefore, the actual path of bill redemption post-1775 should be incorporated into the expectations formed pre-1775 that are used to construct the APV of bills pre-1775. Alternatively, his logic implies that if the Revolution and the monetary havoc that ensued could not be predicted, then nothing can be predicted. Michener's failure to grasp that rational expectations separates futures into predictable futures, based on reasonably known information, and unpredictable futures shows he does not understand rational expectations, how to make it operational, or how it is used in my model of monetary performance.

"I pray thee, bear my former answer back"

—William Shakespeare, *Henry V*, Act 4, Scene 3

Most of the Michener (2020) rejoinder consists of him repeating claims he has already made numerous times in his past publications, in his past published comments on my research, and in his past referee reports on my research submissions. I have already answered those claims in my past published replies to his comments, and in my revisions to manuscripts that satisfied editors. I see no reason to engage in the childish game of simply restating 'you are wrong,' 'no, you are wrong,' back and forth, ad infinitum. Such is just an exercise in trying to have the last word, as though getting the last word somehow makes you right. So, I will simply reply to most of what Michener (2020) trots out again in his rejoinder with, "I pray thee, bear my former answer back."

A few comments, however, should be made here. Michener's (2020, 313–321) lengthy diatribe on how to calculate the market exchange value (MEV) of

were crazy irrational people. Michener's colonial monetary models, if you can call them models, often rest on assuming that colonists were crazy irrational people (see Grubb 2006a, 54–60).

colonial New Jersey's paper money contains three errors that render his discussion both erroneous and irrelevant. The issue turns on establishing what the face value at redemption was for a New Jersey bills of credit. That face value in relation to the current market trade value is used to calculate the current market exchange value as a percentage of that face value (what I call MEV). First, as in his past commentaries, Michener often conflates the current exchange value (what the bills currently passed for) with the value at redemption (the bill's face value). That would be like conflating the current cash-in value of a U.S. savings bond with its face-value redemption amount at maturity. Second, Michener fails to note that the face value on the bills, its redemption value, was expressed on the face of each bill as being so much silver plate in penny-weight and grains of silver (Grubb 2015, 18; Newman 2008, 250–258). This by itself renders most of Michener's discussion on this issue irrelevant.

Third, Michener fails to understand the redemption structure used by New Jersey. Michener's (2020, 317) quotation of Governor Bernard in 1767 actually confirms the correctness of my approach, something Michener does not see. New Jersey set redemption windows for each emission that spanned a number of years. Once the amount or quota of redemptions for a given year was filled, the treasurer could refuse to redeem any more bills that year, in effect saying that the holder got there too late and the quota was filled and so the holder would have to return in a future year within the redemption window of years to redeem his bill. Add to that the facts that New Jersey had ongoing emissions with overlapping redemption windows of years and that legal tender laws made redemption among existing emission and their respective redemption windows fungible, and you get the outcome Governor Bernard describes. This does not mean that if you did get to the treasury in a year within the redemption window of years covered by that bill, and that year's redemption quota had not been filled yet, that you did not get face value or its equivalent for your bill. You did. This is all modelled and explained by me in great detail (Grubb 2015; 2016a; 2016b; 2016c). Michener does not consider this structure and, from what he says, does not understand it or how to model it. In short, Michener erroneously assumes that if you could not always get face value on demand then you never got face value for your bill.

The 1741 exchange rate debate and quotation interpretation

Michener asserts that I distorted the meaning of Governor Morris's August 16, 1741 letter to the Lords of Trade and so misled the reader because I did not

include everything from the letter in the quote but pared it down to make it easier for the reader to understand. Let me start by saying that nothing is hidden here. I encourage the reader to go read the entire letter (Morris 1852, 132–137), or for that matter all of Morris's correspondence (Morris 1852), and decide for themselves who is misleading whom.

Michener (2020, 325–326) adds the clauses that immediately precede and immediately follow the portion I quoted (Grubb 2020a, 83) from Morris's August 16, 1741 letter (Morris 1852, 132–137). He argues that adding that material alters the meaning of what I quoted to be what Michener claims it is and not what I use it to demonstrate. Michener's additions are a tiny portion of what is in the respective paragraphs in Morris' August 16, 1741 letter, and a very tiny portion of what is in Morris' entire correspondence on the issue. Michener does not consider the material written that is antecedent and subsequent to the material he added to the quotation I used and so misses the context of that material, context that casts doubt on Michener's interpretation of the material he added to the quote I used.[6]

The whole point of Governor Morris's August 16, 1741 letter to the Lords of Trade was to explain the acute problems with New Jersey bills of credit. If that acute problem was that the £2,000 New Jersey pounds issued by the New Jersey assembly for the expedition against the Spanish bought far more troops and supplies than expected, namely if New Jersey bills of credit were trading way above par or way above what it had in the recent past (as Michener claims), there would have been no need or reason for Morris to write his letter in the first place. Everyone would have been happy, and the Lords of Trade would have needed no explanation of anything.

Morris, in his letter, is trying to explain to the Lords of Trade the difference between what the New Jersey assembly meant when the assembly asserted that they had made no alteration in the currency and what Morris thought was an alteration of New Jersey's currency. The whole point of what "alteration in our currency" (Morris 1852, 133) meant, was in trying to explain what was behind its acute loss in value. If New Jersey's paper money had acutely gained value, as Michener asserts, again there would have been no need for Morris's discussion or letter. The New Jersey assembly considered the currency unaltered if its face value at redemption remained unaltered. Morris considered the currency altered if the current trade value changed. Much of Morris's letter revolves around explaining the acute alteration in New Jersey's paper money value as understood by Morris versus as understood by the assembly.

In Morris's letter over several pages, up to the quotation that I used and

6. I have shown in the past that Michener manipulates meaning by ignoring context when trotting out quotation evidence as support for his views (see Grubb 2006a, 63–67; 2006b, 254–257).

that Michener augmented, Morris is talking only about New Jersey. He does not mention other colonies' bills of credit other than in the sentence that preceded the quotation I presented. Now let's look at what Morris was doing in that section of his letter. Morris is explaining to the Lords of Trade what he means, as opposed to what the New Jersey assembly means, by an "alteration in our currency." He uses New England as his example where no alteration in the currency occurred as he (Morris) understood it. Then he explains what an alteration in the currency was as he understood it where he was experiencing it (and in a falling-value direction not in a rising-value direction as Michener would have you believe). That he is referring to New Jersey bills and not New England bills in the rest of this paragraph is made clear from the antecedents earlier in this lengthy paragraph and then by what Morris subsequently talks about, and how he talked about it, immediately in the next paragraph. Michener is misleading the reader by throwing the phrase with "New England" in it just before the quotation I presented without giving the entire antecedent and subsequent material that reveals the true context of Morris's discussion. Michener implies that Morris is only talking about New England in this paragraph, which is patently untrue.

In fact, the end of the paragraph in question and the beginning of the next paragraph, namely what I presented, would be incoherent as Michener wants to present and interpret it. The end of the first paragraph quoted has paper money falling in value, whereas the start of the second paragraph, as Michener would have you interpret it, has paper money rising in value. This would make Morris's discussion of currency movements confused and incoherent. My interpretation makes Morris's discussion coherent and consistent.

Finally, Michener's addition to the end of my quotation again misleads the reader by not explaining what Morris is trying to explain to the Lords of Trade. Morris starts by setting out the assembly's explanation for the acute movement in the value of its paper money. The assembly thought that the demand for bills of exchange in New Jersey for exportation increased and that caused the value of paper money to fall, namely it took more paper money to buy a bill of exchange than previously. Because more people wanted to acquire and export bills of exchange, it bid up the paper money price of a bill of exchange. Bills of exchange are denominated in specie, and so paper money was falling in value relative to specie. Morris says, "They [the assembly] would not allow that there was any alteration in our currency, but that bills of exchange had got to a higher rate than they had been, and that the Exportation being Encreas'd, the course of Exchange had fallen to 50 pr cent, & that the Increase of the Exportation was the chief cause thereof" (Morris 1852, 133). Unless you think supply curves slope down and demand curves slope up (as Michener seems to), you cannot interpret the "fallen to 50 pr cent" as anything other than the value of paper money falling relative to

specie (bills of exchange).

Morris tries to offer a different explanation for the fall in the value of New Jersey paper money, and this is the "want of specie" which also means that it takes more paper money to buy a given amount of specie than it did previously, namely the value of paper money falls relative to specie (not rises as Michener thinks). Then Morris continues on to offer an incoherent set of conditions thereafter which is what Michener adds after where I stopped the quotation. Morris says that specie and paper money were both scarce and that no change in the export of bills of exchange occurred (what he goes on to say after Michener stops his quotation). You cannot explain the exchange rate movements Morris states in his letter with these three conditions only.

It is more coherent to interpret this portion of Morris's quotation as saying that the demand for supplies and troops in New Jersey were increased by war requisitions and so prices in New Jersey rose temporarily making the value of paper money relative to specie fall. You cannot have both specie and paper money being (equally) scarce and also have the exchange rate between paper money and specie change dramatically, unless you abandon basic microeconomic theory (something Michener seems to frequently do). Given that the assembly explicitly issued new paper money, it seems likely that specie was relatively scarcer than paper money was scarce, thus explaining the exchange rate movements stated in Morris's letter. Michener never addresses the logic or evaluates the possible coherent meanings in this portion of Morris's statement.

Michener's motivations

Michener (2019, 187–192) spills a lot of ink proclaiming his motivations for writing so many critical comments on my research. Michener (2020, 305) then complains that I did not address his motivation discussion in my reply (Grubb 2020a). Michener (2019, 187–192) claims that somehow my analysis of colonial paper money means you can no longer tell a coherent story about "the causes of the discontent that led to the American Revolution." He insists that my analysis of colonial paper money must be dismissed or, I guess, there would be no reason for the colonists to be discontented enough to take up arms against England. He is not exactly clear, however, how this would be the case. He tells a rambling disjointed narrative that makes almost no mention of the role of colonial paper money, and no mention at all of my analysis of colonial paper money, in terms of what contributed to the discontent that justified Revolution. His discussion makes up in obfuscation what it lacks in style.

As such, Michener's motivation claim is a sham. He does not explain how

my research is threatening to any reasons justifying the Revolution, let alone to the reasons he states. He just alludes to some mysterious threat without any analysis or investigation. In fact, there seems to be a total disconnect between Michener's discussion of the reasons justifying the Revolution and my analysis of colonial paper money, that is, no relationship at all.

As such, Michener's motivation discussion about what caused the discontent that justified the Revolution comes across not as any commentary on my research but only as a way for Michener to repeat and slip into the discourse his cherished myths that he has promulgated for over 30 years. Myths he just repeats over and over and over and over again without evaluating the totality of the evidence, without crafting any hypotheses, and without doing any hypothesis testing—he just asserts them as true. These myths include: (1) the colonies relied on specie as their domestic medium of exchange and money supply, (2) the current exchange rate in the marketplace between specie and paper bills of credit was somehow held constant, (3) the Currency Act of 1764 prevented new emissions of bills of credit, and (4) the bills of credit of each colony flowed willy-nilly across colonial borders (apparently without cost) so that colonial borders are monetarily meaningless. The evidence against these claims is substantial. Michener ignores that and simply repeats his assertions over and over again because he needs them to be true to support his personal monetary ideology. It is just another example of Michener trying to have the last word on a topic. That is all his motivation discussion is really up to.

Any reading of my canon of published research on colonial paper money reveals numerous colonial grievances against the British government for interfering in colonial monetary matters (most notably see Celia and Grubb 2016; Cutsail and Grubb 2017; 2019; Grubb 2015; 2016a; b; c; 2017; 2018a; 2019a; 2020b). In fact, my research identifies with more clarity and with more exactitude what those episodes were that generated colonial grievances, more so than anything Michener has ever identified regarding colonial paper money. Therefore, Michener's motives for his crusade against me lay elsewhere. And with seven published critical comments on my research, it is clearly just that—a crusade, and an ad hominem crusade at that.

That said, I am loath to talk about someone else's motives because you cannot see what is inside someone else's head. All I can say is: observe his behaviors rather than listen to his justifications, and infer his motives as best you can from those behaviors. If Michener's complete correspondence with scholars and editors on the topic of colonial money were laid out for all to see, his motives might appear quite different from what Michener claims they are. In the end, however, Michener's motives do not matter. All that matters are sound arguments and verifiable evidence, not bluff and bluster wrapped in bombastic rhetoric.

Conclusion

I am flattered that Michener (2020, 306, 328) thinks I am so beloved in the profession that scholars and journal editors at prestigious universities would willingly compromise their professional integrity to shield me from Michener's machinations. This reply, along with my prior six replies, reveals that I hardly need shielding from Michener's attacks.

I am also flattered that Michener (2020, 327) thinks that I am such an incredible genius that I could figure out in advance how to artfully manipulate numerous different data series and numerous different measurement methods and then get all those manipulations to magically align in such a way as to achieve a preordained outcome, namely to yield a coincidence between the time paths of APV and MEV in both levels and movement in several colonies (Cutsail and Grubb 2017; Grubb 2016a; 2018a; 2020b). In other words, only a genius could pull off such a grand hoax showing that the present value of bills of credit when hypothesized to be zero-coupon and interest-bearing bonds closely tracked the bills' observed current market trade values when, as Michener claims, they do not. If Michener is correct, what would it take to alter all these data series and measurement methods to eliminate that non-coincidence in levels and movement? How would you start and how could it be done, especially in such a way that no one other than Michener could ferret out? Michener must believe I am an amazing genius.

Alternatively, if all my "errors" as identified by Michener (2020) were done by accident, say randomly, what would be the probability that all these random errors could make APV and MEV coincide when according to Michener they do not in fact coincide? If you take all of the errors that Michener alleges I commit and say they could have gone this way or that, multiply together the probabilities of these random errors yielding the coincidence between APV and MEV (when supposedly none in fact existed), then the probability of me producing the coincidence of APV and MEV outcome that I did would be minuscule—a highly improbable outcome.

Of course there is another explanation, and that is I did a reasonable job of data construction, correction, measurement, and estimation, and while there is always room for error in such empirical work, the overall impact of any errors is small. The coincidence between APV and MEV in numerous colonies is not an accidental outcome of poor research methods or a pre-planned hoax on the profession, but it is in fact a real outcome uncovered by applying a new model of colonial monetary performance.

References

Celia, James, and Farley Grubb. 2016. Non-Legal-Tender Paper Money: The Structure and Performance of Maryland's Bills of Credit, 1767–1775. *Economic History Review* 69(4): 1132–1156.

Cutsail, Cory, and Farley Grubb. 2017. The Paper Money of Colonial North Carolina, 1712–1774. *NBER Working Paper* 23783. National Bureau of Economic Research (Cambridge, Mass.). **Link**

Cutsail, Cory, and Farley Grubb. 2019. The Paper Money of Colonial North Carolina, 1712–74: Reconstructing the Evidence. *Journal of the North Carolina Association of Historians* 27: 1–31.

Farrand, Max, ed. 1966. *The Records of the Federal Convention of 1787*, 4 vols. New Haven, Conn.: Yale University Press.

Grubb, Farley. 2005. State 'Currencies' and the Transition to the U.S. Dollar: Reply—Including a New View from Canada. *American Economic Review* 95(4): 1341–1348.

Grubb, Farley. 2006a. Theory, Evidence, and Belief—The Colonial Money Puzzle Revisited: Reply to Michener and Wright. *Econ Journal Watch* 3(1): 45–72. **Link**

Grubb, Farley. 2006b. Benjamin Franklin and Colonial Money: A Reply to Michener and Wright—Yet Again. *Econ Journal Watch* 3(3): 484–510. **Link**

Grubb, Farley. 2006c. The U.S. Constitution and Monetary Powers: An Analysis of the 1787 Constitutional Convention and the Constitutional Transformation of the U.S. Monetary System. *Financial History Review* 13(1): 43–71.

Grubb, Farley. 2015. Colonial New Jersey's Paper Money Regime, 1709–1775: A Forensic Accounting Reconstruction of the Data. *Historical Methods* 48(1): 13–34.

Grubb, Farley. 2016a. Colonial New Jersey Paper Money, 1709–1775: Value Decomposition and Performance. *Journal of Economic History* 76(4): 1216–1232.

Grubb, Farley. 2016b. Is Paper Money Just Paper Money? Experimentation and Variation in the Paper Monies Issued by the American Colonies from 1690 to 1775. *Research in Economic History* 32: 147–224.

Grubb, Farley. 2016c. Colonial New Jersey's Provincial Fiscal Structure, 1709–1775: Spending Obligations, Revenue Sources, and Tax Burdens During Peace and War. *Financial History Review* 23(2): 133–163.

Grubb, Farley. 2017. Colonial Virginia's Paper Money Regime, 1755–1774: A Forensic Accounting Reconstruction of the Data. *Historical Methods* 50(2): 96–112.

Grubb, Farley. 2018a. Colonial Virginia's Paper Money, 1755–1774: Value Decomposition and Performance. *Financial History Review* 25(2): 113–140.

Grubb, Farley. 2018b. Colonial Maryland's Post-1764 Paper Money: A Reply to Ron Michener. *Econ Journal Watch* 15(2): 168–178. **Link**

Grubb, Farley. 2019a. Creating Maryland's Paper Money Economy, 1720–1740: The Confluence of Political Constituencies, Economic Forces, Transatlantic Markets, and Law. *Journal of Early American History* 9(1): 34–58.

Grubb, Farley. 2019b. Colonial Virginia's Paper Money, 1755–1774: A Reply to Michener. *Financial History Review* 26(3): 401–408.

Grubb, Farley. 2020a. Colonial New Jersey's Paper Money, 1709–1775: Why Ronald

Michener Insists on Using Uncorrected Data—A Reply. *Econ Journal Watch* 17(1): 71–89. **Link**

Grubb, Farley. 2020b. Money and Prices in Colonial America. In *Handbook of the History of Money and Currency*, eds. Stefano Battilossi, Youssef Cassis, and Kazuhiko Yago, 431–454. Singapore: Springer.

Michener, Ronald W. 2019. Re-examination of the Empirical Evidence Concerning Colonial New Jersey's Paper Money, 1709–1775: A Comment on Farley Grubb. *Econ Journal Watch* 16(2): 180–217. **Link**

Michener, Ronald W. 2020. Re-examination of the Theoretical and Historical Evidence Concerning Colonial New Jersey's Paper Money, 1709–1775: A Further Comment on Grubb. *Econ Journal Watch* 17(2): 305–332. **Link**

Morris, Lewis. 1852. *The Papers of Lewis Morris, Governor of the Province of New Jersey, from 1738 to 1746*. New York: George P. Putnam.

Newman, Eric P. 2008. *The Early Paper Money of America*, 5th ed. Iola, Wis.: Krause.

Smith, Paul H., ed. 1976–1994. *Letters of Delegates to Congress, 1774–1789*, 21 vols. Washington, D.C.: Library of Congress.

United States. Continental Congress. 1904–1937. *Journals of the Continental Congress, 1774–1789*, 34 vols., eds. Worthington Chauncey Ford, Gaillard Hunt, John Clement Fitzpatrick, and Roscoe R. Hill. Washington, D.C.: U.S. Government Printing Office.

About the Author

Farley Grubb is professor of economics at the University of Delaware, NBER research associate, and Financial History Series editor for Routledge, Taylor & Francis Group. He earned his Ph.D. in economics at the University of Chicago. He took multiple graduate courses in economics from each of the following: Gary S. Becker, Robert W. Fogel, Robert E. Lucas, Douglass C. North, and George Stigler. He has published numerous articles in refereed journals and edited volumes on the economic history of colonial and Revolutionary-era America, in particular on contract labor (indentured servitude and convict labor) and on monetary institutions and their performance. His email address is grubbf@udel.edu.

Go to archive of Comments section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6029/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 95–126

Against Standard Deviation as a Quality Control Maxim in Anthropometry

Austin Sandler[1]

LINK TO ABSTRACT

Anthropometry is the study of the measurements and proportions of the human body. It is widely accepted that for practical purposes anthropometry is the most useful tool for assessing the malnutrition status of children (WHO 1986). Malnutrition is responsible for 45 percent of all deaths among children worldwide (Black et al. 2013). In 2017, acute malnutrition (wasting) menaced over 50 million young children while over 150 million young children suffered from chronic malnutrition (stunting) (UNICEF, WHO, and the World Bank 2018). Even a small change in child malnutrition rates can have major consequences in terms of lives saved or lost. The financial and human costs associated with the practice of anthropometry can be enormous. In 2014 alone, global donors disbursed nearly $937 million in nutrition-specific programing (KFF 2016). According to Meera Shekar et al. (2017), to achieve the World Health Assembly global nutrition targets, the world needs to invest $70 billion over 10 years in high-impact nutrition-specific interventions.

The two most widely studied expressions of anthropometric indices are weight-for-height (WHZ) and height-for-age (HAZ) z-scores (de Onis and Habicht 1996; UNICEF et al. 2018). These z-scores express anthropometric measurements in terms of standard deviations below or above a reference population value. A z-score is the difference between a particular child's measure-

1. Graduate student, University of Maryland, College Park, MD 20740. I wish to give thanks to Emily Dacquisto, Julie Silva, Laixiang Sun, and three anonymous referees. Their critical and insightful comments improved the paper greatly.

ments and the mean value of comparable children from a reference population, divided by the standard deviation of that reference population (WHO 1995). Z-scores require a well specified reference population with a normal distribution, a condition which would imply that z-score cutoff values for stunting, wasting, or underweight are stable across different reference populations.

However, many practitioners operate under the assumption that the standard deviation (SD) of a survey's anthropometric indices is a necessary and sufficient measurement for quality control (QC).[2] The practice is particularly persistent for anthropometric surveys within the field of childhood malnutrition, with particularly grievous consequences. In one typical article, the quality control maxim for z-scores states, "summary statistics can be compared with the reference, which has an expected mean Z-score of 0 and a SD of 1.0 for all normalized growth indices" (Mei and Grummer-Strawn 2007, 441). Others suggest that if a survey presents with "an excessive standard deviation…the survey results should be rejected" (Grellety and Golden 2016). The maxim is certainly simple, but does its simplicity compensate for its disadvantages?

Suppose you wish to conduct an anthropometric survey across the Karamoja region of northeast Uganda to assess the health of the region's children. Your well-designed survey includes measurements of height, weight, and age from a sample of children. You combine the measurements to make anthropometric indices of health such as weight-for-height and height-for-age. After performing some rudimentary summary analysis, you discover the sample standard deviations of the survey indices are (for example) 1.3 times greater than those of the 2006 World Health Organization (WHO) reference standards, which is not surprising given that the two groups of children come from two distinctly different populations. However, the quality control maxim used by many anthropometric researchers would dismiss your Karamoja survey as low quality, simply because the standard deviations are 1.3 times greater than the 2006 WHO reference standards.

Anthropometric research generally works with z-scores, however, and the practice that I am objecting to is expressed in terms of z-scores, not sample standard deviations. Couched in terms of z-scores, the nature of the putative quality control requirement is a bit harder to understand. But it is really as simple

2. Exactly how many is up for debate and a potential direction for future research. Suffice it to say the number is large. If one is unfamiliar with this particular body of literature or the day-to-day pragmatics of organizations working in this field, then the SD as QC problem might not seem endemic. But much like dust in the air, to borrow a metaphor, SD as QC seems invisible—even if you're choking on it—until you let the sun in. Then you see it's everywhere. A collection of quotes from this search is provided in Appendix A to help illuminate the extent, certainly representing only a small sample of all the potential articles and reports. Not to mention the many unreported, unknown, and unknowable studies that never saw the light of day because of internal or external suppression for having supposedly overly large standard deviations.

as the Karamoja example: when the ratio of standard deviations (of the sample and a reference) is in excess of a fixed threshold (e.g., 1.3) the study fails the quality control test. It can be shown that an anthropometric survey has a z-score standard deviation of 1.3 (or any other arbitrary cutoff value) *if and only if* the sample standard deviation of the anthropometric index is 1.3 times that of the standard deviation of the reference population. From a mathematical standpoint, a claim about the standard deviation of a z-score is equivalent to a claim about the ratio of an index's sample standard deviation to that of a reference population. For a proof, see Appendix B.

The notion that I wish to challenge is the following: Any anthropometric survey and subsequent z-score index (e.g., height-for-age or weight-for-height) not normally distributed with a standard deviation of approximately 1.0 (e.g., 1.3) indicates a serious problem and should be considered unusable.[3] And I suggest there is neither statistical justification nor scientific evidence that supports the SD as QC maxim.

There are, of course, inaccurate surveys that deserve to be dismissed. Garbage in, garbage out. Wariness is appropriate, but tests and conditions other than a standard deviation threshold must be applied. For example, the United States Agency for International Development (USAID) identify 26 potential indicators that could measure anthropometry data quality during fieldwork (Allen et al. 2019). WHO recommends considering several indicators such as population characteristics, sample size, survey design, measurement methods, and missing

3. Although the maxim is widely practiced, it is not always consistent. WHO suggests the z-score "distribution should be relatively constant and close to the expected value of 1.0 for the reference distribution" (1995, 218). De Onis and Blössner, citing WHO (1995), claim good quality SD ranges of HAZ (1.10 to 1.30), WAZ (1.00 to 1.20) and WHZ (0.85 to 1.10) and state these values are "the expected ranges of standard deviations of the z-score distributions for the three anthropometric indicators" (1997, 51). De Onis and Blössner also state that "[a]ny standard deviation of the z-scores above 1.3 suggests inaccurate data" (ibid.). Golden and Grellety suggest "The spread of the standard deviations…was small; ranging from 0.8 to 1.2 in 95% of the surveys" (2002, 5). Grellety and Golden, citing WHO (1995) and Golden and Grellety (2002), state "the SD for Weight-for-Height (WFH) should be between 0.8 and 1.2 Z-score units in all well-conducted surveys, with about 80% between 0.9 and 1.1Z" (2018, 2). Mei and Grummer-Strawn, citing WHO (1995), present the same example z-score table of HAZ (1.10 to 1.30), WAZ (1.00 to 1.20) and WHZ (0.85 to 1.10) and claim these values are a "recommendation from a WHO expert panel" as the "ranges for data quality assessment" (2007, 445). Mei and Grummer-Strawn (2007) also suggest the ranges for data quality assessment should be wider, given by HAZ (1.35 to 1.95), WAZ (1.17 to 1.46) and WHZ (1.08 to 1.50). We are told by USAID "that high quality anthropometric data should be normally distributed with a standard deviation of approximately 1" (2016, 15). But later USAID informs us that "very large standard deviations, for example greater than 2, might be a sign of poor quality" (ibid.). Bilukha et al., citing WHO (1995) and WHO and UNICEF (2019), give the recommendation that "Absent measurement error, distributions are expected to be approximately normal with a SD close to 1" (2020, 2). However, Bilukha et al. choose the exclusion criteria of "greater than 1.8 or lower than 0.8" (2020, 3).

data (WHO 1995). WHO and UNICEF (2019) suggest performing a seven-point data quality assessment, which interprets and reports: completeness; sex ratio; age heaping; height and weight digit preference; and z-score implausibility, standard deviations, skewness and kurtosis. And Nandita Perumal et al. (2020) have implemented this suggestion to its fullest potential.

Emmanuel Grellety and Michael H. Golden (2016) highlight random measurement, digit preference, and rounding error as potential sources of error. David A. Siegel and Jacob S. Swanson (2004) warn against heaping and digit preference. Researchers should also look out for confounding effects, specification error, non-linearity, bias of the auspices, measurement error, experimental error, and sample selection bias. Others point out that there is not even a consensus in the literature as to what constitutes a usable dataset (Crowe et al. 2014; Waterlow et al. 1977; USAID 2016). Shireen Assaf, Monica T. Kothari, and Thomas W. Pullum (2015) say the need for well-defined quality assessment criteria remains unmet, and they recommend more training and better equipment in the meantime.

In their methodological guidelines for assessing nutrition in crisis situations, the SMART (Standardized Monitoring and Assessment of Relief and Transitions) inter-agency initiative recognized that survey samples do *not* follow reference standards, and that even "the standard population is not normally distributed" (2006, 24 n.9). Later, however, the guidelines rely on the SD as QC maxim, claiming bias "can be estimated from examination of the standard deviation of the WFH, which should always be 0.8–1.2 z-scores" (ibid., 38).

Inspection of surveys for small SD remains in many QC recommendations (e.g., Allen et al. 2019; SMART 2006; WHO and UNICEF 2019) as a necessary if not sufficient condition for acceptance, while for others it is even a sufficient condition (e.g., Bilukha et al. 2020; Grellety and Golden 2016; 2018; Mei and Grummer-Strawn 2007). I propose that SD is neither a necessary nor sufficient indicator of QC. Low-quality surveys can have small SD, and high-quality surveys can have large SD. Errors of commission and omission waste precious resources that are already spread thin. The disregarding of surveys with high standard deviation could result in funds and research being syphoned away from the people most in need. It is my aim to illustrate the archival, statistical, logical, theoretical, and practical evidence that standard deviation should serve as neither a necessary nor a sufficient arbiter of quality control.

Unsound beginnings

It was sculptors and painters who first measured the relative proportions of the human form (Tanner 1981). Scientific study of the measurements of the

human body emerged notably with the work of Adolphe Quételet in 1832. Much like contemporary practitioners, Quételet performed a cross-sectional study of the height and weight of newborns and children, and observed a likeness between the distribution of weight and height to a normal (Gaussian) distribution (Quételet 1832; 1835). This Quételet Index, later redubbed Body Mass Index, is still relevant today. Unlike Quételet, however, contemporary practitioners have transposed his observation, and adopted the quality control practice of judging a survey based on its likeness to a standard normal distribution.

The source of the misconception originates in a presentation at the 15th International Congress of Nutrition in 1993 by Ray Yip. Despite its later impact on the literature, the SD as QC proposal does not even appear in the summary of the workshop, including Yip's abstract (Yip 1993). But two years later the WHO issued a technical report titled *Physical Status: The Use of and Interpretation of Anthropometry* that many have cited as the origin of and authority for the SD as QC maxim.

In less than one page of a 463-page report, some of the most recurrent maxims are found. WHO (1995) outlines several steps involved in assessing the quality of anthropometric data, including the observed standard deviation of the z-score distribution. With accurate measurements, the report claims, the "distribution should be relatively constant and close to the expected value of 1.0 for the reference distribution" (WHO 1995, 218). Citing the 1993 conference abstract, the report presents a table of "the standard deviations of the height-for-age, weight-for-age, and weight-for-height z-score distributions" all ranging "within approximately 0.2 units of the expected value" (WHO 1995, 218). The table of values include: HAZ (1.10 to 1.30), WAZ (1.00 to 1.20), and WHZ (0.85 to 1.10). The expected value of 1.0, the range of plus or minus 0.2 units, and the specific table values have all been widely cited as criteria by which to judge a survey's quality (e.g., Blanton and Bilukha 2013; Bilukha et al. 2020; de Onis and Blössner 1997; Grellety and Golden 2018; Mei and Grummer-Strawn 2007; SMART 2006; WHO and UNICEF 2019).

WHO (1995) presents the table of SD ranges only as an *example* that was observed during multiple large-scale Centers for Disease Control and Prevention (CDC) surveys presented once at a conference. The range of plus or minus approximately 0.2 units is merely a generalization they ascribe to the example surveys. In fact, WHO (1995) goes on to say that in some surveys the observed standard deviations ranged from 1.4 to 1.8, even after excluding extreme outliers. The specific SD values were *not* given in WHO (1995) as QC ranges as many have claimed (e.g., Grellety and Golden 2018; Gupta et al. 2020; Castro Bedriñana and Chirinos Peinado 2014; Kwena et al. 2003; Jacob et al. 2016; Mei and Grummer-Strawn 2007; Wijaya-Erhardt 2019).

The report does suggest a SD > 1 *could* be an indicator of inaccuracy, but the

notion was couched in a larger discussion of indicators, including validity of the reference population, the notorious quality of age estimates, errors of rounding and digit bias, number of missing and improbable values, and overall data compilation and documentation. Standard deviation is but one potential indicator, of many, to flag surveys for further inspection, *not* a sufficient measure of quality (WHO 1995). And the report recommends: "Verification of accuracy is best done by remeasurement of a sub-sample of the original sample by individuals who are fully qualified in anthropometric procedures" (WHO 1995, 216). In other words, standard normal SD is certainly not a sufficient QC condition.

Soon after, Mercedes de Onis and Monika Blössner (1997) echoed the SD as QC maxim as a definitive fact of nutrition surveys in their report *WHO Global Database on Child Growth and Malnutrition*, which many have cited as the source of the idea. In particular, de Onis and Blössner claim:

> If the surveyed standard deviation of the Z-score ranges between 1.1 and 1.2, the distribution of the sample has a wider spread than the reference. Any standard deviation of the Z-scores above 1.3 suggests inaccurate data due to measurement error or incorrect age reporting. (de Onis and Blössner 1997, 51)

The first sentence is referring to the *survey* data compared to the *reference* data. It is only making general statements about how variance and spread can be described for any two distributions of data. The second sentence, however, jumps to the conclusion that a z-score standard deviation above 1.3 "suggests inaccurate data."

Without question, z-score summary statistics can indicate community-wide malnutrition; that is their function. As de Onis and Blössner state earlier "if a condition is severe, an intervention is required for the entire community, not just those who are classified as 'malnourished' by the cut-off criteria" (1997, 50). That is to say, when analyzing z-scores, *if* many observed z-scores are well below the reference, *then* one might conclude that the appropriate intervention mechanism should be aimed at the population, and not the individual level. This is a sensible, if tautological, suggestion. But the inverse is not necessarily true. Namely, if you do not observe a standard normal distribution of z-scores shifted in mean only, then you conclude that none of the population has been affected and the sample is simply of low quality.

It seems obvious that a population by definition will *not* move together as a whole. We know that low-income families are more vulnerable to price volatility and uncertainty because they have fewer options, entitlements, and capabilities (Sen 1984). Calorie elasticity is not zero (Subramanian and Deaton 1996). Low-income families spend a large percentage of income on food, making them more vulnerable, thus skewing the distribution asymmetrically.

Larger z-score SD implies larger spread implies inaccurate data: simple but unsatisfying. I have not found substantiating evidence or theoretical justification for the maxim—in de Onis and Blössner (1997) in particular or the literature in general. But what I have found is a history of citations built upon a shaky foundation.

In my estimation there are really only two studies which one could argue have attempted to show evidence or justification for SD as QC, if only tangentially. The first comes from a conference paper presented at the Standardized Monitoring and Assessment of Relief and Transitions (SMART) Workshop, July 23–26, 2002. At the workshop Michael H. Golden and Yvonne Grellety presented a working paper in which they claim to disprove the assertion that "social heterogeneity would lead to changes in the shape of the distribution curve of acute malnutrition when a population is exposed to famine" (2002, 3). And through their analysis they conclude that "there was no change in the spread of wasting within the population as it became more malnourished" (ibid.).[4]

The findings of the Golden and Grellety (2002) working paper rest largely on Kolmogorov-Smirnov tests. In this case, the null hypothesis claim is that heterogeneity of wasting (i.e., z-score distribution curve) is heteroscedastic and the goal of the test is to falsify that claim. Their objective is to prove distributional spread (i.e., SD) is independent, stable, and standard normal (i.e., close to 1.0) as populations are exposed to starvation and famine (i.e., changes in average z-scores). And as an extension of their Kolmogorov-Smirnov test, they suggest SD is a measure of QC, stating:

> If a survey is observed to differ significantly from normality or have a large standard deviation, then we suggest that either two distinctly different populations may have been included in the sample or there is methodological error. All surveys should be checked for normality and any difference investigated. (Golden and Grellety 2002, 10)

But the specific Kolmogorov-Smirnov tests that Golden and Grellety (2002) devise assume the data are normally distributed from the start. In this case the null hypothesis is not heterogeneity, but that z-score distribution curves are in fact normal. Furthermore, Thomas Bayes (1763) taught that it is incorrect to assume $\Pr(\text{Data} \mid H_0) = \Pr(H_0 \mid \text{Data})$. And testing for normality is not equivalent to testing a unit SD. We are also not provided the power of the tests (i.e., the probability of correctly rejecting the null hypothesis), making it difficult for one to judge a null hypothesis false when it is false.

4. Emmanuel Grellety and Michael Golden (2018) stipulate that these findings confirm that SD should be between 0.8 and 1.2 z-score units in all well-conducted surveys.

Finally, in their figures, they purport that mean and standard deviation are uncorrelated. But if two random variables are statistically uncorrelated, that does not imply they are independent—yet it is independence that they seek. In addition, they show that kurtosis varies from −0.75 to 1.75 decreasing as wasting escalates, and skewness varies from −0.5 to 0.75 increasing as wasting escalates, contradicting the claim that malnutrition prevalence remains fixed and normally distributed.

In my estimation, even if Golden and Grellety (2002) had shown what they intended, it is still a great leap to conclude that therefore standard deviations are a necessary and sufficient quality control measure. The link is missing. Many alternative hypotheses still exist. As Deirdre N. McCloskey and Stephen T. Ziliak point out, "Failing to reject does not of course imply that the null is therefore true. And rejecting the null does not imply that the alternative hypothesis is true: there may be other alternatives which would cause rejection of the null" (1996, 102). And elsewhere, Golden concedes: "Most experimental studies do not include the acutely ill children for ethical reasons; the children are studied after they have recovered from acute infections and other major complications" (2009, S280). The esteemed pediatrician James Tanner knew in 1952 that unhealthy populations could be non-Gaussian and skewed; as such, standard deviations may be biased and not locate the right points (Tanner 1952).

The second study comes from an article by Zuguo Mei and Laurence M. Grummer-Strawn (2007). Mei and Grummer-Strawn claim to "assess whether the SD of height- and weight-based Z-score indicators derived from the 2006 WHO growth standards can still be used as data quality indicators," finding "the SD for all four indicators were independent of their respective mean Z-scores across countries" (Mei and Grummer-Strawn 2007, 441). They conclude that "the SD of Z-scores could still be used as a data quality indicator for evaluation of anthropometric data" (ibid., 445).

Again, WHO (1995, 218) presents a table of z-scores with different ranges of distribution values (i.e., HAZ 1.10 to 1.30, WAZ 1.00 to 1.20, and WHZ 0.85 to 1.10). However, as I hope I have illustrated, the table is presented only as an *example* of observed ranges. And the standard deviation z-score ranges were never meant for data quality assessment, nor has SD ever been shown to be a sufficient QC indicator.

But the point is lost in Mei and Grummer-Strawn (2007), who submit that WHO (1995) recommended "standard deviation ranges for data quality assessment" and claim to assess "whether these Z-score ranges still apply." I suggest they never did. Mei and Grummer-Strawn even concede that "the observed ranges of SD for all four indicators from our analysis were consistently wider than those recommended by WHO" (2007, 441). Yet these specific values were never given in WHO (1995) as the acceptable range for good quality surveys.

Citing WHO (1995), Mei and Grummer-Strawn assert that:

> On the basis of the 1978 WHO/National Center for Health Statistics (NCHS) growth reference, WHO has previously indicated that the SD of Z-scores of these indicators is reasonably constant across populations, irrespective of nutritional status, and thus can be used to assess the quality of anthropometric data. (Mei and Grummer-Strawn 2007, 441)

I think it is telling that they point to the 1995 technical report instead of pointing to John C. Waterlow et al., who were the actual developers of the WHO/National Center for Health Statistics (NCHS) growth reference[5] and who warned against universal principles: "Decisions of this kind have to be taken locally, and it is not possible to make international recommendations about them" (Waterlow et al. 1977, 491). Indeed, we need to make judgments backed up by logic, theory, and evidence, and not follow a binary decision rule that lacks contextual nuance. Waterlow et al. affirm that sub-populations are heterogenous, imploring us to make judgments on a case-by-case basis:

> Clearly, if there were differences dependent on different gene distributions, then the target for one population would not be the same as the target for another. … Because the reference population cannot be used as a universal target, the question of what is a realistic goal in any particular situation does become important. (Waterlow et al. 1977, 490)

The purpose of Waterlow et al. was to "present recommendations for the analysis and presentation of height and weight data" (1977, 489), *not* to present ways to exclude such data. All constraints that Waterlow et al. do propose are wholly directed at constructing a *reference* population. Whereas a *standard* represents a desirable target or norm, the sole aim of a *reference* is to be a common basis in order to group, analyze, and compare different populations (WHO 1995). Unfortunately, the distinction between references and standards was, and continues to be,

5. In 1971, as part of a long tradition for child growth references, the Maternal and Child Health Program, the United States Public Health Service, and the American Academy of Pediatrics concurred that more rigorous standards were needed for clinical characteristics of early childhood malnutrition. This decision was the impetus for the Health and Nutrition Examination Survey carried out by the Centers for Disease Control and Prevention's National Center for Health Statistics Task Force. First released in 1977, the National Center for Health Statistics Growth Curves were a combination of data from the National Center for Health Statistics' Health Examination Surveys, the Health and Nutrition Examination Survey, and the Fels Research Institute. Wanting in on the action, a WHO working group on nutritional surveillance made recommendations on the criteria for the anthropometric reference population and presented recommendations for the analysis of data from surveys involving nutrition and anthropometry, thus the "WHO/National Center for Health Statistics" growth reference.

indifferently heeded and left in unclarity.

The 1978 WHO/NCHS growth *reference* is distinct in its purpose and function from the 2006 WHO Multicentre Growth Reference Study (MGRS) growth *standards*. And neither can inform, through comparing standard deviations, whether or not any particular *sample* is of poor quality. But Mei and Grummer-Strawn assert that, "our analysis confirms the WHO assertion that the SD remains in a relatively small range for each indicator" (2007, 445). To do so, however, is to conflate *standards*, *references*, and *samples*.

In 1993, the Expert Committee on Physical Status, convened by WHO, concluded that previous *reference* growth charts had long been misconstrued as a *standard* for growth (de Onis and Habicht 1996). As a result, the WHO Multicentre Growth Reference Study was implemented between 1997 and 2003. The designers of the new Growth Reference were intentionally *prescriptive* rather than *descriptive* (Garza and de Onis 2004). They designed a growth chart for how children *should* grow rather than how children *actually* grow. In other words, it was purposely designed to produce an idealized *standard* rather than a baseline *reference*.

Even the initial sample data for the Multicentre Growth Reference Study did not have small and well-behaved standard deviations. To produce the growth standards, the sample was manipulated to fit specific distributional requirements (WHO 2006). And even though the MGRS sought out the healthiest, most ideal population to measure, 93 percent to 69 percent of the healthy populace were ineligible and did not conform to this ideal.[6] In other words, even in the healthiest

6. The Multicentre Growth Reference Study (July 1997–December 2003) consists of both cross-sectional and longitudinal surveys from six cities: Davis, California, USA; Muscat, Oman; Oslo, Norway; Pelotas, Brazil; in select affluent neighborhoods in Accra, Ghana; and South Delhi, India (WHO 2006). The distributions of children across the different survey countries for the longitudinal component are: 119 USA; 149 Oman; 148 Norway; 66 Brazil; 227 Ghana; and 173 India. The distributions of children across the different survey countries for the cross-sectional component are: 476 USA; 1,438 Oman; 1,385 Norway; 480 Brazil; 1,403 Ghana; and 1,487 India. Prior to constructing the standards, if a child was 3 SDs above the sample median or 3 standard deviations below the sample median they were excluded. For the cross-sectional sample the truncation procedure was even stricter. If a child was 2 SDs above the sample median or 2 SDs below the sample median they were excluded. Children were selected for inclusion based on: no known health or environmental constraints to growth, mothers willing to follow feeding recommendations, no maternal smoking before and after delivery, single term birth, and absence of significant morbidity. Of the 13,741 children screened for the longitudinal survey, less than 7 percent or 882 children (428 boys and 454 girls) were eligible and included in the final study. In addition, of the 21,520 children screened for the cross-sectional survey, less than 31 percent or 6,669 children (3,450 boys and 3,219 girls) were eligible, and included in the final study. In other words, 69 to 93 percent of the populace did not fit the standard. After selective sampling and exclusion, the sample was exceedingly skewed to the right (WHO 2006). To rectify the non-normality, the data were cleaved at the median, and then reflected to create two symmetrical distributions. Each mirrored distribution was used to derive standard deviation cut-off values (e.g., what is the severe wasting cutoff value where a WHZ score is less than 3 SDs from the median) for the respective upper and lower portions of the data.

and most ideal sub-populations, most children do not fit the growth standards, nor are they normally distributed with standard deviations close to one. The MGRS provided a growth standard intended for measuring benchmark distances from an idealized healthy child. It is not the only permittable distribution for a sample dataset nor is it relevant for measuring data quality.

Spurious theory and flawed logic

SD as QC may be believed by some to be loosely related to the seminal concepts of the eminent epidemiologist Geoffrey Rose, whose ideas transformed the strategy of preventive medicine. Central to Rose's analysis was an assumption that the width of the distribution of a variety of biological measures remains similar across different populations even as the mean of the distribution shifts: a mean-centric view of population (Rose 1992). He observed that most risk-factor distributions across populations appear to have uniform displacements, with risk changing the same amount at different parts of the risk-factor distributions. Rose's assumption implies that the mean of a distribution can be used as a proxy for a population's intrinsic traits.

But it is an untenable leap to go from Rose's notion that distributions of biological measures tend to have consistent spread, independent from the central tendency, to the misconception that any distribution of a biological measure that does not have a small and precise spread is invalid, inaccurate, and not insightful. Furthermore, Rose's conceptualization is anchored on the cohesiveness of populations, an assumption that may be violated by differential changes in the BMI distribution occurring globally within populations (Razak, Davey Smith, and Subramanian 2016).

Contrary to theoretical and observational expectations, some have claimed whole population distributions shift equally in the face of malnutrition stressors and that any data set which does not behave that way (i.e., any data set with z-score standard errors not equal to one) must be a low-quality survey (e.g., Blanton and Bilukha 2013; Bilukha et al. 2020; de Onis and Blössner 1997; Golden and Grellety 2002; Grellety and Golden 2016; Grellety and Golden 2018; Mei and Grummer-Strawn 2007). But the assertion remains unsubstantiated. If true, it would follow that whenever there was a famine (malnutrition stressor) anywhere in the world, you sitting at the breakfast table, drinking your coffee, oblivious to the famine, would also become slightly malnourished, too, to maintain a normally distributed population with a standard deviation of one. We all must move together to preserve the spread of the distribution, you see. Now, presumably, the SD as QC crowd would say that interpretation is preposterous, and that mean shifts in z-scores

do not occur for the entire planet but are only applicable to some smaller sub-population. Ah, then, by 'shifts in the population,' they don't really mean the Population. Okay, but they still have to contend with the problems of *sorites* and the fallacy of the transposed conditional (on that fallacy, see **here** or Appendix C).

If the effect is only valid for some sub-population then the boundaries of that sub-population must be defined, and the sub-population is by definition not representative of the whole population. So we should not be casual in talking of 'populations.' The meaningfulness of descriptive statistics depends on how meaningfully a population is defined in relation to the inherent intrinsic and extrinsic dynamic generative relationships by which they are constituted (Krieger 2012).

Prevalence and distributions of z-scores are therefore highly reliant on boundary definitions and cannot be extrapolated out of sample. Remember, too, that the 'reference population' used for judging a child's health is really a *standard* and by design a small sub-population of only the healthiest of healthy children. And, even still, those 'standard' children were not distributed standard normal with an SD of one (WHO 2006). There is no reason to believe that a healthy sub-population should behave the same way a malnourished sub-population does.

Standard deviation is merely the measure of dispersion for a set of values, unlike digit preference (heaping at 0 and 5), incompleteness (missing values), rounding errors (chop vs. nearest), data formatting (short, long, float, double), transposition and transcription errors (obvious typos), or procedural errors (e.g., a child measured lying down when they should have been standing), which are all direct quality control metrics of a specific error. For example, the standard deviation of WHZ only gauges the ratio of the weight-for-height sample standard deviation to that of the weight-for-height standard deviation of a reference population. The reference population (even if it is a *standard*) cannot signify anything qualitative about the sample data, nor should it. A reference population is merely a datum or a fixed point. It is a quantitative scale not a qualitative apparatus.

Measurement errors *might* generate inflated SD. Then again, they might not. Inflated SD does not necessarily imply measurement error (Biehl et al. 2013; Ulijaszek and Kerr 1999). The quality control maxim poses the prior "if the population is distributed normal, then the observed data will be distributed normal," and supposes wrongly "if data is observed, then the population it is drawn from is distributed normal." If *H*, then *O*, does not affirm if *O*, then *H*. It is the same as thinking if a person is hanged, then he will probably die; therefore, if observing a corpse, then one should conclude he was probably hanged (Ziliak and McCloskey 2008, 17).

Random errors lower precision by inflating confidence intervals. Random error is but one of many dozens of errors and seldom the biggest (Ziliak and

McCloskey 2008). It is systematic errors that we should be worried about. They cause bias. Especially when the costs of failure (i.e., child mortality) are high, the choice between low bias or low precision is not really a choice at all. If I can't be precisely right, I would rather be generally right than precisely wrong. More importantly, Ziliak and McCloskey note "sampling precision says nothing about the oomph of a variable or model" (2008, 25).

Systematic errors may even attenuate SD. A small spread in SD is not a necessary condition for a lack of systematic error, making SD a poor metric from which to judge quality. Suppose, for example, I performed an especially erroneous survey of child anthropometry in which instead of actually measuring different weights and heights, I just marked down the exact same value for every survey participant. Is my systematic measurement error captured by an inflated standard deviation? No. Obviously, this is an extreme and absurd example. But there exists a non-zero proportion of the total sample space in which systematic errors diminish rather than inflate standard deviation. Try to imagine the countless number of possible surveys with less extreme systematic error structures, all of which exhibit 'a standard deviation of approximately one.' If it is systematic errors that we are concerned with, SD signifies very little.

The obverse problem with SD as QC remains, too. Since Anscombe's quartet and the more recent Datasaurus Dozen, students of statistics have long known that different datasets with wildly varying graphical distributions can all have the exact same descriptive statistics, including standard deviation (Anscombe 1973; Matejka and Fitzmaurice 2017). Logic dictates SD is neither a necessary nor sufficient indicator of QC.

Informed dissent from the maxim

The debate surrounding standard deviation as a quality control metric is ongoing and unresolved. After two national nutrition surveys in Nigeria exhibited divergent estimates, both USAID and United Nations Children's Fund (UNICEF) staff in-country felt that substantial quality problems must exist in either one or both surveys (USAID 2016). In July 2015, the USAID Nutrition Division convened a technical meeting aimed at resolving the issues of accuracy and comparability of anthropometric data. Participants included representatives from USAID, CDC, UNICEF, WHO, the Pan American Health Organization, and external nutrition experts. The meeting report highlights that the importance of standard deviations for measuring data quality was a major point of contention. The report concludes that "there was no agreement on what is a reasonable standard deviation of z-scores to expect in heterogeneous populations" (USAID

2016, 17).

The meeting report features arguments for the SD as QC maxim given by an unspecified presenter from the CDC. In reference to the Demographic and Health Surveys, the CDC presenter asserted that high-quality anthropometric data will *always* be normally distributed with a standard deviation of approximately one regardless of population heterogeneity, and that a standard deviation greater than one *must* mean the data are of poor quality (USAID 2016, 16). One example they pointed to was the National Health and Nutrition Examination Survey in the United States with a (recent) stable trend of small standard deviations. Furthermore, they claimed the shape of the distribution does not change as a population becomes more malnourished, concluding there is no relationship between the mean z-score and standard deviation. In their estimation, this lack of relationship is sufficient to conclude standard deviation is a quality control metric.

The report suggests, however, that not all participants agreed with the SD as QC maxim. Some participants felt that standard deviations greater than one could reflect heterogeneity in the population. For the Demographic and Health Surveys in particular, they expressed concern regarding the emphasis on standard deviations of height-for-age, weight-for-age, and weight-for-height z-scores close to one as an indication of quality. The report details that other participants noted:

> In Kano state, Nigeria, for example, a majority of the within-cluster standard deviations were below 1, however, the average standard deviation in Kano state was more than 1. If the states are different, it is impossible for the standard deviation to be 1 in every state, and 1 for the country as a whole. (USAID 2016, 17)

Other researchers acknowledged that the Demographic and Health Surveys in particular did show the most variability in parameters such as standard deviation. But they also noted that the Demographic and Health Surveys Program has the largest number of surveys and covers the largest span of time; standard deviations may have changed with time as nutritional status of the populations changed or improved. One meeting facilitator affirmed that it is *not* true that the shape of the distribution does not change as nutritional status of the population changes. While others pointed out that in terms of the factors that influence anthropometric indicators (e.g., water, sanitation, and food security), the United States may be more homogeneous than other countries (e.g., India) (ibid., 16).

Given that standard deviations capture inherent population heterogeneity, there is no reason to assume that the standard deviation will be the same across all surveys. It is true that poor data quality could inflate the standard deviation of anthropometric measures, but given that anthropometric z-scores are biologic parameters, one would anticipate some population heterogeneity both within and

between countries, even in situations of high-quality data collection.

The Joint FAO/WHO Expert Committee on Nutrition (1971) noted that statistical evaluation cannot by itself distinguish between what is normal and abnormal in the biological sense. Even seminal author and pediatric expert Dr. Derrick Jelliffe (1966) emphasized the problems and difficulties of non-sampling errors, which cannot be detected with tests of sampling errors. And Jonathan Gorstein et al. (1994) noted that when the nature of a nutrition problem is unclear, it should be interpreted within the situational context.

Standard deviation is not indicative of quality control for some studies. There are researchers and journals confident enough in the quality of their findings even with standard deviations *not* approximately one. Yirgu Fekadu et al. (2015) found z-score standard deviations of 1.3 (weight-for-height), 1.33 (height-for-age), and 1.06 (weight-for-age) in Ethiopian children. Michel Garenne et al. (2009) found weight-for-height z-score standard deviations of 1.28 and 1.398 for Niakhar, Senegal, and Bwamanda, D. R. Congo, respectively. Afework Mulugeta et al. (2010) observed z-score standard deviations of 1.8 (height-for-age), 1.3 (weight-for-age), and 1.3 (weight-for-height) for children in northern Ethiopia.

In addition, Paul B. Spiegel et al. (2004) performed a meta-analytical quality assessment of anthropometric surveys with no mention of standard deviation. Daniel E. Roth et al. (2017) estimated that across 64 low- and middle-income countries, when mean height-for-age z-scores were zero, the standard deviation was 2.10 (95% CI 2.00 to 2.20), far above most QC thresholds. Examining mid-upper arm circumference (MUAC) for 852 cross-sectional nutritional surveys of children, Severine Frison et al. (2016) found that only 319, or 37.7 percent, follow a normal distribution.

In his survey of famines and economics, Martin Ravallion remarks on the unusual nature of malnourished communities: "I will say that a geographic area experiences famine when unusually high mortality risk is associated with an unusually severe threat to the food consumption of at least some people in the area" (1997, 1205). The phenomenon of malnutrition is by its very nature unusual, i.e., not normal. It would be bizarre to think that measures would behave the same in lean times as in abundance. In their appraisal of different anthropometric indices, André Briend et al. get to the heart of the matter when they observe "for most populations, little information is available on the amount of nutritional change one has to expect in a community and also on the standard deviations of some nutritional indices" (1989, 770).

Eschew the maxim

The SD as QC maxim is built on a history of shaky citations, corroborated with imprudent tests, substantiated by logical fallacies, and endorsed inconsistently by empiricists. It lacks archival, statistical, logical, theoretical, and practical merit. Of course, there are inaccurate surveys and samples that don't deserve our consideration, but other tests and conditions must be adopted.

Once the SD as QC maxim is abandoned, the therapeutic and ameliorative next step is more difficult. But good science is difficult. If it were easy, it would have already been done (Wasserstein, Schirm, and Lazar 2019). Good science embraces the explicable and ineffable (McCloskey 1994). Doing serious scientific inquiries calls for serious thinking about what makes a dataset 'good' or 'bad' and how its 'goodness' may impact the results. We need to consider the dozens of sources of *real* error, and reckon their effects on our results. As Ziliak and McCloskey put it, "After all, reconciling differences of effect, finding the common ground, is the point of statistics. … Most important is to minimize Error of the Third Kind, 'the error of undue inattention'" (2008, 246).

Appendix A
SD as QC in the literature

The practice of SD as QC is pervasive, almost to the point of being a norm or a given first principle of the field were citation and evidence are not required. And I believe that the SD as QC maxim is preventing more studies and surveys from being used and published. In Google Scholar, Mei and Grummer-Strawn (2007) are cited over 170 times, not to mention the over 8,950 articles citing WHO (1995) or the 760 citing de Onis and Blössner (1997). Clearly not all are relevant to the SD as QC discussion. To help illustrate the point I spent an afternoon tracking down articles that explicitly and openly abide by the SD as QC maxim in some form or another. Below are excerpts from a sample of 32 articles citing Mei and Grummer-Strawn (2007) where authors point to the SD as QC maxim. I have put some words in boldface for emphasis.

"Researchers also have analyzed ways in which use of the WHO standards might affect prevalences of wasting, stunting, and underweight worldwide, as well as **the distribution of z scores, a commonly used indicator of data quality** in international surveys" (Grummer-Strawn, Reinold, and Krebs 2010, 13).

"Accepted best practices for field-level quality control were followed. Systematic repeat data entries were done for all anthropometric data. Postanalysis **quality checks compared SDs of anthropometric data by site to WHO standards** and other studies for children <2 y of age" (Remans et al. 2011, 1636).

"There were another 5,010 children whose length-for-age z-scores (LAZs) were flagged in the DHS data files either as missing or as biologically implausible according to the WHO flags (Mei & Grummer-Strawn, 2007). **These children were excluded from the analysis**. We also removed 71 children whose mothers had a height of less than 130 cm, as these were considered to be implausible and likely due to measurement or recording errors" (Krasevec et al. 2017, 2).

"*z* score **SDs were within the valid range** accepted by the World Health Organization (WHO)" (Corvalán et al. 2009, 548).

"Summary statistics showed that **standard deviations** of the three indices Z score (weight for age, height for age and weight for height) were **between 0.92 and 1.03, indicating high quality** data" (El Mouzan et al. 2008, 339).

"The data were subjected to post-hoc methods of quality determination, and, if of suitable quality, included in the adequacy evaluation. … Accepted practices for field-level quality control were followed. However, systematic repeat measures, repeat sampling and inter-lab sampling were not available for quality control of the MICAH data. Therefore alternative, post-hoc methods were used for evaluating the quality of data collected. Some of these methods have been used previously, whereas others were developed for the purpose of this evaluation. … Comparison of magnitude of SDs of continuous variables to SDs in other, well-controlled studies… This method of **comparing SDs with reference populations has been recommended for anthropometrics**. We assume that common levels of variations will exist for other variables. … SDs of continuous variables in MICAH surveys in baseline (1996 or 1997), follow-up (2000) and endline (2004) compared with examples from the literature, for quality control purposes" (Berti et al. 2010, 613, 617, 618).

"In the analysis, plausibility of anthropometric Z scores were checked using the WHO protocol recommendations (2006), which provide **standard deviation cut points for anthropometric Z-scores as a data quality assessment tool**" (Abate and Belachew 2017, 6).

"Mei and colleagues previously reported a lack of a relationship between SD and mean HAZ across DHS surveys; however, they did not quantitatively

assess the change in SD with the age-related decline in mean HAZ, and **they interpreted their findings only as a justification for using SD as an indicator of anthropometric survey quality**" (Roth et al. 2017, e1255).

"Mei and Grummer-Strawn [2007] supported the use of **SD as a quality indicator** for anthropometric data" (Afifi et al. 2012, 2655).

"In our opinion reports from **surveys with an SD of more than 1.2 are unreliable**. … An analysis of DHS and MICS shows **elevated SD values with all of the mean SDs outside the acceptable range; none of mean SDs for any of the surveys was less than 1.0Z**. In agreement with the data from West Africa, the 5th and 95th centiles of the SDs of 51 recent DHS surveys were HAZ 1.35–1.95; WAZ 1.17–1.46, and WHZ 1.08–1.50. Mei & Grummer-Strawn conclude that they 'concur with **the WHO assertion that SD is in a relatively small range**'" (Grellety and Golden 2016, 19).

"Before turning to multivariate regressions, we relate our results to two indicators of measurement error used in previous work. The first step is to compare our December–January gap with the SD of HAZ. The SD of HAZ could reflect genuine dispersion related to health inequality but is **widely used as an indicator of survey errors** in both height and age (Assaf et al. 2015; Mei and Grummer-Strawn 2007)" (Larsen et al. 2019, 716–717).

"**Standard recommendations state that a standard deviation of greater than 1.3 for HAZ reflects excessive random variation** in either height measurements or age estimates. The standard deviation of HAZ in the three DHS greatly exceeds this threshold for data quality; however, this recommendation is based on the use of the old NCHS:CDC:WHO reference population. There is evidence that standard deviations for HAZ greater than 1.3 are common in DHS in other countries and may be normal when using the WHO Child Growth Standard (Mei & Grummer-Strawn 2007)" (Woodruff et al. 2017, 15).

"Many DHS surveys have **standard deviations greatly exceeding the quality criteria** defined by the World Health Organization. … Ranges are then used to describe the overall quality of the survey and arbitrary cut-offs are used to decide whether the data are acceptable or not" (Tuffrey and Hall 2016, 4–5, 14).

"We calculated z-score standard deviations (SD) and analyzed SD disaggregated by age (under and over two years of age) **to determine if the quality of measurements** differed by age. … We can consider z-score standard deviation to illustrate the importance of reaching consensus on interpretation and action. WHO

and the US CDC promote the use of normative ranges of SD to determine if survey quality is acceptable, but the ranges are based on surveys that have evidence of poor data quality. The most recent DHS data quality assessment showed that 30 of 52 countries had HAZ SD greater than 1.5, but only one country suppressed data because of poor quality. According to SMART **data quality is not acceptable if HAZ SD is above 1.2**, and a recent modeling study showed that SD of 1.5 can result in substantial overestimation of stunting prevalence. Meanwhile, the published normative range for HAZ SD that some organizations use to deem data quality acceptable is 1.35–1.95" (Conkle et al. 2017, 5, 10).

"Few studies have assessed the distribution of WFH. Two looked at the standard deviations of the WFH distributions. In 1977, **Waterlow et al. showed that the WFH distributions were skewed at the upper centiles**. Their analysis was performed on data from surveillance or surveys involving nutrition and anthropometry in young children up to the age of 10 years. In 2006, Mei et al. analysed data from 51 DHS surveys representing 34 developing Countries. They found a mean WFH and SD WFH (z-scores) of 0.06 and 1.40 respectively. The mean ranged from −0.91 to 0.83 and the SD range [*sic*] from 1.03 to 1.55. **They concluded that their analysis confirms the WHO assertion that the SD remains in a relatively small range (i.e. close to SD from a standard normal\ distribution)**, no matter the *Z*-score mean although the observed range of SD for was [*sic*] consistently wider" (Frison et al. 2016, 7).

"Summary statistics showed SDs of the 3 indices' *Z* score (weight for age, height for age, and weight for height) between 0.92 and 1.03, **indicating high-quality data**" (El Mouzan et al. 2009, 68).

"Previous research has demonstrated that Z-scores within a population are normally distributed with a SD of approximately 1.0; the shape of the distribution does not vary based on the nutritional status of the population, as measured by the mean Z-score. Based on the finding that SD remains in a relatively narrow range for each indicator regardless of mean Z-score, **WHO guidance recommends that the SD of Z-scores can be used as a data quality indicator** as well as a measure of variability. The introduction of random non-directional errors, such as those introduced when age is estimated rather than calculated or when teams are imprecise in measuring height or weight, can result in wider SD relative to the acceptable ranges outlined by WHO. … We therefore included **SD of the Z-scores to assess the degree to which data quality** in addition to variability impact DEFF in anthropometric surveys. … The SD of WHZ and WAZ were approximately 1.00, as expected in high-quality anthropometry surveys (WHZ

median = 1.03, WAZ median = 1.04)" (Hulland et al. 2016, 2–3, 10).

"Anthropometry **data quality indicators were extremely high (median SDs** for weight-for-length, length-for-age and weight-for-age z-scores 1.01, 0.98, and 1.03, respectively), likely due to extensive training, standardization, and monitoring efforts. … Anthropometry data quality indicators were monitored throughout the study. The medians of monthly standard deviations for weight-for-length, length-for-age, and weight-for-age z-scores were 1.01, 0.98, and 1.03, respectively; **close to the expected value of 1.0 for a reference distribution**. Standard deviations for z-scores varied month-to-month, but never reached the WHO thresholds for measurement error or incorrect age reporting" (Aceituno et al. 2017, 2, 8).

"The standard deviations reported in this study are much lower than the **suggested standard deviations** reported by Mei and Grummer-Strawn estimations in a cross-country analysis" (Sharma et al. 2020, 17).

"We also examined the quality of the 2009 data by **assessing the SD as a quality indicator** for anthropometric data (Mei and Grummer-Strawn 2007) and examining whether or not age heaping was evident. These assessments did not reveal any concerns" (Boylan et al. 2017, 2261).

"Based on the WHO Technical Report, the **SD for Weight-for-Height (WFH) should be between 0.8 and 1.2 Z-score units in all well-conducted surveys**. This has been confirmed empirically with well conducted surveys in both the developed world where large national surveys of heterogeneous populations have been conducted, for example the National Health and Nutrition Examination Survey (NHANES) from USA's National Centre for Health Statistics (NCHS) and the developing world. … The SD of organisation "t" differs significantly from the others (Student's t test < 0.0001), with 69% (53/77) of their surveys for WHZ having **an SD of more than 1.2 Z**. … For most anthropometric measurements the SD from single surveys should lie between 0.8 and 1.2, with about 80% between 0.9 and 1.1Z. For these reasons the **SD has been used as a useful measurement of data quality**" (Grellety and Golden 2018, 2, 3, 10).

"The median SD and range for HAZ were greater overall and across all surveys than for WHZ. The absolute difference in HAZ by MOB of age reporting should be close to 0 if there is no systematic error in age reporting, but was 0.25 (in *z* score units) overall and up to 0.90 in Timor-Leste in 2009. … HAZ SD and WHZ SD had the highest factor loadings in the data quality indices indicating that **SD is an important measure of anthropometric data quality**" (Perumal et al. 2020, 809S, 812S).

"Absent measurement error, distributions are expected to be approximately normal with a SD close to 1. … To exclude surveys with exceptionally poor anthropometry data quality or where data manipulation might be suspected, **we excluded from analysis surveys where the SD for WHZ, WAZ, HAZ, or BMIZ was outside of the following empirically defined cutoffs**: greater than 1.8 or lower than 0.8; or the SD for MUACZ greater than 1.8 and less than 0.7" (Bilukha et al. 2020, 2, 3).

"Anthropometric data collected during the 2008 to 2009 and 2014 Kenya surveys were reanalyzed to assess standard parameters of quality: standard deviation, skewness, and kurtosis of z-score values for 3 anthropometric indicators (weight for height, height for age, and weight for age)… The primary objective of the comparative analysis was to observe the quality of anthropometric variables. The **first metric of quality, standard deviation**, is presented in Table 3. … One key measure is SD of the continuous *z*-score distributions. As noted, previous research suggests that for a given population, Z-scores are normally distributed with an SD of approximately 1.0" (Leidman et al. 2018, 406, 412, 414).

"Careful interpretation is required, as the **standard deviations** for our anthropometric measurements are outside the World Health Organization range **for data quality assessment purposes**" (Bennett et al. 2020, 2038).

"Note that the standard deviations (SD) of WHZ and MUACZ in all rounds are near or even below 1.0, which gives us **confidence in the quality of the anthropometric data** (Grellety and Golden 2016b; Mei and Grummer-Strawn 2007). The average SD—across all four survey rounds—is 1.03 for WHZ and 0.95 for MUACZ" (Ecker et al. 2019, 10).

"Seventeen surveys had large standard deviations (SD) for HAZ, which could result in attenuated regression coefficients when HAZ was used as an explanatory variable in regression analyses. To avoid attenuation, HAZ values for each child were **adjusted to obtain a standard deviation for HAZ of 1.2** for each of these surveys by subtracting the survey mean for HAZ, dividing by the survey SD for HAZ, multiplying by 1.2, and then adding back the survey mean for HAZ" (Frongillo et al. 2017, 3038).

"The World Health Organization (WHO) has recommended the use of Z-score of these indicators to classify nutritional status, given the constancy of their values, independent of nutritional status, and can even be used as **indicators of the quality of anthropometric data**" (Martins et al. 2010, 1106).

"Z-score plausibility was determined using WHO cutoffs. We used the following WHO-defined **standard deviation (SD) ranges to assess the quality** of data (HAZ 1.1–1.3, WAZ 1.0–1.2, and WHZ 0.85–1.1)" (Gupta et al. 2020, 2–3).

"…as per WHO standards. Some individuals may have met >1 **exclusion criterion**" (Varghese and Stein 2019, 1208).

"Protocol used for obtaining data was an adaptation of that published by Lapham et al. and Mei et al." (Samiak and Emeto 2017, 2).

"Studies investigating the quality of the DHS data report the quality to be good (Mei Z and Grummer-Strawn LM., 2007, Mishra et al., 2006)" (Reda and Lindstrom 2014, 1160).

Appendix B
Z-score SD Proof

The aim here is to move away from the convoluted discussion of z-scores and standard deviations of z-scores to simply anthropometric index measurements and standard deviations of anthropometric index measurements. To make this simplification I will show that a z-score standard deviation is equivalent to the ratio of standard deviations of an anthropometric index to that of the reference population. The standard deviation of a given survey's anthropometric index is calculated as:

$$s_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}$$

where:

- s_x: anthropometric index sample standard deviation
- N: is the number of children in the sample
- x_i: is a child's anthropometric index value (e.g., weight-for-height)
- $\overline{x}$: is the anthropometric index sample average given by: $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

A z-score tells you how many standard deviations away an individual data value falls from the mean. It is calculated as:

$$Z_i = \frac{(x_i - \mu)}{\sigma}$$

where:

- Z_i: is a child's z-score
- x_i: is a child's anthropometric index value (e.g., weight-for-height)
- μ: is the reference mean
- σ: is the reference standard deviation

A given survey's z-score standard deviation is calculated as:

$$s_Z = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2}$$

where:

- s_Z: z-score sample standard deviation
- N: is the number of children in the sample
- Z_i: is a child's z-score
- $\overline{Z}$: sample average z-score given by $\overline{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$

Thus, we are left with the question: Is the statement, *if an anthropometric survey has a z-score standard deviation greater than 1.3 it fails the test*, equivalent the statement, *if the sample standard deviation of an anthropometric index is 1.3 times that of the standard deviation of the reference population it fails the test*? Or in other words, is the ratio of the sample standard deviation of (weight-for-height) to the reference population standard deviation of (weight-for-height) equivalent to the standard deviation of (weight-for-height) z-scores.

Claim:

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2} = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}}{\sigma}$$

Squaring both sides and reducing gives:

$$\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

Note x_i is a random variable and μ and σ are constants such that $Z_i = \frac{(x_i - \mu)}{\sigma} = \frac{-\mu}{\sigma} + \frac{1}{\sigma}x_i$ is a linear transformation of the form $Z_i = a + bx_i$.

If $Z_i = a + bx_i$ then,

$$E[Z_i] = E[a + bx_i] = a + bE[x_i] = a + b\overline{x}$$

and

$$Var[Z_i] = Var[a + bx_i] = b^2\sigma_x^2$$

where

$$\frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2 = \sigma_Z^2 = Var[Z_i]$$

and

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

giving

$$\frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2 = \sigma_Z^2 = b^2\sigma_x^2 = b^2\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

Note for our purposes $b = \frac{1}{\sigma}$ such that $b^2 = \frac{1}{\sigma^2}$ giving

$$\frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2 = \frac{1}{\sigma^2}\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

which reduces to

$$\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2$$

QED.

Appendix C
The fallacy

The fallacy of the transposed conditional, also known as confusion of the inverse or the statistical equivalent to the fallacy of affirming the consequent, is the jumbling of the probability of a set of data given a hypothesis, and the probability of a hypothesis given a set of data.

In statistical terms, the fallacy of the transposed conditional is corroborated through Thomas Bayes' (1763) theorem, given by:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

where A and B are two different outcomes or events (i.e., a hypothesis and a data set) and $\Pr(B) \neq 0$. Therefore, we can see $\Pr(A \mid B) = \Pr(B \mid A)$ holds true if and only if $\Pr(A) = \Pr(B)$ at the same time.

It is a fallacy if one claims to test the likelihood of a null hypothesis assuming the data are true, if what is actually tested is the likelihood of the data assuming the null hypothesis is true. It is incorrect to assume $\Pr(\mathrm{Data} \mid H_0) = \Pr(H_0 \mid \mathrm{Data})$.

In terms of rhetoric and logic, the fallacy of affirming the consequent is stated:

$$\frac{P \rightarrow Q, Q}{\therefore P}$$

where one takes the true statement $P \rightarrow Q$ and incorrectly concludes the converse $Q \rightarrow P$ to be true. In plain terms, the fallacy is demonstrated with the simple and absurd statement: All dogs are animals; therefore, all animals are dogs.

References

Abate, Kalkidan Hassen, and Tefera Belachew. 2017. Women's Autonomy and Men's Involvement in Child Care and Feeding as Predictors of Infant and Young Child Anthropometric Indices in Coffee Farming Households of Jimma Zone, South West of Ethiopia. *PloS One* 12(3): e0172885. **Link**

Aceituno, Anna M., Kaitlyn K. Stanhope, Paulina A. Rebolledo, Rachel M. Burke, Rita Revollo, Volga Iñiguez, Parminder S. Suchdev, and Juan S. Leon. 2017. Using a Monitoring and Evaluation Framework to Improve Study Efficiency and Quality During a Prospective Cohort Study in Infants Receiving Rotavirus Vaccination in El Alto, Bolivia: The Infant Nutrition, Inflammation, and Diarrheal

Illness (NIDI) Study. *BMC Public Health* 17: article 911. **Link**

Afifi, Hanan H., Mona S. Aglan, Moushira E. Zaki, Manal M. Thomas, and Angie M. S. Tosson. 2012. Growth Charts of Down Syndrome in Egypt: A Study of 434 Children 0–36 Months of Age. *American Journal of Medical Genetics Part A* 158(11): 2647–2655.

Allen, Courtney K., Trevor N. Croft, Thomas W. Pullum, and Sorrel M. L. Namaste. 2019. Evaluation of Indicators to Monitor Quality of Anthropometry Data During Fieldwork. *DHS Working Paper* 162. ICF International, Inc. (Rockville, Md.). **Link**

Anscombe, Francis J. 1973. Graphs in Statistical Analysis. *American Statistician* 27(1): 17–21.

Assaf, Shireen, Monica T. Kothari, and Thomas W. Pullum. 2015. An Assessment of the Quality of DHS Anthropometric Data, 2005–2014. *DHS Methodological Reports* 16. ICF International, Inc. (Rockville, Md.). **Link**

Bayes, Thomas. 1763. An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53: 370–418. **Link**

Bennett, Aisleen, Louisa Pollock, Khuzwayo C. Jere, Virginia E. Pitzer, Benjamin Lopman, Naor Bar-Zeev, Miren Iturriza-Gomara, and Nigel A. Cunliffe. 2020. Duration and Density of Fecal Rotavirus Shedding in Vaccinated Malawian Children With Rotavirus Gastroenteritis. *Journal of Infectious Diseases* 222(12): 2035–2040. **Link**

Berti, Peter R., Alison Mildon, Kendra Siekmans, Barbara Main, and Carolyn MacDonald. 2010. An Adequacy Evaluation of a 10-Year, Four-Country Nutrition and Health Programme. *International Journal of Epidemiology* 39(2): 613–629. **Link**

Biehl, Anna, Ragnhild Hovengen, Haakon E. Meyer, Jøran Hjelmesæth, Jørgen Meisfjord, Else-Karin Grøholt, Mathieu Roelants, and Bjørn Heine Strand. 2013. Impact of Instrument Error on the Estimated Prevalence of Overweight and Obesity in Population-Based Surveys. *BMC Public Health* 13: article 146. **Link**

Bilukha, Oleg, Alexia Couture, Kelly McCain, and Eva Leidman. 2020. Comparison of Anthropometric Data Quality in Children Aged 6–23 and 24–59 Months: Lessons From Population-Representative Surveys From Humanitarian Settings. *BMC Nutrition* 6: article 60. **Link**

Black, Robert E., Cesar G. Victora, Susan P. Walker, Zulfiqar A. Bhutta, Parul Christian, Mercedes de Onis, Majid Ezzati, Sally Grantham-McGregor, Joanne Katz, Reynaldo Martorell, Ricardo Uauy, and the Maternal and Child Nutrition Study Group. 2013. Maternal and Child Undernutrition and Overweight in Low-Income and Middle-Income Countries. *The Lancet* 382(9890): P427–P451. **Link**

Blanton, Curtis J., and Oleg O. Bilukha. 2013. The Probit Approach in Estimating the Prevalence of Wasting: Revisiting Bias and Precision. *Emerging Themes in Epidemiology* 10: article 8. **Link**

Boylan, Sinead, Seema Mihrshahi, Jimmy Chun Yu Louie, Anna Rangan, Norsal Salleh, Ilham Md Ali, Roseyati Dato Paduka, and Timothy Gill. 2017. Prevalence and Risk of Moderate Stunting Among a Sample of Children Aged 0–24 Months in Brunei. *Maternal and Child Health Journal* 21(12): 2256–2266.

Briend, André, Kh. Z. Hasan, K. M. A. Aziz, Bilqis A. Hoque, and F. J. Henry. 1989.

Measuring Change in Nutritional Status: A Comparison of Different Anthropometric Indices and the Sample Sizes Required. *European Journal of Clinical Nutrition* 43(11): 769–778.

Castro Bedriñana, Jorge, and Doris Chirinos Peinado. 2014. Z-Score Anthropometric Indicators Derived from NCHS-1977, CDC-2000 and WHO-2006 in Children Under 5 Years in Central Area of Peru. *Universal Journal of Public Health* (Horizon Research Publishing, San Jose, Calif.) 2(2): 73–81. **Link**

Conkle, Joel, Usha Ramakrishnan, Rafael Flores-Ayala, Parminder S. Suchdev, and Reynaldo Martorell. 2017. Improving the Quality of Child Anthropometry: Manual Anthropometry in the Body Imaging for Nutritional Assessment Study (BINA). *PloS One* 12(12): e0189332. **Link**

Corvalán, Camila, Ricardo Uauy, Aryeh D. Stein, Juliana Kain, and Reynaldo Martorell. 2009. Effect of Growth on Cardiometabolic Status at 4 y of Age. *American Journal of Clinical Nutrition* 90(3): 547–555. **Link**

Crowe, Sonya, Andrew Seal, Carlos Grijalva-Eternod, and Marko Kerac. 2014. Effect of Nutrition Survey 'Cleaning Criteria' on Estimates of Malnutrition Prevalence and Disease Burden: Secondary Data Analysis. *PeerJ* 2: e380. **Link**

de Onis, Mercedes, and Monika Blössner. 1997. *WHO Global Database on Child Growth and Malnutrition*. WHO/NUT/97.4. Geneva: World Health Organization. **Link**

de Onis, Mercedes, and Jean-Pierre Habicht. 1996. Anthropometric Reference Data for International Use: Recommendations from a World Health Organization Expert Committee. *American Journal of Clinical Nutrition* 64(4): 650–658. **Link**

Ecker, Olivier, Jean-François Maystadt, and Zhe Guo. 2019. Can Unconditional Cash Transfers Mitigate the Impact of Civil Conflict on Acute Child Malnutrition in Yemen?: Evidence From the National Social Protection Monitoring Survey. *Middle East and North Africa Regional Program Working Paper* 17. International Food Policy Research Institute (Washington, D.C.). **Link**

El Mouzan, Mohammad I., Abdullah S. Al Herbish, Abdullah A. Al Salloum, Peter J. Foster, Ahmad A. Al Omar, Mansour M. Qurachi, and Tatjana Kecojevic. 2008. Comparison of the 2005 Growth Charts for Saudi Children and Adolescents to the 2000 CDC Growth Charts. *Annals of Saudi Medicine* 28(5): 334–340. **Link**

El Mouzan, Mohammad I., Peter J. Foster, Abdullah S. Al Herbish, Abdullah A. Al Salloum, Ahmad A. Al Omar, Mansour M. Qurachi, and Tatjana Kecojevic. 2009. The Implications of Using the World Health Organization Child Growth Standards in Saudi Arabia. *Nutrition Today* 44(2): 62–70.

Fekadu, Yirgu, Addisalem Mesfin, Demewoz Haile, and Barbara J. Stoecker. 2015. Factors Associated with Nutritional Status of Infants and Young Children in Somali Region, Ethiopia: A Cross-Sectional Study. *BMC Public Health* 15: article 846. **Link**

Frison, Severine, Francesco Checchi, Marko Kerac, and Jennifer Nicholas. 2016. Is Middle-Upper Arm Circumference "Normally" Distributed? Secondary Data Analysis of 852 Nutrition Surveys. *Emerging Themes in Epidemiology* 13: article 7. **Link**

Frongillo, Edward A., Shibani Kulkarni, Sulochana Basnet, and Filipa de Castro. 2017. Family Care Behaviors and Early Childhood Development in Low- and Middle-Income Countries. *Journal of Child and Family Studies* 26(11): 3036–3044.

Garenne, Michel, Douladel Willie, Bernard Maire, Olivier Fontaine, Roger Eeckels, André Briend, and Jan Van den Broeck. 2009. Incidence and Duration of Severe Wasting in Two African Populations. *Public Health Nutrition* 12(11): 1974–1982. **Link**

Garza, Cutberto, and Mercedes de Onis. 2004. Rationale for Developing a New International Growth Reference. *Food and Nutrition Bulletin* 25: S5–S14. **Link**

Golden, Michael H. 2009. Proposed Recommended Nutrient Densities for Moderately Malnourished Children. *Food and Nutrition Bulletin* 30(3): S267–S342. **Link**

Golden, Michael H., and Yvonne Grellety. 2002. Population Nutritional Status During Famine. Presented at the Standardized Monitoring and Assessment of Relief and Transitions (SMART) Workshop, University of Aberdeen (Aberdeen, UK). **Link**

Gorstein, Jonathan, Kevin Sullivan, Ray Yip, Mercedes de Onis, Frederick Trowbridge, Peter Fajans, and Graeme Clugston. 1994. Issues in the Assessment of Nutritional Status Using Anthropometry. *Bulletin of the World Health Organization* 72(2): 273–283. **Link**

Grellety, Emmanuel, and Michael H. Golden. 2016. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. *PloS One* 11(12): e0168585. **Link**

Grellety, Emmanuel, and Michael H. Golden. 2018. Change in Quality of Malnutrition Surveys Between 1986 and 2015. *Emerging Themes in Epidemiology* 15: article 8. **Link**

Grummer-Strawn, Laurence M., Chris Reinold, and Nancy F. Krebs. 2010. Use of World Health Organization and CDC Growth Charts for Children Aged 0–59 Months in the United States. *Morbidity and Mortality Weekly Report* 59(RR-9): 1–15. **Link**

Gupta, Priya M., Emily Wieck, Joel Conkle, Kristina A. Betters, Anthony Cooley, Selena Yamasaki, Natasha Laibhen-Parkes, and Parminder S. Suchdev. 2020. Improving Assessment of Child Growth in a Pediatric Hospital Setting. *BMC Pediatrics* 20: article 419. **Link**

Hulland, Erin N., Curtis J. Blanton, Eva Z. Leidman, and Oleg O. Bilukha. 2016. Parameters Associated With Design Effect of Child Anthropometry Indicators in Small-Scale Field Surveys. *Emerging Themes in Epidemiology* 13: article 13. **Link**

Jacob, Amita, Leah Thomas, Kezia Stephen, Sam Marconi, Joseph Noel, K. S. Jacob, and Jasmin Prasad. 2016. Nutritional Status and Intellectual Development in Children: A Community-Based Study from Rural Southern India. *National Medical Journal of India* 29(2): 82–84. **Link**

Jelliffe, Derrick B. 1966. *The Assessment of the Nutritional Status of the Community (with Special Reference to Field Surveys in Developing Regions of the World)*. Geneva: World Health Organization. **Link**

Joint FAO/WHO Expert Committee on Nutrition. 1971. *Joint FAO/WHO Expert Committee on Nutrition, Eighth Report: Food Fortification, Protein-Calorie Malnutrition*. Geneva: World Health Organization. **Link**

Kaiser Family Foundation (KFF). 2016. U.S. Funding for International Nutrition Programs. April 25. Henry J. Kaiser Family Foundation (San Francisco). **Link**

Krasevec, Julia, Xiaoyi An, Richard Kumapley, France Bégin, and Edward A. Frongillo. 2017. Diet Quality and Risk of Stunting Among Infants and Young

Children in Low- and Middle-Income Countries. *Maternal & Child Nutrition* 13(Supp. 2): e12430. **Link**

Krieger, Nancy. 2012. Who and What Is a "Population"? Historical Debates, Current Controversies, and Implications for Understanding "Population Health" and Rectifying Health Inequities. *Milbank Quarterly* 90(4): 634–681. **Link**

Kwena, Arthur M., Dianne J. Terlouw, Sake J. de Vlas, Penelope A. Phillips-Howard, William A. Hawley, Jennifer F. Friedman, John M. Vulule, Bernard L. Nahlen, Robert W. Sauerwein, and Feiko O. ter Kuile. 2003. Prevalence and Severity of Malnutrition in Pre-School Children in a Rural Area of Western Kenya. *American Journal of Tropical Medicine and Hygiene* 68: 94–99. **Link**

Larsen, Anna Folke, Derek Headey, and William A. Masters. 2019. Misreporting Month of Birth: Diagnosis and Implications for Research on Nutrition and Early Childhood in Developing Countries. *Demography* 56(2): 707–728. **Link**

Leidman, Eva, Louise Masese Mwirigi, Lucy Maina-Gathigi, Anna Wamae, Andrew Amina Imbwaga, and Oleg O. Bilukha. 2018. Assessment of Anthropometric Data Following Investments to Ensure Quality: Kenya Demographic Health Surveys Case Study, 2008 to 2009 and 2014. *Food and Nutrition Bulletin* 39(3): 406–419. **Link**

Martins, Rita C. B., José E. Corrente, Ana Paula Vioto, Emilia Alonso Balthazar, and Maria Rita Marques de Oliveira. 2010. Body Mass Index as Indicator of the Nutritional Status of School Children from the Public Schools of Piracicaba-SP from 2003 to 2006. In *Proceedings of the 9th Brazilian Conference on Dynamics, Control and Their Applications*, 1106–1112. São Carlos, Brazil: Brazilian Society of Applied and Computational Mathematics. **Link**

Matejka, Justin, and George Fitzmaurice. 2017. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In *CHI'17: Proceedings of the 2017 ACM SIGCHI Conference on Human Factors in Computing Systems*, 1290–1294. New York: Association for Computing Machinery. **Link**

McCloskey, Deirdre N. 1994. Economics: Art or Science or Who Cares? *Eastern Economic Journal* 20(1): 117–120.

McCloskey, Deirdre N., and Stephen T. Ziliak. 1996. The Standard Error of Regressions. *Journal of Economic Literature* 34(1): 97–114.

Mei, Zuguo, and Laurence M. Grummer-Strawn. 2007. Standard Deviation of Anthropometric Z-Scores as a Data Quality Assessment Tool Using the 2006 WHO Growth Standards: A Cross Country Analysis. *Bulletin of the World Health Organization* 85: 441–448. **Link**

Mulugeta, Afework, Fitsum Hagos, Gideon Kruseman, Vincent Linderhof, Barbara Stoecker, Zenebe Abraha, Mekonen Yohannes, and Girmay G. Samuel. 2010. Child Malnutrition in Tigray, Northern Ethiopia. *East African Medical Journal* 87(6): 248–254. **Link**

Perumal, Nandita, Sorrel Namaste, Huma Qamar, Ashley Aimone, Diego G. Bassani, and Daniel E. Roth. 2020. Anthropometric Data Quality Assessment in Multisurvey Studies of Child Growth. *American Journal of Clinical Nutrition* 112(Supp. 2): 806S–815S. **Link**

Quételet, Adolphe. 1832. Recherches Sur Le Poids De L'homme Aux Différens Âges [Research on the Weight of Man at Different Ages]. *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* 7: 1–83.

Quételet, Adolphe. 1835. *Sur L'homme Et Le Développement De Ses Facultés* [*On Man and the Development of His Faculties*]. Paris: Bachelier.

Ravallion, Martin. 1997. Famines and Economics. *Journal of Economic Literature* 35(3): 1205–1242.

Razak, Fahad, George Davey Smith, and Sankaran Venkata Subramanian. 2016. The Idea of Uniform Change: Is It Time to Revisit a Central Tenet of Rose's "Strategy of Preventive Medicine"? *American Journal of Clinical Nutrition* 104(6): 1497–1507. **Link**

Reda, Ayalu A., and David Lindstrom. 2014. Recent Trends in the Timing of First Sex and Marriage Among Young Women in Ethiopia. *African Population Studies* 28(2 Supp.): 1157–1170. **Link**

Remans, Roseline, Paul M. Pronyk, Jessica C. Fanzo, Jiehua Chen, Cheryl A. Palm, Bennett Nemser, Maria Muniz, Alex Radunsky, Alem Hadera Abay, Mouctar Coulibaly, Joseph Mensah-Homiah, Margaret Wagah, Xiaoyi An, Christine Mwaura, Eva Quintana, Marie-Andree Somers, Pedro A. Sanchez, Sonia E. Sachs, John W. McArthur, and Jeffrey D. Sachs. 2011. Multisector Intervention to Accelerate Reductions in Child Stunting: An Observational Study from 9 Sub-Saharan African Countries. *American Journal of Clinical Nutrition* 94(6): 1632–1642. **Link**

Rose, Geoffrey. 1992. *The Strategy of Preventive Medicine*. New York: Oxford University Press.

Roth, Daniel E., Aditi Krishna, Michael Leung, Joy Shi, Diego G. Bassani, and Aluisio J. D. Barros. 2017. Early Childhood Linear Growth Faltering in Low-Income and Middle-Income Countries as a Whole-Population Condition: Analysis of 179 Demographic and Health Surveys from 64 Countries (1993–2015). *Lancet Global Health* 5(12): e1249–e1257. **Link**

Samiak, Louis, and Theophilus I. Emeto. 2017. Vaccination and Nutritional Status of Children in Karawari, East Sepik Province, Papua New Guinea. *PloS One* 12(11): e0187796. **Link**

Sen, Amartya. 1984. *Resources, Values, and Development*. Oxford: Basil Blackwell.

Sharma, Nikita, Madhu Gupta, Arun Kumar Aggarwal, and Mutyalamma Gorle. 2020. Effectiveness of a Culturally Appropriate Nutrition Educational Intervention Delivered Through Health Services to Improve Growth and Complementary Feeding of Infants: A Quasi-Experimental Study from Chandigarh, India. *PloS One* 15(3): e0229755. **Link**

Shekar, Meera, Jakub Kakietek, Julia Dayton Eberwein, and Dylan Walters. 2017. *An Investment Framework for Nutrition: Reaching the Global Targets for Stunting, Anemia, Breastfeeding, and Wasting*. Washington, D.C.: World Bank. **Link**

Siegel, Jacob S., and David A. Swanson. 2004. *The Methods and Materials of Demography*. London: Elsevier Academic Press.

Spiegel, Paul B., Peter Salama, Susan Maloney, and Albertien van der Veen. 2004. Quality of Malnutrition Assessment Surveys Conducted During Famine in Ethiopia. *Journal of the American Medical Association* 292(5): 613–618. **Link**

Standardized Monitoring & Assessment of Relief & Transitions (SMART). 2006. *Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations: SMART Methodology*. Toronto: Action Against Hunger Canada. **Link**

Subramanian, Shankar, and Angus Deaton. 1996. The Demand for Food and Calories. *Journal of Political Economy* 104(1): 133–162.

Tanner, James M. 1952. The Assessment of Growth and Development in Children. *Archives of Disease in Childhood* 27: 10–33. **Link**

Tanner, James M. 1981. *A History of the Study of Human Growth*. Cambridge, UK: Cambridge University Press.

Tuffrey, Veronica, and Andrew Hall. 2016. Methods of Nutrition Surveillance in Low-Income Countries. *Emerging Themes in Epidemiology* 13: article 4. **Link**

Ulijaszek, Stanley J., and Deborah A. Kerr. 1999. Anthropometric Measurement Error and the Assessment of Nutritional Status. *British Journal of Nutrition* 82(3): 165–177. **Link**

UNICEF, WHO, and World Bank. 2018. Levels and Trends in Child Malnutrition: Key Findings of the 2018 Edition. World Health Organization (Geneva). **Link**

USAID. 2016. Anthropometric Data in Population-Based Surveys: Meeting Report, Washington, DC, July 14–15, 2015. Food and Nutrition Technical Assistance III Project (Washington, D.C.). **Link**

Varghese, Jithin Sam, and Aryeh D. Stein. 2019. Malnutrition Among Women and Children in India: Limited Evidence of Clustering of Underweight, Anemia, Overweight, and Stunting Within Individuals and Households at Both State and District Levels. *American Journal of Clinical Nutrition* 109(4): 1207–1215. **Link**

Wasserstein, Ronald L., Allen L. Schirm, and Nicole A. Lazar. 2019. Moving to a World Beyond "$p < 0.05$." *American Statistician* 73(Supp. 1): 1–19. **Link**

Waterlow, J. C., R. Buzina, W. Keller, J. M. Lane, M. Z. Nichaman, and J. M. Tanner. 1977. The Presentation and Use of Height and Weight Data for Comparing the Nutritional Status of Groups of Children under the Age of 10 Years. *Bulletin of the World Health Organization* 55(4): 489–498. **Link**

WHO Working Group. 1986. Use and Interpretation of Anthropometric Indicators of Nutritional Status. *Bulletin of the World Health Organization* 64(6): 929–941. **Link**

WHO Expert Committee on Physical Status. 1995. *Physical Status: The Use of and Interpretation of Anthropometry*. Geneva: World Health Organization.

WHO. 2006. *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. Geneva: World Health Organization. **Link**

WHO and UNICEF. 2019. *Recommendations for Data Collection, Analysis and Reporting on Anthropometric Indicators in Children Under 5 Years Old*. Geneva: World Health Organization. **Link**

Wijaya-Erhardt, Maria. 2019. Nutritional Status of Indonesian Children in Low-Income Households with Fathers That Smoke. *Osong Public Health and Research Perspectives* (Korea Disease Control and Prevention Agency, Cheongju) 10(2): 64–71. **Link**

Woodruff, Bradley A., James P. Wirth, Adam Bailes, Joan Matji, Arnold Timmer, and Fabian Rohner. 2017. Determinants of Stunting Reduction in Ethiopia 2000–2011.

Maternal & Child Nutrition 13(2): e12307. Link

Yip, Ray. 1993. Expanded Usage of Anthropometry Z-Scores for Assessing Population Nutritional Status and Data Quality. Abstract. In *Abstracts Book No. 1, 15th International Congress of Nutrition (Adelaide)*, 279. Adelaide: International Union of Nutritional Sciences.

Ziliak, Stephen T., and Deirdre N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Is Costing Jobs, Justice, and Lives*. Ann Arbor: University of Michigan Press.

About the Author

Austin Sandler is a Ph.D. candidate in the Department of Geographical Sciences at the University of Maryland and Consultant at The World Bank Group. He holds an M.S. in Applied Economics from the University of Minnesota and an M.S. in Agricultural and Applied Economics from the University of Wyoming. His teaching includes principles of economics and human geography, mapping and geographic information science, and economic geography. His research covers economic development and growth, food security and childhood malnutrition, spatial economics and econometrics, and history of scientific thought. His email address is sandlera@terpmail.umd.edu.

Go to archive of Economics in Practice section
Go to March 2021 issue

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 127–155

Adam Smith in Love

F. E. Guerra-Pujol[1]

LINK TO ABSTRACT

> *Is any resentment so keen as what follows the quarrels of lovers, or any love so passionate as what attends their reconcilement?*
>
> —Adam Smith (1980a/1795, 36)

Was Adam Smith speaking from personal experience when he posed those questions?[2] Here I report on my investigations into the matter.

An investigation into someone's love life is not the sort of endeavor that Smith would have ever undertaken. At the same time, if someone had ever produced such a report on, say, Montaigne or Grotius, we can imagine Smith glancing at it. The authors we most admire and learn from are human beings, and their character, personality, and private lives often figure into our understandings of their works. In the present report on Smith's love life, I do not turn to interpreting Smith's works, notably *The Theory of Moral Sentiments*, which contains several substantive passages about romantic love and about lust and licentiousness.[3] Instead, I presuppose that the reader has a natural and healthy curiosity about Smith's personal life, including his love life.

1. University of Central Florida, Orlando, FL 32816. I thank Alain Alcouffe and three anonymous reviewers for their comments, clarifications, and suggestions.

2. This quotation appears in Section 1 of Smith's essay on "The History of Astronomy." Although "The History of Astronomy" was first published in 1795 along with some other writings of Smith, it is more likely than not that Smith first wrote this particular essay during his young adult years prior to his appointment at the University of Glasgow in 1751 (see Luna 1996, 133, 150 n.3).

3. Smith's discussion of romantic love appears in Book 1, Section 2, Chapters 1 and 2 of *The Theory of Moral Sentiments*, while his rebuke of lust and licentiousness appears in his critique of the ideas of Bernard Mandeville in Book 7, Section 2, Chapter 4. It is also worth noting that Smith's philosophical analysis of romantic love has generated significant scholarly commentary. See, for example, Den Uyl and Griswold 1996; Dawson 2013; Harkin 2013b; Tegos 2019. No evidence exists to indicate whether any of these passages are autobiographical in nature; nevertheless, Smith's treatment of romantic love in *The Theory of Moral Sentiments* seems consistent with a man who may have himself once fallen in love.

To begin with, previous scholars have adopted different stances when writing about Smith's love life. Some (Heilbroner 1999; Muller 1993; Ginzberg 1934; Mackay 1896; Haldane 1887) simply avoid the subject altogether. Others (Fay 2011, 144; Phillipson 2010, 136; Ross 2010, 227–228; Weinstein 2001, 8–10) entertain the possibility of love affairs, but do so reluctantly, either relegating their romantic speculations to a footnote (Stewart 1980/1811, 349–350) or merely alluding in passing to the possibility of Smith falling in love during his days as a travelling tutor in France (Ross 2010, 227–228; Buchan 2006, 77–78; Scott 1936, 404; Rae 1895, 212–213). By way of example, Edith Kuiper (2013, 69–70) devotes time and attention to "Smith's romantic relationships" and concludes that information about his love life is "scarce." Gavin Kennedy (2005, 4–5) addresses the possibility of numerous "love interests" in the opening pages of his intellectual biography of Smith—but calls the possibility "speculation." Dennis Rasmussen (2017, 131) refers to reports of Smith's dalliances as "rumors," while Nicholas Phillipson (2010, 136) refers to these reports as "gossip."

I do not fault Smith biographers for treating the gossip, rumors, and speculations lightly. But gossip and rumors might be true, and a careful look at all of the evidence is in order. Although his lifelong devotion to his intellectual life and to his widowed mother Margaret Douglas may have prevented him from getting married and forming his own household, the evidence shows that it is more likely than not that Smith fell in love on multiple occasions.

With a view toward systematizing the available evidence and extending the work of previous scholars, I will first put Smith's love life in historical context by describing the strict ecclesiastical regulation of sex in the Scotland of his youth. Next, I will reassemble the available evidence. Specifically, I shall present the following five pieces of primary evidence regarding Smith's loves:

- an obscure but intriguing end note that was first published in the second edition of Dugald Stewart's biography of Smith's life and writings;
- a private letter dated July 14, 1784, addressed to Stewart;
- a brief anecdote by Henry Mackenzie, a prominent Scottish lawyer and writer and a co-founder (along with Stewart) of the Royal Society of Edinburgh;
- a personal letter dated September 18, 1766, written by one of Smith's closest friends and confidants, containing details about Smith's love life; and
- a letter of introduction dated sometime in October 1766 authored in the hand of a possible love interest, Madame Marie-Jeanne Riccoboni.

In addition to the presentation of this body of evidence, I draw reasonable infer-

ences, make several conjectures, and consider a few hypotheses concerning Smith's sexuality and romantic attachments. I then conclude by speculating about Smith's desire to have his papers destroyed and about the possibility of a lost travel diary from his Grand Tour.

The ecclesiastical regulation of sex in the Scotland of Smith's youth

One of the most regulated aspects of Scottish life during Smith's lifetime was sex (Hardy 1978; Mitchison and Leneman 2001). Smith's world was one in which intellectual life and sexual activity were strictly monitored by Church elders, and nowhere was the regulation of sexual morality more oppressive than in Scotland.[4] During Smith's lifetime, every parish in Scotland had its own ecclesiastic or church court. These parish courts or 'kirk sessions' had jurisdiction over every parishioner's private and public conduct, including over all matters of sexual morality. According to historians Rosalind Mitchison and Leah Leneman (2001), during Smith's lifetime the great majority of these church cases consisted of sexual matters.[5] Mitchison and Leneman have also painted a detailed picture of the repressive nature of Smith's world and of the roving jurisdiction of these parish courts or kirk sessions over sex:

> In the early modern period every parish in Scotland had its own church court (the kirk session) dealing with matters of conduct and morality. Drunkenness, sabbath breaking, slander, riotous behavior—all these came under the aegis of the session. However, partly through a sharper defining line between the roles of lay and of ecclesiastical jurisdictions, by the mid-eighteenth century the great majority of cases were of a sexual nature…. (Leneman and Mitchison 1988, 483)

Leneman and Mitchison (1988, 483) also emphasize "[t]he thoroughness with which these cases were pursued." By way of example:

> The usual train of events was for an unmarried girl to be reported as 'with child' at a meeting of the kirk session and to be cited to appear at the next meeting. At that time she would be asked to name the man who had been

---

4. See generally Mitchison and Leneman 2001; Leneman and Mitchison 1987; 1988. For an overview of the legal and ecclesiastical regulation of sex in Scotland prior to the birth of Smith, see Hardy 1978.

5. Mitchison and Leneman (2001) survey over 8,000 church court records spread across 78 Scottish parishes from the mid-seventeenth to mid-eighteenth centuries.

> guilty with her, and that man would in turn be cited to appear at a forthcoming meeting. Unless a case were in some way unusual, for instance if the man denied fornication with the woman, further enquiry would not normally be made into the circumstances surrounding the act. However, for some unknown reason, certain parishes in the Western Highlands and certain parishes in Fife often went on to ask where, when and how often intercourse had taken place. (Leneman and Mitchison 1988, 483)

Moreover, even marital sex in anticipation of marriage or 'ante-nuptial fornication' was a sin, though there was a disconnect between official Church doctrine and informal social norms on the matter of pre-marital sex.[6]

Thus, the sex lives of parishioners in the Scotland of Smith's youth were strictly monitored by Church elders, and the penalties for fornication, adultery, and other such moral offenses consisted of shaming penalties, or "penance on the pillar" (Leneman and Mitchison 1988, 495). Although Smith's theism and views of religion are unclear,[7] a cautious and careful scholar of Smith's stature would most likely not have wanted to incur such penalties as they would have derailed his prestigious academic career and lucrative private tutoring opportunities. These general observations must thus be kept in mind when exploring the question: Who were Smith's loves?

The evidence

Rasmussen (2017, 131) reports of "occasional rumors, throughout Smith's life, of potential romantic connections" but concludes that "none of them amounted to much."[8] Similarly, Ian Simpson Ross (1995; 2010), the scholar who has painted the most comprehensive picture to date of Smith's possible amorous interludes (2010, 227–228), concludes that "the biographer can do little more with the topic of Smith's sex life than contribute a footnote to the history of sublimation." Information about Smith's romantic encounters is admittedly "scarce" (Kuiper 2013, 62), but it is not non-existent. At least five separate pieces

6. For ordinary people, betrothal was a part of marriage, and as such it made sexual intercourse permissible. Church elders, however, generally did not approve of such 'irregular' marriage. For the Church, a marriage required the public exchange of promises in the presence of the parish minister (Gillis 1985, 52–54; see also Hardy 1978, chs. 4 and 5).

7. Or in the words of Margaret Jacob (2019, 128, 126), "Adam Smith kept his religious beliefs very private" and "[his] private religious beliefs will probably never be known."

8. It is unclear whether the "them" in this passage refers to the rumors of Smith's romantic connections or to Smith's possible romances themselves. Harkin (2013a, 502) notes a "complete dearth of information" about Smith's love life, while Weinstein (2001, 10) describes Smith's romantic life as "virtually non-existent."

of primary evidence mention or refer to Smith's love life.[9]

Dugald Stewart's "Note (H.)"

The earliest published reference to Smith's love life appears in 1811, 21 years after Smith's death, in the very last endnote—"Note (H.)"—of the third and fourth editions of Dugald Stewart's biographical essay "An Account of the Life and Writings of Adam Smith" (Stewart 1811a, 150; 1811b, 552).[10] Stewart's enigmatic note reads in full as follows:

> In the early part of Mr Smith's life *it is well known to his friends*, that he was for several years attached to a young lady of great beauty and accomplishment. How far his addresses were favourably received, or what the circumstances were which prevented their union, I have not been able to learn; but I believe it is pretty certain that, after this disappointment, he laid aside all thoughts of marriage. The lady to whom I allude died also unmarried. She survived Mr Smith for a considerable number of years, and was alive long after the publication of the first edition of this Memoir. I had the pleasure of seeing her when she was turned of eighty, and when she still retained evident traces of her former beauty. The powers of her understanding and the gaiety of her temper seemed to have suffered nothing from the hand of time. (Stewart 1980/1811, 349–350, my emphasis)

Ian Simpson Ross (2010, 227) describes this early love interest as "a Fife lady whom he [Smith] had loved very much," but neither Ross nor Stewart provides any additional evidence about the geographical location of this love affair; nor do they identify this woman by name.[11] Nevertheless, if this love affair occurred in the Kirkcaldy of Smith's youth, a small parish with a population around 1500 at the time (Heilbroner 1999, 46), it should not be impossible to identify the lady. I explore this matter further in the next part of this paper.[12]

---

9. More significantly, these five separate sources of information—these five historical witnesses, so to speak—all knew Smith personally or knew people who travelled in Smith's social circles.

10. An historical aside is in order. Dugald Stewart had originally written his biography of Smith in the early 1790s and had read his "Account of the Life and Writings of Adam Smith" to members of the Royal Society of Edinburgh on January 21 and March 18 of the year 1793. He then published his biographical essay in 1794 in Volume 3 of the *Transactions of the Royal Society of Edinburgh*. A second edition of this essay was then published in 1795 in a book titled *Essays on Philosophical Subjects*. Neither the 1794 edition of Stewart's essay nor the 1795 one, however, contain any notes or any reference to Smith's love life. As a further aside, "Note (H.)" subsequently became "Note (K.)" when additional end notes were added to the 1858 edition of Stewart's biographical essay (see Ross 1980, 265–268, who provides a complete list of the first five editions of Stewart's biography of Smith).

11. Fay (2011/1956, 144) refers to her as the "Maid of Fife."

12. For what it is worth, Alain Alcouffe and Andrew Moore (2018, 15 n.18) identify Smith's "lady of Fife"

Regardless of the question of geographical location, Stewart is a credible witness to an attachment "well known" to Smith's friends. Stewart knew Smith and many of Smith's acquaintances.[13] Also, to give the reader some idea of Stewart's stature and sterling reputation, he co-founded—along with Henry Mackenzie, a Scottish lawyer, novelist, and writer whom we shall re-encounter soon—the Royal Society of Edinburgh in 1783 and held the chair of moral philosophy at the University of Edinburgh for thirty-five years, from 1785 until 1820.[14] Why would Stewart risk sullying his own reputation (and that of his friend Smith) by reporting mere gossip or an unfounded rumor?

James Currie's letter to Stewart in 1794 about a French connection

A next relevant item is a letter dated July 14, 1794, addressed to Stewart (Currie 1831, 317–320). This personal correspondence is signed by one James Currie (1756–1805), a medical doctor who was then residing in Liverpool, and is addressed to "Dugald Stewart, Esq." This letter is important because it contains a second-hand account of a second Smith love affair.

Before proceeding any further, however, why did Currie write this letter to Stewart? Since Dugald Stewart's biography of Smith was read to the members of the Royal Society of Edinburgh in January and March of 1793 and then appeared in published form a year later (1794), it is possible that Currie himself may have obtained a copy of this first edition of Stewart's biography of Smith. Or, Currie may have heard about Stewart's biography from someone who, in turn, had either heard or read Stewart's account.[15] Currie writes to Stewart with the intention of providing further proof of another affair involving Smith. Although this evidence consists of a second-hand report, Currie states that his source of information, a "Captain Lloyd," spent considerable time with Smith and with Smith's student Henry Scott, the Third Duke of Buccleuch,[16] during their three-year Grand Tour in

as Lady Janet Anstruther, who "was renowned for her beauty and for her reputation as a flirt." Dugald Stewart's original "Note (H.)," however, refers to a lost love "in the early part of Mr. Smith's life," while Alcouffe and Moore are referring to "a famous lady of Fife in the 1760's," when Smith would have been in his late 30s and early-to-mid 40s. Also, Lady Janet is said by one source to have died at the age of 76 in 1802 (**link**) so she may not have lived to see her 80th birthday.

13. Or in the words of Alain Alcouffe and Philippe Massot-Bordenave (2020, x): "Dugald Stewart…had the advantage of having been close to both Smith and witnesses to his life."

14. For more information about Stewart's contributions to the Scottish Enlightenment, see Haakonssen 1984; Rashid 1985; Wood 2000.

15. The first edition of Stewart's biography of Smith was published in the third volume of *The Transactions of the Royal Society of Edinburgh* in 1794.

16. For an in-depth biography of the Third Duke of Buccleuch, as well as a portrait and family tree, see

the mid-1760s.[17] Currie writes:

> Another source from which I have heard much of Dr. Smith, was the information of a Captain Lloyd, who was much in [Smith's] intimacy in France; and who passed the whole time that he spent at Abbeville with the Duke of [Buccleuch], in his society. Captain Lloyd was bred a soldier, but left the army early. He is one of the most interesting and most accomplished men I ever knew. (Currie 1831, 317–318)[18]

Currie says further: "I could perceive from many circumstances [that Smith and Lloyd] were on a footing of great intimacy; and many curious particulars of the Doctor's [Smith's] conduct he has related" (Currie 1831, 318). Among these "curious particulars" are the allegations that Smith "was deeply in love with an English lady" during his sojourn in Abbeville.[19] Currie's report reads as follows:

> Dr. Smith, it seems, while at Abbeville, was deeply in love with an English lady there. What seems more singular, a French Marquise, a woman of talents

Alcouffe and Massot-Bordenave (2020, 28–34); for a short biography of Duke Henry, see Valentine 1970, 2:773.

17. For a summary of Smith's travels in France see generally Ross 2010, ch. 13, as well as Rae 1895, chs. 12–14. (For a map and detailed timeline of Smith's extensive travels in the South of France, see also Alcouffe and Massot-Bordenave 2020, xiii–xiv, xviii–xix.) In summary, Duke Henry's stepfather, Charles Townsend, had appointed Smith to be the future Duke's private tutor and chaperone, and Smith personally supervised Duke Henry's Grand Tour from early 1764 through the fall of 1766. Duke Henry, a direct descendant of King Charles II of England and King Henry IV of France, was born into one of the wealthiest and most prestigious families in Scotland, and upon coming of age in September of 1767, Smith's pupil would become one of Scotland's largest landowners (see generally Alcouffe and Massot-Bordenave 2020, 28–34; Valentine 1970, 2:773). For a description of Duke Henry's landholdings and his lifelong friendship with Smith, see Bonnyman 2014.

18. Hirst (1904, 131) speculates that Captain Lloyd was "doubtless on a patriotic visit to the field of Crecy" when he reportedly met Smith and Duke Henry in Abbeville.

19. The precise date of Smith's stay in Abbeville and his reasons for visiting there are unclear. In an unpublished paper, I speculate that Smith may have travelled to Abbeville to witness the execution of the Chevalier de La Barre, who became the last man in Europe to be put to death for the crime of blasphemy. For detailed histories of this case, see Claverie 1992; 1994; Chassaigne 1920. At the time, "l'affaire du Chevalier de La Barre" attracted attention across France—even attracting the sustained notice of the celebrated atheist and free-thinker Voltaire, who wrote not one but two accounts of the young de La Barre's prosecution and sentence (see Voltaire 2000/1766; 2000/1775. (Voltaire's first essay about this case is dated 15 July 1766, but some scholars believe this essay was actually written in 1767 or 1768.) For a summary of Voltaire's involvement in this notorious case, see Claverie 1994; see also Braden 1965, 58–65. Also, this case has become so central to the identity and history of modern France that many streets are named after the Chevalier de La Barre and many monuments were subsequently erected in his honor, including a statue standing at the gates of the famous Sacred Heart Cathedral in the Montmarte neighborhood of Paris. A picture of this particular monument to de La Barre is available online (**link**). Alas, this monument was taken down during the Second World War on orders of Marshal Philippe Pétain and melted down (Caulcutt 2020).

> and *esprit*, was smitten, or thought herself smitten, with the Doctor, and made violent attempts to obtain his friendship. She was just come from Paris, …[and she] was determined to obtain his friendship; but after various attempts was obliged to give the matter up. Dr. Smith had not the easy and natural manner of Mr. Hume…. He [Smith] was abstracted and inattentive. He could not endure this French woman, and was, besides, dying for another. (Currie 1831, 318–319)

Currie then concludes his July 1794 letter by offering to put Stewart in touch with Captain Lloyd. Alas, no evidence exists of further communication between Lloyd and Stewart or between Stewart and Currie.

Henry Mackenzie's recollections of Miss Campbell

The next piece of primary evidence comes from Henry Mackenzie (1745–1831), a distinguished Scottish lawyer and popular novelist who, as I mentioned previously, co-founded, along with Stewart, the Royal Society of Edinburgh.[20] According to Ross (2010, 227), Mackenzie knew Smith personally and "was much in Smith's company when he [Smith] lived in Edinburgh in the last twelve years of his life."

Toward the end of his long and remarkable life, Mackenzie jotted down a series of personal recollections, hoping to have these memories published in a book of "anecdotes and egotisms," as Mackenzie himself referred to them (Fieser 2003, 251). Mackenzie's wide-ranging collection of anecdotes was eventually assembled by Harold William Thompson and published by Oxford University Press in 1927.[21] Among other things, Mackenzie's collection of anecdotes includes an entry with the title of "Smith and Hume in Love." The first part of Mackenzie's brief recollection about Smith is quoted in full below:

> Adam Smith [was] seriously in love with Miss Campbell of ________ (the name is so numerous that to use it cannot be thought personal), a woman of as different dispositions and habits from him as possible. (Mackenzie 1927, 176; reprinted in Fieser 2003, 255, omission and parenthetical remark both appear in the original)[22]

20. For a sketch of Mackenzie's life as well as his contributions to Scottish letters, see Scott 1834; Drescher 2004; see also Fay 2011, 2; Valentine 1970, 2:567–568.

21. Thompson wrote his dissertation on Mackenzie. For an overview of Thompson's life and work, see Caplan, French, and Mineka 1964.

22. About David Hume, Mackenzie goes on to write (1927, 176; reprinted in Fieser 2003, 255): "His friend, David Hume, was deeply smitten with a very amiable young lady, a great friend of mine, Miss Nancy Ord, but the disparity of age prevented his proposing to her, which he once intended. She was a great admirer of his, and he was a frequent guest at her father's, where I met him, and made one of his whist party with

Who was "Miss Campbell of ________"? When exactly did Smith fall in love with her? And what, if anything, became of this romance? In a parenthetical remark, Mackenzie implies that "Campbell" was a common last name—a "name so numerous that to use it cannot be thought personal"—so that he is not giving anything way by identifying "Miss Campbell" as the object of Smith's affections. That said, could Mackenzie's Miss Campbell nevertheless be the same "young lady of great beauty and accomplishment" that Stewart refers to in Note (H.)—now Note (K.)—of his biography of Smith? Or could Mackenzie perhaps be referring to Lady Frances Scott (1750–1817), the daughter of Caroline Campbell Scott and Duke Henry's younger sister? This conjecture is not far-fetched, especially considering Mackenzie's observation that the woman was "of as different dispositions and habits from him as possible"—she being the daughter of a wealthy aristocratic family and he an absent-minded professor. Smith corresponded with Lady Frances on multiple occasions, and both lived at Dalkeith House during the fall of 1767.[23] Further below, however, I explain why it seems unlikely that the Miss Campbell in Mackenzie's anecdote is the young Lady Frances.

Colbert's letter of September 1766: a smoking arrow?

There is a fourth piece of primary evidence, a long French-language letter addressed to Smith and to his teenage pupil, Henry Scott, the Third Duke of Buccleuch.[24] The identity of the letter's author is disguised under an abbreviated and jocular pseudonym: "Le Gr. Vic. Eccossois," which stands for Grand Viccaire Eccossois (Smith 1987, 165). Nevertheless, it is most likely that this French-speaking "Great Scottish Vicar" was none other than Seignelay Colbert de Castle-Hill, also known as Abbé Colbert, a fellow Scotsman and Smith's "chief guide and friend" during his extended 18-month sojourn in the South of France (Rae 1895, 176).[25] In one passage of this intimate letter, dated September 18, 1766,[26] the author

the young lady and some other person. I played well at the time and so did she. D. Hume was vain of his playing whist. That game has much of observation in it, and such games best suit a thinking man."

23. At least three letters by Smith addressed to Lady Frances survive: nos. 97, 98, and 225 in Smith 1987. For details regarding Smith's stay at Dalkeith, see Bonnyman 2014, 58–59.

24. At the time this contemporaneous letter was composed, Smith was serving as a private tutor and chaperone for Duke Henry—and for his younger brother Hew Campbell Scott as well, who had joined them in Toulouse subsequently (see Alcouffe and Massot-Bordenave 2020, chs. 1 and 2; Bonnyman 2014, ch. 2; Ross 2010, ch. 13).

25. It is also worth noting that Colbert would eventually be appointed the Bishop of Rodez (Alcouffe and Massot-Bordenave 2020, 63–64), but at the time of Smith's travels in France, Colbert had been appointed as one of the vicars general of the diocese of Toulouse in the South of France. For an overview of Abbé Colbert's life and career as well as an illuminating summary of his relationship to Smith, see Alcouffe and Massot-Bordenave 2020, 54–66.

26. Private correspondence dated September 18, 1766, located in the National Archives of Scotland,

refers in jest to some of Smith's romantic attachments, including one by name:

> Et tu, Adam Smith, philosophe de Glasgow, heros et idole des high-broad Ladys, que fais tu, mon cher ami? Comment gouvernes tu La duchesse d'Anville et Mad. de Boufflers, ou ton coeur est il toujours epris des charms de Mad. Nicol et des apparent apparens que laches de cette autre dame de Fife, que vous aimees tant? (letter dated 18 September 1766, National Archives of Scotland, GD224/2040/62/3, quoted in Alcouffe and Moore 2018)

Translated, the passage is:

> And you, Adam Smith, Glasgow philosopher, high-broad Ladies' hero and idol, what are you doing my dear friend? How do you govern the Duchess of Anville and Madame de Boufflers, where your heart is always in love with Madame Nicol and with the attractions as apparent as hidden of this lady of Fife that you loved. (as translated in Alcouffe and Massot-Bordenave 2020, 260)[27]

This letter of 1766 is a crucial piece of evidence for two reasons. First of all, it is the first primary source to mention Smith's love interest in France by name, and secondly, it is the first source to pinpoint the geographical location in Scotland of Smith's first love. Does "this hidden lady of Fife" refer to the same woman mentioned in Note (H.) of Stewart's biography of Smith? Does "Madame Nicol" refer to the love interest in Abbeville mentioned in James Currie's hearsay report?[28] Either way, Colbert's testimony is a highly credible source by any measure. He became Smith's closest friend and confidant during Smith's sojourn in Toulouse (March 1764 to November 1765), and he even travelled with Smith and Duke Henry to Bordeaux and to other places in the South of France during this 18-month period (Alcouffe and Massot-Bordenave 2020, 216–217; see also Rae 1895, 179). Colbert got to know Smith the man, for the jocular and intimate tone of his letter suggests camaraderie and close connections, or in the words of Alain Alcouffe and Philippe Massot-Bordenave (2020, 217), the letter "is probably a private correspondence between friends who have established trust."

GD224/2040/62/3. The letter is reprinted in full in Alcouffe and Massot-Bordenave 2020, 260–261. The *Correspondence of Adam Smith* edited by Ernest Campbell Mossner and Ian Simpson Ross incorrectly dates the letter as "February 18, 1766" (letter 91 in Smith 1987). As a result, at least two scholarly sources (Buchan 2006, 77; Kuiper 2013, 76 n.11) incorrectly identify the date of this letter as February 18, 1766, instead of September 18, 1766. Also, Kuiper (2013, 76 n.11) incorrectly attributes the authorship of this letter to David Hume.

27. For slightly different translations of Abbé Colbert's letter of September 18, 1766, see Ross 2010, 227; Buchan 2006, 77; Smith 1987, 111).

28. Ross (2010, 227) reads the two letters in this way.

Riccoboni's letter of October 1766

The last piece of primary evidence is a letter from Madame Marie-Jeanne Riccoboni dated sometime in October of 1766.[29] Although Riccoboni is little remembered today, she was a highly accomplished actress in the Théâtre-Italien, located in the Hôtel de Bourgogne of Paris,[30] and an illustrious *femme de lettres*, one of the best-selling novelists of her day (Darnton 1998, 255). Riccoboni became acquainted with Smith during his extended residency in Paris in 1766.

Recall that Smith, along with Duke Henry and the Duke's younger brother Hew Campbell Scott, had returned to Paris in late 1765 or early 1766. At some point thereafter, most likely in the Parisian salon of the Baron d'Holbach,[31] did Smith and Riccoboni become acquainted (Nicholls 1976, 16). Riccoboni described the impression Smith had made during their first meeting in a private letter dated May 21, 1766:

> Two Englishmen have arrived here. One [David Hume] is a friend of Garrick's; the other is Scottish; my God what a Scot! He speaks with difficulty through big teeth, and he's ugly as the devil. He's Mr. Smith, author of a book I haven't read. I speak to him about Scotland, and especially about mountains. (quoted in Dawson 2018, 6)

Whatever Smith lacked in looks or vocal refinement, however, he must have made up with his intellect and personality, for Riccoboni quickly developed "a schoolgirl crush on the Scot" (Leddy 2013, 11). In a subsequent letter addressed to fellow actor David Garrick and dated sometime in October 1766, she reveals her feelings for Smith thus:

> I am very pleased with myself, my dear Garrick, to offer you that which I miss very sharply: the pleasure of Mr. Smith's company. I am like a foolish young girl who listens to her lover without ever thinking of loss, which always accompanies pleasure. Scold me, beat me, kill me! But I adore Mr. Smith, I adore him greatly. I wish the devil would take all our philosophes, as long as he returns Mr. Smith to me. (quoted in Leddy 2013, 11; Dawson 2018, 10)

---

29. I wish to thank Alain Alcouffe for bringing this correspondence to my attention.
30. For a history of this celebrated theater, see Roy 1995.
31. Baron d'Holbach (Paul-Henri Thiry) was a philosopher, translator, and devotee of the French Enlightenment who played a prominent role in Parisian intellectual circles through his salon. The guest list of his salon included many of the most prominent intellectual and political figures in Europe (see LeBuffe 2020). For an introduction to the institution of the Parisian salon, see Goodman 1994.

Riccoboni's exuberant confession, however, may very well be an example of unrequited love or simply fondness and affection, as we have no further evidence of any affair between the two. Also, the month of October 1766—the month in which Riccoboni wrote the second letter quoted above—was a fateful moment for the Smith party in Paris, for that was the month that Duke Henry's younger brother died of fever in Paris and Smith and Duke Henry decided to cut short their Grand Tour. It thus seems possible that Riccoboni, writing immediately upon Smith's departure, might have been exaggerating her feelings.

As for Smith's words about Riccoboni, he mentions her in a 1766 letter to Hume (Smith 1987, 113), but, more significantly, Smith ranks her as a novelist among illustrious company in material he introduced in the sixth edition of *The Theory of Moral Sentiments*: "The poets and romance writers, who best paint the refinements and delicacies of love and friendship, and of all other private and domestic affections, Racine and Voltaire, Richardson, Marivaux, and Riccoboni, are, in such cases, much better instructors than Zeno, Chrysippus, or Epictetus" (Smith 1976/1790, 143.14).

To sum up, the evidence presented thus far—Stewart's own testimony in his 1811 "Note (H.)"; Mackenzie's brief 1831 anecdote; and Colbert's intimate 1766 letter—all suggest that Smith had fallen in love on at least two or perhaps three occasions during his life, while additional evidence—Riccoboni's letter of October 1766 as well as Currie's secondhand report in his July 1794 letter to Stewart—indicate that Smith was not lacking in admirers during his sojourn overseas. But who were these ladies?

Inferences and conjectures

I will now draw the most reasonable inferences from the evidence presented above and propose several new concrete conjectures. Given my background in law—I graduated Yale Law School (Class of '93) and teach law and ethics at the University of Central Florida—I will borrow the common law's 'more likely than not' or 'preponderance of the evidence' standard used to try facts in civil cases. I argue that it is more likely than not that Smith did, in fact, fall in love on several occasions in his life.

Smith's first love?: The hidden lady of Fife

Stewart, the only biographer who knew Smith when he was alive, reports from his personal knowledge that it was "well-known to [Smith's] friends that he was for several years attached to a young lady of great beauty and accomplishment,"

that this attachment occurred "in the early part of Mr Smith's life," and that he (meaning Stewart) had once met the lady in person "when she was turned of eighty" (Stewart 1980/1811, 349, 350). Stewart's *Account* is a professional and circumspect work, choosing its words carefully. Given these facts and their reputable source, I conjecture that this love, Smith's first romantic attachment, would most likely have occurred or begun during the years 1746 to 1748, when the young Smith returned to his hometown, the small coastal community of Kirkcaldy, and lived with his mother for two years after having completed his formal studies at Oxford.[32] Smith would have been between 23 and 25 years old at that stage in his life.

Further, given the small population in Kirkcaldy during Smith's lifetime as well as the existence of detailed Church records for this small parish, it is my belief that historians should be able to identify the woman with some confidence.[33] For their part, Alain Alcouffe and Andrew Moore (2018, 15 n.18) have recently identified a reference to "this hidden lady of Fife" in Colbert's September 1766 letter as Lady Janet Anstruther (1725–1802), who "was renowned for her beauty and for her reputation as a flirt."[34] Is this the same lady referred to by Stewart (1980/1811, 349–350) in "Note (H.)," now "Note (K.)," in his first-hand account of Smith's life? Alcouffe and Moore (2018, 15 n.18) speak of "a famous lady of Fife in 1760's." Stewart's "Note (H.)," however, dates this love to "the early part of Mr Smith's life." Also, Stewart mentions in "Note (H.)" that he himself had the pleasure of meeting the lady "when she was turned of eighty," but when exactly did this meeting occur? Stewart's "Note (H.)" did not appear in published form until 1811, and Stewart himself states in "Note (H.)" that this lady "survived Mr Smith for a considerable number of years, and was alive long after the publication of the first edition of this Memoir," so the meeting between Stewart and the Smith's first love could have occurred as late as 1810. If she were 80 years old in 1810, then she

32. In the alternative, it is also possible—but in my view less likely—that Smith's first love may have been a Glaswegian, a resident of the port city of Glasgow, where Smith lived for over 15 years—first from 1737 to 1740, when he was a student at the University of Glasgow, and then from 1751 to 1763, when he held a prestigious professorship there. (For a visual outline of Smith's biography, see that produced by Liberty Fund (**link**); see also Wight 2002, App. A, 267–269.) I say, however, 'less likely' because Smith would have been very young during his first residency at the University of Glasgow (1737–1740). During the extended period of his second residency in Glasgow (1751–1763), Smith would have been financially independent and thus less dependent on his mother, so we cannot rule out the remote possibility of a lost love in Glasgow. Towards the end of his life, for example, Smith himself once referred to his years in Glasgow "as by far the happiest and most honourable period of my life" (see letter 274 in Smith 1987, dated November 16, 1787; see also Alcouffe and Massot-Bordenave 2020, 4, 12).

33. According to Heilbroner (1999, 46), Kirkcaldy boasted a population of only 1500 souls at the time of Smith's birth in 1723. See also Jacob (2019, 124), who notes that Edinburgh, the largest city in Scotland during Smith's lifetime, had only about 40,000 residents.

34. For a 1761 portrait of Lady Anstruther by Sir Joshua Reynolds, see **here**. I thank Alain Alcouffe for bringing the existence of this beautiful portrait to my attention.

would have been born in 1730, seven years after Smith's birth in 1723.[35]

For my part, I wonder whether Smith's mother, Margaret Douglas Smith, who was by all accounts a strong-willed and dominating mother (Kuiper 2013),[36] may have objected to any proposed union between her son and this first love. Of course, Smith and the woman may, themselves, one or both, have seen a union as impractical or unwelcome for any number of reasons. But I wish to explore now the specific conjecture of a maternal veto, more or less against the inclinations of the son. Such a conjecture is relevant given what we know about early modern Scottish society as well as Smith's lifelong devotion to his mother.[37]

Margaret Douglas Smith belonged to the landed gentry, descending from a respected landowning family on her mother's side.[38] Adam Smith's complete financial dependence on his mother during this stage of his life must also be noted.[39] Smith was most likely largely financially dependent on the support of his widowed mother until his initial appointment in 1751, at the age of 27, as the Chair of Logic at the University of Glasgow.[40]

Although parental consent was not a legal requirement in early modern Scotland (Leneman 1999, 673; Leneman and Mitchison 1993, 845, 847), it was generally expected that "children should have the consent of their parents, or those 'in loco parentis', to their marriage" (Hardy 1978, 531). The parental consent norm was so pervasive that it "could vary from marriages arranged by parents without consideration being given to the personal wishes of their children to marriages where the child made the selection of marriage partner and the parents were expected to accede to their choice" (ibid.). This parental consent norm makes all the more sense given the economic structure of Scottish society during Smith's lifetime, a neo-feudal and religious society in which property, especially property in

35. Given that Lady Janet Anstruther was reportedly born in 1725, Alcouffe and Moore's conjecture about her might be correct, after all. On the other hand, Dugald Stewart writes in his Note (H.) of 1811 that he "had the pleasure of seeing her [Lady Janet?] when she was turned of eighty," but as mentioned previously, Lady Janet is said to have died in 1802, aged 76.

36. For a portrait of Margaret Douglas, see **here**.

37. Or in the words of John Rae (1895, 4): "His mother herself was from the first to last the heart of Smith's life." Smith's first biographer, Dugald Stewart (1980), also confirms the central role Margaret Douglas played in Smith's life.

38. *See* Özler 2012, 346–347; Kuiper 2013, 64. For his part, Smith's father—also named Adam Smith—had died a few months before his son Adam was born and had accumulated some wealth during his lifetime, having served as "Judge Advocate for Scotland and Comptroller of the Customs in Kirkcaldy" (Rae 1895, 1). By all accounts, Smith's father left a large income and considerable property to his young widow, Margaret Douglas Smith (see Özler 2012, 346; Kuiper 2013, 64).

39. By way of example, one of Smith's own cousins, Lydia Marianne Douglas, found herself in dire financial straits after she married a man against the will of her parents (see Ross 1995, 401).

40. A year or two after this initial appointment, Smith subsequently accepted the Chair of Moral Philosophy at the University of Glasgow (Heilbroner 1999, 46), a position he held until early 1764, when he departed on his Grand Tour with Duke Henry.

land, was held on a family basis.[41]

Smith's second love?: Madame Nicol

Next, I conjecture that Smith may have fallen in love yet again at some point during his Grand Tour (1764–1766) alongside Duke Henry (Henry Scott Campbell), the Duke of Buccleuch.[42] In 1764 Smith was no advanced senior—he was 41 years old—and one of our primary sources—Colbert's letter of 1766—identifies a Madame Nicol as a possible love interest during Smith's travels in France.[43] More recently, Alain Alcouffe and Philippe Massot-Bordenave (2020, 262) have identified this potential love interest as a resident of Toulouse: "Madame Nicol, the wife of Capitoul Nicol."[44] Was this the same woman with whom Smith is reported to have fallen in love in Abbeville in 1766? Is it possible that Smith had already fallen in love with her during his 18-month stay in the south of France?

The main outline of Smith's travels in France is well known (Ross 2010, Ch. 13; Rae 1895, chs. 12–14). After arriving in Paris on February 13, 1764, Smith and his pupil Duke Henry travelled to the south of France and established a base in the tranquil town of Toulouse, where they lived for many months.[45] They arrived in Toulouse in March of 1764, travelled across the South of France during the summer and autumn of 1764, and returned to Toulouse a second time in January of 1765.[46] At some point upon his return to Toulouse in early 1765, Smith wrote a letter to Charles Townsend, Duke Henry's stepfather and the man who was

41. *Cf.* Leneman (1999, 675), explaining why some Scottish couples resorted to clandestine marriages: "usually because the man was (or said he was) financially dependent on relations who would not approve of marrying at that stage in his life, or of his choice of wife."

42. In setting off for France, the father of modern economics and the young duke were following an elite and well-established tradition, for the Grand Tour was a rite of passage of the sons of elite British families as well as the "'crown' of [their] education" (Cohen 2001, 129; Brodsky-Porges 1981, 178). Michèle Cohen (1992; 2001) has explored the educational and cultural ideals of the Grand Tour and has identified many deep "contradictions and ambiguities" of these tours. In addition, the sexual aspect of Grand Tours by young British aristocrats (and their tutors?) during this era should also not go unnoticed (see, e.g., Chapter 5 of Black 2011/1985, which is titled "Love, Sex, Gambling, and Drinking;" see also Black 1981, 660, 666 n.7; Black 1984, 413–414; Cohen 1992, 255–256).

43. In the alternative, could this Madame Nicol refer to the Marie-Louise-Nicole Elizabeth (1716–1794), the duchesse d'Anville? According to Mossner and Ross (in Smith 1987, 111 n.3), Marie-Louise-Nicole—with her son, the young Duc de La Rochefoucauld—met Smith in Geneva at the end of 1765.

44. Although Alcouffe and Massot-Bordenave (2020, 262) provide additional details about Madame Nicol's husband, Jacques Nicol de Montblanc, a wealthy Anglophile Frenchman who presided over the Mont Blanc Estate in the present Croix Daurade district of Toulouse, they do not provide any further details about Madame Nicol.

45. Smith and Duke Henry were subsequently joined by the Duke's younger brother, Hew Campbell Scott.

46. See the timeline in Bonnyman 2014, xiii–xiv; see also chs. 4 and 5.

financing their Grand Tour, requesting permission to relocate to Paris.[47] Was Smith hoping to leave Toulouse to avoid public scrutiny or to arrange a *rendezvous* with Madame Nicol, i.e., away from her husband, Capitoul Nicol?

Either way, Charles Townsend granted Smith's request in a letter dated April 22, 1765,[48] but two points are worth noting. First, Smith and Duke Henry did not leave the South of France for good until the fall of 1765,[49] so for some unknown reason Smith was apparently in no hurry to leave Toulouse, after all (Alcouffe and Massot-Bordenave 2020, 285). Did this change of itinerary have anything to with the aforementioned Madame Nicol?

Secondly, Townsend warns his stepson in his April 22, 1765 letter "against any female attachment" (Ross 1974, 184). The relevant part of Townsend's April 22 letter reads as follows:

> If you go much into mixed company, as I suppose you will, let me warn you against any female attachment. Your rank & fortune will put women of subtle characters upon projects which you should not be the dupe of, for such connexions make a young man both ridiculous & unhappy. Gallantry is one thing; attachment is another; a young man should manifest spirit & decorum even in this part of his character, & preserve his mind entire & free in lesser as well as greater things. (Ross 1974, 184)

Townsend's warning to his stepson Duke Henry was not an academic or abstract admonition, for a fellow contemporary with personal knowledge of Townsend's habits and dispositions, Lady Louisa Stuart (1985/1827, 38), describes Duke Henry's stepfather as "a man of pleasure, a libertine." In any case, could Townsend's warning "against any female attachment" have been meant for Smith as well?

Sex in the City of Light: The Paris theater scene

Previously, we presented evidence regarding Madame Riccoboni. It is still

47. According to Alcouffe and Massot-Bordenave (2020, 283), it was Smith—not his pupils Henry and Hew—who wanted to relocate from Toulouse to Paris. Alcouffe and Massot-Bordenave even speculate that Smith was becoming "impatient."

48. This letter is reprinted in full by Ross (1974, 182–184) and Alcouffe and Massot-Bordenave (2020, 284–285).

49. Smith and Duke Henry, along with the Duke's younger brother Hew Campbell Scott, eventually returned to the City of Lights sometime during the month of December of 1765, where they resided until the month of October of that same year, until Hew's tragic and untimely death. *Cf.* Ross (1995, 209): "it has been assumed that Smith and his pupils [Hew and Henry] travelled to Paris from Geneva in December 1765." I thank Jonathan Wight (2020) for helping me clear up the actual date of Smith's return to Paris. See also the timeline in Alcouffe and Massot-Bordenave 2020, xiii–xiv.

worth mentioning that Riccoboni, an accomplished actress and novelist, and Smith, an admirer of the stage, were by all accounts avid theater and opera fans, especially during Smith's stay in the City of Light.[50] Indeed, "it is very likely Smith took recommendations from Riccoboni as to which theatrical performances to attend" (Dawson 2018, 8), and so it is not far-fetched to imagine to them attending a play or opera or concert together.

Many Smith scholars have failed to mention that these theatrical venues were the center of an elite Parisian sexual marketplace, the famed *dames entretenues* or "kept women" of French high society (Kushner 2013). Famous for their talent, glamour, and beauty, these *femmes galantes* were the most highly sought-after women of pleasure in all Europe, models and actresses who "earned their living by engaging in long-term sexual and often companionate relationships with men from the financial, political, and social elites, known as *le monde* (high society)" (ibid., 3). This sultry scene overlapped directly with the world of the theater.[51] Although not all theater women were kept mistresses or *femmes galantes*, "[i]t was widely understood that any woman in the Opéra, and to a lesser degree the other theater companies, was a *dame entretenue*, or at least wanted to be" (ibid., 31). The world of theater was the center of this high-end sex market because "being on the stage greatly increased…'sexual capital,' the desirability of a mistress and hence the prices she could command for her services" (ibid., 5), and the theater district of the French capital was teeming with high-end brothels and places of ill repute.[52] But there is no evidence to indicate that Smith himself partook of any such transactions.

"Miss Campbell" is probably not Lady Frances Scott Campbell

The most tenuously conjectured love affair would be one during Smith's extended stay at Dalkeith House in late 1767. On this theory Smith may have carried out a short-lived love affair with the younger sister of his former pupil Henry Scott,[53] Lady Frances Douglas (1750–1817), née Campbell Scott, whose

---

50. By way of example, John Rae (1895, ch. 14) and Ian Simpson Ross (2010, ch. 13), scholars who have produced two of the most comprehensive biographies of Smith, both commented on Smith's fondness for the opera during his second sojourn in Paris.

51. Kushner (2013, 4–5): "About a fifth of the kept women under police surveillance at midcentury worked in the theater. Most were in the Opéra or its school, as dancers and singers."

52. In the words of Kushner (2013, 110), "Many brothels were in the center of town, on the rue St. Honoré or nearby, making them convenient for men leaving the Opéra."

53. Recall that Smith was Duke Henry's private tutor during their Grand Tour from February 1764 to October 1766. As a further aside, Edith Kuiper (2013, 70) incorrectly identifies Lady Frances as Duke Henry's elder sister. In fact, Lady Frances was four years younger than her brother (see Alcouffe and Massot-Bordenave 2020, 32–33; Bonnyman 2014, 1).

mother's maiden name was Campbell and whose other brother, Hew, was referred to as Hew Campbell Scott.[54] Lady Frances would have been 17 years old at the time (Stuart 1985/1827, 54); as a result, such an affair, if it really occurred, would be one between a man of 44 and young woman of 17. Lady Frances did not wed until 1783.[55]

Also worth noting is that Smith may have first met Lady Frances nine years earlier, when Duke Henry's stepfather Charles Townsend "visited Scotland in the summer of 1759 with Lady Dalkeith [Caroline Campbell Scott] and her daughter [Lady Frances]" (Ross 1974, 179). Among other places on his Scottish itinerary, Townsend went to Glasgow "to make the necessary arrangements for the period five years ahead when the duke would…complete his studies by travelling on the Continent with his tutor" (Ross 1974, 179).

Of particular relevance to this conjecture is Mackenzie's recollection about "Smith and Hume in Love," in which he partially identifies by name a "Miss Campbell of __________" as the object of Smith's romantic affections. Could this "Miss Campbell" be a veiled or indirect reference to Lady Frances, whose mother's maiden name was also Campbell? If so, her prominence and subsequent marriage in 1783 to Lord Archibald Douglas might explain Mackenzie's reluctance to identify her by name.[56]

Duke Henry, along with his sister Lady Frances and other members of the Buccleuch family, returned to the Buccleuch estates in Scotland and took up residency at Dalkeith House in September of 1767 upon Duke Henry's coming of age.[57] According to one scholarly source (Bonnyman 2014, 59–60), Smith resided

54. The mother of Lady Frances and her brothers Henry and Hew was Caroline Campbell Scott, Lady Dalkeith (Ross 1974, 178). For more details about Caroline Campbell and her family background, see Bonnyman 2014, 9, 12–19.

55. *See* Alcouffe and Massot-Bordenave 2020, 290. When she married in 1783, at the age of 33, Lady Frances became Lord Archibald Douglas' second wife (see Jill Rubenstein's biography of Lady Frances in Stuart 1985/1827, 11). For a short biography of Lord Douglas, see Valentine 1970, 1:258. As a further aside, Lady Frances, upon reaching the age of 21, was entitled to an income from her family of £600 per year (see Stuart 1985/1827, 59). To put this monetary amount in perspective, Smith's compensation for serving as the Duke's private tutor during their Grand Tour from 1764 to 1766 consisted of an annual salary of £500, plus travel expenses, and a pension of £300 a year thereafter (see Mackay 1896, 237; Heilbroner 1999, 48; Bonnyman 2020, 41).

56. Recall that Mackenzie did not jot down this particular passage about "Miss Campbell" until the end of his life—the late 1820s or early 1830s. Also, as an aside, in the very same passage in which Mackenzie mentions "Miss Campbell" as the object of Smith's affections, Mackenzie also writes about David Hume's love life, but instead of partially disclosing or withholding the identity of Hume's romantic attachment, as he does with Adam Smith, Mackenzie fully identifies Hume's love interest by name as "Miss Nancy Ord" (Mackenzie 1927, 176; reprinted in Fieser 2003, 255).

57. Dalkeith House (or Dalkeith Palace) was the Buccleuch family's principal residence in Scotland and is located only four miles south of Edinburgh, where Smith lived the last 12 years of his life. This neo-classical palace was commissioned and built in the early eighteenth century and then refurbished in preparation for

at Dalkeith House in the fall of 1767, a stay that coincides with Lady Frances's residency there.[58] Moreover, Smith's stay at Dalkeith House lasted at least two months, from mid-September to mid-November 1767.[59] Thus the possibility of a love affair between Smith and the young Lady Frances during this time, though unlikely, is not altogether inconceivable.[60] This conjecture, however, is doubtful. In his brief recollection, Mackenzie says that "Miss Campbell" was "a woman of as different dispositions and habits from him as possible." That he would make such a remark without also remarking on other extraordinary aspects of an affair with Lady Frances would be odd, unless he purposely wanted to disguise her identity. In that case, however, why identify her as "Miss Campbell" at all? In a parenthetical remark immediately following the words "Miss Campbell of ____________" (see quoted text in Mackenzie 1927, 176, reprinted in Fieser 2003, 255), Mackenzie says that "the name is so numerous that to use it cannot be thought personal." This comment further suggests that "Campbell" tells little or nothing about the woman's identity. Also worth noting here is the reference to "*Miss* Campbell" instead of *Lady* or *Dame* or some other signifier of aristocratic distinction. Perhaps this reference to "Miss" points us away from Lady Frances.

That said, however, two additional points are worth making regarding the relationship between Lady Frances and Smith. Firstly, among the letters of Smith that were not lost or destroyed are three addressed directly to Lady Frances, dated October 15, 1766, October 19, 1766, and March 17, 1783 (Smith 1987, letters 97, 98, and 225). The two of 1766 precede Smith's visit to Dalkeith, and concern

Duke Henry's first visit to Scotland in honor of his reaching the age of majority (see Bonnyman 2014, 10, 57–58). For an engraving of this palace as it appeared in the early eighteenth century, see **here**.

58. In her memoir of her cousin and close friend Lady Frances, Lady Louisa Stuart (1985/1827) confirms that Lady Frances had been growing up in London under the watchful eye of her stepfather Charles Townsend. Lady Stuart (1985/1827, 48–50) also describes the circumstances surrounding Lady Frances's return to Scotland.

59. For further details regarding Smith's stay at Dalkeith, see Bonnyman 2014, 58–59. According to Ross (1974, 180), Heilbroner (1999, 50), and Kuiper (2013, 76, n.14), Smith may have also visited Dalkeith House and the Buccleuch estates on several subsequent occasions. None of these sources, however, provide any actual evidence in support of this proposition, so it is unclear whether Smith, in fact, ever returned to or stayed at Dalkeith House following his initial two-month stay in late 1767.

60. For a biography of Lady Frances written by a contemporary of hers, see Stuart 1985/1827. Among other things, Lady Louisa Stuart's intimate memoir of Lady Frances's life and circle of family and friends paints a very unflattering picture of Lady Frances's mother; in addition, Stuart (1985/1827, 45–49) describes Lady Frances's stepfather Charles Townsend as extremely possessive: "unwilling ever to have her out of his sight." Stuart's intimate memoir also highlights Lady Frances's worldliness and awareness of adult double standards (see especially ibid., 54, 58). Given these facts, along with her stepfather's (Charles Townsend) sudden death in August of 1767, perhaps it is not far-fetched to imagine a romance or fling between her and Smith. For a portrait of Lady Frances by Sir Joshua Reynolds when she was still a child, see **here**. As a further aside, Lady Frances was an accomplished artist in her own right; a collection of her works is available from Tate (**link**).

the melancholy news of the death, in France, of Campbell Scott, brother of both Lady Frances and the Duke of Buccleuch. The third letter, of 1783, is brief and somewhat curious. It is thought to be authored by Smith, but it refers to Smith in the third person. Specifically, it thanks Lady Frances for having sent to Smith "his paper on upon Italian and English verse" and promises "to send her a more perfect copy as soon as he has compleated his plan" (Smith 1987, 265). It would seem that Smith had shared his essay on Italian and English verse, which is contained in the modern edition of *Essays on Philosophical Subjects*.

Secondly, Lady Louisa Stuart (1985/1827, 90–93)—who wrote an intimate memoir of her cousin and best friend, Lady Frances—mentions toward the end of her memoir that she (Lady Frances) may have had a mysterious lost love of her own. Or in the melodramatic words of Lady Louisa herself: Lady Frances had at one time in her life fallen "prey [to] one of those fixed, deep-rooted torturing passions" (ibid., 90). Could this lost love have been Smith?

For context, at this point in her Lady Frances memoir, Lady Louisa is writing about Frances Scott's decision to marry a widower, Lord Archibald Douglas.[61] According to Lady Louisa, Lady Frances confided to her:

> …at one time I could not doubt that he was extremely inclined to make me his proposals, and would have done so on a very little encouragement. But no—Oh no!—I had just—only just enough reason left to see that this must not be—that we must never marry—it would have been madness to think of it—And I withstood the temptation like a famished wretch refraining from the food he knows to be mingled with poison. (Stuart 1985/1827, 91)

Who is the "he" to whom Lady Frances is referring to here? Could it be none other than Adam Smith? Alas, not only did Lady Frances fail to reveal the identity of this lost love to Lady Louisa; she explicitly forbade Lady Louisa "to guess who it was" (Stuart 1985/1827, 93).[62] Nevertheless, although Lady Louisa reports that Lady Frances never gave her "the least clue to discover who this person was" (ibid., 92), Lady Louisa does identify three possible clues in her memoir of Lady Frances, none of which rule Smith out.

The first clue is a temporal one. Writing from the vantage point of 1781 or 1783,[63] Lady Louisa (1985/1827, 90) says that this love affair originated many years

61. Lord Archibald and Lady Frances wed on May 1, 1783 (see Stuart 1985/1827, 95). Lady Frances was Lord Archibald's second wife.

62. Lady Louisa further states: "I confess I had an eager desire to know [the identity of Lady Frances' lost love]; but so far from gratifying my curiosity, [Lady Frances] besought me…to repress it, desist from enquiring, and forbear to even form a conjecture" (Stuart 1985/1827, 92).

63. Although this part of the memoir deals with "the momentous year 1783" (Stuart 1985/1827, 85), the year Lady Frances wed Lord Archibald Douglas, Lady Louisa also mentions an anecdote from the year

ago but "not less than twelve or fourteen" years. In other words, Lady Frances' mysterious love affair occurred in the late 1760s or early 1770s, right after Smith's stay at Dalkeith house in the fall of 1767.

The next clue involves the wavering religious beliefs of Lady Frances' lost love. According to Lady Louisa (Stuart 1985/1827, 91), this lover was "an unbeliever, almost to the extent of atheism."[64] Although Smith's private religious beliefs are unknown, since he was careful during his life to keep his views about religion to himself, it is not inconceivable that Smith confided his private views to someone he was romantically attached to. (Indeed, perhaps it was Smith's secularism—especially in an age of religious conformity—that prevented his love interests from going any further.)

The last clue approaches closest to Smith. According to Lady Louisa:

> The only thing [Lady Frances] ever let fall which might have led to [a conjecture] was a circumstance mentioned by chance in speaking of Mr. Townshend [Lady Frances's stepfather]. She had told him, she said, that she felt quite sure she should never be in love—the persuasion of most sensible young people before their hour is come—"Yes, you will" answered he gazing at her pensively—"Your brother [Duke Henry] will come from abroad, and amongst his young friends your eye will single out some man destined to be master of your fate"—"Good heavens!" added she with energy, putting her hand to her forehead—"One would actually suppose he had been endowed with the spirt of prophecy." (Stuart 1985/1827, 92–93)

Although Smith could not be considered one of Duke Henry's "young friends," Smith did return with Duke Henry from abroad, resided at Dalkeith House with Duke Henry and Lady Frances in the fall of 1767, and was by all accounts (Bonnyman 2014; Alcouffe and Massot-Bordenave 2020) Duke Henry's closest friend and confidante at the time.

To sum up our conjectures thus far: all the available pieces of evidence, scarce as they are, as well as the most reasonable inferences that can be drawn from these sources, point to the existence of two or perhaps three love interests. One is Stewart's mysterious maiden of Smith's youth, a time and place in which people's love lives were strictly monitored by Church elders and in which parental consent for marriage was the norm. Another is Colbert's Madame Nicol in Old Regime France, the France of Louis XV, literary salons, and the *demimonde*. A third is "Miss

1781 involving Lady Portarlington (ibid., 92).

64. For the reader's reference, here is the complete passage in which this clue appears (Stuart 1985/1827, 91): "One particular [Lady Frances] mentioned [to Lady Louisa], that in conversation which passed between them [between Lady Frances and her lover], he had frankly avowed himself an unbeliever, almost to the extent of atheism."

Campbell," though that could be anyone, including possibly the Lady of Fife.

Additional conjectures

For thoroughness, we may also consider two additional conjectures about Smith's sexuality and love life. One is that Smith was romantically involved with his unmarried cousin Janet Douglas. Miss Douglas had moved into Smith's household as early as 1754 (Özler 2012, 348) and lived under the same roof with Smith and Smith's mother—Douglas's aunt, Margaret Douglas Smith—until her death in 1788 (Kennedy 2005, 5).[65] Given the decades that they lived in the same household, this thought is something that a nosy sleuth must ponder,[66] though there is no evidence in support of the speculation beyond these circumstantial facts. Indeed, the lack of evidence for what would be such a protracted and proximate intimacy constitutes strong evidence against the idea.

A final hypothesis is that Smith was not attracted to women.[67] Daniel Klein has pointed out to me in conversation how some of the surviving correspondence between Smith and Hume (see especially Smith 1987, letters 70, 88, 92, and 121), as well as the remarkable frequency in Smith's works of subtle, playful, inside-joke-like textual connections to Hume's works, suggests a 'bromance' and brotherly tenderness between them, although Klein clarifies to me that he does not mean to suggest a sexual relationship. In a letter dated February 22, 1763, Smith invites Hume to visit him in Glasgow thus: "Tho you have resisted all my Sollicitations, I hope you will not resist this" (Smith 1987, 139). And in a letter dated September 1765, Smith writes to Hume:

> In short I have a very great interest in your settling at London, where, after many firm resolutions to return to Scotland, I think it is most likely I shall settle myself. Let us make short excursions together sometimes to see our friends in France and sometimes to see our friends in Scotland, but let London be the place of our ordinary residence. (Smith 1987, 161)

---

65. Alas, scholars have been unable to confirm Janet Douglas's year of birth (Kuiper 2013, 62).

66. Weinstein (2001, 10) says that Miss Douglas and Smith were "quite close." Weinstein, however, does not suggest that Janet Douglas and Smith were romantically involved—only that Smith "seemed to have good relationships with women as his time and stature would have allowed." For what it is worth, Voltaire, who Smith admired, is alleged to have had an affair with his niece, Madame Denis, during their younger years (see, for example, Alcouffe and Massot-Bordenave 2020, 171 n.8). Nevertheless, such types of incestuous relationships were strictly forbidden by the Church of Scotland (Hardy 1978), and during most of Smith's lifetime the sex lives of Scots were strictly monitored by the local 'kirk sessions' or ecclesiastical courts of each parish (Mitchison and Leneman 1998).

67. Regarding speculation that Smith may have been gay, both Edith Kuiper (2013, 70) and Gavin Kennedy (2005, 4) conclude that it is unfounded; see, also, a fascinating and thoughtful discussion on the Reddit website r/AskHistorians (**link**).

Absence of evidence or evidence of absence? Smith's lost diary and the destruction of his private papers

Smith's letters

Is there any evidence of a love affair written in Smith's own hand? In an early essay Smith wrote on the "Imitative Arts" (1980b/1795, 190), Smith states: "It is a lover who complains, or hopes, or fears, or despairs." Also, some references to romantic love appear in various places in *The Theory of Moral Sentiments*, but on his deathbed Smith specifically instructed his literary executors, Joseph Black and James Hutton, to destroy his unpublished manuscripts, correspondence, and other private papers (Ross 2010, 404–405). In fact, Smith may have insisted on the destruction of his private papers and letters as early as 1773 (Phillipson 2010, 279), when he had made his first will and had appointed his friend David Hume his executor. Nevertheless, despite Smith's desire to have his private papers and personal letters destroyed upon his death, a small sample of Smith's correspondence still survives, including three letters addressed to Lady Frances. According to Alcouffe and Massot-Bordenave (2020, x), in all "only 193 letters written by [Smith] and 129 addressed to him remain."[68]

The key question is, Why did Smith want to destroy his private letters and papers as early as 1773, only a few years removed after his extended sojourn in the South of France (1764–1765), his ten-month residency in the City of Lights (1766), and his two-month stay at Dalkeith House (1767)? Most scholars, like Ross (2010, 405), point to "Smith's prudence" and "his concern for his literary reputation" as the motivating factors.[69] Alcouffe and Massot-Bordenave (2020, 133) write: "it is important to underline also that the place ascribed to the judgement of posterity is found in the arrangements which he [Smith] to make prior to his death; to make his personal papers disappear and thus to control the image which posterity would later preserve of him."

68. Most of these remaining 322 letters (304 to be exact) are reprinted in Smith 1987. Also, according to W. R. Scott (1940, App. II 272, 273): "there is just a possibility that a large body of documents relating to Smith may still be in existence" and "there remain opportunities, even at this late date, for remedying the present meagre knowledge of Adam Smith's life." Tracking down any new Smith letters addressed to Lady Frances, however, will be a daunting task. Lady Frances, for example, eventually married Archibald Douglas in 1783, and the couple had six children (see Jill Rubenstein's biography of Lady Frances, in Stuart 1985/1827).
69. Likewise, a reviewer has suggested to me that Smith simply wanted to avoid any airing or public scrutiny of his private views about "religion, politics, his correspondence with Hume, etc."

The lost travel diary

It has been reported that Smith brought back no less than three trunks of documents from his extended travels in France (Alcouffe and Massot-Bordenave 2020, 204). Also, according to Jeremy Black (1981, 657), the custom of "keeping a travel diary of some form" among the British during their Grand Tours was "relatively widespread." In addition, Smith scholar W. R. Scott (1940, App. II 273) has speculated that Smith or his pupil (or both?) may have kept a travel diary during their extended travels in the South of France and Paris in the mid-1760s (see also Ross 2010, 248 n.2; Rasmussen 2017, 286 n.61). In an unfinished appendix to his survey article titled "Studies Relating to Adam Smith during the Last Fifty Years," Scott (1940) specifically refers to the existence of this lost diary,[70] which was sold in the 1920s to an unknown buyer from an Edinburgh bookshop owned by one Mr Orr:

> Contrary to the report of Dugald Stewart, Mr. Orr, a bookseller of George Street, Edinburgh, maintained that Adam Smith did keep a diary when he was in France, and that he had had it in his possession and had sold it for cash to an unknown customer who was believed to be from one of the Dominions, or perhaps from the United States. (Scott 1940, App. II 273)

Scott (1940, 273) further reports in his survey article that he was personally able to interview the employee who had made the actual sale and that this employee "was clear as to the…particulars": the fact that Mr Orr's bookshop had at one time a copy of Smith's travel diary and had sold it for cash to an unknown buyer.[71] Scott further speculates about the identity of this unknown buyer:

> It may be guessed that the purchaser cannot have been an economist, else he would surely have printed extracts from a manuscript of such interest. It may be he was a collector of autographs, in which case the tracing of the diary must be largely a matter of chance. (Scott 1940, App. II 274)

Perhaps this lost diary, if it exists at all, would contain details of Smith's personal life during his travels in France.

Given the absence of strong evidence, Ross (2010, 228) famously concluded

70. W. R. Scott (1868–1940), the Professor of Political Economy at the University of Glasgow, was a prolific writer and an authority on the life and works of Smith (see Coase 1993, 355). Given these impeccable credentials and his record of scholarship, Scott is a credible source of information.

71. Alas, this employee's memory as to the date of the sale was foggy. Specifically, Scott (1940, 273) reports the employee "was doubtful about the date of the transaction. In 1935 he thought it was over ten years earlier, and last year [1939] he put it back to 'nearly twenty years ago.'"

that "the biographer can do little more with the topic of Smith's sex life than contribute a footnote to the history of sublimation." But is this sparse record evidence of absence? The little available evidence shows that most likely Smith fell in love at least twice, if not three times: that he had a romantic bond with the Lady of Fife and quite possibly one with Madame Nicol. That Mackenzie's "Miss Campbell" accounts for an actual and separate affair is less likely. Beyond that, who knows! Or, as Smith himself is once reported to have asked, "Am I beau to no one but my books?"[72]

References

Alcouffe, Alain, and Andrew Moore. 2018. Smith's Networks in Occitania—March 1764–October 1765. Presented at the 31st Annual Conference of Eighteenth-Century Scottish Studies Society, July 17–21, 2018 (Glasgow). **Link**

Alcouffe, Alain, and Philippe Massot-Bordenave. 2020. *Adam Smith in Toulouse and Occitania: The Unknown Years*. London: Palgrave Macmillan.

Black, Jeremy. 1981. British Travellers in Europe in the Early Eighteenth Century. *Dalhousie Review* 61(4): 655–667.

Black, Jeremy. 1984. France and the Grand Tour in the Early Eighteenth Century. *Francia* 11: 407–416.

Black, Jeremy. 2011 [1985]. *The British and the Grand Tour*. London: Routledge.

Bonnyman, Brian. 2014. *The Third Duke of Buccleuch and Adam Smith: Estate Management and Improvement in Enlightenment Scotland*. Edinburgh: Edinburgh University Press.

Braden, Irene. 1965. *Voltaire and Injustice*. Master's thesis, Kansas State University, Manhattan, Kan. **Link**

Brodsky-Porges, Edward. 1981. The Grand Tour: Travel as an Educational Device, 1600–1800. *Annals of Tourism Research* 8(2): 171–186.

Buchan, James. 2006. *The Authentic Adam Smith: His Life and Ideas*. New York: Norton.

Caplan, Harry, Walter H. French, and Francis E. Mineka. 1964. Harold William Thompson, June 5, 1891–February 21, 1964. Cornell University (Ithaca, N.Y.) **Link**

Caulcutt, Clea. 2010. French Free-Thinking Knight Still a Controversial Figure. RFI.fr (France Médias Monde, Paris), December 16. **Link**

Chassaigne, Marc. 1920. *Le procès du Chevalier de la Barre*. Paris: Librairie Victor Lecoffre.

Claverie, Elisabeth. 1992. Sainte indignation contre indignation éclairée: L'affaire du Chevalier de La Barre. *Ethnologie Française* 22(n.s., 3): 271–290.

Claverie, Elisabeth. 1994. Procès, affaire, cause: Voltaire et l'innovation critique. *Politix* 7(26): 76–85.

Coase, Ronald H. 1993. Duncan Black, 1908–1991. *Proceedings of the British Academy* 82:

72. Quoted in Wight 2002, 292. For a slightly different formulation of this possibly apocryphal quotation, see Smellie 1800, 297 (quoted in Rae 1895, 329; Heilbroner 1999, 45): "I am a beau in nothing but my books."

353–365.

Cohen, Michèle. 1992. The Grand Tour: Constructing the English Gentleman in Eighteenth-Century France. *History of Education* 21(3): 241–257.

Cohen, Michèle. 2001. The Grand Tour: Language, National Identity, and Masculinity. *Changing English: Studies in Culture and Education* 8(2): 129–141.

Colbert, Abbé. 1766. Letter dated 18 September 1766. *National Archives of Scotland*, GD224/2040/62/3.

Currie, James. 1831. Letter 87 ("To Dugald Stewart, Esq., Edinburgh, July 14, '94, Respecting Dr. Adam Smith"). In *Memoir of the Life, Writings, and Correspondence of James Currie, M.D. F.R.S., of Liverpool*, ed. William Wallace Currie, vol. 2, 317–320. London: Longman, Rees, Orme, Brown, & Green.

Darnton, Robert. 1998. Two Paths Through the Social History of Ideas. *Studies on Voltaire and the Eighteenth Century* 359: 251–294.

Dawson, Deidre. 2013. Love, Marriage and Virtue: Mary Wollstonecraft and Sophie de Grouchy, Marquise de Condorcet, Respond to The Theory of Moral Sentiments. *Adam Smith Review* 7: 24–46.

Den Uyl, Douglas J., and Charles L. Griswold Jr. 1996. Adam Smith on Friendship and Love. *Review of Metaphysics* 49(3): 609–637.

Drescher, H. W. 2004. Mackenzie, Henry (1745–1831). In *Oxford Dictionary of National Biography*, ed. David Cannadine. Oxford: Oxford University Press.

Fay, C. R. 2011 [1956]. *Adam Smith and the Scotland of his Day*, reissue ed. Cambridge, UK: Cambridge University Press.

Fieser, James. 2003. *Early Responses to Hume*, vol. 10. London and New York: Thoemmes Press.

Gillis, John R. 1985. *For Better, for Worse: British Marriages, 1600 to the Present*. Oxford: Oxford University Press.

Ginzberg, Eli. 1934. *The House of Adam Smith*. New York: Columbia University Press.

Goodman, Dena. 1994. *The Republic of Letters: A Cultural History of the French Enlightenment*. Ithaca, N.Y.: Cornell University Press.

Haakonssen, Knud. 1984. From Moral Philosophy to Political Economy: The Contribution of Dugald Stewart. In *Philosophers of the Scottish Enlightenment*, ed. Vincent Hope, 211–232. Edinburgh: Edinburgh University Press.

Haldane, R. B. 1887. *Life of Adam Smith*. London: Walter Scott.

Hardy, John R. 1978. *The Attitude of the Church and State in Scotland to Sex and Marriage: 1560–1707*. M. Phil. thesis, University of Edinburgh. **Link**

Harkin, Maureen. 2013a. Adam Smith on Women. In *The Oxford Handbook of Adam Smith*, eds. Christopher J. Berry, Maria Pia Paganelli, and Craig Smith, 501–520. Oxford: Oxford University Press.

Harkin, Maureen. 2013b. Adam Smith on Women: Introduction. *Adam Smith Review* 7: 3–7.

Heilbroner, Robert L. 1999. *The Worldly Philosophers: The Lives, Times, and Ideas of Great Economic Thinkers*, 7th rev. ed. New York: Touchstone.

Hirst, Francis Wrigley. 1904. *Adam Smith*. London: Macmillan.

Jacob, Margaret C. 2019. *The Secular Enlightenment*. Princeton, N.J.: Princeton University

Press.

Kennedy, Gavin. 2005. *Adam Smith's Lost Legacy*. Basingstoke, UK: Palgrave Macmillan.

Kuiper, Edith. 2013. The Invisible Hands: Adam Smith and the Women in his Life. *Adam Smith Review* 7: 62–78.

Kushner, Nina. 2013. *Erotic Exchanges: The World of Elite Prostitution in Eighteenth-Century Paris*. Ithaca, N.Y.: Cornell University Press.

LeBuffe, Michael. 2020. Paul-Henri Thiry (Baron) d'Holbach. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University (Stanford, Calif.). **Link**

Leddy, N. B. 2013. Grave, Philosophical, and Cool Reasoner: Mary Wollstonecraft on the Use of Gender in Adam Smith. *Adam Smith Review* 7: 8–17.

Leneman, Leah. 1999. Wives and Mistresses in Eighteenth-Century Scotland. *Women's History Review* 8(4): 671–692.

Leneman, Leah, and Rosalind Mitchison. 1987. Scottish Illegitimacy Ratios in the Early Modern Period. *Economic History Review* 40(1): 41–63.

Leneman, Leah, and Rosalind Mitchison. 1988. Girls in Trouble: The Social and Geographical Setting of Illegitimacy in Early Modern Scotland. *Journal of Social History* 21(3): 483–497.

Leneman, Leah, and Rosalind Mitchison. 1993. Clandestine Marriage in the Scottish Cities, 1660–1780. *Journal of Social History* 26(4): 845–861.

Luna, Francesco. 1996. From the *History of Astronomy* to the *Wealth of Nations*: Wonderful Wheels and Invisible Hands in Adam Smith's Major Works. In *Inflation, Institutions and Information: Essays in Honour of Axel Leijonhufvud*, eds. Daniel Vaz and Kumaraswamy Velupillai, 131–153. London: Macmillan.

Mackay, A. J. G. 1896. *A History of Fife and Kinross*. Edinburgh and London: William Blackwood and Sons. **Link**

Mackenzie, Henry. 1927 [1831]. Smith and Hume in Love. In *The Anecdotes and Egotisms of Henry Mackenzie, 1745–1831*, ed. Harold W. Thompson, 176. Oxford: Oxford University Press.

Mitchison, Rosalind, and Leah Leneman. 2001. *Girls in Trouble: Sexuality and Social Control in Early Modern Scotland, 1670–1780*, 2nd ed. Dalkeith, UK: Scottish Cultural Press.

Muller, Jerry Z. 1993. *Adam Smith: In His Time and Ours*. Princeton, N.J.: Princeton University Press.

Nicholls, James C. 1976. Introduction to *Mme Riccoboni's Letters to David Hume, David Garrick and Sir Robert Liston, 1764–1783*, by Marie-Jeanne Riccoboni, ed. James C. Nicholls. Oxford: Voltaire Foundation at the Taylor Institution.

Özler, Şule. 2012. Adam Smith and Dependency. *Psychoanalytic Review* 99(3): 333–358.

Phillipson, Nicholas. 2010. *Adam Smith: An Enlightened Life*. New Haven, Conn.: Yale University Press.

Rae, John. 1895. *Life of Adam Smith*. London: Macmillan.

Rashid, Salim. 1985. Dugald Stewart, "Baconian" Methodology, and Political Economy. *Journal of the History of Ideas* 46(2): 245–257.

Rasmussen, Dennis C. 2017. *The Infidel and the Professor: David Hume, Adam Smith, and the Friendship That Shaped Modern Thought*. Princeton, N.J.: Princeton University Press.

Ross, Ian Simpson. 1974. Educating an Eighteenth-Century Duke. In *The Scottish Tradition: Essays in Honour of Ronald Gordon Cant*, ed. G. W. S. Barrow, 178–197. Edinburgh: Scottish Academic Press.

Ross, Ian Simpson. 1980. Introduction to *Account of the Life and Writings of Adam Smith, LL.D.* by Dugald Stewart. In *Essays on Philosophical Subjects*, by Adam Smith, eds. W. P. D. Wightman and J. C. Bryce, 265–268. Oxford: Oxford University Press.

Ross, Ian Simpson. 1995. *The Life of Adam Smith*. Oxford: Clarendon Press.

Ross, Ian Simpson. 2010. *The Life of Adam Smith*, 2nd ed. Oxford: Oxford University Press.

Roy, Donald. 1995. Comédie-Italienne. In *The Cambridge Guide to the Theatre*, ed. Martin Banham, 233–234. Cambridge, UK: Cambridge University Press.

Scott, Walter. 1834. Henry Mackenzie. In *Miscellaneous Prose Works of Sir Walter Scott*, ed. Walter Scott, vol. 4, 1–19. Edinburgh: Cadell. **Link**

Scott, W. R. 1936. New Light on Adam Smith. *Economic Journal* 46(183): 401–411.

Scott, W. R. 1940. Studies Relating to Adam Smith During the Last Fifty Years. *Proceedings of the British Academy* 26: 249–274.

Smellie, William. 1800. *Literary and Characteristical Lives of John Gregory, M.D.; Henry Home, Lord Kames; David Hume, Esq.; and Adam Smith, L.L.D.* Edinburgh: Alex. Smellie.

Smith, Adam. 1976 [1790]. *The Theory of Moral Sentiments*, eds. D. D. Raphael and A. L. Macfie. Oxford: Oxford University Press.

Smith, Adam. 1980a [1795]. The History of Astronomy. In *Essays on Philosophical Subjects*, ed. W. P. D. Wightman and J. C. Bryce, 33–105. Oxford: Oxford University Press.

Smith, Adam. 1980b [1795]. Of the Nature of That Imitation Which Takes Place in What Are Called the Imitative Arts. In *Essays on Philosophical Subjects*, ed. W. P. D. Wightman and J. C. Bryce, 176–209. Oxford: Oxford University Press.

Smith, Adam. 1987. *The Correspondence of Adam Smith*, 2nd ed., eds. Ernest C. Mossner and Ian Simpson Ross. Oxford: Oxford University Press.

Stewart, Dugald. 1811a. *Account of the Life and Writings of Adam Smith, LL.D.* In *Biographical Memoirs of Adam Smith, LL.D., of William Robertson, D.D., and of Thomas Reid, D.D.*, 3–152. Edinburgh: George Ramsay and Company.

Stewart, Dugald. 1811b. *Account of the Life and Writings of Adam Smith, LL.D.* In *The Works of Adam Smith, LL.D. and F.R.S. of London and Edinburgh*, ed. Dugald Stewart, vol. 5, 403–552. London: T. Cadell and W. Davies.

Stewart, Dugald. 1980 [1811]. *Account of the Life and Writings of Adam Smith, LL.D.*, ed. Ian Simpson Ross. In *Essays on Philosophical Subjects*, by Adam Smith, eds. W. P. D. Wightman and J. C. Bryce, 269–351. Oxford: Oxford University Press.

Stuart, Lady Louisa. 1985 [1827]. *Memoire of Frances, Lady Douglas*, ed. Jill Rubenstein. Edinburgh: Scottish Academic Press.

Tegos, Spyridon. 2019. Is Love Ridiculous??? *Adam Smith Works* (Liberty Fund, Inc., Carmel, Ind.), February 11. **Link**

Valentine, Alan. 1970. *The British Establishment, 1760–1784: An Eighteenth-Century Biographical Dictionary*, vols. 1 and 2. Norman, Okla.: University of Oklahoma Press.

Voltaire. 2000 [1766]. An Account of the Death of the Chevalier de La Barre. In *Treatise on Tolerance and Other Writings*, ed. Simon Harvey, 139–148. Cambridge, UK: Cambridge University Press.

Voltaire. 2000 [1775]. Cry of Innocent Blood. In *Treatise on Tolerance and Other Writings*, ed. Simon Harvey, 151–156. Cambridge, UK: Cambridge University Press.

Weinstein, Jack Russell. 2001. *On Adam Smith*. Belmont, Calif.: Wadsworth Publishing Co.

Wight, Jonathan B. 2002. *Saving Adam Smith: A Tale of Wealth, Transformation, and Virtue*. Upper Saddle River, N.J.: Prentice Hall.

Wight, Jonathan B. 2020. Email correspondence with the author, November 8.

Wood, Paul. 2000. Dugald Stewart and the Invention of "The Scottish Enlightenment." In *The Scottish Enlightenment: Essays in Reinterpretation*, ed. Paul Wood, 1–35. Rochester, N.Y.: University of Rochester Press.

About the Author

F. E. Guerra-Pujol received his J.D. from Yale Law School and teaches business law and ethics at the University of Central Florida. His areas of research include markets, property rights, and the history of legal ideas. Dr. Guerra-Pujol is the author of many scholarly papers and book chapters including "Gödel's Loophole," "Buy or Bite?," and "The Poker-Litigation Game," and he is currently writing a book on probability and the law. His email address is fegp@ucf.edu.

Go to archive of Watchpad section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6031/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 156–163

Knowledge and Humanity: The History of Economic Thought as a Refined Liberal Art

Kevin Quinn[1]

LINK TO ABSTRACT

My title refers to two of three components of David Hume's "indissoluble chain"—the other being "industry"—that he delineates in his essay "Of Refinement in the Arts."[2] Hume's essay addresses the civic republican argument against commercial society hinging on the purported undermining of civic virtue fostered by the luxury that commerce brings in its train.[3] In the essay Hume notes: "Another advantage of industry and of refinements in the mechanical arts, is that they commonly produce some refinements in the liberal" (1987/1777, 270). He writes:

> The more these refined arts advance, the more sociable men become: nor is it possible, that, when enriched with science, and possessed of a fund of conversation, they should be contented to remain in solitude, or live with their fellow-citizens in that distant manner, which is peculiar to ignorant and barbarous nations. They flock into cities; love to receive and communicate knowledge; to show their wit or their breeding; their taste in conversation or

1. Bowling Green State University, Bowling Green, OH 43403.
2. Hume (1987/1777, 271): "Thus industry, knowledge and humanity are linked together by an indissoluble chain, and are…peculiar to the more polished, and, what are commonly denominated, the more luxurious ages."
3. A second aim of the essay, on the other hand, is to distinguish the defense of luxury he gives from that given by Mandeville, whom he does not name but clearly has in mind, in the latter's *Fable of the Bees*. Hume's is not, he says, by way of distinction, a defense of "vicious" luxury (Hume 1987/1777, 269). The differences with Mandeville are small, however, compared to his differences with the civic republican condemnation of luxury.

> living, in clothes or furniture. Curiosity allures the wise; vanity the foolish; and pleasure both. Particular clubs and societies are every where formed: Both sexes meet in an easy and sociable manner; and the tempers of men, as well as their behavior, refine apace. So that, beside the improvements which they receive from knowledge and the liberal arts, it is impossible but they must feel an encrease of humanity, from the very habit of conversing together, and contributing to each other's pleasure and entertainment. (Hume 1987/1777, 271)

In this wonderful passage, Hume is describing, I would argue, a new paradigm of *civilization*—sociable, democratic, urbane, scientific, conversational, gender-equal, and, most important, knowledge-engendering—that puts to shame Classical civilization, the lament for the passing of which underlies the civic republican dismissal of commercial and industrial progress. I would argue, moreover, that the knowledge and humanity Hume describes here are not just, *chez* Hume, "another advantage" of the growth of industry and commerce, but the very point of the latter. I mean that it is for the sake of the former that Hume welcomes the latter. It is not the growing prosperity that commercial society produces that justifies the latter, but the knowledge and humanity, the *civilization*, that such prosperity may underpin.[4] Industry is the *means*, while knowledge and humanity are the *ends* in Hume's triad.

In what follows, I argue that the study of the history of economic thought, itself a sort of industry, can exemplify the indissoluble link between knowledge and humanity that Hume describes. I share an experience from my career that serves as well to memorialize a friend whose work and life particularizes that link.

There are different ways of pursuing the history of economic thought, and different answers have been given to the question 'Why study the history of economic thought?'—where the context I have in mind is whether such study should be part of an economics curriculum.

I grew up in Maryland, and when I was in high school and thinking about college I was attracted to the curriculum of St. John's College in Annapolis, only a few miles from my home. The model they use is unique: all students pursue the same curriculum, which involves learning physics, for example, by reading the classic works that defined the field, beginning with Isaac Newton. Mathematics at St. John's starts with Euclid. The curriculum was completely text- and history-centered. When I suggested to my father, a physicist, that this seemed like an interesting approach, he was aghast. Pick up the latest modern physics text, he said, and you will find anything that was correct in the work of earlier physicists. And

4. I say 'may underpin' to indicate that, to the extent that it fails to do so, there would be, on my reading, a Humean rationale for reforms to commercial society that make it better able to underpin this Humean ideal of civilization.

why spend any time at all learning what was erroneous? The history becomes, as far as what it adds in value, only a history of errors, a waste of time. I did not go to St. John's.

Now, my father's Whiggish view has more plausibility when it comes to the natural sciences. It would be absurd, on the other hand, to apply it to the study of literature. Imagine: whatever was correct in Shakespeare you will find in the best contemporary playwright, so why bother with him? The study of literature is essentially connected to the study of texts which define its history, in a way that the study of physics is not.

It might seem that, of the two, economics better fits the physics model. It is a study of the world, not essentially a study of texts. But there is a third model of the relationship between a discipline and its history that I think is a better fit than either of the two—philosophy. It can be argued that the study of philosophy is much more closely tied to the study of the history of philosophy than we see with the natural sciences. Reading Plato on 'the good' adds value beyond just seeing the errors that the best contemporary moral philosopher has wisely avoided; nevertheless, unlike literature, some would say, philosophers would appear, like scientists, to be investigating aspects of the world.

Now many modern philosophers would disagree with this characterization of their discipline. The natural science model—especially in the Anglo-analytic tradition—is quite appealing and carries a great deal of status. There would be more agreement, I think, among philosophers in the Continental tradition. And I chose Plato on 'the good' here on purpose. It seems to me that moral philosophers[5] would be more inclined to agree with my picture than would, say, logicians.

Whether they know it or not, economists have always been in the business of doing moral philosophy insofar as they are arguing, implicitly or explicitly, for a view of the good life and the good society. As you might guess, the poster person for my account is Adam Smith, since he was quite explicitly a moral philosopher—and you do not understand his thought if you neglect *The Theory of Moral Sentiments*. But there are so many others: Hume, John Stuart Mill, Karl Marx, Thorstein Veblen, Friedrich Hayek, John Maynard Keynes, and Amartya Sen come to mind as economists who wear their moral philosophical *bona fides* on their sleeves, as it were.

Moral philosophy is more intimately related to its history than physics, I think, because the questions it is concerned with are essentially contestable and the answers to them essentially plural. My view (Quinn 2019) is that the world contains more than the facts that natural scientist finds; that moral dialogue is an attempt to

5. And political philosophers. Rawls's *Lectures on the History of Moral Philosophy* (2000), for example, is a philosophical work.

'get things right' about such further aspects of the world; and that it is a fact of the matter that values are plural and to some degree incommensurable.[6] It seems to me that an appropriate model for thinking about ethical matters is *conversational*.

On conversing with past thinkers, I cannot resist quoting Niccolò Machiavelli, in a famous letter to Francesco Vettori:

> When the evening has come, I return to my house and go into my study. At the door I take off my clothes of the day, covered with mud and mire, and I put on my regal and courtly garments; and decently clothed, I enter the ancient courts of ancient men, where, received by them lovingly, I feed on the food that alone is mine and that I was born for. There I am not ashamed to speak with them and to ask them the reasons for their actions; and they in their humanity reply to me. And for the space of four hours, I feel no boredom, I forget every pain, I do not feel poverty, death does not frighten me. I deliver myself entirely to them. (Machiavelli 1998/1513, 109–110)

We should listen to the plurality of voices and engage with them, across space and across time. And this in turn makes it understandable why an attention to the past and to the texts of past thinkers is part of what doing moral philosophy amounts to. Finally, then, the same is true of economics to the extent that it has always been, *inter alia*, a branch of moral philosophy.

The competing claims of liberty and utility, of utility and dignity, of price and dignity, and of course equity and efficiency have always been negotiated, either in the foreground or the background, in the discipline. Unlike many past thinkers, most modern economists make such negotiation orthogonal to their economics, and therefore they do those things badly. Most embrace positivist slogans about facts and values—the 'objectivity' of discussion about the former versus the 'subjectivity' of discussion about the latter—slogans embodying a view which itself amounts to a philosophical claim, one which—on positivist grounds—would seem to have no 'objective' or 'factual' standing.

I am defending one way of doing economic thought and do not at all mean to suggest that that is the only way. (Perhaps it would be better denominated economic philosophy.) Many historians of economic thought are comfortable in strictly separating doing economics from doing the history of economic thought—indeed the mode I defend is and has been embattled for some time.[7] I am content

6. Isaiah Berlin is perhaps the best-known exponent of such a view. Charles Taylor is another. For neither does moral pluralism entail subjectivism. Berlin illustrates my thesis in another respect: in doing the history of ideas, he is just as surely doing philosophy as well.

7. Roy Weintraub has been vocal on this front. See, e.g., Yann Giraud's conversation with Weintraub (Giraud 2019), where Giraud refers to "your dissatisfaction with…heterodox economists who are using historical or methodological argument in order to criticize economics." I know Weintraub's views in this

to make the case that it is one way of proceeding, that it is legitimate, and that it belongs in the economics department![8]

This particular way of pursuing the history of economic thought disproportionately attracts heterodox thinkers of every stripe, from libertarians to Marxists. On my account this makes sense, not because, as Roy Weintraub would perhaps have it, it is really not history of thought at all but simply an excuse for criticizing orthodoxy that selectively enlists the authority of past thinkers to bolster a particular ideological position, but rather because heterodox thinkers of whatever stripe are more aware than orthodox thinkers that economics has always been a moral science and that moral thinking is essentially contestable and plural, without being thereby any the less a matter of thinking and reasoning. They are more aware of this because the thinkers they cherish, and whose texts they study—Smith, Mill, Marx, Veblen, Hayek, James M. Buchanan—all pursue economics as a moral science.

To see how deeply positivist attitudes about values go in the profession, and how disabling they can be for appreciating the potential integration of economics and the history of economics I am championing, consider Duncan Foley's book *Adam's Fallacy*, a book that has never been off my syllabus for the undergraduate History of Economic Thought course I teach since it came out. Foley, now at the New School for Social Research, began his career working at the frontiers of general equilibrium theory and became one of the most important contributors to modern Marxist economics. The book demonstrates how to do the history of economic thought as economics along the moral-philosophical lines I have been discussing. In the Preface, however, it is telling how Foley represents his own achievement. He has subtitled the book "A Guide to Economic Theology," which he glosses thus:

> The most important feature of Adam Smith's work is not what it tells us about capitalism…but its discussion of how we should feel about capitalist economic life and what attitude it might be reasonable for us to take toward the complicated and contradictory experience it affords us. These are discussions above all of faith and belief, not of fact, and hence theological. (Foley 2006, xv)

If you are offering reasons for taking certain attitudes, as Foley in the first sentence says Smith does, then are you asking us to take the matter on faith? And is "fact" so disjoined from belief or even faith? Are facts not theory-laden? To call *The Wealth of*

---

respect chiefly through my long-time participation in the History of Economics Society email list. (This comment should not be taken as a criticism of Weintraub's own work in the history of economic thought, which I have long admired.)

8. The eviction of history of thought from the economics curriculum has been going on for some time.

Nations a work of theology seems to stretch "theology" for the sake of polemics.

And while it is not theology, *The Wealth of Nations* and Smith's work generally may well evoke reverence in the practicing historian of thought (Klein 2010). Aristotle taught that moral knowledge can't simply be gained by opening one's eyes to the empirical facts and thinking coherently and consistently. In this realm, as William Fitzpatrick put it, "getting correct results from deliberation depends crucially on having the right starting points" (Fitzpatrick 2008, 168). The starting points will be right only if one has been properly edified. The role of the *phronimos*, someone who can judge appropriately in the ethical realm, is inescapable on the Aristotelian account. We learn to judge properly by apprenticing ourselves, either directly or indirectly, to such a phronimos. When we read Smith, or Mill, Hayek, Keynes—any of the great thinkers who have made our discipline—we are learning to think *with* them, about the good society, the good economy, the good life. We apprentice ourselves to them, take them as authorities—so reverence is well in order. Go back to Hume's ethically rich and complex description of the civilization made possible in a liberal, commercial society I cited at the outset. In reading Hume, aren't we gaining starting points for our own reflection on the worth of such a society, and aren't we properly awed by the wisdom his work exhibits?[9]

Envoi: In memoriam, Don Lavoie

I turned 66 this year, and I have been reflecting lately on my scholarly life (my so-called career!). I went to graduate school in the late 1970s and early 1980s, at American University in Washington, D.C., which at the time was one of handful of schools that specialized in (mostly left-wing) heterodoxy. For us, George Mason University, in the Virginia suburbs, the home of libertarianism, was the enemy. When I became a professor, my research interests and teaching gravitated over time from macroeconomics to the history of economic thought. My own political/economic views skew towards the left—I am a proud Sraffian, an old Keynesian, and a left-wing Smithian. At conferences, I got to know many people from GMU and libertarian circles generally, who were disproportionately interested in the classic texts of the history of economic thought, as I was. At one point, I wrote a paper that took issue with an argument that Don Lavoie and others were making to the effect that there was a special affinity between hermeneutics and libertarian economics. Lavoie, at the time a faculty member at GMU—who died in 2001, tragically, at the age of 51—is a towering figure in libertarian thought, the author of a brilliant book about the "calculation debate," *Rivalry and Central Planning* (Lavoie

9. Regarding Hume's infamous footnote on race, do not rush to judgment before reading Asher 2020a; b.

1985).[10] I sent an abstract to present the paper at the Eastern Economic Association annual conference, and it was accepted.

To my consternation, the assigned discussant was to be Don Lavoie. I was sure I would be eviscerated. Instead, he was enthusiastic about the paper and suggested it I send it to *Critical Review*. I did and it was published (Quinn and Green 1998). That began what I consider to be the greatest intellectual friendship of my life. Don was an intellectual's intellectual, with an astonishingly fertile mind and a passion for ideas. Among other things, we were both interested in Hannah Arendt and put together a panel at an Eastern Economic Association conference on Arendt and economics. He had me come to GMU, into the lion's den as it were, to present work on Smith and Arendt that I was doing.

We never discussed politics per se—we kept our eyes on the text and its author: Hans-Georg Gadamer, Jürgen Habermas, Arendt, Smith, and Marx, among others. Our passion for author and text transcended, and ultimately lessened, our prior differences over their interpretation. I miss him and our on-going conversation immensely.

I know that with Don, across what can sometimes appear as a chasm between clashing schools of thought (as it did for me in graduate school), I felt "an encrease of humanity, from the very habit of conversing together, and contributing to each other's pleasure and entertainment."

References

Asher, Kendra H. 2020a. Rush to Judgement on David Hume? *Past Meets Present* (Institute of Intellectual History, University of St. Andrews, St. Andrews, UK), October 31. **Link**

Asher, Kendra H. 2020b. Interpretations of Hume's Footnote on Race. Working paper. **Link**

Fitzpatrick, William J. 2008. Robust Ethical Realism, Non-Naturalism, and Normativity. In *Oxford Studies in Metaethics*, vol. 3, ed. Russ Shafer-Landau, 159–206. Oxford: Oxford University Press.

Foley, Duncan. 2006. *Adam's Fallacy*. Cambridge, Mass.: Belknap Press.

Giraud, Yann. 2019. "I Had Always Operated on the Outside": A Conversation with E. Roy Weintraub on the History of Economics, Science Studies and Academic Morals. *CHOPE Working Paper* 2019-16. Duke University Center for the History of Political Economy (Durham, N.C.). **Link**

Hume, David. 1987 [1777]. Of Refinement in the Arts. In *Essays: Moral, Political, and*

10. *Rivalry and Central Planning* was illustrative of the type of work I am defending here, being simultaneously a contribution to the history of economic thought and to the study of comparative economic systems.

Literary, 268–280. Indianapolis: Liberty Fund.

Klein, Daniel B. 2010. In Defense of Dwelling in Great Minds. *Society* 47: 83–84.

Lavoie, Don. 1985. *Rivalry and Central Planning*. New York: Cambridge University Press.

Machiavelli, Niccolò. 1998 [1513]. Letter to Francesco Vettori, December 10. In *The Prince*, 2nd ed., trans. Harvey C. Mansfield, 107–111. Chicago: University of Chicago Press. **Link**

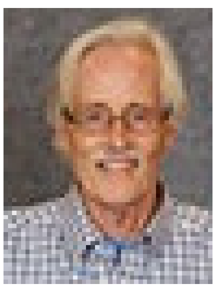
Mandeville, Bernard. 1714. *The Fable of the Bees: or, Private Vices, Publick Benefits*. London: J. Roberts.

Quinn, Kevin. 2019. Was Smith a Moral Subjectivist? *Economic Thought* 8(1): 30–39.

Quinn, Kevin, and Tina R. Green. 1998. Hermeneutics and Libertarianism: An Odd Couple. *Critical Review* 12(3): 207–223.

Rawls, John. 2000. *Lectures on the History of Moral Philosophy*. Cambridge, Mass.: Harvard University Press.

About the Author



Kevin Quinn is a Professor of Economics at Bowling Green State University in Bowling Green, Ohio, where he has taught for 30 years. He holds an undergraduate degree in philosophy from the University of Maryland and a doctoral degree in economics from American University. His research interests are in the history of economic thought, economic philosophy, and economic pedagogy. His most recent work explores moral realist themes in Adam Smith's *The Theory of Moral Sentiments*. His email address is kquinn@bgsu.edu.

Go to archive of Watchpad section
Go to March 2021 issue

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6032/

EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 18(1)
March 2021: 164–191

What 21st-Century Works Will Merit a Close Reading in 2050?: Second Tranche of Responses

Econ Journal Watch

LINK TO ABSTRACT

prologue by Daniel B. Klein

Adam Smith applied the expression "never to be forgotten" to two thinkers he knew personally, born more than 300 years ago, and those two thinkers are not yet forgotten.

We undertook the present query in 2020, looking a mere 30 years ahead. What 21st-century works will merit a close reading in 2050? That is the question asked of *Econ Journal Watch* **authors** (specifically: those who authored material in sections *other than* the Comments section of the journal). The previous issue of this journal provided responses from authors with last names beginning A through K (**link**). Here we present nineteen responses from authors L through Z.

Our invitation clarified the question as follows:

> *If you were to provide a reading list for someone who in 2050 was aged 40 and who had already come to an outlook like your own, what works published 2001–2020 would you include? What 2001–2020 works would you urge such a person to read if he or she hasn't already?*
>
> Clarifications:
>
> - Assume that the person already basically shares your moral and political sensibilities.
> - You may select up to ten works.

- Regarding any of the works you select:
    - The work may be a book, an article, a chapter, or any other written form.
    - It may not be authored or coauthored by yourself.
    - It need not be confined to your own outlook. A listed work may be of whatever flavor.
    - It may be from any discipline, represent any point of view, and may even be fiction or poetry.
    - It may be of any language.
- We also encourage brief remarks or annotations about:
    - your reasons for selecting the works,
    - commentary on the selected works, and/or
    - reflections on making such a list.

We intend to publish the responses and to reveal the identity of the provider of each and every response.

In this tranche we have nineteen responses, from Mitchell Langbert, Andrés Marroquín, Steven G. Medema, Alberto Mingardi, Paul D. Mueller, Stephen R. Munzer, Evan W. Osborne, Justin T. Pickett, Rupert Read and Frank M. Scavelli, Hugh Rockoff, Kurt Schuler, Daniel J. Schwekendiek, Per Skedinger, E. Frank Stephenson, Scott Sumner, Cass R. Sunstein, Slaviša Tasić, Clifford F. Thies, and Richard E. Wagner.

response from Mitchell Langbert

The books that have influenced me most and that have been published since 2000 concern the substitution of liberal with left-wing culture and ideology, and the long-term effects on academic freedom, civil liberties, and economic freedom. The reasons for political intolerance have been dissected in two books, one by **Jonathan Haidt** (*The Righteous Mind*, 2012) and the other by **Jonathan Haidt and Greg Lukianoff** (*The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure*, 2018). Haidt shows why we insist on our own point of view and how culture tends to reinforce group conformity. Haidt and Lukianoff show how the educational system has attenuated students' faculties through emphasis on feelings, coddling, and demonization of those with whom we disagree.

Bryan Caplan, in his *Myth of the Rational Voter* (2007), explains how the kind of emotions that Haidt and Haidt and Lukianoff describe lead to the likelihood of

large numbers of people voting against self-interest, i.e., against liberal principles of freedom of expression and association, in favor of self-image flattering positions or candidates that tap into the feelings. Related to the processes that lead to groupthink and self-righteousness are narcissism and psychopathy. The worst of those who get on top might be psychopaths, and psychopaths understand how to tap into narcissists' self-indulgence to manipulate them. Since the culture has become increasingly narcissistic, it is increasingly prone to such manipulation. Christopher Lasch's *Culture of Narcissism* predates our recent period, but a follow-up book I found useful is **Jean Twenge and W. Keith Campbell**, *The Narcissism Epidemic: Living in the Age of Entitlement* (2009). An excellent book on white collar psychopathy, which I see as a pattern that will increase in frequency, is **Paul Babiak and Robert D. Hare**, *Snakes in Suits: When Psychopaths Go to Work* (2006).

Central to the culture of narcissism and psychopathy that I see emerging in the United States is the media. A useful book that has appeared recently about the media is **Mark Levin**, *Unfreedom of the Press* (2019). The curtailment of press freedom, freedom on the Internet, and freedom in universities have, at their core, the rejection of the liberal, limited state and the imposition of an unfettered fiat money system. **Ron Paul**'s *End the Fed* (2009) will be relevant in 2050.

response from Andrés Marroquín

This is a list of books, articles, and chapters for somebody who, at age 40, has come to an outlook like mine. That imaginary person in 2050 is interested in aspects of economic development and has some interest in complexity and science fiction. By 2050, I hope, poverty will not be a big problem, which it is today. In that case, these readings might be of importance only from the perspective of history, or history of economics analysis.

Acemoglu, D., and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. It differentiates between extractive and inclusive institutions—which is key to understand the wealth and poverty of nations.

Elliott, J. 2006. *Empires of the Atlantic World: Britain and Spain in America 1492–1830*. It explains the historical differences in the way the US and Latin America were colonized. This has important implications for today's economic development.

Coase, R., and Wang, N. 2012. *How China Became Capitalist*. It explains the marginal market-oriented revolutions that happened in China starting in the late 1970s, which led to significant reductions in poverty in that country.

Clemens, M. 2011. **"Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?,"** *Journal of Economic Perspectives*. It explains the high gains that can

come from a free movement of people across borders.

Caplan, B. 2001. **"What Makes People Think Like Economists? Evidence on Economic Cognition from the 'Survey of Americans and Economists on the Economy,'"** *Journal of Law and Economics*. It explains the factors that are correlated with economists' beliefs.

Henrich, J. 2016. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Henrich explains cultural evolution and why that is important for long-term socio-economic success.

Colander, D., and Kupers, R. 2014. *Complexity and the Art of Public Policy*. It explains public policy when we take social complexity seriously.

Chiang, T. 2019. **"The Merchant and the Alchemist's Gate."** This is the first science fiction story in the book *Exhalation*. Chiang writes a beautiful piece on time travel and the viability of changing the past.

Alesina, A., and Giuliano, P. 2014. **"Family Ties."** In *Handbook of Economic Growth*, edited by P. Aghion and S. Durlauf. The authors analyze the link between family ties and economic development.

Snowdon, B., and Vane, H. 2005. *Modern Macroeconomics: Its Origins, Development and Current State*. It describes the recent history of the different schools in macroeconomics.

response from Steven G. Medema

I have two items to offer:

Roger E. Backhouse, *Founder of Modern Economics: Paul A. Samuelson. Volume 1: Becoming Samuelson, 1915–1948*. Oxford University Press, 2017. And volume 2 when it is published sometime in the next couple of years.

Mary S. Morgan, *The World In the Model: How Economists Work and Think*. Cambridge University Press, 2012.

response from Alberto Mingardi

I interpret this "question from the future" as coming from somebody who "already came to an outlook like my own," that is: that she is a 40 year old classical liberal in 2050. But I also assume that she has a special interest in works that helped in shaping the nuances of classical liberal arguments in the 21st century. If the values of such a rich tradition of thought as liberalism survive to 2050, the way they are argued for must adapt to different circumstances. I am therefore suggesting books that I hope will shape the classical liberal sensibility in the coming years.

Alberto Alesina, Carlo Favero, and Francesco Giavazzi, *Austerity: When It Works and When It Doesn't* (Princeton University Press, 2019). Does fiscal consolidation hurt economic growth? Or does it perhaps go well with it, if it focuses on reducing public spending rather than raising taxes? This book is a major accomplishment by the very authors who most contributed to this debate, is filled with historical evidence and will help our grandchildren to make sense of "austerity."

Alan Greenspan and Adrian Wooldridge, *Capitalism in America: A History* (Penguin, 2018). This book is a serious investigation in American history, which, in the understanding of the authors, is characterized by a different relationship between politics and the economy than anywhere else in the world. That relationship, in turn, gave rise to great economic dynamism. Greenspan was crazily idolized and then much maligned in his capacity as a central banker. This book will stand up over time better than his reputation as a central banker.

Yang Jisheng, *Tombstone: The Great Chinese Famine 1958–1962* (Farrar Straus & Girour, 2013). The great Chinese famine is an even less studied subject than Holodomor. The author is a brave Chinese journalist and scholar who understands it also in the light of the economic calculation problem. I hope it will be remembered as a turning point in the historiography of communism.

Jeffrey Friedman and Wladimir Klaus, *Engineering the Financial Crisis: Systemic Risk and the Failure of Regulation* (University of Pennsylvania Press, 2011). Before Covid-19, the event which most defined our generation of liberals seemed to be the financial crisis. This book criticizes the standard account of this event and explains how regulatory hubris "engineered" the crisis. Unlike many other impromptu works on the subject, it will last.

Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Penguin, 2012). By a great social psychologist, this work helps us understand the nature of politics like few others.

Michael Huemer, *The Problem of Political Authority: An Examination of the Right to Coerce and the Duty to Obey* (Palgrave, 2012). This is one of the best works in political theory ever for defending a libertarian position. By 2050 it will be recognized as such.

John Lachs, *Meddling: On the Virtue of Leaving Others Alone* (Indiana University Press, 2014). A reviewer suggested that this work is a contemporary version of *On Liberty* by John Stuart Mill. I think the comparison is not much off the mark. Lachs writes in a beautiful, terse style, with patience and philosophical detachment. He makes arguments that are central for a free society—yesterday, today, tomorrow.

Deirdre N. McCloskey, the *Bourgeois Era* trilogy and in particular ***Bourgeois Dignity: Why Economics Can't Explain the Modern World*** (University of Chicago Press, 2010) and ***Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World***

(University of Chicago Press, 2016). Of the books in this list, I bet no other will change the liberal vocabulary as much as Deirdre McCloskey's work. Because of her style, of her brilliance, of her personality, and of course because of her arguments, McCloskey is helping us all to forge a better classical liberal rhetoric. This ambitious trilogy should define the framework for other ambitious works.

Kenneth Minogue, *The Servile Mind: How Democracy Erodes the Moral Life* (Encounter Books, 2012). How do democracy and the welfare state diminish our moral life? The argument that government intervention is detrimental to societal norms of taking individual choice seriously has never been made more brilliantly than in this book.

Matt Ridley, *The Rational Optimist: How Prosperity Evolves* (Fourth Estate, 2011). Is everything going worse? Ridley is a professional journalist but even the pickiest scholar cannot dismiss this book as 'mere' journalism. It builds on the author's profound understanding of the theory of evolution and of economics to show us how the division of labor and trade actually built a better world.

response from Paul D. Mueller

It's a challenging question you ask. Here is my stab at some books I have found important and may continue to be so in 2050:

Coming Apart by **Charles Murray**. In many ways this book began ringing alarm bells about class divide in the U. S. and about the deterioration of culture among middle income to poorer Americans. A few of the striking takeaways from the book are (1) poor white communities are on the same trajectory as poor black communities in terms of less marriage, more children born outside of wedlock, joblessness, etc. They are just a decade or two behind. (2) Increasingly there is no commonly shared cultural experience or background between wealthy Americans and poor Americans. Murray has a survey to help you see just how differently the two classes live. (3) The educated, wealthy, elite class has abdicated its role as moral exemplars. They say that there is no right way to live, or have a family, or be successful, or be happy, yet overwhelmingly wealthy elites get married and stay married, they have a job and work hard, they save and invest, etc.

The Righteous Mind by **Jonathan Haidt**. This book has framed a lot of conversation about how people form moral judgments and why there is so much disagreement among people on moral issues. Haidt argues that appeals to reason or to one's cognition are limited in how much they can really change people's behavior.

Why Liberalism Failed by **Patrick Deneen**. This book ignited significant debate and controversy regarding whether a liberal order can be sustained. Part of

the controversy stems from the fact that Deneen plays fast and loose with the term 'liberalism,' sometimes meaning the 'high' or 'collectivist' liberalism common in most universities and associated with the Democratic political party in the United States, and other times trying to include classical liberals (i.e., limited government, free market liberals). He suggests that the two types of liberals share enough intellectual DNA to make distinguishing them relatively unimportant—hence he lumps Hobbes, Locke, and Rousseau together—and barely mentions Scottish figures like Adam Smith at all. His book argues that the cultural and political turmoil experienced by the United States in the past two decades is not an aberration, but the result of an internal contradiction between promoting individual freedom and using greater and greater amounts of government intervention to promote that freedom. Deneen questions whether any meaningful liberal-democratic republic can be regained or whether we need new forms of political and social organization altogether.

The Rise of the Creative Class by **Richard Florida**. This is a book on economics and sociology. It studies how the economy is changing and what that means for public policy and for urban life. Florida argues that the Creative Class (in contrast to the working class or the service class) is the wealthiest and happiest class. It is also the most productive. It is also the fastest growing. This book (one of many he has written on the topic) explains what the creative class is, who is a part of it, and how it relates to cities, suburbs, culture, and the economy.

Engineering the Financial Crisis by **Jeffrey Friedman and Wladimir Kraus**. This is my favorite book on the financial crisis. They overview a bunch of different explanations of the 2008 financial crisis, but they also give the most compelling meta-explanation for the crisis: misregulation. The problem wasn't deregulation or low interest rates per se, instead, it was regulations that encouraged and rewarded risky behavior that created systemic fragility and then self-reinforcing negative cycles in the face of debt deflation.

Hidden in Plain Sight by **Peter Wallison**. This is my second favorite book on the 2008 crisis—although it is more about the housing bubble and its connection to the crisis than about the crisis itself. Wallison was on the congressional committee investigating the causes of the 2008 financial crisis. He points out in this book that there was a massive deterioration in home loans/mortgage quality from the early 1990s to 2007—far more deterioration than most people realize. It wasn't simply more subprime loans—it also included more adjustable-rate loans, more loans with little or no money down, etc. He argues that concerted efforts by politicians and regulators caused this deterioration in lending standards, not the market or deregulation.

Desiring the Kingdom by **James Smith**. This book talks about Christian theology and has made a lot of waves in Christian circles. Smith argues that people

are primarily shaped by what they love, rather than by what they think or believe. As such, what people spend their time doing shapes who they are. Do they pray when they first wake up or do they check their phones? Do they go to church or do they go to brunch on Sundays? Do they spend their time watching Netflix every night or reading books and having conversations? He argues that Christians need to actually behave in certain ways if they want to love God and their neighbor well.

response from Stephen R. Munzer

I would recommend **Derek Parfit**, *On What Matters* (Oxford University Press, 2011), 3 volumes. This is an extraordinary work of analytic moral philosophy, and a worthy successor to the same author's *Reasons and Persons* (1984). Parfit's writing style is compressed, precise, and lucid. His sentences are short and pithy. He is immensely creative. In my view he is the greatest British moral philosopher in the period 1984–2011.

response from Evan W. Osborne

To the 2050 lovers of liberty, especially American ones:

I write to you in early 2021, as the United States has just experienced the most fractious of a series of very divisive presidential elections. The list below was compiled under the premise that domestically the classical-liberal spontaneous order has been under threat for some time, and in 2021 is more urgently so from several trends in the US, which will either explain how we got to where you are in 2050 or what we somehow managed to avoid. As an American citizen I choose to limit my recommendations to those primarily involving my country. My ten recommendations are numbered.

The crushed individual, the ascending tribe

The biggest threat is the growing tendency to say, once again, but in a new way, that the collective should dominate our lives so as to overcome societal conflict. While for much of post-Civil War America it was the working class and the capitalists who were pitted against each other, after decades in which the standard of living of the erstwhile proletariat has continued to rise worldwide, in many prosperous societies it is the new holy trinity of race, sex/sexual orientation, and religion that is increasingly seen as conflicts that must be managed by the state.

The decades after the collapse of the Soviet Union were fruitful in terms of

governments relaxing the chokeholds they had on the productive, market-driven energies of people all over the world. But now in some countries we are told demographics is destiny. One of the writers who has found the most success in promoting the thoroughness of racism in particular in U.S. history is **Ibrahim X. Kendi**, currently on the faculty at Boston University. While his biggest-selling book is 2019's *How to Be an Antiracist*, which argues that all people have a duty to confront racism whenever they see it (with the antiracist getting to define whether something is "racist"), his most accomplished book is the award-winning ***Stamped from the Beginning: The Definitive History of Racist Ideas in America*** (2016). It provides a guide to the antiracism ideology, said to be superior to mere nonracism. The appeal of this ideology is growing and, to the extent it implicates politics in whom to hire (and not hire) and even imposes tribal criteria for all human interaction, it is hostile to the spontaneous evolution of society.

The attack on honest income

The dramatic rise in individual opportunity all over the globe has meant a corresponding rise in the reward to having a high marginal revenue product. And this has not sat well with those who feel, rightly or wrongly, that it would be better to live in the simpler America (or Europe) of 1960. The Harvard philosopher **Michael Sandel** has in ***The Tyranny of Merit: What's Become of the Common Good?*** (2020) argued that the very idea that people should get whatever others are willing to voluntarily pay them is a mistake. The idea that allowing freedom today may benefit people in the future much more than rewarding people today based on some other criterion would does not make much of an appearance in this book. But Sandel makes a prima facie argument, certainly not the first of its kind, that things were better before globalization, computerization and other liberal '-ations' made the world a worse place.

In defense of the free society

A case for liberty is made by **Charles Murray**, long a scholar at the American Enterprise Institute and a self-identified libertarian. With respect to the new tribalism, in ***Human Diversity: On the Biology of Gender, Race and Class*** (2020) Murray argues that genes are one factor in what an individual can or cannot do well. This work builds on Murray and Richard Herrnstein's 1996 book *The Bell Curve: Intelligence and Class Structure in American Life* (1996). In *Human Diversity*, Murray continues from the premise that every individual has dignity, regardless of IQ. He makes the case for accepting the significance of genetic inheritance (of all sorts) for understanding differences in life paths and for the unimportance of such

considerations when evaluating an individual morally.

In another of his books, *Coming Apart: The State of White America, 1960–2010* (2013), Murray maintains that the US is in fact increasingly divided not by tribe but by class. Here the idea is not that the rich and powerful are exploiting the poor and powerless, but that a new ruling class is emerging. The success of high-IQ and/or hardworking people is partly due to their adopting lifestyles that in his words they practice but refuse to preach. (He confines the data to whites to avoid entangling issues of race.) Together, *Coming Apart* and *Human Diversity* help us understand the growing cognitive stratification of society, and its threats to freedom. Murray has long been concerned that the breakdown of what were once considered individual virtues is leading to growing pressure among those who flourish for expanding the custodial state. The flourishing increasingly see government as necessary to babysit those who can't flourish, and see themselves as virtuous in calling for such babysitting.

If the 2050 classical liberal wants to read a masterful apologetic for classical liberalism, I suspect he will not be able to do better than to read **George F. Will**'s *The Conservative Sensibility* (2019). Will, a prominent voice in the public conversation for decades, indicates at the beginning that by "conservative" he means "classical liberal." The book is replete with historical events, political and economic theory, all deployed with considerable skill to make the claim that these values best conduce to societal well-being.

And the stakes are great. The improvement in the lot of humanity both materially and ethically is analyzed by **Deirdre N. McCloskey** in *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World* (2016), the third book in her trilogy on the source of modern, continually rising prosperity. Throughout the trilogy she novelly attributes what she calls The Great Enrichment to a change in attitudes toward both wealth accumulation and the people who excelled at it (also emphasized by Murray in *Coming Apart*). But in the final section of this book McCloskey diagnoses a threat to this continued progress, the criticism since the mid-1800s by the increasingly influential clerisy. By this she means the educated-yet-not-knowledgeable intellectual elite who argue that in the name of justice and the superior wisdom of such elites Hayek's "spontaneous order" should be displaced by a planned regime. She describes the advocacy of such politically regimented order despite its repeated failure and the long success of societies built on free competition deriving from respect for bourgeois values, with their natural tendency for dynamic improvement. It provides a compelling sketch of where we are, and where we risk going.

Holy democracy

Another trend worth keeping track of is the worship of democracy, by which I mean merely the choosing of rulers by elections under a universal adult franchise. In 2021 there is a growing belief that people's conflicting desires in an environment of constraints are best resolved collectively, through the astonishingly crude instrument of representative democracy, and that to want to limit government by upholding, say liberty and the rule of law, is self-evidently wrong. A remarkable example of this genre is **Nancy MacLean**'s *Democracy in Chains: The Deep History of the Radical Right's Stealth Plan for America* (2017), which takes public-choice giants, James Buchanan in particular, to task for supposedly sabotaging "democracy" in the interests of big business. In this way of modeling the world Big Capital is sinister, Big Democracy somehow pristine. The book was widely criticized, even by people who are not supporters of limited government (see, e.g., Henry Farrell and Steven Teles in *Vox*) for among other things badly misrepresenting the thought of Buchanan. Nonetheless, it too was highly acclaimed.

And the reasons we should be suspicious of unfettered representative democracy, even ignoring questions of liberty and economic efficiency, are many and profound. The Georgetown philosophy professor **Jason Brennan** has done yeoman's work, partly through bringing the objections of such figures as Montesquieu to light for modern readers, in his *Against Democracy* (2016).

With some of the foregoing works, we get an attack on a "merit" seen as systematically unjust. Freedom and (classical) liberal politics are themselves under siege. A broad analysis of current threats to liberalism was presented by **Francis Fukuyama** in 2020 in *The American Prospect*, **"Liberalism and its Discontents."**

Explaining it all: the hidden hand

The sacralization of elections has in the US followed a century of concentrating more power in the hands of a single person, the president, whose power to enact executive orders (including retracting those of a previous president) and to name appointments to the administrative agencies has made him unusually important. Given this, another thing I ask 2050 readers to understand about this moment is a growing sickness, the assuming of the absolute worst about America when some particular presidential election doesn't go your way. This increasingly takes the form of flat-out conspiracy theories about why things are what they are. Over perhaps two decades we have heard from the left empirically implausible theories that corporations secretly control public policy, and from the right in 2020 we have witnessed the extraordinary spectacle of widespread belief that the Democratic Party managed to steal the presidential election, despite the fact that

in over 50 opportunities the evidence presented by the supposed victim of these machinations, President Donald Trump, did not even minimally satisfy American courts of law. And so, whatever the flavor of your particular conspiracy, its sinister workings have fatally damaged democracy, and the entire liberal order needs to be replaced by something truly reflecting the presumed will of the people.

Conspiracy theories have implications for freedom. Because often believers are convinced that state power is secretly controlled by nefarious actors, they call to use that power to punish such malefactors. One is also unlikely to believe that freedom can be trusted, and that the way things are today is the product not of free interactions but of machinations by the Deep State, Jews, Freemasons, large corporations or what have you. To understand the surprising current prevalence of such thinking in America, I thus recommend a paper in an edited collection on conspiracism, **Adam M. Enders and Steven W. Smallpage**, **"Polls, Plots and Party Politics: Conspiracy Theories in Contemporary America,"** in Joseph E. Ucinski, ed., *Conspiracy Theories and the People Who Believe Them*, Oxford University Press, 2019, 298–318. It documents the prevalence of these beliefs in the US even before the chaos of the 2020 election. (Other papers in the volume are also of interest.)

For the classical liberal, 2021 is a trying time. We hope in 2050 things are better.

response from Justin T. Pickett

Here is the reading list I would suggest:

Daniel Kahneman, *Thinking Fast and Slow*

Steven Pinker, *The Better Angels of Our Nature*

Stuart Ritchie, *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*

James Stimson, *Tides of Consent: How Public Opinion Shapes American Politics*

Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect*

Sarah Igo, *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*

Steven Pinker, *Enlightenment Now*

Peter Enns, *Incarceration Nation: How the United States Became the Most Punitive Democracy in the World*

Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*

Ronald Weitzer and Steven Tuch, *Race and Policing in America*

response from
Rupert Read and Frank M. Scavelli

Samuel Alexander, *Entropia*. *Entropia* by Samuel Alexander provides a worthwhile, semi-novelized account of the process of both industrial-civilization breakdown, and attempts to build a new civilization in the aftermath. Given the fact that we, as a global civilization, continue to deplete our rapidly diminishing 'carbon budget,' amidst a more general ecocide, without any real net movement toward anything genuinely worth calling 'sustainability', Alexander's book may be as it were all too practically relevant to someone in 2050. The key reason for reading Alexander's book is the power of the astonishing 'reveal' near the end, about what is actually going on in the semi-utopia of Entropia. We won't reveal it here, but let's just say that anyone reading this book in 2050 will have good textual reason to think back to the present time.

Bruno Labour, *Down to Earth*. Latour's *Down to Earth* should be regarded as among the truly worthwhile and pathbreaking contributions to political theory and philosophy of this century, and is the most important book he has yet written. Latour's core argument concerns the need to become 'terrestrial' again (hence coming, socially and politically, 'down to earth'), in contrast to the view, characteristic of modernity, of the Earth as a reified scientific object—a planet among planets, and wild nature, near and far, a thing to be studied under microscope and telescope. As opposed to: our home, our place. What is truly brilliant about Latour's book, however, is his interweaving of this apt philosophical analysis with the emergence, and lately, breakdown of globalization and its 'modernizing' drive, as well as his corresponding explanation of the emergent neo-fascist politics of Trump et al. *Down to Earth* will likely be *de rigueur* for anyone wishing to understand how, for good or ill, we arrived at the world of 2050.

Richard Powers, *The Overstory*. Among the very greatest of all novels yet written in the 21st century, *The Overstory* lends real emotional resonance to a visceral conjuring of our ongoing destruction of the natural world, by way of being centred as much in tree-lives as in human lives. Moreover, it contains a remarkable, vivid depiction of ecological nonviolent direct action in the USA, of roughly the EarthFirst! variety, which one of us (Read) has had personal involvement in: such NVDA in defence of trees will be essential for the next generation, if there is to be a habitable world by 2050.

Kim Stanley Robinson, *Aurora*. Another astonishing novel, and perhaps the greatest from the rich KSR stable, *Aurora* is that rarest of things: a *realistic* sci-fi story. It tells the story of a group of would-be colonists en route to colonize an 'Earth analog' planet in the Tau Ceti system. The journey is fantastically long

and difficult. Unsurprisingly, the minimal biological life on the planet, in the form of hard-to-detect and rapid-acting prions, proves totally disastrous for any hope of colonization, after generations of space travel. This novel serves as a reminder that dreams of space-faring and interstellar colonization are just that—dreams, mere whimsies with no prospect whatsoever of being more than that—for the *long-foreseeable* future. Perhaps we would be better off concentrating on making the best of things on the world we, in fact, evolved on? A lesson for those on an even more depleted Earth, decades from now, than our own.

David Fleming, *Lean Logic: A Dictionary for the Future and How to Survive It.* Fleming's magisterial *Lean Logic* stands in a class of its own. Written over the course of thirty years, Fleming's book is an integrated dictionary of topics relevant to a rebuilding of human society in the approaching post-fossil fuel, climate change-ravaged world. It would be impossible both to canvas the sheer variety of subjects covered, as well as the implacable scholarly rigour and lively rhetoric Fleming provides in their discussion. To name just one of the hundreds of subjects, Fleming details the workings of a localized and communal, 'old Common Law'-based legal system, modelled on those extant throughout Britain up to the enclosure period. *Lean Logic* will be an indispensable practical guide, as well as philosophical resource, for those coping with the world of 2050.

Angela Nagle, *Kill All Normies.* A key question for the denizens of 2050 will likely be: why was so much time wasted, while there remained a chance to make significant changes, the impact of which would have exponentially improved the lot of contemporary (2050) humanity? Why did the system remain unaltered for decades, despite full knowledge of myriad, existentially-threatening problems? Angela Nagle's *Kill All Normies* is a worthwhile read in this regard. Nagle chronicles the embrace of the 'culture war' by both the Right and Left, a facile version of (a non-)politics which seeks to portray an identity rather than effect social change. *Kill All Normies* is and will be key reading for anyone wishing to overcome this mode of discourse, with the goal of radical and unificatory social change in mind.

Mark Fisher, *Capitalist Realism: Is There No Alternative?* Fisher's *Capitalist Realism* is one of the more well-known works in contemporary political-economic philosophy, and rightly so. Fisher's adroit blending of film, music, and other aspects of 'pop culture' into his analysis makes for a fascinating read. More importantly, his analysis of 'Capitalist Realism,' his observation that it is 'easier to imagine the end of the world than the end of capitalism,' is a vastly important one for understanding the seemingly-never-ending predicament we are in. What's more, Fisher's contention that the mental health crisis and ecological crises represent the two forces most likely to 'break through' Capitalist Realism and send people on a genuine search for an alternative is, in our view, crucial for those, now and decades from now, looking for ways of making the necessary step in linking the

social and ecological crises in their search for an alternative.

Suzanne Collins, *The Hunger Games*. It is possible that some readers of *Econ Journal Watch* look down their noses at the likes of the incredibly popular, 'young adults' *Hunger Games* trilogy. If so, that is a serious error. These books are a profound, startling wake-up call—a warning of a possible human future, of extreme inequality in a post-climate-ravaged world: of the most vicious of bread and circuses. And, a tremendous tale of how such a future might be resisted. By 2050, readers will know what aspects of the trilogy were prophetic—and what were, perhaps, a successful raising of the alarm. (It is worth reading Collins's books against the background of René Girard's similarly salient *Violence and the Sacred*, which is the most brilliant analysis we have of the dynamics of the kind of scapegoating mechanism that is the central topic of *The Hunger Games*, and could well rear its ugly head much more than it has already done in the hard times which lie before us.)

Iain McGilchrist, *The Master and His Emissary: The Divided Brain and the Making of the Western World*. This weighty tome by the polymathic McGilchrist is an absolutely vital read for anyone wanting to understand the biases of the academic world, the real meaning of the Industrial Revolution, or the vast intellectual obstacles which stand in the way of our transforming our society into something less ecocidal. It draws richly on philosophy (offering a novel take on Wittgenstein and Heidegger, among others), literature (Wordsworth is a hero of the work), and history (across a jaw-dropping sweep of millennia) as well as neurology to account brilliantly for our predicament. It is one of few nonfiction works of our time that we are confident will endure for generations.

Tyson Yunkaporta, *Sand Talk: How Indigenous Thinking Can Save the World*. Yunkaporta is an aboriginal intellectual and practitioner with a (literally) wildly provocative take on our world. Coming, in effect, from outside the dominant (terminal) global civilisation, he can see clearly things that most of us can, at best, barely bear to contemplate. If our civilisation is to transform into something viable within the next 20–30 years—if it is to be more butterfly (or maybe, phoenix) than dodo—it will be because it has learnt from the likes of this book. (If such wise transformation does not occur, then a vengeful Gaia will likely force it upon us by way of collapse; in which case, it is likely that anyone reading any of these books by mid-century will be doing so by candlelight—*at best*.)

response from Hugh Rockoff

Here is a list of readings in my field of economic history, especially American economic history, published since 2000 that I think will still be worth reading in

2050. I considered important papers and book chapters; but my 10 top choices are books.

A good textbook in American Economic History. For example, **Hughes, Jonathan R. T., and Louis P. Cain**. *American Economic History*. 6th ed., Addison Wesley, 2003. I must emphasize, of course, that there are other fine textbooks. There will be revised versions of this and the other fine textbooks that that will be issued before the target date of 2050. However, it is also useful to read older textbooks to find out 'what were they thinking.'

Boustan, Leah Platt. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton University Press, 2017. The 'Great Migration,' the migration of African Americans to the North after the Civil War, is one of the most important events in American history. Boustan carefully and convincingly analyzes the successes and, sadly, failures of the Great Migration to produce the hoped-for gains in wellbeing.

Eichengreen, Barry J. *Hall of Mirrors: the Great Depression, the Great Recession, and the Uses—and Misuses—of History*. Oxford University Press, 2015. When the financial crisis of 2008 erupted, economists and economic historians immediately reached back to the Great Depression for lessons about what to do and not do. Eichengreen explores this analogy in detail and reflects on the general question of whether historical analogies can provide useful guides for meeting current problems.

Goldin, Claudia Dale, and Lawrence F. Katz. *The Race Between Education and Technology*. Belknap Press of Harvard University Press, 2008. Goldin and Katz examine the role that education has played in determining the growth and distribution of income in the United States. Their creativity and thoroughness in using economic ideas and empirical data provides a model for research in economic history.

Irwin, Douglas A. *Clashing over Commerce: A History of US Trade Policy*. The University of Chicago Press, 2017. Irwin explores U.S. trade policy from colonial times to the early twentieth century. He analyzes the economic and political factors that shaped trade policy and the economic consequences of changes in trade policy. *Clashing over Commerce*, I predict, will become a cornerstone of future research on trade policy, much like Friedman and Schwartz's *A Monetary History of the United States* has proved to be for monetary policy.

Khan, B. Zorina. *The Democratization of Invention: Patents and Copyrights in American Economic Development, 1790–1920*. NBER and Cambridge University Press, 2005. Through careful empirical studies, Khan shows that the American patent and copyright system of the nineteenth century was more successful than its European counterparts were because it provided clearly enforced property rights awarded on a democratic basis.

McCloskey, Deirdre N. *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World*. University of Chicago Press, 2016. This book is the final volume of a trilogy on bourgeoisie values. The title is a brilliant summary of the thesis. McCloskey then makes a convincing case.

Meltzer, Allan H. *A History of the Federal Reserve*. University of Chicago Press, 2003. A critical source for anyone who wants to understand the role of monetary policy. In several important ways, it enriches and continues through the mid-1980s the story told in Friedman and Schwartz's classic *A Monetary History of the United States*, which covered up to 1960.

Mokyr, Joel. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press, 2011. Mokyr argues based on a wide range of evidence that in the eighteenth and nineteenth centuries, a fruitful interactive relationship developed between technology and science that produced the industrial revolution and importantly an acceleration of long run economic growth.

Piketty, Thomas. *Capital in the Twenty-First Century*. Harvard University Press, 2014. Although some of its claims have been disputed on factual or theoretical grounds, there is no denying the book's tremendous impact on the discussion of economic issues by the public and on the research agendas of economists, historians, and other academics.

response from Kurt Schuler

Robert Alter, *The Hebrew Bible: A Translation with Commentary* (New York: W. W. Norton and Company, 2018). A new translation of an old book. Read the commentary.

Ibn Khaldûn, *The Muqqadimah: An Introduction to History*, translated by Franz Rosenthal, edited by N. J. Dawood (Princeton: Princeton University Press, 2005). The full three-volume translation was issued in 1958, but it and an abridged hardcover were long out of print when the abridged paperback was issued, so I am counting this book as being from the 21st century. Ibn Khaldûn was the first real social scientist, that is, he applied theoretical ideas to patterns in human history. This book, written in 1377, contained insights that European thinkers would not independently rediscover until centuries later. The book was not translated into any European language for 500 years.

Jacques Barzun, *From Dawn to Decadence: 500 Years of Western Cultural Life, 1500 to the Present* (New York: HarperCollins, 2000, paperback 2001). Barzun completed this book when he was in his early 90s, and it is the fruit of a lifetime of reading and reflection, as well as personal acquaintance with some of the figures he discusses.

Napoleon A. Chagnon, *Noble Savages: My Life among Two Dangerous Tribes—The Yanomamö and the Anthropologists* (New York: Simon and Schuster, 2013). Anthropologists are like economists in that their idea of a tragedy is a beautiful theory killed by inconvenient facts. Chagnon's memoir recounts his 1960s fieldwork with the Yanomamö tribe of the Amazon basin, one of the last primitive peoples having little contact with the outside world. His research into their often violent behavior, which included kidnapping and raping young women from other bands to impregnate them, drew intense criticism from his fellow anthropologists distraught that he had upset the ideal of peaceful primitive man. After decades of resistance, including an ideologically motivated smear campaign, they conceded that he was right.

J. D. Crouch II and Patrick J. Garrity, *You Run the Show or the Show Runs You: Capturing Professor Harold W. Rood's Strategic Thought for a New Generation* (Lanham, Maryland: Rowman & Littlefield, 2015). A bracing view of how great-power politics works.

Charles C. Mann, *1491: New Revelations of the Americas Before Columbus* (New York: Alfred A. Knopf, 2005) and *1493: Uncovering the New World Columbus Created* (New York: Alfred A. Knopf, 2011). Brilliant popularizations of the work of scholars not known to a general readership, along with the author's own thoughts. The books made me think of the history of the Americas and the world in a different way.

Carmen M. Reinhart and Kenneth S. Rogoff, *This Time Is Different: Eight Centuries of Financial Folly* (Princeton, N.J.: Princeton University Press, 2009). The book in my own field of interest from this century to date that I would recommend if I could only recommend one. The first truly comprehensive study of financial crises, and well written, for a book by economists.

response from Daniel J. Schwekendiek

I recommend only one source: **Van Zanden JL, Baten J, Mira D'Ercole M, Rijpma A, Smith C, Timmer M**, editors. *How Was Life? Global Well-being Since 1820*. Paris: OECD Publishing; 2014. It includes socio-economic and bio-demographic data of the world over the last recent centuries, which is very rare to get.

response from Per Skedinger

I read books in recurring cycles, by subject. The first subject is Economics,

then History, followed by Music and Art. Each cycle is concluded with Other, which is anything else, including fiction. I have followed this procedure for about ten years now and to keep it going it helps to have a large library, which I'm fortunate to have. I read every day, but not much, 10–20 pages.

Not only does the cycle provide variety in perspective, it also helpful for understanding linkages between the different subjects. For example, much of History is economics and much of Economics is history. The subject for which it's most difficult to find good books is Economics. To use the Economist's classification of books in this field, many 'Greek' books are either irrelevant or incomprehensible—often both—and almost every 'Airport' book fails to make a coherent argument. The best books in Art and Music make you think about other things than art and music. So rather than recommending specific books for a young person of my own inclination, I suggest a reading procedure and will illustrate how this could work in terms of unexpected associations with examples from mainly Art and Music.

Let's start with Art. Asger Jorn (1914–1973) was a Danish artist central to the development of expressive, abstract painting. He was also a founding member of the politically engaged Situationist International. ***Asger Jorn: Restless Rebel***, edited by **Dorthe Aagesen and Helle Brøns**, and published jointly by Statens Museum for Kunst in Copenhagen and Prestel in 2014, provides an overview of his life and work. The front cover shows his painting "The Sun P***** Me Off" (original title "Le soleil m'emmerde"; my asterisks). If you infer that this was a man with some grievances, you are entirely correct. But his anger was directed more toward his colleagues in the avant-garde than against impersonal heavenly bodies. The problem with the other artists, according to Jorn, was that they paid too little attention to art's political potential. Art should be explicitly used for the betterment of a society befouled by capitalist alienation.

To his credit, Asger Jorn also shows a great sense of humor in his paintings, like his younger brother, the less well-known artist Jørgen Nash (1920–2004). This may have followed from Jorn's objective to "give the psyche free rein in an intuitive, improvising way of working where the brush was allowed to run freely," as Aagesen explains in one of the essays in the book. I have the brothers hanging on my wall, but not side-by-side.

Asger Jorn is the only major artist I know of who had a serious interest in economics. So serious, in fact, that he wrote a book on the subject, ***Værdi og økonomi*** (*Value and Economy*, Borgen, 1962). In it, Jorn tried to improve on the Marxian theory of value. But, ultimately, Jorn believed that images were more fundamental than words for the transformation of society. While his book probably never was sold at airports, it nevertheless fails to make a coherent argument.

Let's turn to Music. The best example of cutting-edge originality combined

with mainstream popularity is perhaps the Beatles. **Mark Lewisohn**, the world's leading Beatles historian, has embarked on a long journey writing the definitive biographical account of the group in the three-volume series *The Beatles: All These Years*. So far, only the first volume, *Tune In*, has appeared (Crown Archetype, 2013). The 900-page tome covers the evolution of the group up to the end of 1962, i.e., just before their breakthrough in the UK.

Lewisohn makes it clear how close it was that the Beatles, as we know them, never happened. After toiling on the club circuit for years with little success, the group were on the verge of breakup when they were suddenly approached by Brian Epstein, who, unlike their previous managers, immediately saw their potential and had the business acumen to help them navigate the dangerous waters of the recording industry. Another important coincidence was the abolition of military conscription in the early 1960s. According to Lewisohn, the Beatles "escaped the call-up ordeal by the skin of their teeth" and "were the first generation in Britain since pre-1939 not to be forced into army duty, and the first and only teenagers to have their own rebellious music—rock and roll."

It may seem far-fetched to relate the political economy of Asger Jorn to the political economy of the Beatles—it's not obvious that they have anything in common—but I will do just that. As far as political *commitment* is concerned, Jorn wins over the Beatles hands down. It's true that George Harrison displayed his frustration over excessive marginal taxes in "Tax Man" and that John Lennon distanced himself from political violence in "Revolution," but overt political statements were hardly the cornerstone of the Beatles' lyrics.

It's a different story altogether when it comes to political *influence*. As argued in *How the Beatles Rocked the Kremlin: The Untold Story of a Noisy Revolution*, by **Leslie Woodhead** (Bloomsbury, 2013), the subversive power of the group's music played a major, psychological role in the downfall of the Iron Curtain. Many Russians, including a researcher at the Institute of Russian History, told Woodhead that the Beatles helped a generation of free people to grow up in the Soviet Union. But "communist leadership were determined to block the Beatles virus," Woodhead explains. Beatles were banned from release on the one and only state-controlled record label, Melodiya, in the Soviet Union—contrast this with the myriad of regional labels in the United States at the time.

Due to the pressure of popular demand and circulating tapes from illegal imports, the music could not be held back entirely, so the state's eventual response was to allow some Russian-language cover versions. A contact in Moscow has supplied me with some of these. One of the first covers to appear was "Drive My Car," by Vesyolye Rebyata (Jolly Fellows), a sanitized version of a rock group, in 1970. An odd choice from the Beatles catalog for a regime with a genuine aversion to private car ownership. The covers, of course, were less effective than the real

thing as an antidote against communist alienation in the young generation.

The contribution of the art and culture of the Beatles to the dismantling of the Soviet system was wholly unintentional, in contrast to Asger Jorn's deliberate attempts to transform Western society through political art. He was critical of some aspects of "real socialism," but remained a member of the Danish Communist Party until his death. While the message of the Beatles was one of true liberation, Jorn's rested ultimately on heavy-handed state coercion—the antithesis of liberation.

Jorn's conviction that images can transform society seems to be widely held among artists and their entourage even today. The late, great **Assar Lindbeck**—a colleague of mine at the IFN and himself an accomplished artist—reflected on this view in his autobiography *Ekonomi är att välja* (*Economy Is to Choose*, Bonniers, 2012): "[I]mages can propagate a certain perception of reality and a certain position. But unlike research, images obviously do not explain the *mechanisms* behind societal problems" (my translation and emphasis). For that, you may need to consult an economist. A real economist.

response from E. Frank Stephenson

In the last five years, several authors have published books documenting the good times in which we were living in the years immediately preceding the Covid-19 pandemic. These books include **Johan Norberg**'s *Progress* (2016), **Steven Pinker**'s *Enlightenment Now* (2018), **Hans Rosling**'s *Factfulness* (2018), **Gregg Easterbrook**'s *It's Better Than It Looks* (2018), and **Michael R. Strain**'s *The American Dream Is Not Dead* (2020). Their books marshal strong evidence that we were living in perhaps the best time to be alive—growing affluence in the U.S. and other developed countries, large drops in severe poverty in poor countries, falling violence, increasing environmental quality—yet the authors felt it necessary to document modern well-being. Why they felt the need to do so will be an interesting question to revisit with the benefit of hindsight.

Even now one can posit several conjectures underlying the 2010s unhappiness. Was it the 2008 financial crisis and the tepid economic recovery afterwards? Was it growing income inequality, perhaps from rapid technological change, a "China shock" to blue-collar manufacturing jobs (**David H. Autor, David Dorn, and Gordon H. Hanson**, **"The China Syndrome,"** *American Economic Review*), or entrenched elites who have "captured" the economy (**Brink Lindsey and Steven Teles**, *The Captured Economy*)? Was it rising "deaths of despair" from opioids or alcohol abuse (**Anne Case and Angus Deaton**, *Deaths of Despair and the Future of Capitalism*)? Was it declining social connectedness through

activities such as volunteerism and church attendance (**Timothy P. Carney**, *Alienated America: Why Some Places Thrive While Others Collapse*)? Was it a sense that education, health care, and immigration policies, all front burner issues in recent years, have been poorly handled by policy makers? Was it a predatory legal system which incarcerated too many people (**Chris W. Surprenant and Jason Brennan**, *Injustice for All*), especially minorities, and seized people's assets without proof of criminal activity? Was it the 2016 presidential campaign in which eventual winner Donald Trump and socialist Bernie Sanders, who nearly captured the Democratic nomination, ran angry campaigns against the status quo? Was it the growth of social media which seemed to coarsen public discourse? Of course, many of these factors were related and many phenomena are not mono-causal. Which of these factors or others motivated Pinker, Rosling, Norberg, Easterbrook, and Strain to try to persuade people that the world before Covid was indeed a great moment in human history will be an important question.

response from Scott Sumner

It is difficult to choose appropriate books for a 2050 reader, because what works today might not be as effective in 29 years. Thus David Reich wrote an excellent book on how genetics can help us to understand the pre-history of man, and Tyler Cowen has an outstanding book on how modern societies are stagnating in many dimensions. But both books will probably be superseded by events, and neither is likely to be state of the art by mid-century. With that in mind, here is my list:

Strange Rebels: 1979 and the Birth of the 21st Century, by **Christian Caryl**. This book looks at a key moment in modern history, when the world shifted from a revolutionary phase to a sort of reactionary/revolutionary mode. The book examines five countries at important turning points in their history: Great Britain, China, Poland, Iran and Afghanistan. In each case there was a revolutionary ferment that looked both forward and backward for inspiration. In a sense, 1979 was the beginning of the 21st century.

The Bourgeois Virtues: Ethics for an Age of Commerce, by **Deirdre McCloskey**. In the popular imagination, capitalism is often associated with selfishness. In fact, a market economy requires certain degree of virtue. Fortunately, a market economy also tends to make people more virtuous. McCloskey explains why with a very engaging style and an impressive range of knowledge. This is just the first of a series.

The Japan Journals: 1947–2004, by **Donald Richie**. For those of us that live in the West, it's easy to overlook the fact that there are other ways of living. In Richie's

entertaining memoir, postwar Japan appears like some sort of strange planet—very different from our world but not necessarily inferior. By 2050, this world will probably appear even stranger.

Shadow of the Silk Road, by **Colin Thubron**. Thubron may be our finest modern travel writer, and this is a wonderful account of his journey from China across central Asia. As with the Donald Richie memoir, he is describing a world that will probably be lost by 2050.

If the past is another country, then great novels might be the best way to explore those countries. Here are six brilliant novels that involve places as diverse as Germany, Japan, Spain, Turkey, Mexico and Norway at the end of the 20th century. Most of them have interesting things to say about politics:

Austerlitz, by **Max Sebald**
1Q84, by **Haruki Murakami**
Your Face Tomorrow, by **Javier Marias**
Snow, by **Orhan Pamuk**
2666, by **Roberto Bolano**
My Struggle, by **Karl Ove Knausgaard**

response from Cass R. Sunstein

As they say, predictions are hard, especially about the future, but here are some of my favorites, and (I hope) potential candidates.

Sarah Conly, *Against Autonomy* (2012). Martha Nussbaum describes this as the best book on liberty since Mill; she might be right. It's piercingly clear and highly original. It also repays spirited engagement.

Daniel Kahneman, *Thinking, Fast and Slow* (2011). This is an obvious choice. What's less obvious is how much the book repays rereading. It's dense in places, and there are so many revelations there to be discovered.

Sendhil Mullainathan and Eldar Shafir, *Scarcity* (2013). Economists deal of course with scarcity, but Mullainathan and Shafir focus on limited mental bandwidth—on cognitive scarcity. If you are hungry, busy, lonely, or poor, you will have less ability to think and act outside of your "tunnel." The book is full of implications for both theory and practice, and its lessons will endure.

Amartya Sen, *The Idea of Justice* (2009). Sen offers the most important discussion of justice since Rawls' *A Theory of Justice*. In some ways, Sen's account can be seen as a significant revision of Rawls' account, respectful of pluralism and the idea that the search for justice is a work in progress. Sen also has a lot to say about well-being.

Edna Ullmann-Margalit, *Normal Rationality* (2017). Ullmann-Margalit was

a brilliant social scientist and philosopher, whose work has numerous implications for economics, political science, psychology, and law. This is a collection of some of her best work. She's a bit like William Blake or Jane Austen, in the sense that she will be far better known after her (tragically short) life. (Disclosure: I am a coeditor of this book.)

response from Slaviša Tasić

In 2050 we may be thinking along very different ideological lines and across dimensions we rarely give any thought to today. We have been witnessing ideological readjustment in the Western world for more than a decade. But the rise of various brands of populism, duly noticed and much discussed, may only be a materialization of more profound technological and epistemological changes that several authors have observed.

Martin Gurri. 2014. *The Revolt of the Public and the Crisis of Authority in the New Millennium*. Technological changes, including the rise of social media, have led to the flattening of authority. As Gurri first suggested in his belatedly acclaimed book, the vertical sources of authority in the form of the established media, corporations, government or academia have precipitously lost their standing. The horizontal flows of information and ideas have suddenly started to play a much larger role in our political economy.

Iain McGilchrist. 2009. *The Master and His Emissary: The Divided Brain and the Making of the Western World*. It was not simply that technology upended the means of communication and changed the ways we interact. As Iain McGilchrist daringly proposed in 2009, we may also be witnessing a historic epistemological shift from the dominance of the general, abstract, theoretical, and deductive style, towards concrete, mundane, inductive, common-sense and practical thinking. McGilchrist's traces the shift to neurological causes, but even if the causality is challenged, his mere identification of the shift is a precious accomplishment. One example of such shift in economics is the increasing appreciation for the empirical work at the expense of high theory. The pages of academic journals are filled with empirical findings, scholars are perfecting methodologies of casual inference, and Harvard's introductory economics course is purging theory to become empirical. The growing availability of data has also helped the flight from theory.

Gerd Gigerenzer. 2007. *Gut Feelings: The Intelligence of the Unconscious*. The best-known attack on theory fundamentals in economics has been the one coming from behavioral economists. While behavioral economics in the tradition of Kahneman and Tversky has been widely publicized and rewarded, Gerd Gigerenzer's work on common sense and ecological rationality epitomizes the

true break. Gigerenzer is a prolific author, but the 2007 book is a particularly useful summary of his ideas relevant for economists and others concerned with knowledge and decision making.

Nassim Taleb. 2012. *Antifragile: Things That Gain from Disorder*. Published in 1998, James C. Scott's *Seeing Like a State* is just outside the limits set for this assignment. Scott's book is a forerunner of the newest wave of literature that attempts to reestablish vernacular, experiential, and circumstantial learning, against overly rigid, linear, and abstract reasoning. Scott, McGilchrist's and Gigerenzer provide the epistemological foundations, but Nassim Taleb's *Antifragile*—and really his entire *Incerto* (2001–2018) series—puts flesh on the bones by providing insights valuable in economics, policy, finance, and everyday life.

Jeffrey Friedman. 2019. *Power Without Knowledge*. Jeffrey Friedman, a longstanding critic of the possibility of expertise in economics, political, and social affairs, takes epistemic skepticism even further in his most recent book. The consistency of his doubt in our ability to know, understand and manage may well be vindicated with time.

Patrick Deneen. 2018. *Why Liberalism Failed*. Liberalism from the title of this book is not the modern American political liberalism, nor is it classical liberalism alone. Deneen uses it to denote the entire Lockean individualistic constitution, and then argues that such liberalism is failing under its own weight. If the technological and epistemic changes I pondered above are important at all, and if it is true that they may be changing the landscape of political ideas, Deneen's 2018 book may turn out to be especially prescient.

response from Clifford F. Thies

I have three:

Angus Maddison, *The World Economy: Historical Statistics*, 2004. Maddison weaves varied measures into a history of GDP, population and perforce GDP per capita of the world and its regions and of countries as the data allow, going back to 'year 0.' His effort is continued by the Maddison Project. Measurement differentiates the scientific from the discursive disciplines.

Georg W. Oesterdiekhoff, *Kulturelle Evolution des Geiste* (*Cultural Evolution of the Mind*), 2006. Using a variety of methods, Oesterdiekhoff argues that the mass of Europeans of the 17th century had low IQs. Fits nicely with the Flynn Effect, Thomas Sowell's observation of the AFQT scores of white ethnics in WWI and WWII, and the progress of IQ in Poland and elsewhere in eastern Europe following the fall of the Berlin Wall.

Al Gore, *An Inconvenient Truth* (adapted from his PowerPoint presentation

and published in book form), 2006. A useful statement of the understanding of the global climate system of the left shortly after the turn of the century. So far, the forecasts of imminent disaster have proven incorrect. But hope springs eternal.

response from Richard E. Wagner

The ten books I list here reflect a type of balanced portfolio of books published between 2000 and 2020. The balancing principle is that this selection is *not* dominated by books on economics. Very strongly do I believe that economic principles should occupy the core of the humane studies. In light of that belief I have included several books by non-economists that present paths along which economic principles can potentially infuse themselves into the entirety of the humane studies.

Andrew Moore. *Shakespeare Between Machiavelli and Hobbes: Dead Body Politics*. Lanham, MD: Lexington Books, 2016. Machiavelli and Hobbes are surely the premier political theorists of our contemporary age. William Shakespeare not only bridges these thinkers chronologically but, and more significantly, his plays reflect themes that both Machiavelli and Hobbes explored in their political theories. Where political theory is abstract, drama deals with concrete situations. Andrew Moore explores the consilience among Shakespeare's drama and the political theories of Machiavelli and Hobbes. Moore explains that throughout his varied body of work, Shakespeare explored, through concrete settings, the relation between consent and force in politics.

Ruth E. Kastner. *Understanding Our Unseen Reality: Solving Quantum Riddles*. London: Imperial College Press, 2015. Ruth Kastner sets forth a transactional interpretation of quantum theory which brings alive the ideas of invisible hands and unseen realities which have pretty much disappeared from economics over the past century. Quantum theory makes heavy use of imaginary numbers, in stark contrast to the data-dominated character of contemporary economics. Imaginary numbers are the domain of plans and hopes that are looking to enter into history, and Kastner offers crisp insight into how quantum theorists have been negotiating this gap between what is observable in history and what is knocking on history's door.

Gregory Chaitin, Newton da Costa, and Francisco Antonio Doria. *Gödel's Way: Exploits into an Undecidable World*. London: CRC Press, 2012. Economists often labor under the presumption that all questions they ask have answers. Chaitin, da Costa, and Doria explain that there are many questions that can't be answered, at least within reasonable lengths of time. These undecidable situations become particularly noteworthy when multiple ingredients must be

combined before deciding. We are accustomed to illustrating our theories with models that have only a few moving parts. Yet reality often faces us with undreamt-of moving parts, in which case we fool ourselves by pretending we engage what we represent as models on whiteboards.

Bruno Latour. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press, 2005. Most social theorists treat societies as sets of data. In contrast, Bruno Latour examines societies from inside an analytical framework he denotes as actor-network theory. Within this framework, societies are subject to continual contestation where that contestation continually generates societal change depending on the path that such contestation takes. Just because a particular regime is in play at one moment is no guarantee that that regime will remain in force because contestation is ineradicable. Constitutions are unavoidably living constitutions.

Lorenzo Infantino. *Infrasocial Power: Political Dimensions of Human Action*. London: Palgrave Macmillan, 2020. While the principle of gains from trade is pivotal for economic theory, Lorenzo Infantino draws especially heavily upon Georg Simmel and Max Weber to explain how the operation of power is also unavoidable in the ordering of human society. The very scarcity that can be mitigated through trade also promotes the insatiable striving after power. Infantino explains that there are institutional arrangements that might soften the force of that striving though not eradicate it.

Jane Jacobs. *Dark Age Ahead*. New York: Random House, 2004. Most economists seem to be philosophical materialists, at least as judged by the emphasis they give to material growth as a solvent for nearly all social maladies. Jane Jacobs was not a philosophical materialist, and she recognized that contemporary society could be on the brink of a new dark age even as measures of GDP per capita continue to grow. As a counterfactual conditional, suppose a highly collective society could generate higher per capita income than a liberal society. Would this situation favor collectivism over liberalism? Jacobs reminds us that there is more to life than material comforts.

Matthew B. Crawford. *Shop Class as Soulcraft: An Inquiry into the Value of Work*. New York: The Penguin Press, 2009. Like Jane Jacobs, Matthew Crawford is no philosophical materialist. He recognizes that one of the main challenges people have in living well is cultivating meaning within their lives. There are many ways this might be done, and the professional classes surely are not reasonable guides. Crawford left a professional life behind to engage in repairing motorcycles where he found himself more deeply engaged in his daily activities than he did in his former professional career. Too much is the agenda of public life set by people who mostly believe that bowling alleys are museums of primitive culture.

Mikayla Novak. *Inequality: An Entangled Political Economy Perspective*. London:

Palgrave Macmillan, 2018. Inequality now seems to be on nearly everyone's agenda, rarely with anything either new or sensible being said. Mikayla Novak's *Inequality* is a luminous exception. She does not treat inequality as a simple matter of measuring Gini coefficients. Nor does she treat redistribution as an obvious remedy. Most significantly, Novak does not embrace the common mode of thinking, where markets generate inequality and politics alleviates it. To the contrary, Novak explains that politics and markets are thoroughly entangled, which leaves inequality as something often intensified through politics.

Manuela Mosca. *Monopoly Power and Competition: The Italian Marginalist Perspective*. Cheltenham, UK: Edward Elgar, 2018. Orthodox economic discussion of competition and monopoly has mostly construed the state as some god-like entity that would restrict private action to maintain a competitive organization of society. In this marvelous book, Manuela Mosca examines the thinking of four creative Italian economists, Vilfredo Pareto, Enrico Barone, Maffeo Pantaleoni, and Antonio de Viti de Marco, who recognized that political action was a common source of monopoly and who thought in terms of entanglement between market and political entities. These four theorists were thoroughgoing realists the better part of a century before the Chicago orientation to antitrust came into play.

Michael McLure. *The Paretian School and Italian Fiscal Sociology*. London: Palgrave Macmillan, 2007. Several Italian economists during the late 19th and early 20th centuries fashioned an orientation toward public economics that transcended the orthodox presumption of benevolent despotism, which still characterizes most of public economics. Those Italians, while friends and colleagues, did not speak with one voice. Where such thinkers as di Viti de Marco and Pantaleoni sought to reduce collective activity to the form of market activity, Vilfredo Pareto assembled a set of theorists who thought the gap between polity and market could not be bridged so easily. Michael McLure supplies an absorbing presentation of the central issues those Italian theorists debated.

Go to archive of Watchpad section
Go to March 2021 issue