



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

Econ Journal Watch

Scholarly Comments on Academic Economics

Volume 19, Issue 1, March 2022

Editor's Notes: Acknowledgments, April 2020 through March 2022 1–3

COMMENTS

Invalid Methods and False Answers: Physics Education Research and the Use of GREs
Michael B. Weisman 4–29

Compared to What? Does Benford's Law Really Detect Corporate Fraud?
Stephen Walker 30–46

On Whether the Size of Government Belongs in Economic Freedom Indices
Ryan H. Murphy 47–57

Leave Size of Government Out of the Measurement of Economic Freedom—
Put Quality of Government In
Jan Ott 58–64

Misrepresenting Mises: Quotation Editing and a Rejection of Peer Review at
Cambridge University Press
Phillip W. Magness and Amelia Janaskie 65–84

Long-Run Determinants of Economic Growth: Putterman and Weil Revisited
Jason Briggeman 85–108

From Hume to Smith on the Common Law and English Liberty: A Comment
on Paul Sagar
Jacob R. Hall 109–123

Comment on Measuring the Size of the Shadow Economy Using a Dynamic
General Equilibrium Model with Trends
Manuel A. Gómez and Adrián Ríos-Blanco 124–132

Response to “Comment on Measuring the Size of the Shadow Economy Using a
Dynamic General Equilibrium Model with Trends”
Mario Solís-García and Yingtong Xie 133–141

CHARACTER ISSUES

Liberalism in Colombia
Sebastián Rodríguez and Gilberto Ramírez 142–165

WATCHPAD

The Market for Goods and the Market for Ideas
Ronald H. Coase 166–175



Editor's Notes: Acknowledgments, April 2020 through March 2022

We are grateful to our institutional home and friend the [Fraser Institute](#), Canada's leading think tank, in particular Jason Clemens, Chris Howey, Cheryl Rutledge, and Venia Tan. For generous support we thank the Charles Koch Foundation, the John William Pope Foundation, and Gerry Ohrstrom (through Donors Trust). I am very grateful to Rich and Mary Fink and the Fink family for the generous Mercatus Center JIN chair, which I hold and that helps to support EJW.

I would like to extend thanks to co-editors Brendan Beare, George Selgin, and Larry White, editorial advisor Jane Shaw Stroup, web designer and master John Stephens, and my chief partner in the project, managing editor Jason Briggeman.

We are grateful to our authors for generously contributing their creativity, craftsmanship, and industriousness, as well as for their patience and cooperation in our editorial process.

We thank the following individuals for generously providing intellectual accountability to EJW:

Refereeing April 2020 through March 2022

Robert Adcock	American University
Alain Alcouffe	Toulouse Capitole University
Jeffrey Bergstrand	University of Notre Dame
Bernard Black (2x)	Northwestern University
Zeljka Buturovic	Institute for Social Sciences, Belgrade
Bryan Caplan	George Mason University
Eric Chyn	Dartmouth College
John Cochrane	Hoover Institution, Stanford University
Julio Cole	Universidad Francisco Marroquín
Tyler Cowen	George Mason University
Douglas Den Uyl	Liberty Fund
Daniel Duarte	Econ Americas
Ceyhun Elgin	Bogazici University
Zachary Elkins	University of Texas, Austin
Nils-Erik Forsgård	Magma, Helsinki

KLEIN

Richard Frankel	Washington University, St. Louis
Andrew W. Gelman	Columbia University
Peter Gordon	University of Southern California
Ian Gow	University of Melbourne
Alexey Guzey	guzey.com
Björn Hasselgren	Timbro, Stockholm
Miguel Hernan	Harvard University
Karen Horn	Nous Network
Jimena Hurtado Prieto	Universidad de Los Andes
Garett Jones (2x)	George Mason University
Robert Kaestner (2x)	University of Chicago
Petri Karonen	University of Jyväskylä
Karras Lambert	George Mason University
Lars-Folke Landgrén	University of Helsinki
David Larcker	Stanford University
Simon Lee	George Mason University
Guy Madison	Umeå University
Erik Matson	Mercatus Center, George Mason University
John McCusker	Trinity University, San Antonio
Evan Osborne	Wright State University
Sam Peltzman	University of Chicago
Rodrigo Pinto	University of California, Los Angeles
Niklas Potrafke	University of Munich
Kevin Quinn	Bowling Green State University
David Prychitko	Northern Michigan University
Daniel Raisbeck	Reason Foundation
Mario Rizzo	New York University
Daniel E. Roth	University of Toronto
Aldo Rustichini	University of Chicago
Margaret Schabas	University of British Columbia
Friedrich Schneider	Johannes Kepler University, Linz
John J. Siegfried	Vanderbilt University
Max Skjösberg	University of Liverpool
Craig Smith	University of Glasgow
Nico Stoeckmann	Paderborn University
Shirley Svorny	California State University, Northridge
Clifford Thies	Shenandoah University
Diana Weinert	Creighton University
Alex Young (2x)	Hofstra University
Leonidas Zelmanovitz	Liberty Fund

**Authors replying to comments in EJW,
published September 2020–March 2022**

Yang Bao	Shanghai Jiao Tong University	(article link)
Jamie Bologna Pavlik	Free Market Institute, Texas Tech Univ.	(article link)
Raj Chetty	Harvard University	(article link)
Tatyana Deryugina	University of Illinois, Urbana-Champaign	(article link)
Archelle Georgiou	University of Minnesota	(article link)
Jacob Goldin	Stanford University, Law School	(article link)
Farley Grubb	University of Delaware	(article link)
Nathaniel Hendren	Harvard University	(article link)
Pinar Karaca-Mandic	University of Minnesota	(article link)
Lawrence F. Katz	Harvard University	(article link)
Bin Ke	National University of Singapore	(article link)
Ithai Lurie	U.S. Department of Treasury	(article link)
Bin Li	Wuhan University	(article link)
Estefania Lujan Padilla	Free Market Institute, Texas Tech Univ.	(article link)
Janet McCubbin	Internal Revenue Service	(article link)
Sarah Miller	University of Michigan	(article link)
David Molitor	University of Illinois, Urbana-Champaign	(article link)
Jan Ott	Erasmus University	(article link)
Benjamin Powell	Texas Tech University	(article link)
Soumya Sen	University of Minnesota	(article link)
Mario Solis-Garcia	Macalester College	(article link)
Laura R. Wherry	New York University	(article link)
Yingtong Xie	Carleton College	(article link)
Y. Julia Yu	University of Virginia	(article link)
Jie Zhang	Nanyang Technological University	(article link)



Invalid Methods and False Answers: Physics Education Research and the Use of GREs

Michael B. Weissman¹

[LINK TO ABSTRACT](#)

Education is one of the largest sectors of our economy, and one that affects the performance of other sectors. Awareness is growing that modern causal inference methods are needed to help educators predict the consequences of different choices in teaching methods and in educational policy (Imbens and Rubin 2015; Murnane and Willett 2011). In this paper I will show how incorrect methods for predicting consequences are used even in a leading journal of physics education research, and how the major methodological errors allow misleading conclusions about one of the most contentious current educational policy issues, the use of standardized tests in admissions decisions.

Especially in the last four decades, sophisticated methods of causal inference have been developed to estimate potential outcomes from observational data (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Greenland, Pearl, and Robins 1999). Field-specific introductions to these methods are now available for economics (Varian 2016), education (Murnane and Willett 2011), psychology (Foster 2010; Rohrer 2018), epidemiology (Hernán and Robins 2020), biology (Glymour, Zhang, and Spirtes 2019), public health (Glass et al. 2013), sociology (Gangl 2010), political science (Keele 2015), and other fields (Imbens and Rubin 2015).

One might expect that physics, the most consistently mathematical of the natural sciences, would take the lead in employing valid new methods of

1. University of Illinois at Urbana-Champaign, Urbana, IL 61801. I thank Jamie Robins for crucial help in understanding causal inference. I thank the PRPER editors for initially suggesting this project.

quantitative reasoning to address its educational challenges. In a recent paper (Weissman 2021) I showed, however, that several papers published before 2021 in the leading journal of physics education, *Physical Review Physics Education Research* (PRPER), made major causal inference errors in particular applications. My previous paper focused on instances in which particular causal patterns were assumed without regard to the wide variety of plausible causal patterns equally consistent with the data.

Different specific issues are raised by three more recent PRPER papers, published in 2021 (Walsh et al. 2021; Young and Caballero 2021; Verostek, Miller, and Zwickl 2021). These papers include not only examples but also explicit instructions for methods whose use would lead to seriously biased causal estimands. The issues raised include the rules for which variables should be included or excluded from regression models to obtain unbiased causal estimands, the ways to identify causal mediators, the specification of potential outcomes, and less specifically causal issues such as the ways to impute missing data. The errors seriously distort causal implications for at least one major policy issue, the use of Graduate Record Exams (GRE) in admissions decisions, on which other seriously flawed papers have previously appeared, as discussed elsewhere (Weissman 2020b; 2020a).

Although the project of finding systematic problems in PRPER was initially suggested to me by its editors, PRPER has expressed a reluctance to publish any further work that describes some of its papers as incorrect. Therefore, I am publishing in this journal, which has been noted for featuring critical commentary (Gelman 2021).

I will start with some background material that some physics education research workers have suggested would be helpful, but will not repeat my previous very brief introduction (Weissman 2021) to the diagrammatic methods by which causal relations are often represented (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Rohrer 2018). I will then discuss an explicit causal methods paper (Walsh et al. 2021) which makes fundamental errors that, among other things, would justify techniques that could show any chosen causal effect to be zero. I then discuss a paper (Young and Caballero 2021) on use of GREs in admissions procedures that makes a similar explicit causal methods error as well as another major methods error, interpreted to justify dropping GRE use. I will conclude with a discussion of a paper (Verostek, Miller, and Zwickl 2021) on the implications of using GREs that adds a variety of other statistical errors, all tending to minimize the predictive power of GREs. The structure is thus to trace how confusion about methods builds a base for erroneous practical conclusions. My arguments will involve no new methods but are just applications of the standard beginning methods of the field.

Background on mediation, bias from including inappropriate variables, and treatment of missing data

In the potential-outcome approach to causation, the average causal effect of a change in one variable on another variable is the average change in the second variable that would occur if one were to change the first variable without intervening to change anything else (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020). This definition has an obvious correspondence to part of what one needs to know to decide whether to make the change in question, although for a more accurate estimation of the utility of some action one should know the full change in the probability distribution for the outcome rather than just its mean. A causal estimand is a formula estimating that causal effect from various correlations among other variables including the cause being studied (*ibid.*).

A causal estimand is said to be biased if it systematically mis-estimates an effect. Bias can arise when the estimand either contains terms that do not correspond to results of intervening on the cause or omits relevant terms that do correspond to such changes. (An excellent easily readable introduction to this issue can be found in Rohrer 2018.) The different coefficients found in different predictive models employing different predictive variables are not themselves ‘biased’; they are simply the coefficients of different models that would be used with different available predictive information. They become biased only if they are used to estimate something else—the causal effects. It is meaningless to say which coefficient is ‘true’ and which is ‘biased’ without specifying a causal question and the true causal model.

Descriptions of bias in causal estimands often focus on the role of omitted variables that have effects on both the cause in question and the outcome (Rohrer 2018). In a diagram, the simplest such confounder would be one with arrows to both the cause and the outcome. Simple cause-outcome correlations in models omitting the confounders give biased estimates of the actual causal effects because they include correlations between the cause and the outcome induced by the confounders. Controlling for confounders, e.g., by including them in an appropriate regression model, can remove this bias.

A causal mediator is a variable which appears as a node on a path from a cause to an outcome. Tar deposits in lungs provide a classic example, because tar appears on an important pathway from smoking to lung cancer. A more topical example would be inflammation as a mediator on one path from SARS-CoV-2

infection to death (RECOVERY Collaborative Group 2021). Operationally, a mediator can be identified because the portion of the causal effect that occurs via the mediated path can be shut down if the mediator can somehow be prevented from changing despite changes in the earlier cause.

This operational definition of mediation is a reminder of why it can sometimes be important to identify mediators. Sometimes it is easier to change the mediators by some intervention than it is to change the preceding cause. For example, intervening on the SARS-CoV-2→death pathway by suppressing the inflammation mediator with dexamethasone is fairly effective (RECOVERY Collaborative Group 2021), although not nearly as effective as suppressing the novel coronavirus itself with vaccines would have been. The effect of this intervention provides strong evidence that inflammation is indeed an important mediator, not just a marker for actual causes—a result that could not be determined simply from previously observed correlations. For a contrasting example, the weaker effectiveness of steroids in reducing mortality from sepsis shows that the inflammation accompanying sepsis is not such an important mediator in that analogous case (Yao et al. 2019).

When it is suspected that one of two variables mediates the other's effects on the outcome, the time order of the two potential causes can eliminate one of the two possible mediation relations, since the mediator must come after its cause. Mere time order does not, however, establish that a mediation relation exists. When, as often happens, the observed variables are not themselves on the causal path but rather are measurements of underlying latent variables, the time order of the measurements does not even constrain the direction of any possible causal relation between those latent variables. For example, if someone measures that they have a fever and then gets a PCR test that is positive for SARS-CoV-2, it would not be correct to use the time order of the variables temperature measurement and viral measurement to conclude that the fever caused the viral infection. The time order of the onsets of fever and infection is opposite to the time order of their measurements, and in this instance the latent variables' time order gives the true causal direction.

The portion of a causal effect that occurs via a measured mediator is often termed an 'indirect effect.' Any remaining causal effect for which no mediator has been identified is then termed the 'direct effect.' Despite this terminology, it is important to realize that there is no fundamental difference between these types of causation. On the classical scale, all causation proceeds through time-like event paths. Therefore any classical effect can be converted from 'direct' to 'indirect' by recording and controlling for a big enough set of mediators. Thus, despite the infelicitous standard terminology, the description of a causal effect as 'indirect' should not be interpreted to mean that it is any less real than one described as

‘direct.’

In general causal estimates of the magnitude of, e.g., smoking→cancer or SARS-CoV-2→death will be severely biased if they are based on models in which one controls for mediators, e.g., by including them in a linear regression model (Rohrer 2018). The effect of smoking cigarettes on lung cancer rates should not be estimated by excluding the large portion of the effect that occurs via tar deposits, since the tar deposits are themselves usually caused by the smoking. We say SARS-CoV-2 causes death, even though if one controlled for known mediators such as low blood oxygen, lung scarring, blood clots, and systemic inflammation, the ‘direct’ effect would be small.

Mediators are not the only variables that should not be controlled for if one wishes to use model coefficients to obtain an unbiased causal estimand, although they may be the most obvious such type of variable. Estimands will also be biased by controlling for variables that are affected both by the suspected cause and by other unmeasured causes that have effects on the outcome. Such variables are called colliders because in a causal diagram arrows from the suspected cause and the unmeasured causes collide on them. Controlling for a collider, e.g., by inclusion in a regression model, thus introduces collider stratification bias to the estimand (Pearl, Glymour, and Jewell 2016; Greenland 2003; Rohrer 2018).

One of the best-known illustrations of collider bias is the low birth weight paradox (VanderWeele 2014). In a model controlling for birth weight, which is a risk factor for infant mortality, maternal smoking becomes *negatively* associated with infant mortality. This does not mean that smoking is protective, but rather that the model was biased. Out-of-model causes of low birth weight (e.g., certain diseases or types of drug use) cause low birth weight, as does smoking or some variable associated with smoking. Thus birth weight is a collider between smoking and those out-of-model causes. A non-smoker with a low birth weight infant is more likely to have those other causes than is a mother whose smoking already created a likelihood of low birth weight. It turns out that the other causes are more dangerous than smoking, so among low birth weight infants those with non-smoking mothers are less likely to survive. Omitting birth weight from the model gives a positive association between smoking and infant mortality, which gives a more accurate representation of the causal effect (VanderWeele 2014).

In many real-world observational studies, some of the data are missing. It is often advantageous to impute the probability distributions of the missing data using the other data, both to avoid throwing out information and to avoid biased estimates of the true correlations that can arise when the data are not missing completely at random. An easily readable introduction to such imputation has been published in PRPER (Nissen, Donatello, and Van Dusen 2019).

Incorrect imputation methods can introduce bias even when the data are

missing completely at random. To explain the key reason, we can consider a toy example, so simple that one would not be tempted to use imputation in practice because it adds nothing to the information available from the complete-case data set. For simplicity I will describe the logic assuming large- N samples of bivariate normal variables, X and Y with correlation coefficient r and each with distribution $N(0,1)$. Say 25% of the X values are missing completely at random, so that the correlation inferred on the 75% of cases that are complete provides an unbiased estimator of the true correlation. Replacing the missing values with values imputed solely from what is known of X just means using imputed values from $N(0,1)$ that are uncorrelated with Y . Including them will be expected to reduce the correlation coefficient to $0.75r$ in the sample that includes imputed points, i.e., to give an estimate biased toward zero. If instead the missing X values are each replaced with a value imputed from the complete-case probability distribution conditional on *both* X and Y then the imputed x is of the form $ry + (1-r^2)^{1/2}e$, where e is an $N(0,1)$ random variable uncorrelated with X and Y . The correlation on the resulting sample is then expected to be r , an unbiased estimate of the true value.

In a more relevant case there are several predictors X_i with some correlations, not just one X . For simplicity we still assume that data are missing completely at random so that the correlations found in the complete cases are unbiased estimators of the true correlations, and assume multivariate normality so that the covariance matrix contains all the information. Imputing a missing x_1 via linear regression using the probability distribution for X_1 conditioned on both Y and the other X_i predictors will still not bias the coefficient estimates since the imputed distribution shares all covariances with the complete-case distribution, by construction of linear regression. Imputing the missing x_1 via a probability distribution conditioned only on the other X_i predictors would reduce the magnitude of the projection of Y onto the X predictor space, i.e., reduce R^2 of the model fit, just as in the toy one-dimensional case. Thus omitting outcome variables from the conditioning will bias estimates, causing them to be less predictive even in the simplest, least problematic cases.

Turning to the three PRPER 2021 articles

I turn now to the three articles which I shall criticize. I abbreviate the author groups as follows and treat the articles in the order shown:

- **WSTSH:** Cole Walsh, Martin M. Stein, Ryan Tapping, Emily M. Smith, and N. G. Holmes, “Exploring the Effects of Omitted Variable Bias in Physics Education Research,” *Physical Review Physics Education Research*,

- 2021.
- **YC:** Nicholas T. Young and Marcos D. Caballero, “Physics Graduate Record Exam Does Not Help Applicants ‘Stand Out,’” *Physical Review Physics Education Research*, 2021.
 - **VMZ:** Mike Verostek, Casey W. Miller, and Benjamin Zwickl, “Analyzing Admissions Metrics as Predictors of Graduate GPA and Whether Graduate GPA Mediates Ph.D. Completion,” *Physical Review Physics Education Research*, 2021.

Choosing variables to include in causal models

WSTSH (2021) attempt to introduce PRPER readers to some of the issues involved in inferring causation from correlations, focusing on bias that can result from omitting variables. Although much of the verbal introduction is correct, the rules about which variables should be controlled for are stated incorrectly and the examples used undercut the message that one needs to carefully define causal questions before determining the correct combination of correlations needed to infer a causal effect with minimal bias. The title already conveys that bias arises from *omitting* variables from a model, but as we have seen bias can equally well arise from *including* inappropriate variables in a model (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Greenland 2003); here and below, ‘including’ means controlling for by inclusion in a regression model, the usage of WSTSH.

The following general rule is stated by WSTSH (2021, 3):

A variable included in a model will be biased by an omitted variable if the following two conditions are met:

- (1) the correlation between the omitted variable and the included variable is nonzero, and
- (2) the “true” effect of the omitted variable on the dependent variable is nonzero.

Although following that rule could succeed in removing simple confounders, the rule is untrue in general, since it would often imply that a causal model should include mediators and colliders (Rohrer 2018). *Following this rule would allow one to show that essentially any causal effect is zero by including enough mediators in the model.*

WSTSH’s introduction promises to give examples using “explanatory” models, as distinct from “predictive” models, for “testing causal hypotheses.” Nevertheless, in practice WSTSH’s examples consistently contrast different “predictive” models without specifying causal patterns. Coefficients in some

models are described as “biased” as compared to coefficients in others. This reflects a misunderstanding of what “bias” means. To say that a properly calculated predictive model is biased means that some of its coefficients are being used to represent something that they do not in fact actually represent. In this context that something else is the causal effects of *intervening* on some variable. Whether a coefficient from a predictive model gives a biased estimate of that causal effect depends on the relations among the different variables in a causal diagram (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Rohrer 2018). Which, if any, of the coefficients are unbiased cannot be judged without knowing the true causal model.

WSTSH go on to discuss various cases in which adding a variable to a multiple regression model makes either large or small changes in the coefficients of other variables. Throughout, the assumption is made that the extra variables should always be used when they make large changes, in order to approximate the “true” model (quotation marks as in WSTSH). The question investigated is “whether omitting a particular variable will lead to bias” (WSTSH 2021, 6), but the other half of the question, whether *including* a variable will lead to bias, is simply not mentioned. As we have seen, this approach is fundamentally mistaken because inclusion of colliders or mediators as covariates often systematically biases the relevant model coefficient away from the desired causal coefficient (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Glymour, Zhang, and Spirtes 2019; Glass et al. 2013; Varian 2016; Foster 2010; Gangl 2010; Keele 2015; Greenland 2003; Rohrer 2018).

The first example given by WSTSH (2021, 2) follows the same logical pattern as the smoking→tar→cancer example. It concerns improved outcomes for males, compared to females, when labs are switched from real-world to virtual reality (VR), i.e., the difference between what I will call ‘VR_benefit’ for males and for females.

The wording of WSTSH gives the impression that the causal question is the effect of gender on VR_benefit. It is not entirely clear to what interventions knowing the causal effect of gender on VR_benefit would be relevant since it is unlikely that anyone deciding on some gender-determining action would be much concerned with the effects of their decision on optimal physics lab modes. Nevertheless it is meaningful to consider gender as a cause, an exogenous random variable that at the time of the study was almost always determined by a quasi-random conception lottery, ignoring selection effects involved in being included in the cohort studied. When considering gender as a cause, however, it is important to realize that its effects are very strongly dependent on social context. For example, it is not a universal biological truth that being male leads to video gaming.

As a purely predictive relation the difference in VR_benefit for males and

females could be relevant to deciding which students might benefit from which lab mode, but for purely predictive relations one needs only conditional probabilities, not a causal model. A raw comparison of the distributions of VR_benefit for males and females suffices for predicting the effects on these groups. The paper claims that the model should include another variable, video game experience, which correlates very strongly with gender and strongly with VR_benefit (WSTSH). The model including gaming gives results that are contrasted with any “conclusions about biological or sociological differences in men’s and women’s abilities to learn from VR” (WSTSH 2021, 2), although it is unclear why gaming is not considered a sociological difference.

The diagram in Figure 1 shows a common-sensical causal pattern, in which being male leads, in the particular social context, to video gaming. Being male can lead to VR_benefit by the video game path or by other paths so the causal effect of being male on VR_benefit is just given by the unconditional regression coefficient in this simple linear model. Inclusion of gaming as an independent predictor biases the estimate of the socially conditioned causal effect of maleness on VR_benefit because it eliminates one of the causal paths for that effect. WSTSH suggest “running the analysis with video game experience in lieu of gender” (2021, 2), which corresponds to erasing not only the arrow from gender to gaming but also the arrow from gender to VR_benefit, in effect treating gender as an irrelevant marker for the gaming.

Which coefficients in the common-sense model are relevant depends on what action is being considered. Is the potential action choosing whether to offer VR labs at all, choosing which students to encourage to take them, encouraging students to play video games, or something else? To give a not very serious example, if one wants to know whether students should be encouraged to play video games to prepare for VR labs, then one would want to know to what extent gaming is an actual mediator on the gender→VR_benefit path or just a marker for gender. Because the correlation between maleness and video gaming is said to be very high, that may be hard to disentangle from the observational data. Nevertheless, clarity about what actions are under consideration would help to guide which coefficients in which diagrams are relevant.

Although WSTSH are not proposing anything that would actually affect choices about interventions on gender, the method used for estimating the effects of gender would consistently give incorrect causal coefficients for treatments on which interventions are possible. The causal question addressed is precisely analogous to the more serious one of finding whether inflammation is an actual mediator of SARS-CoV-2→death, which determines whether intervening with dexamethasone is useful. For an educational policy example, inclusion of time spent viewing lectures in a multiple regression model would bias the estimand

of the effects of two different video lecture styles on exam scores, because that variable would be a mediator on the path from style to scores. Viewing time would become relevant only if there were another way of intervening on it.

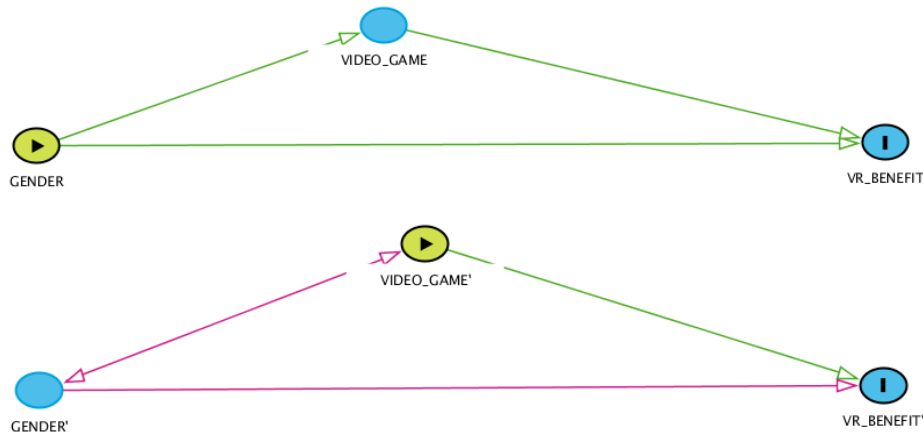


Figure 1. The top figure represents a common-sense view of a causal relation between gender, gaming, and response to virtual-reality labs, in a particular social context. The Markov-equivalent second graph is the one implicitly used to obtain the Gender→VR_benefit coefficient in which gaming and gender are treated as correlated variables neither of which causes the other. One interpretation given by WSTSH would also drop the arrow from gender to VR_benefit, giving a nonequivalent graph that would in principle be distinguishable via observed correlations.

Admissions criteria: Identifying mediators and specifying potential outcomes

In the second article criticized here, YC (2021) examine the causes of admissions to selective graduate programs in physics, focusing on the roles of Graduate Record Exams in physics (GRE-P) and undergraduate grade point averages (UGPA), employing a structural equation model (SEM) analysis (Bollen and Pearl 2013) of the causal role of these variables and others on admissions. The causal theme is to see what effects dropping GRE-P would have on admissions, and in particular whether GRE-P is helping some students be accepted who might otherwise be overlooked. I shall focus here on problems in the translation of statistical claims to causal claims, though also mentioning some in the conversion of data to statistical claims. The causal inference problems include an incorrect rule for identifying causal mediators and unrealistic framing of the potential outcomes

for the interventions under consideration.

YC's informal statistical discussion of the effects of including GREs on admissions has some statistical flaws. Although students appear multiple times in the admissions data because they applied to multiple departments, the paper does not consider the major effects of those clusters of dependent data points on the statistical uncertainties of calculated parameters. Although YC tell us that the data sets included enough parameters to allow these single-student clusters to be identified with good reliability, that information is not used. The justification given is that eliminating duplicates has little effect on distributions of single parameters such as UGPA, but we are not told what effect it would have on the more important causal estimates. When statistical confidence intervals are given, we are told only that they were obtained by bootstrap methods (YC 2021, 6), but the standard forms of those assume that data points have independent errors, which is far from the case for single-student clusters.

YC's initial informal causal discussion looks at acceptance rates in small rectangles of the UGPA–GRE-P plane, without estimates of statistical uncertainty. When a rectangle in the low-UGPA, high GRE-P range shows a high acceptance rate, we are told that the high GRE-P was probably not responsible for that because a nearby rectangle shows a low acceptance rate (YC 2021, 9). Causal interpretations in which one picks and chooses which data points to take seriously are scarcely constrained by data.

The form of the more quantitative SEM analysis has a peculiarity. The causal coefficients for, e.g., GRE-P→admission are calculated separately for four distinct SEMs for different initial causes (race, gender, undergraduate selectivity, and undergraduate institution size), with some substantial variation in the four SEMs. The paper includes no comment on why the variables are not included in a single consistent model with a single value for each such coefficient.

YC's SEM analysis emphasized testing for mediation. Although the verbal definition of causal mediation is presented correctly, the mathematical method presented for identifying mediators treats one causal variable (e.g., GRE-P) as mediating a path from another cause (e.g., UGPA→admission) whenever the 'mediator' can be expressed in terms of the other cause plus a random component, as explicitly shown in their Equation 2. The logistic regression SEM analysis assumed that GRE-P mediates UGPA→admission and concluded that the mediation effect was large because the dependence on UGPA of the logit for the admissions probability changed substantially when GRE-P is held constant, an implication of YC's Figure 13.

Nevertheless, it is evident that the choice of which correlated variable, if either, to call a mediator cannot be based just on the correlation itself, since that relation is essentially symmetrical: the least-squares estimate of each normalized

variable is simply the correlation coefficient r times the other normalized variable, plus a variable that is orthogonal to the predictor variable and has variance $(1-r^2)$. This familiar symmetry is just the simplest case in which different causal graphs have exactly equivalent predictions for correlations (Pearl, Glymour, and Jewell 2016; Richardson 2003). Following YC's procedure would produce evidence of 'mediation' in all cases where two causal variables are correlated, regardless of whether any actual causal mediation exists. That procedure confuses the estimation of coefficients for a known causal diagram with the discovery of the proper causal diagram.

To give a graphical illustration of the issue, I turn to the SEM on the effects of "Race," Figure 13 of YC (2021), reproduced here as Figure 2. "Race" in YC (2021) stands for an aggregate approximately meaning the same as 'underrepresented minority.' The correlations were interpreted to show that GRE-P mediates the causal effects of GPA on admission, and thus also the effects of Race on admission via UGPA. What the correlations themselves show, however, is that Race has very little correlation with GRE-P (not statistically significant in the sample) when UGPA is held constant. YC include no discussion as to why the relation $UGPA \rightarrow GRE-P$ is assumed.

The mistaken mathematical conception of mediation leads to at least one strange implication. A literal reading of the original SEM, taking its causal diagram seriously, would say that the way to get rid of the negative effect of Race on Admit would be *to eliminate UGPA*—not just to drop it as an admissions factor, which would still allow a negative Race effect to flow through it to GRE-P, but to actually eliminate it, perhaps by abolishing grades. That would almost entirely block the path for the negative causal impact of Race on Admit, since almost none of that impact bypasses UGPA in that SEM.

A more conventional view might be that UGPA and GRE-P are correlated because both are affected by a variety of shared factors, e.g., diligence in studying physics. This alternative is illustrated by the second graph in Figure 2, which is Markov equivalent to the first graph and therefore exactly equally consistent with the correlations (Pearl, Glymour, and Jewell 2016; Richardson 2003). This modified SEM does not have any peculiar (presumably unintended) implications about what effect eliminating UGPA would have on GRE-P. Although for the most part UGPA is obtained before GRE-P, that has no implications for the causal order, if any, between the latent variables underlying them.

The correlations encoded in the original SEMs have other, less peculiar, policy implications than the SEMs themselves. The original SEM and the revised one are Markov equivalent, so they share the same correlations between Race and the variables UGPA and GRE-P, i.e., -0.48 and -0.31 , respectively (-0.31 is simply the sum of the effects on the direct and mediated paths from Race to GRE-

P in the original SEM, $-0.094 - 0.480 \cdot 0.445$.) That means that replacing GRE-P with more weight on UGPA could plausibly have *decreased* the admission of underrepresented minorities in this cohort, a conclusion that matters for the policy choices under consideration. This simple conclusion is not included in the paper. For gender, in contrast, such a substitution would favor females, as can be seen from the coefficients in Figure 12 of YC (2021).

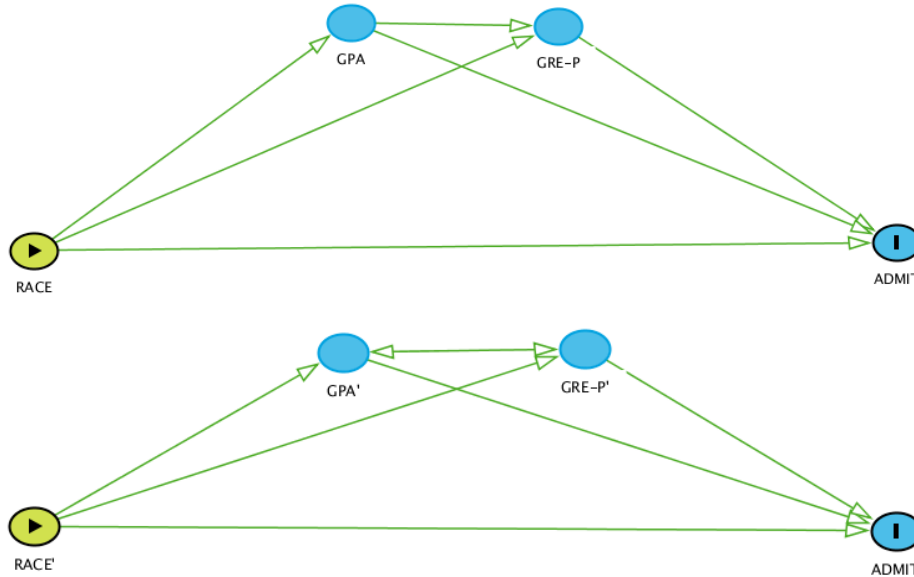


Figure 2. The SEM graph used to describe the causal effect of “Race” on admissions, as given in YC (2021) on top, and using a Markov equivalent and seemingly more plausible graph on the bottom. Bidirectional arrows stand for causation by exogenous unmeasured variables, not two-way causation. The latter graph avoids the false implications that eliminating UGPA would eliminate effects of Race on GRE-P or that dropping GRE-P would reduce the effects of racial differences in GPA on admissions.

Another fundamental issue arises in the framing of the causal effects, i.e., potential outcomes, of changing admissions criteria. Unsurprisingly, YC show that admission probability tends to increase with higher UGPA and GRE-P. The paper emphasizes that “more applicants could be penalized for having a low physics GRE score despite a high GPA than could benefit from having a high physics GRE score despite a low GPA” (YC 2021, 8). The net effect on acceptance of changes in acceptable levels for different criteria, however, depends on the ratio of the hypothetical incremental changes for those two criteria and thus cannot be determined without specifying that ratio. YC do that specification via arbitrary definitions. More fundamentally, framing the outcome as the total number of accepted applicants does not consider the actual potential outcomes, in which the

total number of accepted applicants who can attend graduate school is not set by admission criteria but by limits on funding, mentorship, and job openings. Any change of admission criteria may be judged on a variety of grounds, but these do not include effects on the total number of new graduate students, which can only be changed by other methods. Changing criteria changes which students are accepted, not how many. The comparison of net acceptance rates for different criteria may, however, make sense if the implicit goal is to increase the number of domestic students by reducing the number of international students, who were omitted from the analysis (YC 2021, 4).

GRE scores as predictors: collinearity, collider bias, and imputation bias

Many physics and astronomy departments have dropped or are considering the possibility of dropping use of GREs for admission, dropping either just GRE-P or both GRE-P and the more widely taken quantitative GRE, GRE-Q (Chawla 2020; Young 2020). Estimating the causal effects of these actions is therefore worthwhile. Some of the pathways by which these choices might affect various outcomes (e.g., by effects on both graduate and undergraduate curricula and grading) are essentially unmeasurable without trying the experiment. Other pathways, such as the potential performance of students who are not currently admitted, may be roughly estimated via model-based extrapolation from results on how well GREs predict graduation in the currently admitted cohort (Miller et al. 2019; 2020; Weissman 2020b; 2020a; VMZ 2021).

This latter approach is used by the third article criticized here, VMZ (2021). Although VMZ use somewhat vague language to describe their aims, e.g., asking which metric “offers the most insight” (VMZ 2021, 1), in context (Miller and Stassun 2014; Chawla 2020; Young 2020) there is no doubt that the treatments under consideration are to drop use of GREs in graduate physics admissions. Unlike for the two papers I discussed above, each of which was characterized by a small number of core errors in causal reasoning, for this paper I will discuss more conceptually disparate errors, united by sharing a sign of their effect in biasing causal estimates.

A causal diagram approximately representing the actual admissions policy choices to be made is shown in Figure 3. Different criteria may be used or not used in choosing whom to admit. These criteria include the ones that are relatively easy to tabulate and study (UGPA, GRE-Q, and GRE-P) and various other factors that are used by most departments (Potvin, Chari, and Hodapp 2017) but are harder to

quantify and tabulate across schools and thus comprise out-of-model-predictors (OOMPs). The effect of choosing to use or not use a criterion is conditional on choices for the other criteria, since inclusion of a highly redundant predictor has little effect, and redundancy depends on what else is used.

To determine the effects of dropping both GRE-Q and GRE-P on the predictability of graduation, one should compare predictions using those two GREs with predictions excluding them but otherwise the same. That direct comparison for graduation rates is not made in VMZ (2021) or its predecessors (Miller et al. 2020; 2019), but we shall see that it may be estimated from the data presented.

Figure 4 shows a causal diagram roughly representing the key points of the model for what causes an individual student to be more or less likely to graduate. Coefficients from this model can then be used implicitly or explicitly to estimate the effects of using different metrics in the policy diagram of Figure 3. Figure 4 encodes some assumptions, probably not quite exactly true but not especially controversial. One is that whether a student is admitted and if so to what rank of department depends on the explicit metrics available to the admissions committees, not on the traits that they partially measure. Another is that the metrics only partially reflect some unspecified traits of the applicants that make them more or less likely to graduate. Correlations between metrics arise because they measure overlapping traits. I arbitrarily use four such traits in Figure 4, enough to allow linear independence of their different combinations in the four metrics.

A causal diagram is included as Figure 5 in VMZ (2021), but since it includes only a single metric, it does not capture that each causal effect in our Figure 3 is conditional on which other metrics are also used. That dependence approximately corresponds to the dependence of the predictive coefficients on which other predictors are included in a model for individual students like Figure 4. Thus the diagram used by VMZ obscures the key point—that dropping either GRE substantially raises the coefficient and the statistical significance of the other GRE for graduation probability (Weissman 2020b; 2020a). The effects of dropping one predictor on the other's coefficient are later given (VMZ), but only for a different outcome, graduate GPA (GGPA), for which the effect is small because GRE-Q adds little to its prediction, unlike for graduation.

To estimate the causal coefficients of the policy-choice model one needs first to correctly estimate the coefficients and statistical uncertainties of the predictive coefficients of the individual-level model. VMZ (2021) give coefficients and confidence intervals for the coefficients of GRE-P and GRE-Q for predicting graduation within the stratum of enrolled students within a model including UGPA and various demographic predictors. Within the main text, VMZ state at least 12 times that neither of the separate coefficients meets the conventional cutoff for

statistical significance. In the Supplement, however, we find that when a somewhat less biased method is used for imputing missing data, both coefficients actually do pass the standard “significance” criterion (VMZ 2021, Supplement Section V). (Here I do not mean to endorse the common null-hypothesis-significance testing approach of making qualitatively dichotomous interpretations of small differences in coefficient p-values (Amrhein, Greenland, and McShane 2019), but merely go along with it to focus on other issues.)

In VMZ as well as in its predecessors (Miller et al. 2020; 2019) a substantial amount of missing predictive data was filled in using a previously unspecified (Miller et al. 2019) multiple imputation method. VMZ now point out that in its main text as well as in the preceding work the imputation model used omitted outcome variables, claiming that “the imputation approach presented here is theoretically sound” (2021, 6). VMZ’s Supplement expresses surprise that the “counterintuitive” inclusion of outcome variables results in less biased estimates, based on concern that “Employing a model of data imputation that uses the outcome variable to predict missing values of the independent variable may seem like a self-fulfilling prophecy, guaranteeing a relationship to exist between them” (VMZ 2021, Supplement p. 13). Nevertheless, the Supplement concedes that “research suggests that including all variables, including the outcome variable, in the imputation model in fact tends to produce less biased results.” We have seen, however, that one needs only algebraic inspection of the simplest models to reveal that imputation methods that omit outcome variables from the conditioning are not theoretically sound because they introduce systematic bias. (This issue may provide a clue as to how proficiency in beginning algebra, as measured by GRE-Q, could help predict research performance.)

VMZ’s Supplement now gives imputations including conditioning on GGPA. Even that partial correction suffices to make GRE-P and GRE-Q statistically significant predictors of graduation in the U.S. cohort. It appears, however, that graduation itself was still not included in the imputation method of the Supplement, so the results would still be biased toward weakening the model’s predictive power, especially for that outcome. Estimating from blow-ups of figures in VMZ’s main text (Figure 4) and Supplement (Figure 5) as well as from the decrease in p-values and the partial data given on effect size for GRE-P (VMZ 2021, Table 5 and Supplement section V.B), just conditioning on GGPA raised the sum of the standardized coefficients of GRE-P and GRE-Q for graduation by about 8 percent. The standardized coefficient of UGPA went up about 50 percent.

Insufficient data are given to tell how much the coefficients for graduation would go up if that imputation were also conditioned on the more relevant outcome, graduation itself. One may get a rough idea from the effect of GGPA conditioning on the coefficients for predicting GGPA. When GGPA was used in

the imputation conditioning (VMZ 2021, Supplement Table V compared to article Table III) R^2 for the model increased from 0.11 to 0.17 and the coefficients of UGPA and GRE-P for GGPA went up about 35 percent (VMZ). (The coefficient of GRE-Q for GGPA was very small and not significant in either imputation.) It would be reasonable to guess that the sum of the GRE coefficients for graduation would also go up another 10 to 30 percent if graduation were properly included in the imputations for that model. The actual value could be calculated easily by the original authors.

Some of the graduation outcome data were also missing because at the end of the data window some students had been enrolled for five years but had not yet graduated. Although it was stated that roughly 95 percent of these were likely to graduate, the model treated their graduation rate as 100 percent, in effect substituting five-year survival for graduation as the outcome (VMZ 2021, 4). A less biased result would be obtained by using imputation based on the 95-percent estimate. The probable effect of this fairly small unnecessary bias, like that of the larger errors, is to understate the predictive power of GREs (Weissman 2020b; 2020a).

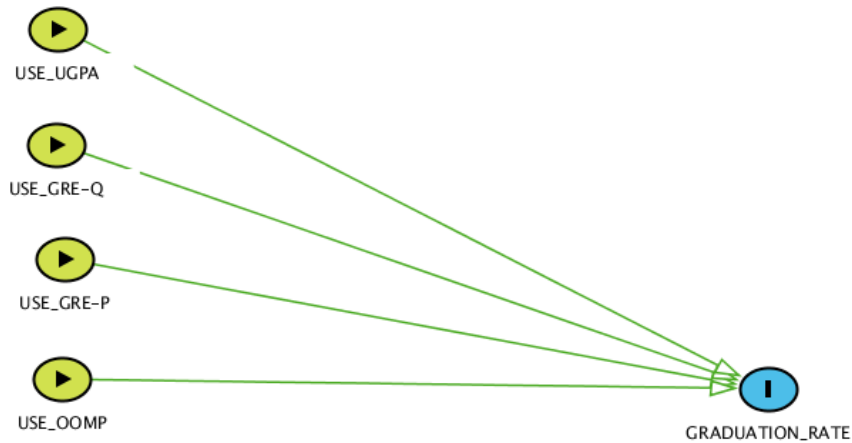


Figure 3. This graph illustrates the effects of possible choices of admissions criteria on the net graduation rate for a particular program. Out-of-model predictors (OOMP) are available to admissions committees but not to subsequent modelers.

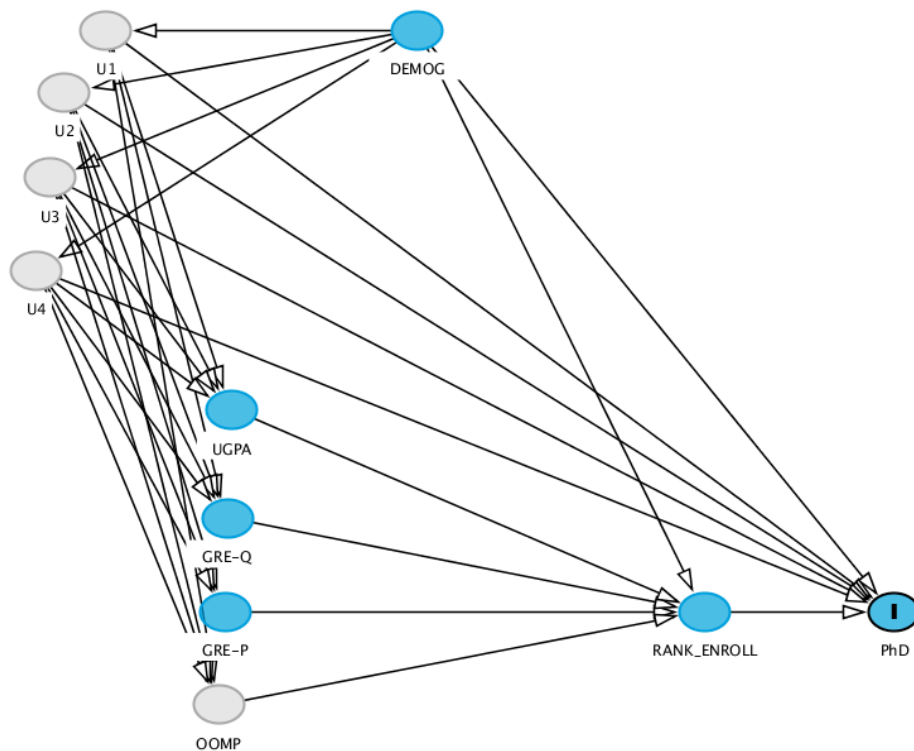


Figure 4. This graph illustrates how the variables available to the admissions committees and the subset available for the overall model are related to the probability of graduation. It contains some explicit assumptions, e.g., that the causal effects of the variables with no direct arrows to PhD are negligible except as mediated by the necessary step of getting into at least one program, even though those variables can be important markers of traits that do affect graduation rates. The correlations among the predictor variables arise because they share underlying causes. Since there are no data on students who were not admitted to any programs, some bias from stratifying on the collider “rank_enroll” is unavoidable. Stratification into rank tiers or into the narrower strata of individual programs increases that bias.

Once the coefficients relevant to Figure 4 have been established, one must estimate their implications for the coefficients of the policy choices of Figure 3. Since GRE-Q and GRE-P are highly correlated, collinearity effects make the separate point estimates and confidence intervals for their coefficients not directly relevant to the question of whether dropping both predictors would significantly reduce predictive power. *Estimating the effects of dropping one or both predictors requires a calculation that uses both the coefficients and the correlations among the predictors.* A previous analysis (Weissman 2020b; 2020a) showed unambiguously that the predictive power for graduation would fall significantly if both GREs were dropped, an effect

in the overall cohort of more than four times the standard statistical error even using the improperly imputed estimates. We shall see that consideration of the information in VMZ (2021) will give the same conclusion but with larger effect sizes.

Even correct use of within-sample predictive coefficients would still give somewhat biased estimates of how much predictive power would be lost by dropping each predictor, because the data are necessarily restricted to those students who were admitted and then enrolled (Weissman 2020b; 2020a). As we see in Figure 4, graduate admission is a collider between the model variables (UGPA, GRE-P, GRE-Q, demographics) and OOMP, a summary of other admission criteria used by almost all programs (Potvin, Chari, and Hodapp 2017). Some collider stratification bias is thus inevitable.

In the initial paper on GRE predictions from the Miller group (Miller et al. 2019) the inevitable collider bias was amplified by further stratifying the sample into three narrow “tiers” of graduate program rank. In response to my objection that this procedure increased collider stratification bias (Weissman 2020b), the authors responded with a reanalysis that eliminated the word “tier” but replaced the three strata with individual programs, i.e., 20 *narrower* strata (Miller et al. 2020).

Table V of VMZ now allows an estimate of the magnitude of the collider effects since it gives the results of the most recent analysis and that of the previous three-tier analysis (Miller et al. 2019). (The comparison is very slightly complicated because the new estimates appear to use a probit link function and the old ones use a logit link function, but that should have very minor effect on this comparison. Also the new numbers use only about 80 percent of the old sample, but VMZ do not give any indication that that change affects the comparison.) The new less-stratified graduation predictive effect sizes for UGPA, GRE-P, and GRE-Q were increased from those found in the more-stratified model by factors of 1.7, 1.2, and 1.6, respectively. This substantial change illustrates the importance of collider bias. These factors are far more precise than one might guess from a raw combination of the separate confidence intervals on the more- and less-stratified model results, because both models use the same set of partially random outcomes. The actual narrow confidence intervals for the factors could be determined by a pairwise bootstrap calculation using the actual data sets.

We may apply these correction factors to my previous calculation of the effect sizes for the logit of graduation probability as a function of UGPA and equal-weight GRE sum (Weissman 2020b; 2020a). Based on the highly stratified results, the logit effects on going from the 10th to 90th percentile among the U.S. test-taking cohort, holding other predictors constant, were previously found to be 0.60 and 0.72 for UGPA and equal-weight GRE sum, respectively (Weissman 2020b; 2020a). After correction for the excess stratification of that model each logit

effect is 1.0, coincidentally. I suspect that the remaining stratification bias from the necessary absence of OOMP in the newer less-stratified results is not as big as the difference between them and the highly stratified earlier results, in part because the remaining stratification bias is likely to be partly balanced by some confounding bias (Greenland 2003; Weissman 2020b; Rohrer 2018).

Using the new imputation results, partially repaired by including GGPA conditioning, would raise those estimates to 1.5 and 1.1, respectively, for UGPA and GRE sum. Fully repairing the imputation method would likely raise both effects substantially more. It would be good to see the actual result of that straightforward calculation, preferably on the full data set used in Miller et al. (2019).

Estimating the causal effects of admitting students who are now excluded because of low GRE scores or because they did not take any GREs requires extrapolation of results from students who are currently admitted to that broader cohort. Extrapolation involves use of some model link function, e.g., either logit or probit, to convert predictors to probability. This extrapolation is uncertain, especially for GRE-Q, for which the range among all the students who take the test and are interested in going on in physics is more than twice as large as in the enrolled group (Miller et al. 2020). Range restriction is much less severe for GRE-P and UGPA (*ibid.*). Data on the enrolled group can provide at least some hint as to whether that extrapolation is problematic, e.g., by seeing if outcomes for the bottom quintile of those enrolled fall below expectations from the overall model fit. Since extrapolation to the full U.S. cohort, even via the linear logistic model, predicts bigger effects for GRE-Q than for GRE-P (Miller et al. 2019), it would be particularly important to see a quintile breakdown for GRE-Q. The Supplement to VMZ provides such graduation results for quintiles of UGPA and GRE-P but omits the potentially more important results for GRE-Q.

We can now turn to the question of whether dropping just the GRE-P while retaining GRE-Q would significantly reduce predictive power. According to VMZ's Supplement, GRE-P has a significant predictive coefficient in the full model. From results (Miller et al. 2020) for the Akaike Information Criterion (Akaike 1974), dropping both GREs would give a much worse model overall and within the U.S. cohort but dropping just GRE-P only gives a marginally worse U.S. model. The loss of predictive power should be somewhat less than the point estimate of the GRE-P predictive coefficient might lead one to expect (Weissman 2020b; 2020a). The reason again is that once one predictor is dropped, the coefficients of the other predictors change. GRE-Q and UGPA, both correlated with GRE-P, would take up some of the slack left by dropping GRE-P (Weissman 2020b). That's also why if one GRE is dropped, the incremental predictive power obtained from the remaining one is increased, so that dropping the second GRE

gives a *larger* loss than its coefficient in the full model would suggest (Weissman 2020b; 2020a).

The primary motivation for examining the effects of using GREs on admissions is stated to be concern for underrepresented minorities (URM, approximately corresponding to “Race” in YC) and for females. The effect of dropping GREs on URM admissions will depend on the balance of which other factors are used to replace them. Although VMZ do not specify what combination would be used, they emphasize the superiority of UGPA as a predictor. We have seen, however, that in the cohort studied in YC the standardized prediction coefficient for Race→UGPA is *more* negative than that for Race→GRE-P. (The coefficient for GRE-Q was not given.) Thus not only would dropping GREs have a greater negative effect on graduation probability than suggested in VMZ, but it could also have a negative effect on URM admission if more weight is placed on UGPA.

On the other hand, the GREs (especially GRE-P), unlike UGPA, do systematically underpredict both female graduation rate and GGPA (VMZ 2021, 14). Female admissions could be increased by dropping or de-emphasizing the GREs. The same goal could also be achieved by gender-norming those scores, without losing the significant information provided by the GREs. It seems likely that something like that norming was already being done in the period for which data were collected, since the overall female GGPA is insignificantly less than the male GGPA (*ibid.*). Without something like gender-normed admissions, the female GGPA would have been expected to be higher than the male GGPA, since it is significantly higher than predicted by a UGPA-GRE model (VMZ). Figure 12 of YC (2021) confirms that the standardized coefficient for predicting the admissions probability logit for females in a model controlling for GRE-P and UGPA was in fact large and positive.

Much of VMZ (2021) is devoted to a “mediation analysis” in which GGPA is treated as a potential mediator on the paths from UGPA, GRE-Q and GRE-P to graduation. The predictive effects of UGPA and GRE-P for graduation are described as almost entirely “mediated” by GGPA because controlling for GGPA removes almost all of the dependence of graduation probability on UGPA and GRE-P in the model. Just as in the mediation cases we examined above, however, that result does not show whether GGPA actually mediates the causal effects on graduation of the traits measured by UGPA and GRE-P or merely serves as another marker for those traits. VMZ do distinguish between the possibilities that maintaining GGPA can be a roadblock to staying in a program or can be a marker for other difficulties, but lump these together as “mediation.” Only to the extent that GGPA actually mediates the causation or serves as a marker for true mediators, as opposed to being a marker for pre-existing traits, would interventions

that raise it (e.g., better classroom teaching or easier grading) increase graduation rates. (Interestingly, controlling for GGPA scarcely changes the dependence of graduation probability on GRE-Q in the model, suggesting that completing a PhD partly depends on some traits measured by GRE-Q that would not be changed much by changing classroom teaching.)

UGPA and GRE-P are described as providing only “indirect prediction,” an unusual cross between causal and predictive terminology. The use of causal terms, “mediation” and “indirect,” in describing a purely predictive relation, is somewhat misleading. The pre-admission predictors are no more or less predictive of graduation than they would be if post-admission GGPA had not been subsequently recorded.

The methodological errors in VMZ are important in themselves, but in this case, the substantive conclusions are also important, so a summary of the key corrected conclusions may be useful. Unlike UGPA, GREs remain useful in the full cohort, not just the domestic subset. Within the group of all applicants, GREs provide more incremental predictive power for Ph.D. attainment than does UGPA. Within the cohort of U.S. applicants’ GREs provide nearly the same incremental predictive power as UGPA. *Within the cohort of domestic students who are interested in physics graduate school and who have the same UGPA, dropping both GREs would lose an odds ratio of at least a factor of three in estimating how the probability for graduation varies between high-scorers and low-scorers.* Dropping only GRE-P would lose substantially less predictive power than dropping both GREs.

Discussion

We have seen several major errors concerning causal inference in recent PRPER papers, including in explicit methods sections. The errors are directly consequential for estimating the effects of policy choices, including a major policy choice under active consideration—the use of GREs.

With respect to the specific GRE issue, my conclusion concerning physics graduation rates should not be generalized to other outcomes and especially not to other fields, for which results will vary. It has been known for some time that GREs can be highly predictive of success in learning quantitative social science methods (Huitema and Stein 1993). On the other hand, one biomedical program found evidence that incremental predictive power of GREs for a variety of outcomes was very small (Sealy et al. 2019).

The question of how much physics departments should use the predictive effects of GREs involves non-technical considerations including a variety of not always parallel social goals. Nevertheless, finding the best selection method is trivial

in one limiting case. If the papers discussed here arguing against the use of GREs represent what can result from devaluing the sort of basic mathematical and logical reasoning that some GREs are meant to test, they tend to be self-refuting. If we do not try to maintain minimal standards of competence and transparency or even basic logic in our treatment of data, then the optimum group of students whom we should be educating is the empty set.

On the more general methodological questions, in all the papers discussed here and in my previous articles (Weissman 2021; 2020b; 2020a), there was no clear specification of which actual interventions and consequences would be considered. Instead, implicit rather than explicit assumptions were used to justify applying more or less causal language to collections of correlations. Policy recommendations were then intuited or implied, leaning on causal interpretations of some correlations conveniently chosen to support the desired conclusions. Although a similar description might apply to much other social science research, especially from before about 1980, many fields have now moved well beyond that (Pearl, Glymour, and Jewell 2016; Hernán and Robins 2020; Glymour, Zhang, and Spirtes 2019; Glass et al. 2013; Varian 2016; Foster 2010; Gangl 2010; Keele 2015; Murnane and Willett 2011; Rohrer 2018; Imbens and Rubin 2015). Physics education research should join them.

Perhaps the most important lesson for those who wish to do valid causal reasoning is that one should start with an explicit question about reasonably well-defined potential consequences of reasonably well-defined choices of potential actions (Robins and Weissman 2017). Even when a randomized controlled trial is infeasible, just trying to imagine one can help to clarify what question is being asked (Hernán 2021). Although explicit assumptions are still needed to pare down the possibilities to something tractable, that starting clarity can go a long way toward guiding the type of causal graphs that are needed. Graphs then help to see what combinations of correlations provide the best estimate of the effects of possible actions.

Some procedural changes might help reduce the current problems. One might be to preregister protocols for studies (Chambers 2019). That allows some refereeing of methods before data are gathered and before particular methods are closely tied to particular conclusions. That helps limit cherry-picking methods and data, reducing p-hacking and reverse p-hacking (Chuard et al. 2019). Even with pre-registration, however, one still needs a community of reviewers and readers who can distinguish between valid and invalid methods and are motivated to do so. Establishing that may require help from outside physics education research.

It is crucial for scientists to be able to correct errors. It is fortunate that journals that welcome critiques exist as a backup venue for times when a subfield becomes resistant to that process. Although it is painful for me to resort to a journal

whose political and economic philosophy I find to be neither realistic nor humane:
Flectere si nequeo superos, Acheronta movebo.

References

- Akaike, H.** 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19: 716–723. [Link](#)
- Amrhein, Valentin, Sander Greenland, and Blake McShane.** 2019. Retire Statistical Significance. *Nature* 567: 305–307.
- Bollen, Kenneth A., and Judea Pearl.** 2013. Eight Myths About Causality and Structural Equation Models. In *Handbook of Causal Analysis for Social Research*, ed. S. L. Morgan, 301–358. Dordrecht, Netherlands: Springer.
- Chambers, Chris.** 2019. What’s Next for Registered Reports?. *Nature* 573: 187–189. [Link](#)
- Chawla, Dalmeet Singh.** 2020. Graduate Departments Move Away from GREs. *Physics Today*, July 9. [Link](#)
- Chuard, Pierre J. C., Milan Vrtilek, Megan L. Head, and Michael D. Jennions.** 2019. Evidence That Nonsignificant Results Are Sometimes Preferred: Reverse P-Hacking or Selective Reporting? *PLoS Biology* 7(1): e3000127. [Link](#)
- Foster, E. M.** 2010. Causal Inference and Developmental Psychology. *Developmental Psychology* 46(6): 1454–1480. [Link](#)
- Gangl, Markus.** 2010. Causal Inference in Sociological Research. *Annual Review of Sociology* 36(1): 21–47. [Link](#)
- Gelman, Andrew.** 2021. “While Critique Is Certainly an Important Part of Improving Scientific Fields, We Do Not Want to Set a Precedent or Encourage Submission of Articles That Only Critique the Methods Used by Others.” *Statistical Modeling, Causal Inference, and Social Science*, October 30. [Link](#)
- Glass, Thomas A., Steven N. Goodman, Miguel A. Hernán, and Jonathan M. Samet.** 2013. Causal Inference in Public Health. *Annual Review of Public Health* 34(1): 61–75. [Link](#)
- Glymour, Clark, Kun Zhang, and Peter Spirtes.** 2019. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics* 10: 524. [Link](#)
- Greenland, Sander.** 2003. Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias. *Epidemiology* 14(3): 300–306.
- Greenland, Sander, Judea Pearl, and James M. Robins.** 1999. Causal Diagrams for Epidemiologic Research. *Epidemiology* 10(1): 37–48.
- Hernán, Miguel A.** 2021. Methods of Public Health Research—Strengthening Causal Inference from Observational Data. *New England Journal of Medicine* 385: 1345–1348. [Link](#)
- Hernán, Miguel A., and James M. Robins.** 2020. *Causal Inference: What If*. Boca Raton, Fla.: Chapman & Hall/CRC.
- Huitema, Bradley E., and Cheri R. Stein.** 1993. Validity of the GRE Without Restriction of Range. *Psychological Reports* 72: 123–127.
- Imbens, Guido W., and Donald B Rubin.** 2015. *Causal Inference in Statistics, Social, and*

- Biomedical Sciences*. Cambridge, UK: Cambridge University Press.
- Keele, Luke**. 2015. The Statistics of Causal Inference: A View from Political Methodology. *Political Analysis* 23: 313–335. [Link](#)
- Miller, Casey W., and Keivan Stassun**. 2014. A Test That Fails. *Nature* 510: 303–304. [Link](#)
- Miller, Casey W., Benjamin M. Zwickl, Julie R. Posselt, Rachel T. Silvestrini, and Theodore Hodapp**. 2019. Typical Physics Ph.D. Admissions Criteria Limit Access to Underrepresented Groups but Fail to Predict Doctoral Completion. *Science Advances* 5(1): eaat7550. [Link](#)
- Miller, Casey W., Benjamin M. Zwickl, Julie R. Posselt, Rachel T. Silvestrini, and Theodore Hodapp**. 2020. Response to Comment on “Typical Physics Ph.D. Admissions Criteria Limit Access to Underrepresented Groups but Fail to Predict Doctoral Completion.” *Science Advances* 6(23): eaba4647. [Link](#)
- Murnane, Richard J., and John B. Willett**. 2011. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford, UK: Oxford University Press.
- Nissen, Jayson, Robin Donatello, and Ben Van Dusen**. 2019. Missing Data and Bias in Physics Education Research: A Case for Using Multiple Imputation. *Physical Review Physics Education Research* 15(2): 020106. [Link](#)
- Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell**. 2016. *Causal Inference in Statistics—A Primer*. Chichester, UK: Wiley.
- Potvin, Geoff, Deepa Chari, and Theodore Hodapp**. 2017. Investigating Approaches to Diversity in a National Survey of Physics Doctoral Degree Programs: The Graduate Admissions Landscape. *Physical Review Physics Education Research* 13: 020142. [Link](#)
- Richardson, Thomas**. 2003. Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics: Theory and Applications* 30(1): 145–157. [Link](#)
- Robins, James M., and Michael B Weissman**. 2017. Commentary: Counterfactual Causation and Streetlamps: What Is to Be Done? *International Journal of Epidemiology* 45(6): 1830–1835. [Link](#)
- 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](#)
- Sealy, Linda, Christina Saunders, Jeffrey Blume, and Roger Chalkley**. 2019. The GRE Over the Entire Range of Scores Lacks Predictive Ability for PhD Outcomes in the Biomedical Sciences. *PLoS One* 14(3): e0201634. [Link](#)
- RECOVERY Collaborative Group**. 2021. Dexamethasone in Hospitalized Patients with Covid-19. *New England Journal of Medicine* 384: 693–704. [Link](#)
- VanderWeele, Tyler J**. 2014. Commentary: Resolutions of the Birthweight Paradox: Competing Explanations and Analytical Insights. *International Journal of Epidemiology* 43(5): 1368–1373. [Link](#)
- Varian, Hal R**. 2016. Causal Inference in Economics and Marketing. *Proceedings of the National Academy of Sciences* 113(27): 7310–7315. [Link](#)
- Verostek, Mike, Casey W. Miller, and Benjamin Zwickl (VMZ)**. 2021. Analyzing Admissions Metrics as Predictors of Graduate GPA and Whether Graduate GPA Mediates Ph.D. Completion. *Physical Review Physics Education Research* 17(2): 020115.

[Link](#)

- Walsh, Cole, Martin M. Stein, Ryan Tapping, Emily M. Smith, and N. G. Holmes** (WSTSH). 2021. Exploring the Effects of Omitted Variable Bias in Physics Education Research. *Physical Review Physics Education Research* 17: 010119. [Link](#)
- Weissman, Michael B.** 2021. Policy Recommendations From Causal Inference In Physics Education Research. *Physical Review Physics Education Research* 17(2): 020118. [Link](#)
- Weissman, Michael B.** 2020a. Do GRE Scores Help Predict Getting a Physics Ph.D.? A Comment on a Paper by Miller et al. Working paper, June 8. [Link](#)
- Weissman, Michael B.** 2020b. Do GRE scores Help Predict Getting a Physics Ph.D.? A Comment on a Paper by Miller et al. *Science Advances* 6(23): eaax3787. [Link](#)
- Yao, Yang-Yang, Lu-Lu Lin, Hui-Yun Gu, Jun-Yi Wu, Yu-Ming Niu, and Chao Zhang.** 2019. Are Corticosteroids Beneficial for Sepsis and Septic Shock? Based on Pooling Analysis of 16 Studies. *Frontiers in Pharmacology* 10: 714. [Link](#)
- Young, Nicholas T.** 2020. Eliminating the GRE. *Physics Today* 73(12): 11. [Link](#)
- Young, Nicholas T., and Marcos D. Caballero** (YC). 2021. Physics Graduate Record Exam Does Not Help Applicants “Stand Out.” *Physical Review Physics Education Research* 17(1): 010144. [Link](#)

About the Author



Michael Weissman is a Professor Emeritus of Physics at the University of Illinois at Urbana-Champaign and a Fellow of the American Physical Society. He has worked mostly on condensed matter experiments, especially on low-frequency noise in disordered materials. He has worked in botany and chemistry departments. His undergrad degree is in math. He stumbled into looking at physics education research papers by accident. He did time for Vietnam draft resistance and was an originator of the scientists’ boycott of the Strategic Defense Initiative. He thinks that libertarian philosophy is factually detached from the realities of our species, implicitly cruel, and will lead to environmental catastrophe. He supports unions but not the designated-hitter rule. His email is mbw@illinois.edu.

[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at JournalTalk:
<https://journaltalk.net/articles/6042>



Compared to What? Does Benford’s Law Really Detect Corporate Fraud?

Stephen Walker¹

[LINK TO ABSTRACT](#)

The Netflix series *Digits* devoted an episode in 2020 to Benford’s law. Benford’s law possesses fascinating and mysterious properties that have captured the imaginations of the viewing public, but it is not magic and its uses are limited. Benford’s law measures the relative frequency of the first digit of a set of numbers and suggests a non-uniform distribution where the digit 1 “will appear as the first digit 30.1% of the time, 2 will appear 17.6% of the time, and so forth” (Amiram et al. 2015).

In their paper “Financial Statement Errors: Evidence from the Distributional Properties of Financial Statement Numbers,” authors Dan Amiram, Zahn Bozanic, and Ethan Rouen (ABR) develop a unique measure based on Benford’s law that applies the distribution of first digits obtained from publicly available financial statements, which include the income statement, balance sheet, and cash flow statement. They call their measure the Financial Statement Divergence Score (FSD Score), which they make available for free for non-commercial use.² In the abstract to their paper, the authors say that their measure “predicts material misstatements as identified by SEC Accounting and Auditing Enforcement Releases and can be used as a leading indicator to identify misstatements” (ABR 2015). The paper won the 2017 Deloitte Foundation Wildman Medal Award and was featured in the financial press including the *Wall Street Journal* (McGinty 2014). The paper had 131 Google Scholar citations as of 27 February 2022. I seek to

1. I thank anonymous referees for their helpful commentary and suggestions for improvement.

2. The data can be downloaded directly ([link](#)) or through a form on Professor Bozanic’s website ([link](#)).

analyze both the degree to which their measure really predicts material misstatements and the extent to which it can serve as a leading indicator for them. While the ABR claim is essentially a yes/no question as to *whether* their measure works, I instead ask the question as to *how well* it works.³ In other words, I ask a question inspired by one of the greatest social theorists and economists of the last half century, Thomas Sowell: Compared to what? Sowell is the Rose and Milton Friedman Senior Fellow on Public Policy at the Hoover Institution, and he frequently asks “Compared to what?” in public discourse.

Analysis of predicting material misstatements

To start this analysis, I construct a dataset based on the following. First, I obtained the SEC Accounting and Auditing Enforcement Release (AAER) database from Professor Patricia Dechow of the USC Marshall School of Business (see Dechow et al. 2011). Next, I obtained company financial statement variables from the Compustat database and constructed the variables described in the ABR paper;⁴ for purposes of brevity, I do not describe the other variables here, but these definitions can be found in the original paper (ABR 2015). I then downloaded from Professor Bozanic’s website the “FSD Score,”⁵ which is based on Benford’s law. I filled missing values at the means of each variable by industry and year in order to keep observation drops to a minimum. The combined file contains 103,289 unique firm-year observations spanning 1990–2011, with 353 unique AAERs, or 0.34 percent of the sample.

Replicating results from the original paper and considering an alternative timeframe

I start by reconstructing the results from Column 1 of Table 10 in ABR (2015) and successfully replicated their results shown here in Column 3 of Ta-

3. I thank an anonymous referee for suggesting this clarification.

4. Compustat variables that were recoded zero if missing include: *at act cbe lct dlc t:cp inao lt lct dlrt inst pstk rect invt ppent sale ib sstk*.

5. The authors restrict their analysis to companies that had at least 100 account line items though they say that an unrestricted sample did not alter their inferences (ABR 2015, 1551). For consistency, I followed this approach.

TABLE 1. Replicating ABR (2015) and applying the same specification to the period immediately preceding their study

	Years 2001–2011			Years 1990–2000		
	(1)	(2)	(3)	(4)	(5)	(6)
FSD_Score	-33.219*** (-3.15)	-34.683*** (-3.28)	-34.021*** (-3.18)	7.835 (0.97)	4.221 (0.52)	5.138 (0.62)
ABS_JONES_RESID			-1.564* (-1.93)			0.541 (1.22)
STD_DD_RESID			0.754 (1.36)			0.292 (0.61)
MANIPULATOR			0.217 (0.86)			0.048 (0.25)
F_SCORE		0.437*** (5.47)	0.225* (1.75)		0.559*** (10.49)	0.158 (1.60)
ABS_WCACC			1.405 (1.46)			0.977 (1.55)
ABS_RSST			0.414 (1.10)			0.335 (1.17)
CH_CS			0.093 (0.80)			0.166* (1.91)
CH_ROA			0.124 (0.27)			-0.432 (-1.14)
SOFT_ASSETS			1.108*** (2.58)			1.695*** (4.20)
ISSUE			0.340 (0.86)			1.420** (2.42)
MTB			0.034* (1.88)			0.059*** (4.98)
AT			0.000 (0.27)			0.000*** (2.96)
Constant	-5.050*** (-16.69)	-5.424*** (-17.44)	-6.520*** (-11.51)	-5.643*** (-21.93)	-6.161*** (-23.41)	-8.708*** (-12.72)
Observations	55,875	55,875	55,840	47,414	47,414	44,208

Notes: Coefficients are based on a logistic regression and reported in log odds. Following ABR, AAER is an indicator variable equal to 1 for the first year of the misstatement. For other variable definitions, see ABR 2015. $AAER_{it} = a + \beta_1 FSD_SCORE_{it} + \beta_2 ABS_JONES_RESID_{it} + \beta_3 STD_DD_RESID_{it} + \beta_4 MANIPULATOR_{it} + \beta_5 F_SCORE_{it} + \beta_6 ABS_WCACC_{it} + \beta_7 ABS_RSST_{it} + \beta_8 CH_CS_{it} + \beta_9 CH_ROA_{it} + \beta_{10} SOFT_ASSETS_{it} + \beta_{11} ISSUE_{it} + \beta_{12} MTB_{it} + \beta_{13} AT_{it}$. Z-statistics in parentheses. ***p<0.01, **p<0.05, *p<0.1.

ble 1. In this specification, there were a total of 55,840 observations.⁶ I also report additional specifications in Columns 1 and 2 that use the measure as a standalone

6. For robustness, alternative specifications included (1) dropping the missing constructed variables based on the zero recodings for the Compustat variables and (2) dropping the missing constructed variables that did not include these recodings. Including the zero recodings returns 39,713 observations and not including them returns 34,663 observations, a quantity which is closer to ABR's reported 27,805 observations. For these alternative specifications, inferences are similar.

and as an incremental variable to the F-Score, which is the prior detection model from the accounting literature. The FSD Score's coefficients are similar across Models 1, 2 and 3, and are directionally consistent with the -40.691 reported in the ABR paper. Evaluating the marginal effect in Model 3 from the mean to the 25th percentile yields a 0.057 percent change in probability, which was similar to the 0.046 percent estimate from their original regression (I use this range because it is the range provided by ABR's descriptive statistics table). I also considered an alternative period in similar length immediately preceding their study period. Evaluating the same regressions for the years 1990–2000, the coefficient on the FSD Score turns positive and insignificant, which raises questions as to the validity of their FSD Score measure to AAER events outside their study period.

Compared to what? How the FSD Score measures up to the F-Score and a naive sales measure for predicting material misstatements

While the logistic regression shows that there was significance for the FSD Score for ABR's preferred timeframe, interpreting the predictive performance is difficult. For this section, I compare predictive performance in terms of standard metrics from the prediction literature including positive predictive value (precision), AUROC, and NDCG, which were not evaluated in the original paper; for an explanation of these measures, please see the appendix to this paper.

I compare three measures to the FSD Score. The first measure is the *F-Score*, which is constructed based on the coefficients from a material misstatement detection model created by Dechow et al. (2011). The second measure takes Column 2 from Table 1, which simplifies the logistic regression from the ABR (2015) paper taking the only two variables that were reported to be significant at the 0.01 level, the F-Score and the FSD Score; this measure I call *Combined*, and it measures the additive predictive power of the FSD Score relative to the F-Score. I checked the robustness of this simplification to the full ABR regression (from Column 3 in Table 1), and inferences are similar. The third and last measure calculates a four-year geometric sales growth rate, which represents a naive measure. The sales growth rate, a relatively simple measure, was suggested to me in conversation by Professor Howard Schilit, an expert in forensic analysis and author of *Financial Shenanigans*, a popular book among industry professionals.

In their paper, ABR (2015, 1557) acknowledge that their results show that divergence from Benford's law is associated with a *decrease* in likelihood of material misstatement, which was opposite of what they expected, represented by the

negative coefficient from Table 1 for the main sample period (but not by the alternative timeframe). They address the reasons why they believe they found this result. While their argument is up for debate, I do not address it here because my analysis is a predictive exercise, where the ‘why’ is less important than the ‘what.’ The ‘what’ here is the predictive output. Therefore, when comparing Tables 2 and 3, examine Decile 1 for the FSD Score while comparing it to Decile 10 of the others. For Table 2, the dependent variable is an indicator variable coded “1” for the first year of the misstatement (but not for the duration for multi-year misstatements), following ABR’s approach.

TABLE 2. Misstatement year =1 only in the first year

Decile	Years 2001–2011				Years 1990–2000			
	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales
1	0.4%	0.1%	0.1%	0.1%	0.4%	0.1%	0.1%	0.2%
2	0.3%	0.1%	0.1%	0.2%	0.5%	0.1%	0.2%	0.3%
3	0.2%	0.1%	0.2%	0.1%	0.3%	0.2%	0.2%	0.2%
4	0.4%	0.1%	0.2%	0.3%	0.3%	0.3%	0.3%	0.4%
5	0.3%	0.2%	0.2%	0.2%	0.5%	0.4%	0.4%	0.3%
6	0.2%	0.3%	0.2%	0.4%	0.4%	0.2%	0.3%	0.4%
7	0.1%	0.3%	0.2%	0.3%	0.6%	0.6%	0.4%	0.7%
8	0.2%	0.3%	0.4%	0.2%	0.5%	0.8%	0.8%	0.5%
9	0.2%	0.4%	0.4%	0.3%	0.4%	0.6%	0.5%	0.6%
10	0.1%	0.5%	0.6%	0.4%	0.5%	1.3%	1.3%	1.1%
Other Metrics:								
AUROC	0.57	0.64	0.65	0.60	0.52	0.71	0.69	0.66
Precision at 1%	0.5%	0.9%	1.1%	0.7%	0.6%	2.4%	1.9%	0.9%
NDCG	0.381	0.390	0.397	0.387	0.418	0.460	0.460	0.448
NDCG@10%	0.073	0.104	0.110	0.088	0.053	0.162	0.167	0.131
NDCG@1%	0.015	0.027	0.032	0.022	0.009	0.038	0.036	0.012
<i>Notes:</i> The proportion of material misstatements for the first year in the sample from 2001–2011 was 0.2%. From 1990–2000 it was 0.4%. Deciles of probabilities for each of the four models were created by year. The highest risk decile for the FSD Score is Decile 1. For the other measures, it is Decile 10.								

The percentage values reported by decile represent the likelihood of finding a material misstatement and can be compared to the unconditional likelihood, which is the prevalence of material misstatements in the sample. These values in the sample were 0.2 percent for 2001–2011 and 0.4 percent from 1990–2000, which illustrate just how rarely these events occur. Overall, improvements do not move the needle very much. On a relative basis, the FSD Score, as a standalone, performs the worst among the models. Oddly, the standalone scenario was not evaluated in

the ABR paper, as only a fully loaded model with 13 variables was estimated. The next measure, the F-Score, beats the FSD Score in both periods. The FSD Score marginally improves the F-Score in the Combined variable for 2001–2010, but not in the previous period. Interestingly, the naive sales growth screen performs similarly to the complex methodologies at the highest risk decile. While the other measures seem to work in the previous period, the standalone FSD Score appears to fail entirely. In addition to examining these deciles, I produce Precision at 1%, which reports the fraction of true positives in the top 1 percent of each measure by year, similar to what was done in the decile analysis. In addition, I provide other pooled metrics including AUROC, NDCG, and NDCG@k, which are explained in more detail in the Appendix. Regardless of measure, inferences are similar.

TABLE 3. All misstatement years = 1

Decile	Years 2001–2011				Years 1990–2000			
	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales
1	1.1%	0.4%	0.3%	0.5%	0.7%	0.2%	0.2%	0.2%
2	1.2%	0.7%	0.4%	0.6%	1.0%	0.2%	0.4%	0.6%
3	0.9%	0.6%	0.7%	0.8%	0.8%	0.5%	0.5%	0.5%
4	1.0%	0.5%	0.9%	0.7%	0.6%	0.8%	0.7%	0.5%
5	1.0%	0.7%	0.7%	0.7%	0.9%	0.6%	0.7%	0.6%
6	0.9%	0.8%	0.9%	0.8%	0.7%	0.6%	0.6%	0.7%
7	0.8%	1.3%	1.0%	0.9%	0.9%	0.8%	0.7%	1.1%
8	0.6%	1.2%	1.4%	1.0%	1.0%	1.3%	1.3%	0.8%
9	0.8%	1.5%	1.4%	1.3%	0.7%	1.1%	1.0%	1.1%
10	0.6%	1.0%	1.2%	1.6%	0.7%	2.0%	2.0%	2.1%
Other Metrics:								
AUROC	0.55	0.59	0.59	0.60	0.50	0.67	0.66	0.67
Precision at 1%	1.1%	1.1%	1.4%	1.3%	0.8%	2.8%	2.6%	1.7%
NDCG	0.508	0.509	0.513	0.520	0.479	0.520	0.521	0.518
NDCG@10%	0.080	0.074	0.078	0.129	0.053	0.1620	0.171	0.174
NDCG@1%	0.011	0.014	0.016	0.010	0.008	0.037	0.037	0.019
<i>Notes:</i> The proportion of material misstatements for all years in the sample from 2001–2011 was 0.9%. From 1990–2000 it was 0.8%. Deciles of probabilities for each of the four models were created by year. The highest risk decile for the FSD Score is Decile 1. For the other measures, it is Decile 10.								

There is nothing wrong with ABR’s approach to only consider the first year of the material misstatement, but the frequency of the dependent variable is significantly reduced as it only captures the first year, thus making it harder to detect. It is also unclear why investors would not also be interested in detecting an ongoing multi-year misstatement. Therefore, I also analyze material misstatements

including the additional years for the multi-year cases, shown in Table 3. This time, the dependent variable is coded '1' for every year there is a material misstatement. The story remains the same, and sales growth shows a slight edge over the other models in terms of top-decile performance. As to the other measures, AUROCs are similar across F-Score, Combined, and Sales models. For both dependent variables, Tables 2 and 3 show that the FSD Score as a standalone performs the worst, and it is not clear if it really adds much value, if at all, as an additional variable to the existing F-Score model.

As a leading indicator?

So far, the results I have provided reflect the contemporaneous relationship between financials and material misstatements; the relationship attempts to predict fraud given a known set of public financial statements. ABR (2015) say that their measure could serve as a leading indicator in the detection of material misstatements. As evidence of such, they show significant coefficients on one-year and two-year lagged FSD Scores; such significance implies that a material misstatement could be detected prior to the year of occurrence based on these leading values. Curiously, they did not lag the control variables, so these published models could never serve as leading indicators as currently specified. Despite this misspecification, to analyze whether the FSD Score could be used as a leading indicator, I repeat the previous analysis for each measure according to their lagged values. Table 4 shows evidence that the FSD Score does not work as a standalone and it does not improve the F-Score in the combined metric. The sales growth screen still performs well in comparison for the top decile. Since positive coefficients were reported on the lagged variables by ABR, the comparable decile for the FSD Score is now Decile 10. As for the other measures, sales outperforms in terms of NDCG, Precision at 1%, AUROC, NDCG@10%, and NDCG@1% for their main sample period of 2001–2011.

TABLE 4. One-year leading indicator (misstatement year = 1 only for the first year)

Decile	Years 2001–2011				Years 1991–2000			
	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales
1	0.2%	0.2%	0.1%	0.2%	0.5%	0.1%	0.1%	0.1%
2	0.2%	0.1%	0.2%	0.0%	0.2%	0.1%	0.2%	0.2%
3	0.1%	0.1%	0.1%	0.1%	0.5%	0.2%	0.1%	0.2%
4	0.3%	0.2%	0.2%	0.3%	0.5%	0.2%	0.3%	0.5%
5	0.2%	0.2%	0.3%	0.3%	0.5%	0.4%	0.3%	0.3%
6	0.3%	0.2%	0.3%	0.2%	0.4%	0.4%	0.5%	0.4%
7	0.1%	0.4%	0.2%	0.2%	0.4%	0.5%	0.4%	0.6%
8	0.3%	0.2%	0.3%	0.3%	0.4%	0.7%	0.5%	0.5%
9	0.2%	0.4%	0.4%	0.2%	0.3%	0.6%	0.8%	0.5%
10	0.2%	0.3%	0.3%	0.4%	0.4%	1.1%	1.0%	0.9%
Other Metrics:								
AUROC	0.53	0.61	0.59	0.61	0.50	0.70	0.69	0.65
Precision at 1%	0.5%	0.5%	0.4%	0.9%	0.3%	2.6%	2.0%	1.0%
NDCG	0.355	0.368	0.364	0.371	0.396	0.443	0.441	0.423
NDCG@10%	0.051	0.071	0.072	0.085	0.059	0.144	0.145	0.122
NDCG@1%	0.006	0.013	0.010	0.018	0.010	0.044	0.043	0.015
<i>Notes:</i> The proportion of material misstatements for all years in the sample from 2001–2011 was 0.2%. From 1991–2000 it was 0.4%. Deciles of probabilities for each of the four models were created by year. Decile 10 shows the highest risk measure for all four measures. Since one-year lagged values are required, the second period starts in 1991 instead of 1990.								

Table 5 reports the two-year lagged values for the key measures and inferences are similar to the one-year lagged analysis. For the main sample period of 2001–2011, the F-Score outperforms in terms of top decile, top 1 percent, AUROC, NDCG, and NDCG@10%.

TABLE 5. Two-year leading indicator (misstatement year = 1 only for the first year)

Decile	Years 2001–2011				Years 1992–2000			
	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales	(1) FSD Score	(2) F-Score	(3) Combined	(4) Sales
1	0.1%	0.1%	0.1%	0.2%	0.3%	0.1%	0.0%	0.3%
2	0.2%	0.1%	0.2%	0.1%	0.7%	0.0%	0.2%	0.3%
3	0.1%	0.2%	0.2%	0.2%	0.4%	0.2%	0.3%	0.2%
4	0.3%	0.2%	0.1%	0.2%	0.3%	0.3%	0.2%	0.3%
5	0.3%	0.1%	0.1%	0.3%	0.4%	0.4%	0.3%	0.6%
6	0.3%	0.3%	0.2%	0.2%	0.4%	0.6%	0.6%	0.3%
7	0.2%	0.2%	0.4%	0.3%	0.3%	0.5%	0.6%	0.6%
8	0.2%	0.3%	0.2%	0.1%	0.4%	0.5%	0.4%	0.5%
9	0.2%	0.3%	0.3%	0.2%	0.4%	0.7%	0.8%	0.5%
10	0.4%	0.5%	0.4%	0.4%	0.4%	0.8%	0.7%	0.6%
Other Metrics:								
AUROC	0.56	0.64	0.61	0.58	0.51	0.67	0.66	0.61
Precision at 1%	0.2%	0.3%	0.2%	0.0%	0.6%	0.9%	1.4%	0.0%
NDCG	0.350	0.363	0.361	0.362	0.384	0.411	0.408	0.396
NDCG@10%	0.065	0.103	0.092	0.097	0.062	0.115	0.099	0.092
NDCG@1%	0.000	0.005	0.006	0.013	0.007	0.024	0.012	0.000
<i>Notes:</i> The proportion of material misstatements for all years in the sample from 2001–2011 was 0.2%. From 1992–2000 it was 0.4%. Deciles of probabilities for each of the four models were created by year. Since two-year lagged values are required, the second period starts in 1992 instead of 1990.								

Evaluating the predictive margins from the original paper

While ABR (2015) did not report predictive margins—that is, to what extent the probability changes from extreme changes in the underlying metric—their readers could apply a back-of-the-envelope method using the output from their reported logistic regression and their summary statistics table and come to the conclusion that their model does not really move the needle in absolute terms. Their logistic regression from their first model is reproduced here in Table 6, Column 1. Their output was reported in log odds, and, to the layperson, it might seem that all others are beaten by the FSD Score, which has the greatest absolute coefficient value. The real effect on probability is tiny, which can be evaluated by calculating the predictive margins for the coefficient of interest. The table walks through the calculation evaluating the change in probability for a move in the

FSD Score from its mean value to the 25th percentile based on ABR’s descriptive statistics table. A change from the mean to the 25th percentile is the best the reader could do since ABR’s descriptive statistics table does not report values beyond this level. The result is that the change in FSD Score increases the probability of material misstatement detection by 0.046 percent according to their sample—indeed an extremely small change.

TABLE 6. Evaluating the predictive margins from the original logistic regression

	Model 1 coefficients from ABR	Means of variables	FSD Score at 25th percentile; means otherwise	Log odds at means	Log odds varying FSD Score only
	(1)	(2)	(3)	(1) × (2)	(1) × (3)
FSD Score	-40.691	0.030	0.023	-1.204	-0.952
ABS_JONES_RESID	-1.078	0.184	0.184	-0.198	-0.198
STD_DD_RESID	0.011	0.123	0.123	0.001	0.001
MANIPULATOR	0.122	0.143	0.143	0.017	0.017
F_SCORE	1.980	0.401	0.401	0.793	0.793
ABS_WCACC	-1.233	0.054	0.054	-0.067	-0.067
ABS_RSST	0.401	0.138	0.138	0.055	0.055
CH_CS	0.004	0.146	0.146	0.001	0.001
CH_ROA	1.339	-0.002	-0.002	-0.003	-0.003
SOFT_ASSETS	-0.121	0.545	0.545	-0.066	-0.066
ISSUE	-0.341	0.915	0.915	-0.312	-0.312
MTB	0.166	1.360	1.360	0.226	0.226
AT	0.000	3228.380	3228.380	0.000	0.000
Constant	-5.686			-5.686	-5.686
Sum of Log Odds				-6.442	-6.189
Odds Ratio (OR): e^(log odds)				0.002	0.002
Probability (OR / 1+OR)				0.159%	0.205%
Change in Probability				0.046%	
<i>Notes:</i> Column 1 reports the original logistic regression coefficient values from ABR (2015). Columns 2 and 3 are values sourced from ABR’s descriptive statistics. Coefficients are reported in log odds and probabilities are evaluated accordingly. For the logistic regression, AAER was an indicator variable equal to 1 for the first year of the misstatement. For other variable definitions, see ABR 2015. $AAER_{it} = a + \beta_1 FSD_SCORE_{it} + \beta_2 ABS_JONES_RESID_{it} + \beta_3 STD_DD_RESID_{it} + \beta_4 MANIPULATOR_{it} + \beta_5 F_SCORE_{it} + \beta_6 ABS_WCACC_{it} + \beta_7 ABS_RSST_{it} + \beta_8 CH_CS_{it} + \beta_9 CH_ROA_{it} + \beta_{10} SOFT_ASSETS_{it} + \beta_{11} ISSUE_{it} + \beta_{12} MTB_{it} + \beta_{13} AT_{it}$.					

Conclusion

Benford's law has captured the imaginations of fraud researchers worldwide. The results here should curb some of the enthusiasm, particularly for the detection of material misstatements based on publicly available financial statements. Our best models in accounting research, applying financial statement variables do very little to move the needle in terms of the predictive margins based on what is observable in the financial statements. Anecdotally, I've heard that industry professionals have found little use with the models published in accounting research and the evidence here explains why this is the case. ABR (2015, 1541), as motivation for their research, noted that previous methods for analyzing firm-level financial statements "have deficiencies that limit their usefulness." Given the evidence here, their FSD Score joins that club for the purpose of detection of material misstatements. The present critique started with: Compared to what? The FSD Score based on Benford's law does not meaningfully detect material accounting misstatements nor does it perform as a leading indicator as a standalone measure. As an additive measure to the F-Score, the evidence here shows a range of outcomes from very little benefit to potential worsening of predictive power. The naive sales growth screen performs surprisingly well compared to advanced statistical methods. Finally, the FSD Score does not appear robust to an alternative period of time immediately preceding the sample period chosen by ABR.

Appendix

The purpose of this appendix is to explain the available choices of metrics used in classification modeling and to provide the reasoning for the selection of metrics presented in the present paper. To start this discussion, Table A1 shows an example of a classification matrix typically used when analyzing model performance. For example, the logistic regression can be used to estimate a fitted probability for each observation. However, the researcher must choose at which point along this probability continuum to classify cases in the positive or negative. This cutoff choice can either be made at a specific value or standardized through examining top deciles or percentiles. The cutoff choice can be optimized by weighting the relative cost of false positives and false negatives. For an e-commerce company, declining a potentially fraudulent transaction must be weighed against the possibility of lost margin from a sale. For the purposes of corporate fraud detection, there are real-world constraints. Investigative hours are limited and consumed by existing investigations. Detection models, if they work, would act as

a screen or as a watch list to launch new investigations at the margin. Therefore, when analyzing performance, it is accuracy at the top that matters the most, which is why I presented the decile performance for each model considered. This measure is also known as Precision@k, which is a common metric used to analyze true positives out of the predicted positives at some threshold k . In this case, the top decile would be precision at 10 percent. I also added precision at 1 percent to summarize results at the more extreme level of the top percentile. The downside with going more extreme is that most of the corporate fraud cases will be missed entirely as so few cases have fitted probability values at that level. However, for corporate fraud detection, reliable investigative tips typically come from other channels. Therefore, these models may be useful if precision is high enough at the margin for the highest risk cases. In addition to precision, other measures can be calculated from the classification matrix. Table A1 shows a classification matrix which includes the four quadrants of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) in terms of the dependent variable in this paper (AAERs).

TABLE A1. Example of a classification matrix

Prediction based on threshold (e.g., top 1 percent)			
	YES: AAER	NO: AAER	Total
AAER	True Positive (TP)	False Negative (FN)	Total AAERs
Non-AAER	False Positive (FP)	True Negative (TN)	Total Non-AAERs
Total	Total Predicted Positive	Total Predicted Negative	Total Sample (N)

From this classification matrix, the following values can be calculated:

- Prevalence (Total AAERs / N): These are the AAERs in the sample that is known before the classification exercise.
- Classification accuracy (TP+TN)/N: These are the correct classifications of true positives (both AAER and predicted positive) and true negatives (both non-AAER and predicted negative) out of the total observations in the sample.
- Sensitivity (TP/Total AAERs): Also known as the true positive rate, these are the true positives out of the total AAERs in the sample.
- Specificity (TN/Total Non-AAERs): Also known as the true negative rate, these are the true negatives out of the total non-AAERs in the sample.
- Precision (TP/Total Predicted Positive): These are the true positives out of those that were predicted positive.

Classification accuracy, while highly intuitive, is a poor measure to analyze model performance for fraud detection because fraud is a rare event. For rare events, classification accuracy is essentially defined by the true negative rate since non-fraud events dominate. To see why, we can write classification accuracy as the sum of the proportions that were true positive and true negative:

$$\text{Classification Accuracy} = \frac{TP}{N} + \frac{TN}{N}$$

These terms can be re-written in terms of prevalence, sensitivity, and specificity:

$$\text{Classification Accuracy} = [\text{Prevalence} * \text{Sensitivity}] + [(1 - \text{Prevalence}) * \text{Specificity}]$$

When prevalence is extremely low, for example 0.5 percent, the other term would be weighted accordingly, in this example 99.5 percent. Furthermore, the measure is hard to interpret for rare events because it can be arbitrarily maximized through an uninformative decision rule. For rare events, the strategy would be to classify all cases in the negative. In this case, sensitivity would be zero as no AAERs would be captured, but specificity would be 100 percent as all negatives would be classified correctly. Therefore, for an event occurring 0.5 percent of the time, classification accuracy would equal 99.5 percent—which sounds high, but contains no information beyond what was known before the classification exercise, which was the prevalence rate.

Precision, on the other hand, essentially follows from Bayes' rule. Bayes' rule is defined as:

$$P(A | B) = \frac{P(A) * P(B | A)}{P(B)}$$

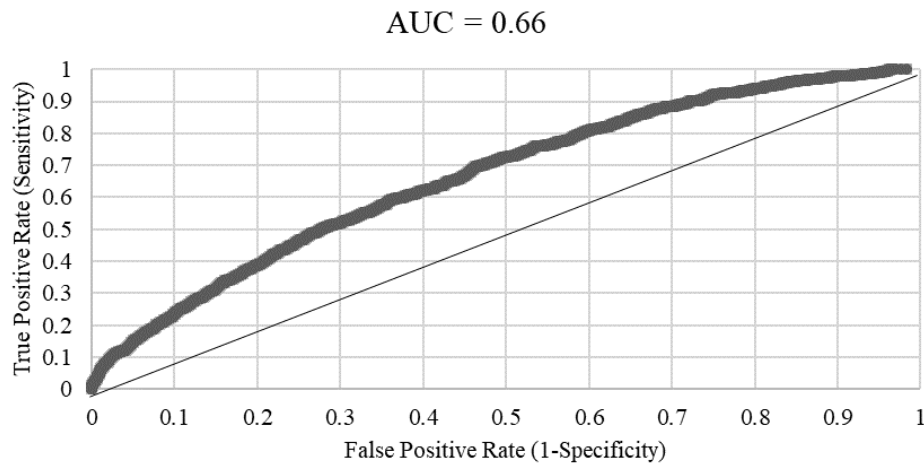
Bayes' rule states that the prior probability $P(A)$ can be updated to the posterior probability $P(A | B)$ through learning about $P(B)$ and $P(B | A)$. In the context of the AAER sample, the prior $P(A)$ is the prevalence of AAERs in the sample. By analyzing financial statement values through complex modeling, we can learn both the probability of a positive classification $P(B)$ as well as the probability of a positive classification given that an AAER occurred $P(B | A)$. This equation can be rewritten in terms of prevalence, sensitivity, and specificity:

$$\text{Precision} = \frac{\text{Prevalence} * \text{Sensitivity}}{\text{Prevalence} * \text{Sensitivity} + (1 - \text{Prevalence}) * (1 - \text{Specificity})}$$

The second alternative metric presented was AUROC, or area under the receiver operating characteristic curve, which has been described as the de facto

standard for measuring classification performance (Fawcett 2006). Some prefer this rule because it measures the area under a curve generated by mapping sensitivity and $(1 - \text{specificity})$ across all possible cutoff points. The AUROC is a summary statistic that is standardized between 0.5 (where the model fails completely) and 1.0 (which would represent a perfect classifier). Figure A1 shows an example of an AUROC.

Figure A1. Area under the curve example



While AUROCs provide a standard statistic, they are more difficult to interpret. A Google machine-learning crash course says that “one way of interpreting AUROC is as the probability that the model ranks a random positive example more highly than a random negative example” ([link](#)).

The AUROC measures the entire area below the curve. If the curve went up in a straight line to where sensitivity is one and the false positive rate is zero (upper left of the graph), the AUROC value would shade the entire graph and equal 1.0 which would imply a perfect classifier (not likely in the real world). The minimum AUROC possible is 0.5, which is represented by the diagonal. Where sensitivity equals the false positive rate, the classifier contributes no information and is not different from a random guess in the sample. The reason why the AUROC cannot be below the diagonal is that the decision rule where the classifier does worse could simply be reversed.

The third measure reported comes from the information retrieval literature and applies a log discounting to the rank order. This measure is called normalized discounted cumulative gain (NDCG) and can be measured across the entire sample, or at a certain cutoff k (NDCG@ k). It takes the following form:

$$DCG@k = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

$$NDCG@k = \frac{DCG@k}{IdealDCG@k}$$

$rel_i = \text{RelevanceScore in rank}_i$

While relevance offers a way to weight observations, for fraud detection this value simplifies to the binary indicator for fraud in the dependent variable. Therefore, the numerator $2^{rel_i} - 1$ simplifies to zero or one matching the AAER indicator value. Table A2 illustrates a toy example for NDCG@k where there are 10 observations and 3 positive cases. Column 1 ranks the probability outcome in rank order from 1 to 10. Column 2 identifies the AAERs, which are shown to be in the second, fourth and fifth rank. DCG@k sums the discounted cumulative gain values applying the discount factor of $1 / \log_2(i + 1)$ as shown in the formula above. To normalize this value, an ideal ranking must be computed. To do so, the cumulative gain column is sorted in descending order. The final column discounts this ranking applying the same factor as before. With the DCG@k and the Ideal DCG@k values, the normalized value NDCG@k can be calculated. In this example, the value is 0.68. A perfect classifier would have the value of 1.0. For more theoretical background on NDCG@k, see Yining Wang et al. (2013).

TABLE A2. Normalized discounted cumulative gain example (NDCG@k)

Rank _i	Cumulative gain	Discount factor ⁷	Discount cumulative gain (DCG@k)	Ideal ranking	Ideal DCG@k
(1)	(2)	(3)	(2) × (3)	(4)	(3) × (4)
1	0	1.00	0.00	1	1.00
2	1	0.63	0.63	1	0.63
3	0	0.50	0.00	1	0.50
4	1	0.43	0.43	0	0.00
5	1	0.39	0.39	0	0.00
6	0	0.36	0.00	0	0.00
7	0	0.33	0.00	0	0.00
8	0	0.32	0.00	0	0.00
9	0	0.30	0.00	0	0.00
10	0	0.29	0.00	0	0.00
Total	3		1.45	3	2.13
NDCG@k	0.68				

7. Discount factor is $1/\log_2(\text{rank}_i+1)$.

NDCG@k is similar to precision without the log discounting. Observe that Column 2, labeled Cumulative Gain, is the sum of the number of true positives above the cutoff threshold k . This value divided by the total observations in the subgroup is 30 percent for this example. Precision@k would also be 30 percent. So, NDCG@k simply adds a discount factor to the rank position for observations in this subgroup. The benefit to NDCG@k is that it adds information as to the order within the group, but at the cost of a loss to interpretability.

Data and code

Data and code used in this research is available from the journal website ([link](#)).

References

- Amiram, Dan, Zahn Bozanic, and Ethan Rouen** (ABR). 2015. Financial Statement Errors: Evidence from the Distributional Properties of Financial Statement Numbers. *Review of Accounting Studies* 20(4): 1540–1593.
- Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan**. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28(1): 17–82.
- Fawcett, Tom**. 2006. An Introduction to ROC Analysis. *Pattern Recognition Letters* 27(8): 861–874.
- McGinty, Jo Craven**. 2014. Accountants Increasingly Use Data Analysis to Catch Fraud. *Wall Street Journal*, December 5. [Link](#)
- Wang, Yining, Liwei Wang, Yuanzhi Li, Di He, and Tie-Yan Liu**. 2013. A Theoretical Analysis of NDCG Type Ranking Measures. Presented at the 26th Annual Conference on Learning Theory (Princeton, N.J.), June. *Proceedings of Machine Learning Research* 30: 25–54. [Link](#)

About the Author



Stephen Walker earned his Ph.D. at the University of California Haas School of Business in May 2021. Prior to his Ph.D. studies, Stephen worked in equity research at Sanford C. Bernstein in New York City. He also holds an MBA from Columbia Business School. He currently works as an economic consultant for shareholder litigation and can be reached via his personal website at stephenwalker.me.

[Go to archive of Comments section](#)

[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6043>



On Whether the Size of Government Belongs in Economic Freedom Indices

Ryan H. Murphy¹

[LINK TO ABSTRACT](#)

Jan Ott (2018) proposes that measures of economic freedom can be “improved” by ceasing to include measures of the size of government. Applying the concept of convergent validity, he shows that including size of government when measuring economic freedom yields a value of Cronbach’s alpha that he finds disturbingly low. Convergent validity and Cronbach’s alpha, which evaluate how well different measures seem to be measuring the same thing, are both taken from the broader literature on construct validity. In Ott’s analysis, lopping off the size of government yields a far improved evaluation of economic freedom indices as measuring a single thing. If it were true that we ought to lop off size of government, it would lead to a considerable shift in evaluating which countries are the most economically free.

This paper criticizes Ott’s suggestion. First, the value of Cronbach’s alpha that Ott finds is based on an older version of economic freedom data² and is based on very specific years of data. Second, the construction of a metric of economic freedom is a conceptual matter, and it is unclear why a purely statistical criterion such as Cronbach’s alpha should dictate how we select variables; the diagnostic and approach Ott uses are applicable to what are known as reflective constructs, not formative constructs. Measures of economic freedom are formative constructs.

1. Bridwell Institute for Economic Freedom, Cox School of Business, Southern Methodist University, Dallas, TX 75275.

2. While Ott (2018) considers both the Heritage *Index of Economic Freedom* and the Fraser Institute’s *Economic Freedom of the World* (Gwartney et al. 2021), we will focus on the Fraser index.

Third, if we were to apply Ott's methodology to other sets of institutions and other indices in use, we would quickly reach untenable conclusions.

Ott's article was published in the journal *Social Indicators Research*. I submitted a comment along the lines of the present article to that journal, but it was summarily turned away.

While Ott (2018) will be the focus of this analysis, it is in a long line of a family of criticisms and re-framings of data on economic freedom. Some research has analyzed the economic freedom data and suggested that the size of government should be treated differently because of its weak correlation with the rest of the index (Heckelman and Stroup 2005; Rode and Coll 2012; Bjørnskov 2016) and others have raised similar issues (Heckelman and Stroup 2000; Carlsson and Lundstrom 2002; Justesen 2008). Summarizing this position, Christian Bjørnskov (2016, 15) states that "spending and revenue components tend to only be weakly associated with other elements and are therefore a separable dimension" of institutions. Even if there is more than one dimension underlying economic freedom, it is inappropriate to remove a dimension from a formative construct, as we will see.

Others have suggested that weighting or manipulating the data differently will improve its in-sample explanatory power; these concerns often focus on the size of government.³ While not all of this literature presents itself as opposed to the current choices in methodology for economic freedom indices, taken together, that literature constitutes perhaps the most substantial academic criticism of the EFW index. That literature has also begun leaking into the public policy world as a criticism (see, e.g., Hammond 2018, 7ff.). What will be explored therefore speaks to the literature on economic freedom indices generally, not only to Ott (2018).

Convergent validity and economic freedom

The *Economic Freedom of the World* (EFW) index (Gwartney et al. 2021)

3. Some of the emphasis found in Ott (2018) is also on simple correlations between institutional measures and outcomes (especially happiness). Reconstructing the index using such correlations have previously been attempted by Heckelman and Stroup (2000), who create weights (including negative weights) from a hedonic regression. Applying weights in this manner is the primary methodology used in the subnational index, *Freedom in the 50 States* (Ruger and Sorens 2018), although they do not apply any negative weights. Huskinson and Lawson (2014) apply clustering analysis to the data, and found that countries with high degrees of economic freedom are bifurcated between those with small governments and large governments. On the surface, countries with high economic freedom and large governments perform better in measures of economic development. While there may be some use in performing these exercises, the analysis found here will suggest limits on what they can tell us. Diagma and Vallee (2021) use principal component analysis for suggesting a different way of weighting the Heritage index.

measures economic freedom for up to 165 countries, going back as far as the year 1970 in the main dataset. The data provides scorings once every five years from 1970–2000 and yearly from 2000–2019. There are five “areas” of economic freedom: the (limited) size of government, the quality of the legal system and property rights, sound money, the freedom to trade internationally, and (limited) regulation. Underlying each area are numerous components and subcomponents. All data is placed on a zero to ten scale with ten always corresponding to “more freedom.” The index has been shown to be robustly related to growth (De Haan et al. 2006) and generally correlates with positive social outcomes (Hall and Lawson 2014). The size of government area, in its present form, is composed of government consumption as a percentage of total consumption, transfers and subsidies as a percentage of GDP, government investment as a percentage of total investment, the top marginal tax rate and the level at which it applies, and government ownership of the economy.

Ott (2018) has a trio of research questions that he hopes to answer at the intersection of happiness research and economic freedom. The second of these questions is what is of interest—whether economic freedom indices can be improved, primarily by using the concept of convergent validity as its method. More broadly, validity in the statistical sense is concerned with whether a statistical construct is measuring what it purports to be measuring. While Ott assesses both EFW and another measure, that from the Heritage Foundation, I focus on EFW, which is more significant in the academic sphere and with which I am associated. Ott’s ultimate conclusion is that measures of the size of government should be omitted from economic freedom indices because they greatly reduce Cronbach’s alpha, an assessment of convergent validity: lopping it off would yield a value of over 0.8 (which Ott classifies as “good”), whereas the value for the index as-is, as measured by Ott, is below 0.7 (the cutoff Ott gives as being “acceptable”).

For all facets of his analyses, Ott (2018, 482) uses the years 2010–2012 in considering his hypotheses. Ending in year 2012 implies that the version of the data report published in 2014 was used. Additionally, Ott apparently omits about⁴ 25 countries from the analyses, as he applies a consistent set of countries across the different exercises he performs in the paper. The country list has not been reproduced, but results with the full list of countries differs immaterially from what Ott observes. These findings are reproduced in Table 1.

My first concern about Ott’s findings is that if the sample is not restricted to 2010–2012 and instead the sample from 1970–2012 is used, Cronbach’s alpha for EFW immediately rises to 0.73, clearly crossing the first threshold given by Ott.

4. Under Table 3 in Ott (2018) it says the analyses in the table use “125–127 nations.” In this version of the Fraser data, 152 countries were assigned economic freedom ratings.

Restricting the sample to more recent years is a defensible position, but it is not explicated or justified by Ott (besides it simply matching the sample of countries and years from his other analyses in the paper). Truncating the sample may purge a large number of data points where large governments were observed to coincide with bad institutional quality, particularly those countries pursuing socialism in the period before the fall of the Soviet Union. Because of data issues, there were only so many data points before the fall of the Soviet Union in this version of the data set, so if anything, the historical instances of large governments coinciding with other measures of bad institutions remain underemphasized in the data.

TABLE 1. Ott (2018) result for Cronbach's alpha, as reported and as reproduced

	EFW	EFW, no size of government	EFW, size of government reversed
Ott (2018)	0.66	0.85	0.76
As reproduced	0.67	0.83	0.73

Since the 2014 publication of the 2012 data, there have been several methodological changes to the data. An adjustment was put in place for the legal system and property rights area, which adjusts countries downwards when there is disparate treatment of women and their economic rights (Fike 2017). Another important update was the inclusion of data on government ownership of the economy, which addresses particularly conspicuous interventions in otherwise economically free Hong Kong and Singapore. The variable is slotted into the size of government area, with the data's source originating in the *Varieties of Democracy* dataset (Coppedge et al. 2021). The legal system and property rights area has been buttressed with historical data from *Varieties of Democracy* as well as data on the rule of law from Drew Linzer and Jeffrey Staton (2015), making ratings for this area before 2000 far more robust. These and further changes have improved the precision of the index, especially for years prior to 2000.

If we simply calculate Cronbach's alpha for all years 1970–2019 using the current index, as is found in Table 2, it is now 0.77, comfortably higher than Ott's "acceptable" threshold of 0.7 and not that distant from the "good" threshold of 0.8. As for Ott's narrow period of 2010–2012, with the revised data it is 0.75. If we look at the full period 2000–2019 (with 2000 as a cut-off when data had strongly improved in quality), Cronbach's alpha holds steady at 0.76. If we look at just the data for the year 2019, which is the most recent at the time this writing, it is also 0.76.

TABLE 2. Further recalculations of Cronbach's alpha, under various assumptions

	EFW	EFW, no size of government	EFW, size of government reversed
Current EFW data, all years	0.77	0.85	0.72
Current EFW data, 2010–2012	0.75	0.86	0.74
Current EFW data, 2000–2019	0.76	0.87	0.75
Current EFW data, 2019	0.76	0.87	0.78

Ott (2018) asks whether the size-of-government data could be reversed because doing so also would improve Cronbach's alpha, though he doesn't conclude this change should be implemented. With the current version of the data, reversing the sign of size of government no longer improves Cronbach's alpha, and actually reduces it for all of the periods in question except for the analysis using only the data for the year 2019. It remains true that eliminating the size of government causes Cronbach's alpha to creep higher still, but when the full data set is examined, EFW is above the first threshold and approaching the second. One would only conclude to exclude the size of government if one prefers improved convergent validity lexically above any other concern we may have in constructing an index.

Validity and economic freedom

'Validity' asks whether a statistical construct is measuring what it claims to be measuring. 'Reliability' asks how well indicators are zeroing in on a single concept and about the internal consistency among those indicators (Coaley 2010, 128–129). Cronbach's alpha is primarily about measuring reliability, with convergent validity being one method among many for assessing validity. "Although [Cronbach's alpha] is a useful tool for summarizing the internal consistency of items on a scale as a measure of reliability, reliability is necessary but not sufficient evidence of validity" (Flake et al. 2017, 6).

Jessica Flake, Jolynn Pek, and Eric Hehman have criticized the use of Cronbach's alpha as a means of choosing which variables to include as one of the main misuses of the tool in the literature: "The heavy reliance on [Cronbach's alpha] also suggests that researchers are using it as a criterion for scale use and even item selection. Indeed, we noted numerous instances in which [Cronbach's alpha] was reported to justify item removal. Reliability is important to consider in construct validation, but *it should not be maximized at the expense of other evidence*" (Flake et al. 2017, 6, emphasis added). There are several other dimensions of validity to consider besides convergent validity, among them content validity, which importantly asks whether every dimension of a theory is being captured by the construct.

In any case, Ott misapplies these methodologies because he implicitly assumes that *Economic Freedom of the World* is a reflective construct. EFW is a formative (causal) construct. In a reflective construct, it is assumed that there is some concept out there which has a set of observable effects. The reflective construct aggregates these effects into a single construct. In a formative construct, this relationship is reversed. Each of the individual components of the concept *define* or *cause* the concept being measured. In the case of EFW, those components are the size of government, the quality of the legal system and property rights, sound money, the freedom to trade internationally, and regulation. (See Bollen and Diamantopoulos 2017a for a lengthy discussion of formative models.)

Indicators in a reflective model are all expected to be rather strongly correlated with one another. They are all effects of the same latent cause. In contrast, “causal indicators of the same concept can have positive, negative, or no correlation” (Bollen and Lennox 1991, 307). In a reflective model, one can think of different items to include as being a sample of the causal indicators. But one *cannot* do this in a formative construct. The list of enunciated concepts defining the construct is better thought of as a census. “Omitting an indicator is omitting part of the construct” (ibid., 308). To use other terminology, convergent validity is irrelevant and content validity is essential for formative constructs. What Ott is asking us to do is analogous to applying convergent validity to the human body, finding that the human head is poorly correlated with the rest of the body, and concluding we should chop it off because it doesn’t fit with the rest of the body.

Ott at one point implies that there is in fact no theory underlying the concept of economic freedom. “It is however debatable to use taxation and government spending as positive indicators for economic freedom. There is no theoretical justification; government spending, consumption and transfers and subsidies included, can be directed at very different policies, liberal or less liberal!” (2018, 485). Ott’s claim that there is “no theoretical justification” is a dismissal of the entire classical liberal tradition of social philosophy (see Rothbard 1973; Friedman 1962, 177–195; White 2012, 332–359, 382–412). And if a certain level of government spending is needed to support the liberal society and free markets, we can note that the index does not punish government consumption below six percent of total consumption, which seems to be a sufficient level for the provision of basic law, courts, and national defense (cf. Gwartney et al. 1998).

Economic freedom, the size of government, and other measures of institutions

We should also think about what Ott's method implies, and ask: What if we apply the strictures of convergent validity to other measures of institutions? For one, we could apply it to the five components of the size of government itself. Cronbach's alpha for the five size-of-government EFW components is 0.220. Does this mean we should say that certain kinds of government spending, taxation, or ownership do not count as the size of government because convergent validity indicates that they are not measuring the same concept?

As another example, consider the *Varieties of Democracy* dataset. Its baseline definition of democracy is "electoral democracy," which it then combines with other democratic or democracy-adjacent concepts to create measures of liberal democracy, participatory democracy, deliberative democracy, and egalitarian democracy. Electoral democracy is in turn composed of five mid-level indices which are meant to capture the core of the idea of democracy—freedom of expression, freedom of association, universal suffrage, free and fair elections, and elected officials. In the 2020 version of the data, the Cronbach's alpha for these five variables is 0.827. If we include all their historical data, it rises to 0.885. But if we remove the variables that are less tightly correlated with the other democracy variables, universal suffrage and elected officials, Cronbach's alpha rises to 0.940. Have we improved the measurement of democracy by increasing Cronbach's alpha by ignoring whether there is universal suffrage in a country?

The answer to these questions is no. The reason why the answer is no is because arbitrarily removing certain kinds of spending from the size of government or universal suffrage from a measure of democracy means that you are picking apart necessary, definitional dimensions of a formative construct and no longer are adequately addressing content validity of what we have defined theoretically at the outset of examining the question. The same is true for the size of government and economic freedom.

Discussion and conclusion

Ironically, a careful consideration of *Economic Freedom of the World's* nature as a formative construct actually demands that we do *not* remove size of government from the index, because doing so would leave us a set of measures that is missing an

important aspect of classical liberal economic institutions—limited government. The point applies not only to Ott (2018), but to others who have suggested that the size of government should be removed from economic freedom because it behaves somewhat differently than the rest of the index (e.g., Heckelman and Stroup 2005; Rode and Coll 2012; Bjørnskov 2016; cf. Heckelman and Stroup 2000; Carlsson and Lundstrom 2002; Justesen 2008). It also raises questions concerning what relationship the results of applications of principal component analysis, factor analysis, or clustering analysis even mean in relation to economic freedom (as in Caudill et al. 2000; Sturm et al. 2002; Rode and Coll 2012; Huskinson and Lawson 2014; Diagla and Vallee 2021).

That is not to say that one cannot disaggregate and restructure the data; Scott Sumner (2010) previously adjusted economic freedom data in a similar manner to Ott (2018), but chose to call economic freedom data without the size of government “neoliberalism” and economic freedom data with government spending inverted “egalitarian neoliberalism.” Whether such simple adjustments to the index conceptually capture all the dimensions of (i.e., perform a ‘census’ for) “neoliberalism” and “egalitarian neoliberalism” would then be a further question to pose. But if it is simply argued that neoliberal institutions are better for human welfare than classical liberal economic institutions (economic freedom), that is a hypothesis to be investigated and debated. It isn’t a reason to use convergent validity in order to cease testing of economic freedom as it has been conventionally defined.

In fact, if we are using these methods as bases for improving *Economic Freedom of the World*, content validity would suggest that several elements of regulation data are currently lacking.⁵ There is no data in the index on environmental or antitrust regulation; and “credit market regulation,” the closest variables to financial regulation in the index, mostly features rather weak proxies for such kinds of regulation. One could plausibly argue that these kinds of regulations, unlike other elements of intervention, have clearer pathways for having positive impacts on economic performance. However, that is not a reasonable rationale for excluding them (though that is not the reason why they are excluded; there simply is not good data available that is sufficiently dense in its coverage).

However, as alluded to above, disaggregating may be desirable for an entirely separate reason—to investigate the origins of economic freedom. The existing literature on the origins of economic freedom is principally concerned with the

5. While these are areas for improvement in the index, Bollen and Diamantopoulos (2017b, 606–607) do caution not to take “census” too literally. The issue is whether lacking a dimension will bias the construct in one direction or another. Leaving out the size of government clearly changes the results and therefore its exclusion would lead to bias.

effects on the overall index (Lawson et al. 2020). The origins of property rights, a high-quality legal system, sound money, free trade, and light regulation may be the same as general measures of institutional quality. The origins of the size of government may differ. We may well be concerned with the net, overall effects on economic freedom, but we may add a certain amount of nuance for this specific question. The main point is to not remove the size of government when one believes the hypothesis being tested concerns economic freedom.

Ultimately, the relationships among the size of government, other dimensions of economic freedom, and other variables remain opaque and understudied. There was a point in time that simply looking at the size of government was thought of as a reasonable first approximation of how much economic freedom there was in a country (e.g., Peltzman 1980), and it wasn't understood this was not the case until the projects to measure economic freedom were well underway (Gwartney 2009). Yet it is wholly inappropriate to remove the size of government from our definition of economic freedom merely because the world is complex.

Data and code

Data and code used in this research is available from the journal website ([link](#)).

References

- Bjørnskov, Christian.** 2016. Economic Freedom and Economic Crises. *European Journal of Political Economy* 45: 11–23.
- Bollen, Kenneth A., and Adamantios Diamantopolus.** 2017a. In Defense of Causal-Formative Indicators: A Minority Report. *Psychological Methods* 22(3): 581–596.
- Bollen, Kenneth A., and Adamantios Diamantopolus.** 2017b. Notes on Measurement Theory for Causal-Formative Indicators: A Reply to Hardin. *Psychological Methods* 22(3): 605–608.
- Bollen, Kenneth A., and Richard Lennox.** 1991. Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin* 110(2): 305–314.
- Carlsson, Fredrik, and Susanna Lundstrom.** 2002. Economic Freedom and Growth: Decomposing the Effects. *Public Choice* 112: 335–344.
- Caudill, Steven B., Fenderando C. Zanella, and Franklin G. Mixon.** 2000. Is Economic Freedom One Dimensional? A Factor Analysis of Some Common Measures of Economic Freedom. *Journal of Economic Development* 25(1): 17–40.
- Coaley, Keith.** 2010. *An Introduction to Psychological Assessment and Psychometrics*. London:

SAGE.

- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Allen Hicken, Anna Lührmann, Seraphine F. Maerz, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundtröm, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, and Daniel Ziblatt.** 2021. *V-Dem Codebook v11.1*. Gothenburg, Sweden: V-Dem Institute.
- De Haan, Jakob, Susanna Lundstrom, and Jan-Egbert Sturm.** 2006. Market-Oriented Institutions and Policies and Economic Growth: A Critical Survey. *Journal of Economic Surveys* 20(2): 157–191.
- Diagla, Isaaka, and Thomas Vallee.** 2021. The Index of Economic Freedom: Methodological Matters. *Studies in Economics and Finance* 38(3): 529–561.
- Fike, Rosemarie.** 2017. Adjusting for Gender Disparity in Economic Freedom and Why It Matters. In *Economic Freedom of the World*, ed. James Gwartney, Robert Lawson, and Joshua Hall. Vancouver: Fraser Institute.
- Flake, Jessica K., Jolynn Pek, and Eric Hehman.** 2017. Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychology and Personality Science* 8(4): 370–378.
- Friedman, Milton.** 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gwartney, James.** 2009. Institutions, Economic Freedom, and Cross-Country Differences in Performance. *Southern Economic Journal* 75: 937–956.
- Gwartney, James, Randall Holcombe, and Robert Lawson.** 1998. The Size of Government and the Wealth of Nations. *Cato Journal* 18(2): 163–190.
- Gwartney, James, Robert Lawson, Joshua Hall, and Ryan H. Murphy.** 2021. *Economic Freedom of the World*. Vancouver: Fraser Institute.
- Hall, Joshua, and Robert Lawson.** 2014. Economic Freedom of the World: An Accounting of the Literature. *Contemporary Economic Policy* 32(1): 1–19.
- Hammond, Samuel.** 2018. The Free-Market Welfare State: Preserving Dynamism in a Volatile World. May. Niskanen Center (Washington, D.C.). [Link](#)
- Heckelman, Jac C., and Michael D. Stroup.** 2000. Which Economic Freedoms Contribute to Growth? *Kyklos* 53(4): 527–544.
- Heckelman, Jac C., and Michael D. Stroup.** 2005. A Comparison of Aggregation Methods for Measures of Economic Freedom. *European Journal of Political Economy* 21(4): 953–966.
- Huskinson, Toby, and Robert Lawson.** 2014. Clusters of Economic Freedom. *Applied Economics Letters* 21(15): 1070–1074.
- Justesen, Morgens K.** 2008. The Effect of Economic Freedom on Growth Revisited: New Evidence on Causality from a Panel of Countries 1970–1999. *European Journal of Political Economy* 24(3): 642–660.
- Lawson, Robert, Ryan H. Murphy, and Benjamin Powell.** 2020. The Determinants of Economic Freedom: A Survey. *Contemporary Economic Policy* 38(4): 622–642.
- Linzer, Drew A., and Jeffrey K. Staton.** 2015. A Global Measure of Judicial Independence,

1948–2012. *Journal of Law and Courts* 3(2): 223–256.

Ott, Jan. 2018. Measuring Economic Freedom: Better Without Size of Government. *Social Indicators Research* 135: 479–498.

Peltzman, Sam. 1980. The Growth of Government. *Journal of Law and Economics* 23(2): 209–287.

Rode, Martin, and Sebastian Coll. 2012. Economic Freedom and Growth: Which Policies Matter the Most? *Constitutional Political Economy* 23: 95–133.

Rothbard, Murray. 1973. *For a New Liberty: The Libertarian Manifesto*. New York: Macmillan.

Ruger, William P., and Jason Sorens. 2018. *Freedom in the 50 States*, 5th ed. Washington, D.C.: Cato Institute.

Sturm, Jan-Egbert, Erik Leertouwer, and Jakob de Haan. 2002. Which Economic Freedoms Contribute to Growth? A Comment. *Kyklos* 55(3): 403–416.

Sumner, Scott. 2010. The Great Danes: Cultural Values and Neoliberal Reforms. Working paper, June 24. [Link](#)

White, Lawrence H. 2012. *The Clash of Economic Ideas: The Great Policy Debates and Experiments of the Last Hundred Years*. New York: Cambridge University Press.

About the Author



Ryan Murphy is a research associate professor at the Bridwell Institute for Economic Freedom at Southern Methodist University. He received a Ph.D. in economics from Suffolk University in September 2013. His research interests are diverse and include institutional economics, public policy, and macroeconomics. His email address is rhmurphy@smu.edu.

[Jan Ott's reply to this article](#)
[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6044>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 58–64

Leave Size of Government Out of the Measurement of Economic Freedom— Put Quality of Government In

Jan Ott¹

[LINK TO ABSTRACT](#)

Ryan Murphy's approach is popular in many debates about social-economic questions. I am glad that Murphy's statistical results (Murphy 2022) differ only immaterially from my results for the years 2010–2012, with a different sample of nations (Ott 2018). At some points he is perhaps too pessimistic about my measurement of economic freedom, but let me first say a few words about two fundamental issues: the *conceptualization* and the *measurement* of economic freedom.

Conceptualization of economic freedom

Freedom is usually defined and interpreted as the actual possibility to choose. This possibility depends on two dimensions: the opportunity to choose and the capability to choose (Sen 1999). The opportunity is a characteristic of the social environment and depends on the absence of inhibitions and the availability of options. When using the word 'freedom,' options and inhibitions are supposed to be *man-made*. Capability is a characteristic of individuals. It is not unusual to ignore capabilities and to concentrate on opportunities in the definition of freedom. The combination of capabilities and opportunities in one concept is difficult to handle.

1. World Database of Happiness, Erasmus University Rotterdam, 3062 PA Rotterdam, Netherlands.

It is not realistic, however, to define opportunities exclusively in a negative way, as just the absence of inhibitions, or exclusively in a positive way, as just the availability of options. Options and inhibitions always go together; they are always intertwined and complementary. Legislation and policies create combinations of inhibitions and options simultaneously.

Specific institutional arrangements facilitate and regulate specific types of freedom in different domains, like going to school, getting married and divorced, religion, and political participation. An interesting outcome of empirical research is that different types of freedoms in nations are correlated.² This is not self-evident because different types of freedom have different priorities for different groups of people. We may conclude that individual freedom is an actual phenomenon and cultural standard in different domains.

Economic freedom is in my view about the opportunity to choose in economic decisions, like buying and selling goods and services, hiring and firing staff, international trade, and starting and maintaining a business. The Fraser Institute, however, rejects this concept of freedom as the opportunity to choose and prefers a more specific concept, as expressed in different formulations in the annual *Economic Freedom of the World* reports. In many reports, and in the first sentence of the Executive Summary of the annual report of 2021, we find the following formulation:

The index published in Economic Freedom of the World measures the degree to which the policies and institutions of countries are supportive of economic freedom. The cornerstones of economic freedom are personal choice, voluntary exchange, freedom to enter markets and compete, and security of the person and privately owned property.

In the annual report of 2016, Chapter 1, page 1, we find an interesting additional explanation:

Conceptually, economic freedom is present when economic activity is coordinated by personal choice, voluntary exchange, open markets, and clearly defined and enforced property rights. People are economically free when they are permitted to choose for themselves and engage in voluntary transactions as long as they do not harm the person or property of others. The Economic Freedom of the World (EFW) measure might be thought of as an effort to identify how closely the institutions and policies of a country correspond with the ideal of limited government, where the government protects property

2. Well-known types of freedom in this context are global freedom by political rights and civil liberties, press freedom, personal autonomy, and freedom to make life choices. The first three are measured by Freedom House and the last one is measured by the Gallup World Poll. See also Veenhoven 2008.

rights and arranges for the provision of a limited set of “public goods” such as national defence and access to money of sound value, but little beyond these core functions. To a large degree, a country’s EFW summary rating is a measure of how closely its institutions and policies compare with the idealized structure implied by standard textbook analysis of microeconomics.

This idealized structure is a structure with many consumers and producers, each of them without substantial market power, who operate in a rational way on the basis of adequate information. The differences between this concept and the usual concept of freedom, as the possibility or opportunity to choose, are very modest at first sight. The individual possibility or opportunity of personal choice is always a key point. There is, however, a lot of attention for inhibitions by government activities, while options created by governments, and options and inhibitions created by other agents, get less attention. In view of the conceptualization, one would at least expect some appreciation for antitrust legislation.

The representatives of the Fraser Institute, Murphy included, argue, however, that they use the concept of economic freedom in a *formative* way, meaning that the constituent dimensions define what it is, and that they are free to decide what these dimensions are. The implication is that there is never any inconsistency between the concept of economic freedom and measurement, because the meaning is implied in the measurement. Validity issues are irrelevant in this approach.

Everybody is free to do that, but I believe it is still an interesting and informative exercise to look at the validity of the measurement by the Fraser Institute, assuming that economic freedom is about the opportunity to choose in economic decisions, without specification of the origin of limitations or options.

The measurement of economic freedom by the Fraser Institute, if interpreted as the opportunity to choose, is reasonable—but better without size of government

The Fraser Institute has selected five items to measure economic freedom. Three of them are assumed to contribute to economic freedom: *rule of law and protection of property*, *sound money* and *freedom to trade internationally*. The other two items, *regulation* and *size of government*, are assumed to diminish economic freedom. We can assess the validity of this measurement, assuming that economic freedom is about the opportunity to choose, even if the Fraser Institute prefers a more specific

meaning. We can have a look at the convergent validity, the underlying observations or content validity, and the predictive validity in view of previous research.

The Cronbach's alpha (CA) is—or can be used as—a statistical measure for the convergent validity of the measurement. This CA is high if the items used measure the same phenomenon. The CA of the measurement of Economic Freedom by the Fraser Institute is very reasonable, but higher if we leave out the size of government as one of the items.³ The reason is that the size of government has a positive correlation with the first three positive items, instead of a negative one as assumed.

Murphy is somewhat pessimistic about my application of the CA, because I seem to apply this measure in a formulaic way. It is indeed important to apply this measure carefully. As a general rule we may say that 0.7 is reasonable and that 0.8 is good, but it is necessary to compare the CA of different sets of items, and to combine information about the differences with additional information about the underlying observations, and with information about the predictive validity. I should have been more explicit at this point.

We can specify this outcome by having a look at the underlying observations, occasionally denoted as content validity. The item *size of government* is the average of four sub-items: *government consumption as a percent of national consumption*, *transfers and subsidies*, *government enterprises*, and *top tax rate*. The first two items have a positive correlation with the three positive items: *rule of law and protection of property*, *sound money*, and *freedom to trade internationally*. This is not consistent with the assumption that the *size of government* has a negative correlation with economic freedom, as measured with these positive items.

The conclusion is that we can improve the measurement of Economic Freedom (as the opportunity to choose!) by leaving out the *size of government* as an item. Leaving *government consumption* and *transfers and subsidies* out of *size of government* is less appropriate, since these sub-items are more representative for the actual size of government than the more specific items *government enterprises* and *top tax rate*. Reversing the sign of *size of government* from negative to positive is less appropriate, since the selection of the sub-items is disputable.

This conclusion is supported by a higher predictive validity if we leave *size* out. In previous research we find that different types of freedom have a positive mutual correlation (Veenhoven 2008), and we find a positive correlation between freedom in nations and average happiness (Ott 2018). If we leave out the *size of government* in the measurement of economic freedom, we see that the correlation

3. The CA with size of government is 0.66 with size of government and 0.85 without size of government. (Ott 2018, data 2010–2012; 127 nations). With the data used by Murphy (2022), this is 0.67 and 0.83 respectively for the years 2000–2019 with 152 nations.

between Economic Freedom and other types of freedom and average happiness goes up.⁴ This is consistent with, and predicted by, previous research. Predictive validity is, just like the convergent validity, not a simple recipe to be applied in an axiomatic way. Predictive validity can be low because previous research may have produced disputable results. There are no indications, however, that the research results just mentioned are disputable.

Another result of previous research is that the correlation between the *size of government* and average happiness depends heavily on the quality of governments (Helliwell and Huang 2008; Ott 2010a). This correlation is positive if the quality of governments is good. This is a trivial but important conclusion. Interesting examples are presented by the Nordic countries, with high levels of government quality, government consumption and taxation, happiness, and economic freedom.⁵ Employers in these countries may even lay off employees without substantial limitations or conditions. The ‘secret’ of these countries is that their governments organize many provisions free of charge, like education, health care, training, and care for children and the elderly. In this way they optimize their investments in human capital.⁶ Their income tax is high, and their governments are relatively big, but they achieve high levels of employment and productivity with high levels of happiness. The correlation between *size of government* and happiness is negative if the quality of government is bad. Governments which are oversized as a result of corruption, rent-seeking, and patronage are regrettable examples.

Conclusion and discussion

The Fraser Institute prefers a more specific interpretation, but the conclusion is that its measurement of economic freedom, if interpreted as the opportunity to choose, is very reasonable. The measurement can be further improved by leaving out the *size of government*. Just reversing the sign is less appropriate, because the selection of sub-items is inadequate.

The relation between economic freedom and the *size of government* depends

4. The positive correlations of economic freedom—size of government included—with global freedom, personal autonomy, and press freedom, as measured by Freedom House, are 0.51, 0.57, and 0.50. Without size of government these are 0.59, 0.66, and 0.59 respectively. The correlation with happiness goes up from 0.46 to 0.55. Murphy (2022) pays no attention to these differences.

5. The Nordic countries always ranked high on the Ease of Doing Business index compiled by the World Bank ([link](#)).

6. The Nordic countries avoid the negative impact of inequality on the education and training of poor people, women, and young people (Jensen 2021). This is, as explained by Andersen (2021), an effective way to mitigate a fundamental market failure.

on the *quality of the government*. It is therefore also an attractive option to substitute the *size of government* as an item of economic freedom with the *quality of government*. The World Bank presents excellent data about this quality, with the Worldwide Governance Indicators (Kaufmann and Kraay 2021).

Putting items in an index is also a way to create more attention for issues. There are arguments to pay more attention to the quality of governments than to their size. One argument is that size as such is usually irrelevant for the quality and availability of goods and services to be delivered. The quality of government is more important. Intelligent governments can organize goods and services themselves, but may also do so with legislation, in cooperation with private organizations. Another argument is that it is easier to improve the quality of governments than to change their size. Attention to the quality is also more productive, because there are usually many practical and non-controversial options to improve this quality (Ott 2010b).

The quality of governments also deserves more attention in view of the current worldwide problems, like climate change, pandemics, and aggression by dictators. Our discussions about freedom are interesting, but I am sure we are all equally motivated to put an end to dictatorships and bad governments in general.

References

- Andersen, Torben M.** 2021. *Economic Performance in the Nordic World*. Odder, Denmark: Aarhus University Press and University of Wisconsin Press.
- Helliwell, John F., and Haifang Huang.** 2008. How's Your Government? International Evidence Linking Good Government and Well-Being. *British Journal of Political Science* 38: 595–619.
- Jensen, Carsten.** 2021. *Equality in the Nordic World*. Odder, Denmark: Aarhus University Press and University of Wisconsin Press.
- Kaufmann, Daniel, and Aart Kraay.** 2021. Worldwide Governance Indicators. World Bank (Washington, D.C.). [Link](#)
- Murphy, Ryan H.** 2022. On Whether the Size of Government Belongs in Economic Freedom Indices. *Econ Journal Watch* 19(1): 47–57. [Link](#)
- Ott, Jan.** 2010a. Good Governance and Happiness in Nations: Technical Quality Precedes Democracy and Quality Beats Size. *Journal of Happiness Studies* 11(3): 353–368.
- Ott, Jan.** 2010b. Greater Happiness for a Greater Number: Some Non-Controversial Options for Governments. *Journal of Happiness Studies* 11: 631–647.
- Ott, Jan.** 2018. Measuring Economic Freedom: Better Without Size of Government. *Social Indicators Research* 135(2): 479–498.
- Sen, Amartya.** 1999. *Development as Freedom*. New York: Anchor Books.
- Veenhoven, Ruut.** 2008. Freedom and Happiness: Comparison of 126 Nations in 2006. Presented at Legatum Prosperity Workshop (London), June 21–22. [Link](#)

About the Author



Jan Ott studied sociology and law at the Erasmus University in Rotterdam, specializing in social economic policy, labour relations, constitutional law, and public administration. He worked as a policy adviser at the Dutch Ministry of Social Affairs and Employment. Nowadays he works as a social researcher at the World Database of Happiness at the Erasmus University. In 2020 he finished his book *Beyond Economics: Happiness as a Standard in our Personal Life and in Politics* (Palgrave Macmillan). His email address is jan.ott@kpnmail.nl.

[Go to archive of Comments section](#)

[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6045>



Misrepresenting Mises: Quotation Editing and a Rejection of Peer Review at Cambridge University Press

Phillip W. Magness¹ and Amelia Janaskie²

[LINK TO ABSTRACT](#)

A properly functioning peer-review process flags potential problems before they appear in print. During the process, an anonymous referee might notice complications to a thesis that an author failed to account for, prompting a round of revisions to improve the piece. If an author misrepresents evidence for a claim, an anonymous referee might alert the journal editor to the problem. Usually the author will be asked to address the issue in a revision. If the problem is severe or intentional, the piece might be rejected outright.

But what if an anonymous referee flags serious problems in an article, such as misrepresented evidence or basic errors of fact, yet the journal's editor chooses to let the errors stand and run the piece anyway? What happens when the same problems are then noticed by other scholars after the article appears in print? Surely, a formal correction would be in order.

Factual corrections used to be a regular practice of most scholarly journals, whether in the form of a short comment or an exchange over the contested claim.

1. American Institute for Economic Research, Great Barrington, MA 01230. The present article proceeds with a first-person singular authorial voice, representing Phillip Magness. Amelia Janaskie worked to recast, integrate, and revise preceding materials by Magness, which are included in the references list below (especially Magness 2021). We thank Fiona Harrigan, Dietmar Georg, and Kevin Hill for their assistance with translations of Mises's works from German to English.

2. American Institute for Economic Research, Great Barrington, MA 01230.

But a growing number of academic venues no longer see a need to attend to basic standards. Factual errors, even egregious ones, are allowed to stand unchallenged, provided that the errors align with a political view. This was my own experience after a frustrating year and a half long effort to seek basic factual corrections to an unambiguous error about the relationship between the Austrian school of economics and the racial dimensions of imperialism.

In 2019, Wellesley College historian Quinn Slobodian published in scholarly journals a pair of articles containing a charge against Ludwig von Mises. Writing for the journal *Cultural Politics*, Slobodian (2019a, 379) alleged that “race theory has an ambiguous place in Mises’s work,” an allegation that encourages modern-day racists to claim inspiration from the free-market economist. Slobodian repeated and elaborated upon the charge in an article for *Contemporary European History* (CEH), stating that “libertarians who scour [Mises’s] writings to validate their divergent positions on migration can claim fairly to find confirmation of both sides of the argument.” One side of the story, he continued, derived from Mises “the realist, who saw race as a quasi-permanent category of global social organisation. Despite his liberal principles the Habsburg polyglot never became the radical anti-racist” (2019b, 155).

While Slobodian acknowledged in both articles that Mises adhered to a broad liberal philosophy that clashed with the racist and imperialist ideologies of his day, his argument held that Mises’s works contained a “parenthetical opening to the possibility of race theory,” which is a reference to pseudoscientific concepts that purport to link race and intelligence (Slobodian 2019a, 380). In turn, the posited “parenthetical opening” allegedly establishes Mises as a historical progenitor of later defenses of race theory and imperialism. Elsewhere, Slobodian (2018b) extends the argument into modern politics by blaming Mises and so-called neoliberals for inspiring anti-immigration and race theory arguments which, he says, attained popularity among Alt-Right and Trumpian political movements in the 2010s.

Slobodian builds his interpretation on textual misrepresentation

Careful textual analysis reveals that Slobodian’s representation of Mises is often in direct contradiction to Mises’s actual position. Mises devoted substantial energy in his 1927 book *Liberalism* to attacking the then-popular field of eugenics. His later works, such as 1944’s *Omnipotent Government*, contained a philosophical broadside against Nazism, singling out the errors and evils of Nazi racial theory

in particular. This reality contradicts Slobodian’s findings that Mises’s works give nourishment to racism.

By omitting pertinent context, Slobodian’s use of excerpts from Mises’s work results in an interpretation of Mises’s position that is factually incorrect and often opposite of Mises’s actual position. Many of these issues arise from Mises’s practice of offering a detailed explication of the position he is attacking or rebutting before stating his own opposing view. As a result Mises’s texts contain detailed descriptive accounts of beliefs that are not his own. Compounding this confusion, Slobodian makes extensive use of snippet quotations of Mises’s descriptions to depict him as sympathetic, or at least open, to a variety of racist and imperialist beliefs, when in fact he was condemning them. In the months that followed their publication, other scholars began to notice the same patterns in Slobodian’s depictions of Mises as well as his use of quotations (Ebeling 2018; Gordon 2021).

In his *CEH* article, Slobodian claims that Mises saw violence as potentially necessary means to grow a global trade network. However, a comparison of the *CEH* passage to Mises’s original text from his 1919 book *Nation, State, and Economy* reveals that Slobodian’s account omits the entire second half of the pertinent sentence. The omitted text not only alters Mises’s position from Slobodian’s attribution, but it also directly undermines Slobodian’s interpretation of Mises’s views in the article.

Slobodian, <i>CEH</i>	Mises, <i>Nation, State, and Economy</i>
When necessary, the opening of world markets had to be achieved through violence. Though ‘one can think only with shudders and anger of the fearful mass murders that prepared the basis for many of the colonial settlements flourishing today’, [Mises] wrote in a book published the year after the First World War, the net gain made it worthwhile; in the end, ‘ <u>all other pages of world history were also written in blood.</u> ’ Violence in the project of expanding the space of foreign investment, wage labour and commercial exchange was not only acceptable, it was necessary. (Slobodian 2019b, 148, underlining added)	It is true that those colonies were not taken with smooth talk, and one can think only with shudders and anger of the fearful mass murders that prepared the basis for many of the colonial settlements flourishing today. But <u>all other pages of world history were also written in blood, and nothing is more stupid than efforts to justify today’s imperialism, with all of its brutalities, by reference to atrocities of generations long since gone.</u> (Mises 2006/1919, 63, underlining added)

Textual edits of this type, usually inverting Mises’s meaning in a light that erroneously casts him as sympathetic to racism or colonialism, are unfortunately common in Slobodian’s writings on the subject.

In the *CEH* article, Slobodian turns next to Mises’s 1944 book *Omnipotent Government* for evidence of him condoning racism. Although Slobodian allows that in this work Mises “distanced himself from people who opposed non-white immigration,” he then notes how Mises “conceded” that

we must not close our eyes to the fact that such views meet with the consent of

the vast majority. It would be useless to deny that there exists a repugnance to abandoning the geographical segregation of various races. Even men who are fair in their appraisal of the qualities and cultural achievements of the colored races and severely object to any discrimination against those members of these races who are already living in the midst of white populations, are opposed to a mass immigration of colored people. There are few white men who would not shudder at the picture of many millions of black or yellow people living in their own countries. (Mises 1944, 107; quoted in Slobodian 2019b, 152)

From this supposed concession Slobodian concludes: “By the 1940s Mises partially legitimised closed borders for non-white migrants as a near-permanent feature of the world order” (Slobodian 2019b, 152).

Slobodian repeats this charge in the second article, for *Cultural Politics*, citing both Mises and his own article in *CEH*:

And yet, in his 1944 work, Mises conceded the difficulties of racial integration, writing in a phrase, often cited by latter-day Mises Institute Austrians, about immigration barriers that “there are few white men who would not shudder at the picture of many millions of black or yellow people living in their own countries” (1944: 107; Slobodian 2019). (Slobodian 2019a, 380)

Additionally, Slobodian claims that the passage was “often cited by latter-day Mises Institute Austrians” (2019a, 380), but without evidence. I know of no such evidence—indeed, a Google search for the passage in question revealed no specific references to this passage on the Mises Institute’s website. By contrast, an exhaustive search for it using Google Scholar produced only two references to it outside of Mises’s original text. Both were written by Slobodian.

A turn to Mises’s original passage demonstrates that Slobodian has again omitted surrounding context that points to an opposite position. Rather than leaving an ambiguous opening for racism, Mises described the racist position and condemned it in clear terms:

We must not close our eyes to the fact that such views meet with the consent of the vast majority. It would be useless to deny that there exists a repugnance to abandoning the geographical segregation of various races. Even men who are fair in their appraisal of the qualities and cultural achievements of the colored races and severely object to any discrimination against those members of these races who are already living in the midst of white populations are opposed to a mass immigration of colored people. There are few white men who would not shudder at the picture of many millions of black or yellow people living in their own countries. **The elaboration of a system making for harmonious coexistence and peaceful economic and political coöperation among the various races is a task to be accomplished by coming generations.**

But mankind will certainly fail to solve this problem if it does not entirely discard etatism. Let us not forget that the actual menace to our civilization does not originate from a conflict between the white and colored races but from conflicts among the various peoples of Europe and of European ancestry. (Mises 2011/1944, 121, underlining and boldface added)

The sentences in boldface, coming right after what Slobodian used, directly undermine Slobodian’s claim that “Mises partially legitimised closed borders for non-white migrants as a near-permanent feature of the world order.” Far from extending near-permanent legitimacy to the racist positions he described, *Mises explicitly called on coming generations to solve the problem of racism*, and he pointed out that—contrary to the race theorists’ recurring predictions of a civilizational war between people of different skin colors—the then-present World War II originated from a white European state invading other white European states.

Another example of a truncated quotation appears in Slobodian’s *Cultural Politics* article, though it is more difficult to detect since it involves his own translations of excerpts from a German-language book by Mises published in 1940. Slobodian’s text is presented next to our translation of the original passage.

Slobodian, <i>Cultural Politics</i>	Mises, <i>Nationalökonomie</i>
<p>In the original German-language version of the book that would later become <i>Human Action</i>, written during his time at the Graduate Institute for International Studies in Geneva, Mises (1940: 157) granted even more ground to race science, writing that “<u>we may take as given that the racial element plays a role among the factors that form the personality and, with it, our values and understanding.</u>” What he objected to was not the possible truth content of race theory but its misuse. “In the doctrine of National Socialism and its derivative teachings in Italian fascism,” he wrote, “there is an unbridgeable gap between the statements of the founders of racial biology and their application to propaganda and use for practical policies.” Yet the fascist politicization of race theory should not discredit it permanently. “Because the keywords of race theory are used to justify measures with which it has nothing to do,” he wrote, “does not free scientific thought from the responsibility to think through to the end the problem of human races (<i>Menschenrassen</i>) in its praxeological significance” (158). (Slobodian 2019a, 380, underlining added)</p>	<p><u>We may take as given that the racial element plays a role among the factors that form the personality and, with it, our values and understanding</u>, i.e., everything with which a man is born, his physical endowment, the hereditary qualities derived from his ancestors. But in the present state of our knowledge, we know nothing about the connection between the physical and the mind, and therefore cannot make any statement as to whether and in what way the physical is capable of influencing <i>Verstehen</i>. Some have attempted to assign certain value judgments (types of <i>Verstehen</i>, <i>Verstehen</i> types) to specific peoples; these attempts failed because it is easy to prove that every attempt to group people according to types of <i>Verstehen</i> thwarts the classification according to ethnicity. (Mises 1940, 157, our translation, underlining added)</p>

Slobodian there portrays Mises as making only a narrowly qualified objection to eugenic theory based on its “misuse” by the Nazis. Such “misuse,” Slobodian claims, did “not discredit it permanently” in Mises’s eyes. To reach this conclusion,

however, Slobodian had to ignore the context of the passages he quotes. The omissions disconfirm his depiction, including the passage where Mises states that attempts to validate eugenic theory had failed on account of inherent impediments that the concept of *Verstehen* creates for grouping people according to their physical or ethnic characteristics.³

The severity of Slobodian's error is compounded by additional liberties that he takes with Mises's text. Slobodian's assertion that "fascist politicization of race theory should not discredit it permanently [in Mises's eyes]" is premised upon a rearrangement of Mises's original sentence structure and presentation. In *National-ökonomie*, Mises concludes his statement about the praxeological implications of race theory with the quoted passage: "Because the keywords of race theory are used to justify measures with which it has nothing to do, does not free scientific thought from the responsibility to think through to the end the problem of human races (*Menschenrassen*) in its praxeological significance." Slobodian depicts this conclusion as following from the passage about Fascist and Nazi deployments of racial eugenics. In reality, Mises's discussion of Nazism is lifted from a subsequent paragraph that specifically attacks Hitler's attempts to classify Jewish people by physical characteristics.

Mises's passage was in fact written to make a very different point—one that likened Marxist ideological theory to its "its racial-biological counterpart" in their shared polylogism and oppositional stance to his own concept of praxeology. Per Mises, race theory falsely asserted "that there are instead just as many disciplines of logic, mathematics, economics, and natural sciences as there are human races" (Mises 1940, 157, our translation). Thus when he calls upon scientists to "think through to the end the problem of human races," it is not intended as an "opening" to revive the race theorists' ideas in the wake of the Nazis, but rather as a plea for a robust and oppositional challenge to those ideas.

Similar quotation practices recur time and again in Slobodian's argument. In another example from the *CEH* article, Slobodian suggests that Mises opposed racial discrimination against groups of white people, as in Nazi anti-Semitism, but drew a line of distinction between anti-white and anti-black discrimination. Slobodian's depiction is presented here along with Mises's original text for comparison.

3. Slobodian shows unawareness of Mises's use of *Verstehen*, employed here in reference to the Weberian sociological concept, and instead translates it literally as the German word for "understanding." The word is more complex in meaning. The Merriam-Webster Dictionary defines it as "an intuitive doctrine or method of interpreting human culture especially in its subjective motivational and valuational aspects through the understanding of symbolic relationships."

Slobodian, <i>CEH</i>	Mises, <i>Omnipotent Government</i>
<p>Yet Mises proved incapable of extending a similar cosmopolitan attitude to populations of colour. Even as he argued emphatically that <u>‘there are today no pure stocks within the class or race of white-skinned people,’</u> he did so by pointing out the difference with black populations. <u>‘Negroes and whites differ in racial—i.e., bodily—features’</u>, he wrote, <u>‘but it is impossible to tell a Jewish German from a non-Jewish one by any racial characteristic.’</u> Mises’s rejection of anti-Semitism was premised on an affirmation of white–black race difference. (Slobodian 2019b, 155, underlining added)</p>	<p>For more than a hundred years anthropologists have studied the bodily features of various races. The undisputed outcome of these scientific investigations is that the peoples of white skin, Europeans and non-European descendants of emigrated European ancestors, represent a mixture of various bodily characteristics. Men have tried to explain this fact as the result of intermarriage between the members of pure primitive stocks. Whatever the truth of this, it is certain that <u>there are today no pure stocks within the class or race of white-skinned people.</u></p> <p>Further efforts have been made to coordinate certain bodily features—racial characteristics—with certain mental and moral characteristics. All these endeavors have also failed.</p> <p>Finally people have tried, especially in Germany, to discover the physical characteristics of an alleged Jewish or Semitic race as distinguished from the characteristics of European non-Jews. These quests, too, have failed completely. (Mises 2011/1944, 192, underlining added)</p>

Note how Slobodian omits text from the immediate surrounding of the excerpted passage, thereby altering its meaning. He then pairs the first excerpted passage with a second passage contrasting Nazi persecution of Jewish persons with African-Americans in the United States. One would not know it from Slobodian’s presentation, but that second passage comes several paragraphs after the first—and the intervening paragraphs are key.

In those intervening paragraphs Mises takes up the question of how Nazi racial ideology differed from other forms of discrimination. One distinguishing factor of the Nazis was their attempt to impose discrimination on the grounds of establishing a genealogical link to Judaism. Within the paragraphs Slobodian excludes are the following sentences, which reveal Mises’s true position:

The Nazis have chosen a different way. They say, it is true, that they want to discriminate not against people professing the Jewish religion but against people belonging to the Jewish race. Yet they define the members of the Jewish race as people professing the Jewish religion or descended from people professing the Jewish religion. The characteristic legal feature of the Jewish race is, in the so-called racial legislation of Nuremberg, the membership of the individual concerned or of his ancestors in the religious community of

Judaism. (Mises 2011/1944, 193)

Mises then offers a point of contrast for how discriminatory institutions operate. He uses the example of Jim Crow in the United States, where discrimination was based primarily on appearances and skin color. It is from the midst of this example that Slobodian extracts his second passage:

If Americans want to discriminate against Negroes, they do not go to the archives in order to study the racial affiliation of the people concerned; they search the individual's body for traces of Negro descent. Negroes and whites differ in racial—i.e., bodily—features; but it is impossible to tell a Jewish German from a non-Jewish one by any racial characteristic. (Mises 2011/1944, 194, underlining added)

Contra Slobodian, there is no evidence that Mises condones discrimination against African-Americans. The evidence shows quite the opposite: he condemns discrimination itself. Neither does his objection to anti-Semitism depend upon “an affirmation of white–black race difference,” as Slobodian claims. Rather, he is simply analyzing how the Nazi racial laws turned to characteristics other than skin color as a means of effecting their persecution of Jewish people. Since skin color was not an available means for differentiation, Nazis instead traced the genealogies of their victims. As Mises pointed out in the very next paragraph, “The Nazis have claimed that they were fighting the decisive war between the Nordic master race and the human underdogs,” yet the very existence of such a “master race” was itself a myth that the Nazis selectively invoked to rationalize their persecution of Jewish people and other disliked groups (Mises 2011/1944, 194).

The quotation practices illustrated here do not appear to be unique to Slobodian's two articles on Mises. In related works on the history of “neo-liberalism,” Slobodian's arguments frequently rely for support on fragmentary quotations that lead the trusting reader into serious misconceptions about original texts. In 2015, Slobodian delivered a lecture about his then-forthcoming book *Globalists* (Slobodian 2018a) in which he misrepresents the meaning of a sentence grabbed from Mises's *Liberalism*.

Slobodian, 2015 lecture	Mises, <i>Liberalism</i>
<p>For Mises, the demands of the world economy trumped all other political claims. In discussing colonialism, for example, he remarked that “<u>no chapter of history is steeped further in blood than the history of colonialism</u>” but still insisted that keeping the colonies was the first priority once Europe became dependent on the empire for raw materials. Self-determination might be thinkable but only under the control of a muscular super-state that could ensure the continuation of free trade. (Slobodian 2015, 9, underlining added)</p>	<p><u>No chapter of history is steeped further in blood than the history of colonialism.</u> Blood was shed uselessly and senselessly. Flourishing lands were laid waste; whole peoples destroyed and exterminated. All this can in no way be extenuated or justified. The dominion of Europeans in Africa and in important parts of Asia is absolute. It stands in the sharpest contrast to all the principles of liberalism and democracy, and there can be no doubt that we must strive for its abolition. The only question is how the elimination of this intolerable condition can be accomplished in the least harmful way possible. (Mises 2005/1927, 93–94, underlining added)</p>

Note how Slobodian argues that Mises believed that “keeping the colonies” was a necessary condition for the survival of the European economy. Yet, while the snippet seems to fit into Slobodian’s narrative, in Mises’s original text it is part of a denunciation of colonialism—within three sentences of the one extracted by Slobodian we find Mises saying that European dominion in Africa and Asia is “intolerable” and that his readers “must strive for its abolition.” Reading on, we see that to Mises self-determination for the colonies was not merely “thinkable” and not “only” if limited by a larger authority as Slobodian (2015, 9) has it, but imperative and ultimately unconditional:

The application of these principles would mean, in effect, that all the overseas territories of the European countries would at first be turned into mandates of the League [of Nations]. But even this would have to be viewed only as a transitional stage. The final goal must continue to be the complete liberation of the colonies from the despotic rule under which they live today. (Mises 2005/1927, 97)

In another section of the same talk, Slobodian (2015) gleans certain quotations from Mises’s *Liberalism* to indict Mises for being in support of state coercion. He again uses selectively curated excerpts of Mises’s original text to make this point. As a comparison of the passages reveals, Slobodian’s excerpts omit surrounding passages that undermine his interpretation, and they are presented as if Mises intended them to be connected points.

Slobodian, 2015 lecture	Mises, <i>Liberalism</i>
<p>After 1918, Mises returned to the Chamber of Commerce. The building remained adorned with the double-headed eagle of the Dual Monarchy holding the bundled rods of the Roman fasces (slide). Mises may have been thinking of this seal when he wrote admiringly of the Roman symbol in 1927, that it did not obscure “<u>the truth of the matter</u>” that “<u>the state is the apparatus of compulsion and coercion.</u>” Contrary to his embrace by latter-day libertarians, Mises did not mean this in a negative sense. The application of state power was often necessary. (Slobodian 2015, 5, underlining added)</p>	<p><u>The state is the apparatus of compulsion and coercion.</u> This holds not only for the “night-watchman” state, but just as much for every other, and most of all for the socialist state. Everything that the state is capable of doing it does by compulsion and the application of force. To suppress conduct dangerous to the existence of the social order is the sum and substance of state activity; to this is added, in a socialist community, control over the means of production.</p> <p>The sober logic of the Romans expressed this fact symbolically by adopting the axe and the bundle of rods as the emblem of the state. Abstruse mysticism, calling itself philosophy, has done as much as possible in modern times to obscure <u>the truth of the matter</u>. For Schelling, the state is the direct and visible image of absolute life, a phase in the revelation of the Absolute or World Soul. It exists only for its own sake, and its activity is directed exclusively to the maintenance of both the substance and the form of its existence. For Hegel, Absolute Reason reveals itself in the state, and Objective Spirit realizes itself in it. It is ethical mind developed into an organic reality—reality and the ethical idea as the revealed substantial will intelligible to itself. The epigones of idealist philosophy outdid even their masters in their deification of the state. To be sure, one comes no closer to the truth if, in reaction to these and similar doctrines, one calls the state, with Nietzsche, the coldest of all cold monsters. The state is neither cold nor warm, for it is an abstract concept in whose name living men—the organs of the state, the government—act. All state activity is human action, an evil inflicted by men on men. The goal—the preservation of society—justifies the action of the organs of the state, but the evils inflicted are not felt as any less evil by those who suffer under them.</p> <p>The evil that a man inflicts on his fellow man injures both—not only the one to whom it is done, but also the one who does it. Nothing corrupts a man so much as being an arm of the law and making men suffer. The lot of the subject is anxiety, a spirit of servility and fawning adulation; but the pharisaical self-righteousness, conceit, and arrogance of the master are no better. (Mises 2005/1927, 34–35, underling added)</p>

By reading the passage in full, we find that once again Slobodian is unjust to Mises. The passage reveals that Mises’s opinion of the state was indeed negative: “All state activity is human action, an evil inflicted by men on men.” Mises dislikes the state for the very reason that Slobodian depicts as something affirmed by Mises: for its use of “compulsion and coercion.”

Mises is not the only free-market economist whose arguments on race and colonialism have been misinterpreted by Slobodian in this fashion. In *Globalists*, Slobodian criticizes William H. Hutt’s alleged position on race in *The Economics of the Colour Bar*. He presents Hutt’s words to suggest that he disapproved of the idea that black Africans, who comprised the majority of the South African population, would have majority rule in a fully representative parliament. Hutt’s original text,

however, shows that he was taking issue with the abuses of black Africans by the white parliamentary majorities that existed in 1964, even as whites constituted a minority of the overall South African population.

Slobodian, <i>Globalists</i>	Hutt, <i>Economics of the Colour Bar</i>
<p>The problem of democracy was the central theme of Hutt's writings. What he described as "<u>the most vital point of my whole thesis</u>" in <i>The Economics of the Colour Bar</i> was not an economic but a political argument: a warning about the "tyranny of parliamentary majorities" under systems of universal suffrage. The fact that blacks were the majority population in South Africa made the situation exceptionally perilous, in his view. The apparent solution of the universal suffrage would only "mean the transfer of power to a new political majority, with no constitutional limitations to prevent retaliatory abuse." Hutt mentioned that he sympathized with those who feared "black supremacy (a mere turning of the tables)." Hutt explained on these ideas in a piece solicited for the <i>National Review</i>, likely by Buckley himself on his state-funded trip to South Africa in the winter of 1962–1963. (Slobodian 2018a, 173–174, underlining added)</p>	<p>It may be thought that residential segregation and separate facilities in respect of universities, libraries, learned societies, beaches, public gardens, the Post Office, theatres, sport, transport, etc., are hardly examples of injustices of <i>economic</i> origin; for in these cases any discriminatory treatment appears to originate in political or socio-psychological factors rather than economic causes. But it is an important part of my case that all forms of what are felt to be injustices—economic or otherwise—can be seen to have their ultimate source in an abuse of parliamentary majorities. For majorities, even if based on 'one man, one vote', may withdraw from minorities or from classes not represented the opportunity to win equal dignity and respect as well as the right to economic opportunities.</p> <p>This is, I feel, <u>the most vital point of my whole thesis</u>. Unless parliaments are restrained by iron-clad constitutional entrenchments, political majorities, as defined and distorted by electoral laws and voting procedures, will almost always be tempted to exploit their power tyrannically, that is, without true regard for the rights or feelings of political minorities or for those who lack effective political representation. The majorities in the white constituencies, to which the authors of the Group Areas programme have successfully appealed, have <i>wanted</i> to humiliate the non-Whites. Let us be under no illusion on that point. This is partly because one of the parties found it profitable deliberately to foster racial pride, racial prejudices, racial fears and racial hatreds. (Hutt 2014/1964, 115–116)</p>

Hutt was making a subtler point about the need for "constitutional entrenchments" that would establish and secure a system of rights that a simple majoritarian vote within a parliament could not overturn or violate. The necessity of such a system, in Hutt's mind, was demonstrated by the example of Apartheid, where a white *parliamentary* majority imposed humiliating and discriminatory restrictions on

a black *population* majority.

These examples show a recurring pattern of textual representations that omit important accompanying wordings and passages, and that frequently alter the ordering of the original text’s presentation of its argument. In each case, the omissions and reorderings alter the reader’s understanding of the point being made in the original passage by the free-market economist. It is not insignificant that these quotation practices build toward a particular thesis about the relationship between “neoliberalism” and racial colonialism. In each case, the questionable quotation practices are used to imply that the excerpted author harbored sympathies for racism, imperialism, eugenics, and similar discredited concepts. Slobodian then uses the excerpts to question their reputations as philosophical liberals on matters of race, even though the unedited original texts display clear condemnations of the positions that he assigns to them.

An attempted correction

After seeing these numerous questionable quotation practices in Slobodian’s *CEH* and *Cultural Politics* articles, I attempted to bring them to his attention and ask for a clarification or correction. This occurred in late 2018 after Slobodian, on Twitter, identified Mises as a source for anti-immigration positions as advanced by Hans-Hermann Hoppe.⁴ I suggested that, contrary to Slobodian’s insinuations, Hoppe did not appear to draw any of his immigration claims out of Mises’s texts and certainly not the passages that Slobodian misquoted. Hoppe, in fact, identifies Mises as an expositor of the “classical” economic argument for open immigration—and rejects the same as an antiquated view from an earlier time. As Hoppe writes in a 2001 text, “The problem with the above argument is that it suffers from two interrelated shortcomings which invalidate its unconditional pro-immigration conclusion and/or which render the argument applicable only to a highly unrealistic—long bygone—situation in human history” (Hoppe 2009/2001, 138).

My alternative thesis, which I outlined to Slobodian in December 2018, called attention to Hoppe’s use of discursive philosophical reasoning to build up a theory which he then used in turn to justify state action to exclude immigrants from a country (Brennan 2013). As I noted at the time, the intellectual genesis of Hoppe’s approach came not from Mises but from Hoppe’s own academic training under the German philosophers Juergen Habermas and Karl-Otto Apel. Hoppe wrote his dissertation directly under Habermas, and at least in the earliest iterations

4. Alluding to Hoppe, Slobodian wrote on December 28, 2018 ([link](#)), that his article (later published as Slobodian 2019b) was “framed by ongoing fights between right- and left-libertarians for [Mises’s] legacy.”

of his arguments, specifically claimed to be adapting their discursive analysis in a rightward direction. Murray Rothbard (1990) wrote that “Hoppe is a libertarian extension” of “the Habermas-Apel doctrine.”

Slobodian was skeptical of my counter-thesis and maintained at the time that he had correctly represented Mises’s texts. Slobodian wrote in a 29 December 2018 Twitter message to me: “Far from convinced based on this summary that Habermas is more important for him than Rothbard but would be interested to see this written up with references. You should do it” ([link](#)).⁵ Shortly thereafter, I began drafting a paper that traced the influence of Habermas and Apel on the Austrian economic debates of the 1980s, identifying their influence on Hoppe’s later work. In doing so, I engaged Slobodian’s pair of articles where he proposed Mises as the alternative origin. This afforded an opportunity to explore Slobodian’s abuse of Mises’s text, calling attention to several examples of altered and out-of-context quotations in both articles. I completed a working draft of that paper by the fall of 2019, and I presented an early version of it at an economics conference (Magness 2019a).

After incorporating feedback, on January 8, 2020, I submitted the paper to *Contemporary European History* for consideration. I submitted it to *CEH* because Slobodian’s *CEH* article had appeared in print a few months earlier, and my piece corrected several unambiguous errors in Slobodian’s depiction of Mises’s texts while also presenting an alternative thesis for the events that Slobodian described. Although no journal is obliged to accept every submission it receives, debates of this sort are the essence of scholarly exchange, particularly when they involve substantive corrections to recent material in the same journal.

Instead of considering my piece or responding to the problems it raised with Slobodian’s quotation practices, an unsigned “editor” of *CEH* sent a prejudicial desk rejection in a matter of days. My article, the journal claimed, had “barely any primary source base or methodology” even though it brought several previously unknown sources to light, such as the aforementioned Rothbard material. This was an odd position for them to take seeing as Slobodian’s article had no clear methodology and relied almost entirely on heavily truncated quotations from secondary sources, such as Mises’s published books. Furthermore, the editor objected that I had uploaded an earlier working draft of the paper for the aforementioned conference presentation—an entirely common and accepted academic practice—and claimed that this draft meant that my article had already been “pub-

5. On August 5, 2019, Slobodian derisively referenced the December 2018 conversation in a subsequent tweet ([link](#)), stating “I once suggested to a long-muted troll on here that he write about this as he loved to claim Hoppe was a secret member of the Frankfurt School. He demurred when it became clear he’d actually have to read some critical theory.”

lished” in another journal. I sent the editors of *CEH* the following inquiry on January 25, 2020:

Dear Editors,

Thank you for your response below. While I am disappointed by your decision, I understand that differences of focus and method carry limitations for a venue such as yours.

Since my argument does contain specific factual corrections to claims made in a recently published article in *Contemporary European History*, I am writing out of curiosity to ask what an appropriate corrective for the readers of your venue (e.g. letter, comment, or note) might entail.

Sincerely,

Phil Magness

Even if they declined to run a lengthy response to Slobodian’s argument, the journal should recognize the errors arising from misrepresenting Mises and permit a short note or correction. On January 31, 2020, I received a short reply from one of *CEH*’s editors, Victoria Harris: “Thanks for this. I sent your first email on to the editor responsible for your piece.” I never received another response, despite sending follow-up inquiries.⁶

Meanwhile, scholars relying on Slobodian’s account began to repeat and amplify his characterization of Mises as a racist. Building upon the altered and out-of-context quotations, Lars Cornelissen (2021; 2020) accused Mises of mounting a “racialized attack” on democracy. Robin De Bruin claimed that Slobodian had shown that “many right-wing European founding fathers,” the author’s categorization of Mises and the Austrian school—“basically had the same core beliefs as the Alt-Right supporters” (2021, 9).

At the same time, other scholars began to address the problems with Slobodian’s textual misrepresentations, including the passages I documented above. David Gordon (2021) posted one such list, comparing Slobodian’s article with Mises’s texts and finding clear evidence of edits that altered the meaning of the original passages. With this list of examples in hand and having never received a response from *CEH*, I wrote the journal again on March 15, 2021, to inquire about an appropriate correction to Slobodian’s article. I received the following answer from the journal on March 24:

Dear Dr Magness,

Our publication process involves a rigorous peer-review system of multiple reviewers for each piece, a process which ensures the high-quality

6. After over a year of deflection, *CEH*’s editors eventually claimed that their unresponsiveness to email was caused by Covid-19.

of our research articles. Scholarship, of course, evolves, and interpretations differ. This is why we host roundtables to discuss different viewpoints on broad topics (the Spanish Civil War, most recently).

However, in this specific case, we are confident about the academic rigour of Slobodian's piece.

We hope you and yours are keeping well in these difficult times.

All Best,

Contemporary European History

The response not only failed to address any of the factual problems with Slobodian's quotation practices, it also unintentionally related to an unforeseen twist to our now year-long exchange.

As I learned shortly thereafter, the problems I pointed out about the Mises quotation edits were not unknown to the editors of *CEH* when I brought the problems to their attention. These issues had in fact been flagged over a year prior during the blinded peer review of Slobodian's submission to *CEH*. The editors of *CEH* nevertheless accepted Slobodian's article and published it without addressing their referee's concerns about the quotations. Seeing as those concerns raised a direct challenge to Slobodian's thesis, the decision to ignore the referee carried substantial implications for the acceptance and publication of the article.

Approving misrepresented quotations

When Slobodian first submitted his article "Perfect Capitalism, Imperfect Humans" to *CEH*, it contained a very different thesis than the final published version. Echoing his 2018 book *Globalists*, Slobodian's approach in the original draft exhibited a generally adversarial stance toward Mises's economics, but it did not charge Mises with harboring a qualified tolerance for racism and imperialism (Magness 2019b). Instead it argued that Mises's writings in the interwar era revealed his attempts to navigate the tumultuous political realities of a post-Habsburg central Europe, as those realities challenged his commitments to a free-market, free-immigration liberal philosophy. Slobodian argued that Mises looked back to the prewar Austro-Hungarian Empire as an institutional framework for a liberal political order in matters of immigration. He also originally argued that Mises's interwar writings maintained that normative liberal preference but found it increasingly intruded upon by post-World War I restrictions on movement and, eventually, the rise of illiberal totalitarian regimes on the European continent in the prelude to World War II.

During the first round of peer review—as I learned from an inquiry with *CEH*'s publisher Cambridge University Press—Slobodian's original article was

sent to four referees.⁷ Two referees recommended minor revisions to the piece, a third recommended major revisions, and a fourth recommended rejection (albeit on the grounds that the article was a repetition of Slobodian’s book *Globalists*, not realizing it was the same author due to blind review). The journal’s editors invited Slobodian to revise and resubmit his article, attending to the three referees who suggested revisions.

By all appearances, Slobodian then made substantial changes to his article and thesis. Most of the problematic Mises quotations were added to the article in preparing the revision for resubmission. The argument was reworked to include the new contention that Mises shifted to an ambiguous position on racism, imperialism, and related matters during and after World War II. He worked this new thesis into the article’s conclusion, even charging that Mises’s rejection of Nazi anti-Semitism “was premised on an affirmation of white–black race difference,” as if to imply that the economist would not extend similar objections to anti-black racism (Slobodian 2019b, 155).

The editors of *CEH* then sent Slobodian’s substantially revised article to two of the original four referees. At some point during this second round, the basic mechanisms of peer review collapsed. One of the referees recommended accepting the revised paper as-is. But the other referee noticed the substantial additions to the article, including its new and altered thesis. When investigating those additions, the referee observed the aforementioned pattern of altered and misrepresented quotations.

I first learned directly of this chain of events in early 2021 after expressing my displeasure with *CEH*’s handling and rejection of my own manuscript challenging Slobodian’s thesis and calling attention to the quotation issues. My comments were shared by chance with another scholar who works on the history of economic thought and, who I soon learned, happened to be one of the referees on Slobodian’s original piece. I learned from this scholar that similar concerns about the misquotation of Mises were brought to the attention of *CEH*’s editors during peer review, only to be ignored.

I quote here with permission directly from the second-round referee report that flagged Slobodian’s changes and alerted the editors of *CEH* to a substantial problem with the revised and ultimately published version of the article:

Having read the revised article and then reread the original version, I am not sure what happened here. It seems that the author has changed his mind on the fundamental meaning of Mises’ work, which he now expresses in the

7. The details of the referee process were confirmed by the author in correspondence with Helen Hardy of Cambridge University Press on August 23, 2021.

introduction and conclusion, while much of the body of the article is left intact. This leads to an underdeveloped, but very serious accusation that Mises is a racist, as well as a stretched attempt to make a link with contemporary political developments. This makes the article substantially weaker than it originally was, and makes one unsure about the central claim of the article.⁸

The referee further noted that Slobodian’s revisions took an initially impartial and scholarly argument and turned it “openly political.” The referee warned the editors that “the article is now weaker than in the original version, and based on dubious historical scholarship: quoting out of context, partial reading of the relevant material, and ascribing views to others that they did not hold.”

The referee specifically called attention to the “out-of-context quote from Mises’ *Omnipotent Government*” and alerted the editors to the surrounding passages, which stated Mises’s objections to the racial dimensions of fascism. Anticipating my own later criticisms of the published article, the referee pointed out that Slobodian was conflating Mises’s descriptions of his adversaries’ positions (i.e., the Nazis) with Mises’s own position, omitting adjacent paragraphs where Mises clearly laid out his own position. The report continues:

In the book ‘Omnipotent Government’ this passage comes after a paragraph that presents the views of his opponents, with which he clearly disagrees, for he sees political-economy causes where others see racial causes. But even from this passage itself it is perfectly clear that he is making an argument about political feasibility at that time (1944!), not a principled argument. And Mises goes on to conclude that the wars of the past years were not at all between races, but instead were conflicts within Europe and Asia, caused by ‘etatism’. Yet the author feels confident enough to conclude that Mises: “By the 1940s, Mises partially legitimized closed borders for non-white migrants as a near permanent feature of the world order.” No, that is what others sought to do, and what Mises sought to overcome, given the political reality of his time.

In the concluding remarks, the referee writes: “I strongly suggest that the article is not published in its revised form.”

A failure of editorial integrity

The editors of *CEH* declined to take any action to address the second-round referee’s warnings. They do not appear to have even asked Slobodian to respond to the issues raised about altered and out-of-context passages from Mises’s works.

8. The original referee report is quoted with permission from its author.

Then, oddly enough, Slobodian himself was named as a new co-editor of *CEH* in the spring of 2020—shortly after the start of my own exchanges with the journal. Spanning the course of a year, each of my attempts to bring specific instances of erroneous quotations to their attention were either brushed aside with a generic form letter or ignored entirely.

After the March 2021 exchange, I contacted Cambridge University Press’s publishing ethics committee about the irregularities and the outcome. Despite initial assurances that they would investigate the matter, I received little more than a formulaic response. The committee declined to act, or even to receive the evidence that I offered to provide them. After unspecified consultation with the editors of *CEH*, they simply stated that they “view these allegations as rooted in scholarly disagreement rather than in problems of research integrity.” After I registered my concerns about the lack of transparency in the process, they simply responded that the editors of *CEH* “reached an informed decision that the article had met the standards and criteria of *CEH* and there was no need for further revision.” When pressed on the specific matter of quotation editing, the committee and/or editors responded that the “quotations chosen do not seem to have been distorted in the article, and the reading of Mises remains a matter of interpretation.”

When reading that conclusion, it helps to revisit an example that was brought to their attention. Here is what Mises wrote in 1919:

But all other pages of world history were also written in blood, and nothing is more stupid than efforts to justify today’s imperialism, with all of its brutalities, by reference to atrocities of generations long since gone. (Mises 2006/1919, 63)

Here is how Slobodian depicted that same passage:

[Mises] wrote in a book published the year after the First World War, the net gain made it worthwhile; in the end, ‘all other pages of world history were also written in blood’. Violence in the project of expanding the space of foreign investment, wage labour and commercial exchange was not only acceptable, it was necessary. (Slobodian 2019b, 148)

The changes may be called “a matter of interpretation,” but Slobodian’s interpretation is obviously irresponsible. It also services his revised thesis, which seeks to impart a tolerance or even sympathy for racial colonialism into Mises’s philosophical worldview. The other documented textual issues perform a similar function.

Given that these issues were flagged by an anonymous referee during peer review, independently raised in my own attempts to submit a rebuttal to *CEH*, and

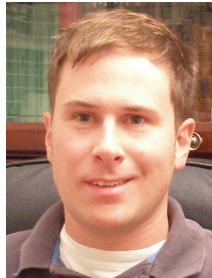
subsequently presented to *CEP's* editors as well as Cambridge University Press, it is not possible for those involved in the journal to claim ignorance of the problems that they present for Slobodian's thesis. The journal's violation of basic standards of truthfulness appears to be willful.

References

- Brennan, Jason.** 2013. Hoppe's Argumentation Ethics Argument Refuted in Under 60 Seconds. BleedingHeartLibertarians.com, December 12. [Link](#)
- Cornelissen, Lars.** 2020. Neoliberalism and the Racialized Critique of Democracy. *Constellations* 27(3): 1–13. [Link](#)
- Cornelissen, Lars.** 2021. Savage Economics: Race, Futurity, and Civilizational Hierarchy in Early Austrian Neoliberalism. *Global Perspectives* (University of California Press, Berkeley, Cal.) 2(1): 23667. [Link](#)
- De Bruin, Robin.** 2021. European Union as a Road to Serfdom: The Alt-Right's Inversion of Narratives on European Integration. *Journal of Contemporary European Studies* 30(1): 52–66. [Link](#)
- Ebeling, Richard M.** 2018. Quinn Slobodian and the Academic Attack on Mises and Hayek. October 6. American Institute for Economic Research (Great Barrington, Mass.). [Link](#)
- Gordon, David.** 2021. The Anti-Imperialist Ludwig von Mises. March 21. Mises Institute (Auburn, Ala.). [Link](#)
- Hoppe, Hans-Hermann.** 2009 [2001]. *Democracy—The God That Failed: The Economics and Politics of Monarchy, Democracy and Natural Order*. New Brunswick, N.J.: Transaction Publishers.
- Hutt, William H.** 2014 [1964]. *The Economics of the Colour Bar*. Salt Lake City, UT: Large Print Liberty.
- Magness, Phillip W.** 2019a. Racial Determinism and Immigration in the Works of Ludwig von Mises: A Critique of Slobodian's Alt-Right Thesis. Working paper, November 20. [Link](#)
- Magness, Phillip W.** 2019b. What Does 'Neoliberalism' Really Mean? *Reason*, January. [Link](#)
- Magness, Phillip W.** 2021. How Activist Academia Destroyed Scholarly Peer Review. September 4. American Institute for Economic Research (Great Barrington, Mass.). [Link](#)
- Magness, Phillip W.** 2022 (forthcoming). Racial Determinism and Immigration in the Works of Ludwig von Mises: A Critique of Slobodian's Alt-Right Thesis. *Journal of Private Enterprise*.
- Mises, Ludwig von.** 1940. *Nationalökonomie: Theorie des Handelns und Wirtschaftens*. Geneva: Editions Union.
- Mises, Ludwig von.** 2005 [1927]. *Liberalism*. Indianapolis: Liberty Fund.
- Mises, Ludwig von.** 2006 [1919]. *Nation, State, and Economy*. Indianapolis: Liberty Fund.

- Mises, Ludwig von.** 2011 [1944]. *Omnipotent Government*. Indianapolis: Liberty Fund.
- Rothbard, Murray N.** 1990. Hoppephobia. *Liberty* (Port Townsend, Wash.) 3(4): 11–12.
- Slobodian, Quinn.** 2015. Let the World Economy Rule: How Neoliberals Imagined the World After Empire. Presented at the Institute for New Economic Thinking Young Scholars Initiative “The Nation-State and the World Economy Between Two Eras of Globalization, 1913–1975” conference (Scuola Superiore Sant’Anna, Pisa, Italy), July 14. [Link](#)
- Slobodian, Quinn.** 2018a. *Globalists*. Cambridge, Mass.: Harvard University Press.
- Slobodian, Quinn.** 2018b. Neoliberalism’s Populist Bastards: A New Political Divide Between National Economies. *Public Seminar* (New School for Social Research, New York), February 15. [Link](#)
- Slobodian, Quinn.** 2019a. Anti-’68ers and the Racist-Libertarian Alliance: How a Schism Among Austrian School Neoliberals Helped Spawn the Alt Right. *Cultural Politics* 15(3): 372–386.
- Slobodian, Quinn.** 2019b. Perfect Capitalism, Imperfect Humans: Race, Migration and the Limits of Ludwig von Mises’s Globalism. *Contemporary European History* 28(2): 143–155.

About the Authors



Phil Magness is a Senior Research Fellow and interim Director of Research and Education at the American Institute for Economic Research. He holds a Ph.D. and MPP from George Mason University. His research focuses on the economic history of the United States, with an emphasis on the 19th Century. His email address is phil.magness@aier.org.



Amelia Janaskie is a Research Associate at the American Institute for Economic Research. She graduated from the College of Charleston Honors College in May 2020 with a B.S. in Economics and minor in English. Her email address is amelia.janaskie@aier.org.

[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6046>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 85–108

Long-Run Determinants of Economic Growth: Putterman and Weil Revisited

Jason Briggeman¹

[LINK TO ABSTRACT](#)

Some researchers have tried to investigate whether there is a statistical relationship between people’s ancestries and the economic performance of their countries. Genealogical research mostly ceases to be reliable past 200 years ago, but these researchers who surmise that economic development is partially owed to ‘deep roots’ tend to presume that today’s ethnic groups correspond to the geographical location of ancestors.

Louis Putterman and David Weil (hereafter PW) published in 2010 a “World Migration Matrix” data set in which for the most part they infer from a person’s year-2000 ethnicity the year-1500 location of the person’s ancestors. PW (2010) then treat the forms of government that prevailed between the years 1 CE and 1500 within the boundaries of a year-2000 country as a personal characteristic that was passed from the people who lived within those boundaries in the year 1500 to their descendants in the year 2000. PW announced in the *Quarterly Journal of Economics* a finding that this supposed characteristic has a positive relationship with current income, sharing regression results indicating a strong linear relationship for the year 2000: countries whose populations were low in “ancestry-adjusted state history” purportedly had lower GDP per capita than did countries whose populations were high in the characteristic.

PW (2010) is widely cited and the “World Migration Matrix” has been used

1. Austin Community College, Austin, TX 78752. Errors are mine. I thank Nathaniel Bechhofer, Adam Gurri, and Mark Koyama for their kind help.

in other research. The paper is acknowledged as important among those papers that offer similar claims about possible effects of ‘deep roots.’ Its publication in a top-five economics journal has no doubt contributed to some belief in its central finding.

Less known is an article coauthored by Putterman that was published years later in the *Journal of Economic Growth*. Oana Borcan, Ola Olsson, and Putterman (hereafter BOP) conducted a similar analysis, but treated the forms of government that prevailed between the years 3500 BCE and 1500 CE as the personal characteristic that was passed down from year-1500 persons to year-2000 persons. BOP announced a finding that once we assume 5,000 years of history shaped our year-1500 ancestors rather than merely 1,500, “the current level of economic development across countries has a hump-shaped relationship with accumulated state history” (2018, 1).

Our situation then seems to be that the reader of PW (2010) alone would presume that ancestry-adjusted state history has a linear relationship with current development, and the reader of BOP (2018) would have the more nuanced understanding that there is a linear relationship when only 1,500 years of history is used while there is a hump-shaped relationship when 5,000 years of history is used.

Neither is true. I show here that *there is a hump-shaped relationship when only 1,500 years of ancestry-adjusted state history is used*. Because of data errors, this hump-shaped relationship would not have been found by PW (2010) had they looked for it, which they did not; meanwhile, BOP (2018) only showed that there is not a hump-shaped relationship when looking at 1,500 years of state history *unadjusted for ancestry*; BOP neglected to provide regression results using the crucial ancestry adjustment. Furthermore, I show here, at length and in various ways, that *the linear relationship PW (2010) claimed to find is much less substantial than their presentation indicates*.

The findings here are important because in so many ways the hump-shaped relationship is less compelling, its interpretation is less clear, and its implications less obvious than those of a linear relationship. When he launched his research into ‘deep roots,’ Putterman hypothesized a linear effect.² I argue here that the PW (2010) evidence for this hypothesis is not strong. BOP (2018) characterize themselves as putting forward a hypothesis of a hump-shaped effect and then

2. “An index of state antiquity was developed by Brown University Professor of Economics Louis Putterman and then Brown University undergraduate Valerie Bockstette circa 1999–2000 to test the proposition that present-day countries that had been the site of nation-states, kingdoms or empires over longer spans of history have achieved more rapid economic development in recent decades. This proposition or conjecture was suggested to us by the observation that countries in East Asia, with long histories of nationhood, have done much better economically in the late 20th century than have countries in sub-Saharan Africa on which the nation-state was imposed by 19th century colonization” (Putterman 2017, 1).

confirming that hypothesis.³ I think we should disregard BOP's claim that they tested "a theoretical framework" with data.

Literature on persistence

Morgan Kelly (2019; 2020) assesses studies of "persistence" including PW (2010). Kelly notes the often very large regression coefficients found in such studies:

A substantial literature on deep origins or persistence finds that many modern outcomes such as income or social attitudes strongly reflect the characteristics of the same places in the more or less distant past, often centuries or millennia previously. ...

Naturally, such findings are open to various charges of *p* hacking, of publication bias, of answers in search of questions, of scepticism about mono-causal and largely atheoretical explanations of complex phenomena, about the mechanisms driving persistence, and so on. However, all of these objections crumble into irrelevance in the face of one blunt fact: the unusual explanatory power of these persistence variables. While a judicious choice of variables or time periods might coax a *t* statistic past 1.96, there would appear to be no way that the *t* statistics of three, four, or even larger that appear routinely in this literature could be the result of massaging one's regressions, no matter how assiduously. (Kelly 2020, 2)

Kelly proceeds, however, to demonstrate that the inevitable presence of spatial autocorrelation could have caused many of these very high *t*-statistics, and thus the effect of persistence variables is likely greatly overstated throughout the literature. Kelly (2019) carried out simulations of PW (2010) to find whether ancestry-adjusted spatial noise would 'explain' year-2000 income. In Kelly's preferred set of simulations, not only was the ancestry-adjusted noise significant at the .001 level in over half of the simulations, but the noise outperformed the actual ancestry-adjusted state history variable 35 percent of the time (Kelly 2019, 21, 35).

I find Kelly's efforts sufficient to cast some doubt on 'persistence' results generally and those of PW in particular. But I also would deny Kelly's notion that the "various charges" that economists normally raise against weak empirical

3. "We outline a theoretical framework where accumulated state experience increases aggregate productivity in individual countries but where newer or relatively inexperienced states can reach a higher productivity maximum by learning from the experience of older states. The predicted pattern of comparative development is tested in an empirical analysis where we introduce our extended state history variable" (BOP 2018, 1).

research must “crumble into irrelevance” whenever one beholds a t-statistic of “three, four, or even larger.” After all, Kelly himself first calls for the use of...geographic control variables (2020, 3ff.). Yes, innocent misspecification can result in the generation of huge t-statistics, and run-of-the-mill “objections” regarding omitted variable bias (yawn!) therefore can be sufficient to undermine findings that rest upon those whopper t-stats.

And PW (2010) *did* carry out regressions with geographic controls. Kelly (2019) does show that the *uncontrolled* regression result of PW has a good chance of being simple noise, but Kelly (2020) in general advocates controls of a form similar to those PW use. Since the key controlled regression in PW chalks up a t-statistic of nearly three, quite possibly our post-Kelly view of PW (2010) should be more skeptical than before but not (yet?) to the point of dismissal. And if that’s true of PW (2010), perhaps it is true of the other persistence studies. In short, I think while Kelly has brought the question into serious doubt, there is further work to do before we can feel strongly one way or the other about persistence findings.

The contribution of the current paper then can be considered to be doing that dull work of raising the usual objections to which economists resort when they doubt certain findings. I locate particular problems that undermine particular findings in particular studies, and in that way this study merely takes a place among those studies. I don’t purport here to level the entire field of persistence studies. But I find it intriguing that what I’ve found in replicating and assessing PW (2010) aligns with what Kelly (2019; 2020) has to say about persistence even though most of the work reported on below was done prior to my seeing Kelly’s papers.

The organization of the remainder of the current paper is as follows. Section 3 describes the data errors in PW (2010) that (as will be shown in Section 5) would have prevented them from finding the hump-shaped relationship had they looked for it. Section 4 shows that robustness checks, including ones advocated in other work by Putterman, indicate a much less substantial linear relationship than PW (2010) claimed. Section 5 shows that (with the data corrections presented in Section 3) the hump-shaped relationship is present using 1,500 years of ancestry-adjusted state history as opposed to only with 5,000 years as claimed by BOP (2018). Section 6 assesses the central claim of PW (2010)—that the linear relationship they found is “surprisingly” strong. Section 7 concludes.

Corrections and updates to Putterman and Weil’s data

Putterman and Weil’s data set has substantial errors, including needlessly

missing values that caused many countries to be dropped from regressions. For the key regression with PW's geographic controls presented in Table 1 below, the corrected and updated⁴ data includes 38 countries that were excluded by PW because of missing values for the latitude, climate, landlocked, state history, or GDP variables, and these values simply should not have been missing.⁵ Among the countries erroneously dropped by PW were all of the former Soviet republics except Latvia and Georgia, plus several other countries previously under communist rule, which perhaps accounts for some of the substantive differences in the results. The corrections and updates cause the t-ratio on the ancestry-adjusted state history variable to fall from 5.24 to 5.05 in the simple regression and from 2.93 to 2.75 in the regression with PW's geographic controls. R-squared falls from 0.22 to 0.19 in the simple regression and from 0.59 to 0.50 in the regression with PW's geographic controls.⁶

	B	OX	OY	
1	World Bank C	Absolute	Climate, F-O	
2	Afghanistan			
3	Angola	8,843		1
4	Albania			
5	United Arab E	23.39		0
6	Argentina	36,676		2
7	Armenia			
8	Australia	32,219		2
9	Austria	48,231		3

Figure 1. Albania and Armenia are two of the 38 countries that would have been included in Putterman and Weil's key controlled regression were it not for needlessly missing values, seen here in the original data file shared by Putterman.

4. The 'updates,' as opposed to the 'corrections,' primarily consist of my using Putterman's latest revisions to his state history data (see BOP 2018) in the construction of the ancestry-adjusted state history variable. A second sense in which the data might be considered 'updated' is that since I newly sourced the GDP data for the year 2000 from the World Bank, any improvements that have been made to that data since 2010 would be reflected in my results but not in PW's.

5. Box A1 lists these mistakenly excluded countries (as distinct from countries that are simply absent from PW's data set, some of which are listed in Box A2). The reader may note that Table 1's corrected and updated regression with controls has 37 more cases than does PW's regression, not 38. The discrepancy is because one case included by PW is excluded from my regressions because that country (Syria) has a missing value in my source for GDP data. I could not locate a source with the same year-2000 GDP per capita figures that PW use. The correlation between their dependent variable and mine, for the countries included in PW's Table II column (2) regression, is 0.961 (n=135). The correlation between their dependent variable and Angus Maddison's year-2000 GDP per capita variable, for the same set of countries, is 0.966 (n=132), and that between my dependent variable and Maddison's variable is 0.951 (n=134).

6. It must be noted that Putterman and Weil present two different results for a simple regression of year-2000 GDP per capita on ancestry-adjusted state history. The first is their Table II col. (2), which I replicated and reproduced in my Table 1, with n=136, and those are the results described above in the text that have a t-ratio for ancestry-adjusted state history of 5.24 and an r-squared of 0.22. The second is their Table IV col. (1), which they present alongside their controlled regression results. This second simple regression has n=111, just like their controlled regression does, i.e., it excludes the countries with missing values for the latitude, climate, and landlocked variables. In that second simple regression, which I also exactly replicated but do not reproduce in a table here, the t-ratio for ancestry-adjusted state history is 5.92 and the r-squared is 0.29.

TABLE 1. Putterman and Weil's original OLS results vs. OLS results with corrected and updated data

Dependent variable: ln(GDP per capita 2000)				
	PW Table II col. (2)	PW Table IV Panel A col. (6)	Corrected and updated data	
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history	2.01*** (0.38)	1.24*** (0.42)	2.01*** (0.40)	1.01*** (0.37)
Absolute latitude		0.0337*** (0.0084)		0.0318*** (0.0060)
Landlocked		-0.558*** (0.172)		-0.750*** (0.196)
Eurasia		-0.327 (0.247)		-0.169 (0.218)
Climate		0.235* (0.121)		0.191* (0.113)
constant	7.61*** (0.17)	6.99*** (0.20)	7.92*** (0.17)	7.47*** (0.20)
N	136	111	148	148
R squared	0.219	0.593	0.194	0.496
<i>Notes:</i> Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.				

Robustness checks of the Putterman and Weil findings

Today's division of the world into countries, with regard both to borders and to the sheer number of countries, is endogenous to history, 'deep' or otherwise, and can have potentially profound effects on regression results when country is used as the unit of observation, as BOP (2018, 3) note. Just to illustrate such effects, I'll use regions—also endogenous and contingent, but a different division of the world—rather than countries as the unit of observation. PW (2010, 1637–1639) define 11 regions of the world and provide some data and analysis by region, but they do not assess whether their main finding holds when using those regions as the unit of observation. Computing GDP per capita and the ancestry-adjusted state history for those regions is straightforward—each is equal to the population-weighted average across countries—so it is easy to show that their finding does not hold up at all. Even in a simple regression the t-ratio is only 0.55, and it is possible to flip the sign of the coefficient by removing just one observation, that for sub-Saharan Africa (Table 2).

Figure 2. Regions scatterplot

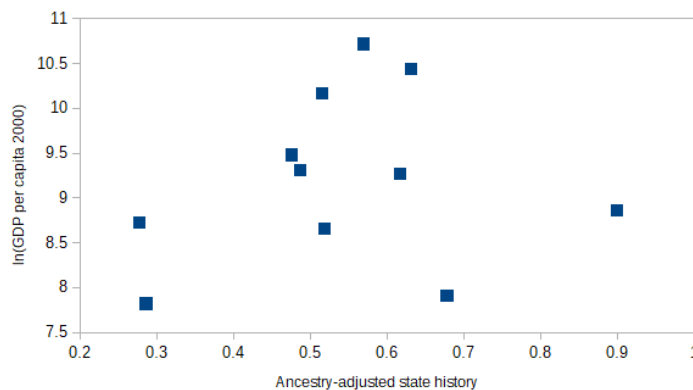


TABLE 2. OLS results with corrected and updated data, for the 11 regions defined by Putterman and Weil

	Dependent variable: ln(GDP per capita 2000)	
	all regions	all regions except sub-Saharan Africa
	(1)	(2)
Ancestry-adjusted state history	0.983 (1.791)	-0.390 (1.926)
constant	8.68*** (1.01)	9.57*** (1.13)
N	11	10
R squared	0.032	0.005
<i>Notes:</i> Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.		

Weighting countries by population is another obvious and easy method to address endogeneity or contingency of country borders, as Bryan Caplan (2016a) notes. Caplan (2016b) finds when that weighting by population is applied using PW’s uncorrected data the coefficient on ancestry-adjusted state history becomes very small. Using the corrected and updated data plus PW’s controls I find that the sign on that coefficient actually *flips*, at the same time that *every one of PW’s geographic controls is statistically significant at the 1-percent level* (Table 3, column 2), in notable contrast to the unweighted regression where only two of the four controls are even significant at the 5-percent level.

One might object to Caplan that year-2000 population is itself endogenous to migration and thus population weighting is not an ideal way to address endogeneity of borders. BOP (2018, 3), in fact, assert that the best unit of observation would be not countries but rather “equal-sized grid cells” of land; they say however that undertaking such an analysis would be very difficult. Here again though there is

TABLE 3. WLS results with corrected and updated data

Dependent variable: ln(GDP per capita 2000)						
	Weighted by population 2000		Weighted by land area 2000		Weighted by arable land area 2000	
	(1)	(2)	(3)	(4)	(5)	(6)
Ancestry-adjusted state history	0.0350 (0.3475)	-0.167 (0.303)	0.921** (0.363)	0.950*** (0.281)	0.0692 (0.3948)	0.617** (0.275)
Absolute latitude		0.0313*** (0.0065)		0.0388*** (0.0047)		0.0438*** (0.0051)
Landlocked		-0.887*** (0.292)		-0.916*** (0.249)		-0.636*** (0.235)
Eurasia		-0.863*** (0.182)		-0.867*** (0.154)		-1.28*** (0.13)
Climate		0.405*** (0.118)		0.126 (0.110)		0.309*** (0.104)
constant	8.62*** (0.24)	7.92*** (0.21)	8.73*** (0.20)	7.78*** (0.22)	9.04*** (0.23)	7.60*** (0.22)
N	148	148	147	147	146	146
R squared	0.000	0.452	0.043	0.518	0.000	0.628
<i>Notes:</i> Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.						

an obvious and easy method to address the concern: we can weight the country observations by land area. When applying land-area weights (Table 3, columns 3–4), the coefficient on the ancestry-adjusted state history variable is much smaller in the simple regression (t-ratio of 2.54, about half as with the unweighted data), but is about the same in the regression with PW’s geographic controls (t-ratio of 3.38).

But it makes little sense to attribute vast spaces of unoccupied, untouched, unbothered-with land to the nominal rule of year-1500 governments. Putterman and his coauthors of a 2014 *American Economic Journal: Macroeconomics* article said so, when they questioned Daron Acemoglu, Simon Johnson, and James A. Robinson (2002) regarding the “appropriateness” of year-1500 population density estimates: “The major conceptual problem is that in most countries, the large majority of the people are found in a small subset of the territory, often including river valleys, coastlines, and fertile plains, and the ratio of largely uninhabited to inhabited territory varies among countries as defined by their modern borders in a fashion that may reflect less on the level of development of the society than on geographic happenstance” (Chanda, Cook, and Putterman 2014, 5 n.4). Therefore a likely better alternative to weighting by land area would be to weight by only *arable* land area, since that represents so much of the space that human societies occupy or live adjacent to.⁷ Using arable land area to weight the data (Table 3, columns 5–6),

7. While there could be some endogeneity concern regarding the particular areas that are cultivated or not in the year 2000 as compared to 1500, I suspect that concern should be judged minor.

the coefficient on ancestry-adjusted state history almost disappears in the simple regression (t-ratio of 0.18) and is significant but smaller in the regression with PW's geographic controls (t-ratio of 2.24); meanwhile, *all four of PW's geographic controls are significant at the 1-percent level* (absolute t-ratios ranging from 2.70 to 9.50).

I now provide two further robustness checks of PW (2010)—checks that are suggested in published work by Putterman and his colleagues, but that PW did not provide themselves. First, Putterman and his coauthor Matthias Cinyabuguma were in a 2011 paper “concerned with what accounts for the poor economic performance of many countries in Africa,” and they argued: “If factors alleged to account for the African difference in global samples perform quite differently within Africa, their relevance to African and other development policy makers would be called into question” (Cinyabuguma and Putterman 2011, 219). As the analysis of regions above suggests, and as Table 4 below confirms, sub-Saharan Africa drives the linear relationship between ancestry-adjusted state history and year-2000 GDP per capita found by PW.⁸ And therefore the Cinyabuguma and Putterman (2011) argument applies: Before we declare the PW results to have relevance for “development policy makers,” we had better examine whether ancestry-adjusted state history is linearly related to year-2000 GDP per capita *within* sub-Saharan Africa.⁹ It isn't, as shown in Table 4.

The second robustness check suggested in Putterman's work that I pursue here is to examine whether there is a relationship between ancestry-adjusted state history and year-1960 GDP per capita, rather than year-2000 GDP per capita. This analysis is readily conducted thanks to Putterman's 2014 *American Economics Journal: Macroeconomics* article coauthored with Areendam Chanda and C. Justin Cook, because it provides a “World Migration Matrix” for the years 1500–1960.¹⁰ The reason to conduct this robustness check is simply that if ‘deep’ history matters a lot, then there seems little reason to think it would affect specifically the year 2000 but not other recent years. And in fact Chanda, Cook, and Putterman's purpose

8. If we were to drop the sub-Saharan African countries from the unweighted simple regression (Table 1, column 3) we would cause the coefficient on ancestry-adjusted state history to drop from 2.01 to 0.83, its t-ratio to fall from 5.05 to 1.66, and R-squared to collapse from 0.19 to 0.03 (as shown in Table 4, column 3).

9. Readers may wonder if there is substantial variation in ancestry-adjusted state history across the sub-Saharan African countries. There is; the mean value is 0.19 and the standard deviation 0.24 (n=44). For all other countries the mean is 0.54 and the standard deviation 0.21 (n=110). One in four countries in sub-Saharan Africa has an ancestry-adjusted state history value above 0.3; over half of all other countries have a value between 0.4 and 0.7.

10. Chanda, Cook, and Putterman perhaps should not have bothered, because their 1500–1960 matrix is hardly different from PW's 1500–2000 matrix. For countries in my data set I find that the correlation between state history as adjusted by the two matrices is 0.9997 (n=154). For more on this see note 19 below.

in creating a 1500–1960 Migration Matrix was exactly that—“to check whether there is anything unusual about the year 1995 as a representation of recent incomes”—and when they found similar results using data for the year 1960 as they had for the year 1995 they took it as a demonstration of robustness (Chanda, Cook, and Putterman 2014, 16).¹¹ For my regressions I used Angus Maddison’s GDP per capita figures ([link](#)) as the dependent variable, both for 1960 and 2000.¹² The results with unweighted data are shown in Table 5, and all told the relationship between ancestry-adjusted state history and GDP per capita in 1960 appears to be about half as strong it was in 2000—while the relationships between PW’s geographic controls and GDP per capita are hardly different. The t-ratio on the ancestry-adjusted state history variable in the simple regression on year-2000 GDP per capita is 6.67, but for that on year-1960 GDP per capita it is only 4.41, and the corresponding t-ratios for the regressions with PW’s geographic controls are 4.25 and 2.10.

When weighting the data by arable land area there is no relationship at all between ancestry-adjusted state history and GDP per capita in 1960, while *every one of PW’s geographic controls is significant at the 1-percent level*, as shown in the first two

11. Putterman (2000), meanwhile, argued the exact opposite: that state antiquity *only* matters for modernization “under late 20th Century conditions” (Putterman 2017), when state antiquity would have helped developing countries cope with the imposition of large-scale institutions by colonizers: “While traditions of bureaucratic and state-level organization pre-date colonial encounters in such countries as India, China, or Japan, such traditions were typically absent or present only on smaller social scales in most parts of Africa” (Putterman 2000, 8). Putterman’s (2000) empirical strategy was to use developing countries’ 1960 levels of “population density, farmers per cultivated hectare, and the prevalence of irrigation” as his measure of their “pre-modern development,” and to see if there was a relationship between that and their GDP growth 1960–1990. He controlled for per capita GDP in 1960, but labeled that as “initial” per capita GDP, not “pre-modern” per capita GDP. The coefficient he found on 1960 GDP per capita was negative (“suggesting that poorer countries grew more rapidly, *ceteris paribus*,” p. 14), while those on 1960 population density, 1960 farmers per cultivated hectare, and 1960 prevalence of irrigation turned up positive (“Each indicator has the predicted sign, and their addition to the equation adds to the explanatory power of the regression,” p. 16). He acknowledged that these results are, to say the least, open to interpretation: “A more general problem with these variables is that while their use to proxy for levels of pre-modern development is consistent with the frameworks of Boserup and other evolutionary theorists, it is not yet possible to rule out alternative interpretations of the correlations between them and economic growth that are reported below. Thus, agricultural intensification could be argued to be conducive to later economic development simply because it generates reserves of labour and other resources that can be drawn upon for the industrialisation process. Population density may encourage growth due to scale economies that bear no relation to the ‘broad human capital’ conception offered above. With a variable set too generic to exclude alternative hypotheses, only further research can determine whether the arguments of this article in fact account for the phenomena observed” (Putterman 2000, 13).

12. Maddison’s year-2000 GDP per capita variable is for whatever reason more favorable to the PW thesis: There is, both with unweighted data and when weighting by arable land area, a stronger relationship when regressing ancestry-adjusted state history on Maddison’s year-2000 figure than when regressing it on my year-2000 figure from the World Bank.

columns of Table 6. The relationship for ancestry-adjusted state history remains intact for the year 2000, as shown in the third and fourth columns. So there may be something substantially unusual about the year 2000 as a representation of recent incomes.

TABLE 4. OLS results with corrected and updated data, for sub-Saharan Africa and for all other countries

Dependent variable: ln(GDP per capita 2000)				
	sub-Saharan Africa only		all other countries	
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history	0.145 (0.642)	0.0149 (0.5537)	0.825 [*] (0.497)	0.812 ⁺ (0.432)
Absolute latitude		0.0341 ⁺ (0.0191)		0.0274 ^{***} (0.0063)
Landlocked		-0.617 ^{**} (0.232)		-0.685 ^{**} (0.265)
Eurasia				-0.528 ^{**} (0.190)
Climate		0.104 (0.187)		0.168 (0.138)
constant	7.61 ^{***} (0.17)	7.36 ^{***} (0.42)	8.81 ^{***} (0.25)	8.13 ^{***} (0.23)
N	43	43	105	105
R squared	0.002	0.191	0.029	0.331
<i>Notes:</i> Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.				

TABLE 5. OLS results with corrected and updated data, Maddison's GDP per capita data for 1960 and 2000

	ln(Maddison's GDP per capita 1960)		ln(Maddison's GDP per capita 2000)	
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history			2.18 ^{***} (0.33)	1.48 ^{***} (0.35)
1960 ancestry-adjusted state history	1.17 ^{***} (0.26)	0.611 ^{**} (0.290)		
Absolute latitude		0.0318 ^{***} (0.0051)		0.0357 ^{***} (0.0063)
Landlocked		-0.616 ^{***} (0.186)		-0.544 ^{***} (0.158)
Eurasia		-0.630 ^{***} (0.224)		-0.450 ^{**} (0.214)
Climate		0.159 (0.135)		0.150 [°] (0.085)
constant	6.97 ^{***} (0.12)	6.62 ^{***} (0.19)	7.19 ^{***} (0.15)	6.66 ^{***} (0.19)
N	127	127	148	148
R squared	0.124	0.463	0.245	0.512
<i>Notes:</i> Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.				

**TABLE 6. WLS results with corrected and updated data,
Maddison's GDP per capita data for 1960 and 2000**

Dependent variable:	ln(Maddison's GDP per capita 1960)	ln(Maddison's GDP per capita 2000)		
	Weighted by arable land area 2000 ¹³			
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history			0.843** (0.382)	1.56*** (0.27)
1960 ancestry-adjusted state history	-0.373 (0.479)	0.00290 (0.36251)		
Absolute latitude		0.0369*** (0.0060)		0.0456*** (0.0048)
Landlocked		-0.684*** (0.249)		-0.375 (0.228)
Eurasia		-1.46*** (0.14)		-1.35*** (0.13)
Climate		0.329*** (0.099)		0.205** (0.098)
constant	7.89*** (0.30)	6.85*** (0.20)	8.06*** (0.22)	6.67*** (0.21)
N	125	125	145	145
R squared	0.005	0.729	0.033	0.644

Notes: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

The hump-shaped relationship

BOP (2018, 2) claim there is a quadratic relationship between deep history and present GDP per capita, and they say they discovered the quadratic relationship only because their new study looked 3,500 years further back in time: “We show that the relationship between state history and current income per capita across countries is hump-shaped rather than linear, and that this is due to the inclusion of state experience before the Common Era.” But they in fact showed that only for state history that is *not adjusted for ancestry*:

Taken together, our estimation results so far [that is, using the unadjusted *Statehist*] are consistent with our predicted pattern [that is, concavity of per capita GDP with respect to state history]. Moreover, this becomes evident only when we employ the new extended *Statehist* index. Are the estimates

13. The World Bank, my source for arable land area, does not have data for 1960. It does have data for 1970 (for example), but there are numerous missing values. The correlation between the data for 1970 and for 2000 on the percentage of land that is arable, for countries included in the simple regression for 1960, is 0.951 (n=124).

improved by accounting for the state histories of the ancestors of present-day populations, instead of the state histories of places? To investigate this, we estimate the model for per capita GDP above using the ancestry-adjusted *Statehist* index. The results are displayed in Table 6, where we use the *Statehist* index in 1500 CE adjusted by the migration matrix (as in previous studies, but for the first time including full state history before 1 CE). (BOP 2018, 32–33)

In other words, BOP (1) showed a quadratic relationship between per capita GDP and unadjusted *Statehist*; (2) showed that that quadratic relationship was only present when including pre-CE history; and (3) then “improved” the estimates by using ancestry-adjusted *Statehist*, with those estimates shown in their Table 6 (BOP 2018, 30) and the fitted quadratic regression curve from the uncontrolled regression (their Table 6 Column 2) is displayed in my Figure 3. But what they did not show is whether the quadratic relationship between per capita GDP and ancestry-adjusted *Statehist* is only present when including pre-CE history. Here in my Table 7, I provide the estimates showing that a quadratic relationship exists in the corrected and updated PW data, which is to say that a quadratic relationship exists even without the pre-CE history, and I display the fitted quadratic regression curve from the uncontrolled regression in Figure 4. Tables 7 and 8 together confirm that there is a “hump-shaped” relationship per the standard applied by BOP (2018, 21), that of there being a negative- or downward-sloping portion of the fitted quadratic curve (Table 7) and that there are positive and negative linear relationships holding respectively below and above the optimal value of the independent variable (Table 8). And so the claim by BOP (2018) regarding the supposed effect of state history before the Common Era does not hold when using ancestry-adjusted state history—and of course they affirm that the ancestry adjustment is vital.¹⁴

Table 9 presents the quadratic relationship using data weighted by arable land area. The absolute coefficients on ancestry-adjusted state history and its squared term are large, but the combined effect size is what yields readily to interpretation. Table 10 thus presents combined effect sizes for different values, which are similar in size to the effects found in the linear regressions. Beyond that, the implications of the quadratic relationship are far different than those of a linear relationship. For example, as shown in Table 10 the quadratic relationship indicates that having the highest possible value of ancestry-adjusted state history is actually worse for GDP per capita in 2000 than having the lowest possible value would be. Not only that, but at the weighted first-quartile value of ancestry-adjusted state history (0.375) the combined effect size is 1.83, while at the weighted third-quartile value (0.687)

14. “Population flows after 1500, when the era of colonization began, are instrumental in mapping the impact of historical events to today’s economic performance” (BOP 2018, 22).

the combined effect size is 1.54—so having the third-quartile value of ancestry-adjusted state history is worse for GDP per capita in 2000 than having the first-quartile value would be.

TABLE 7. OLS results for quadratic relationship

Dependent variable: ln(GDP per capita 2000)				
	PW original data		Corrected and updated data	
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history	5.24*** (1.12)	2.51** (1.03)	6.68*** (1.11)	3.68*** (0.95)
Ancestry-adjusted state history squared	-3.86*** (1.43)	-1.52 (1.23)	-5.61*** (1.43)	-3.17*** (1.15)
Absolute latitude		0.0310*** (0.0087)		0.0266*** (0.0060)
Landlocked		-0.502*** (0.164)		-0.686*** (0.186)
Eurasia		-0.263 (0.256)		-0.0994 (0.2148)
Climate		0.243* (0.125)		0.202* (0.109)
constant	7.20*** (0.19)	6.85*** (0.23)	7.34*** (0.17)	7.20*** (0.21)
N	136	111	148	148
R squared	0.273	0.600	0.291	0.524

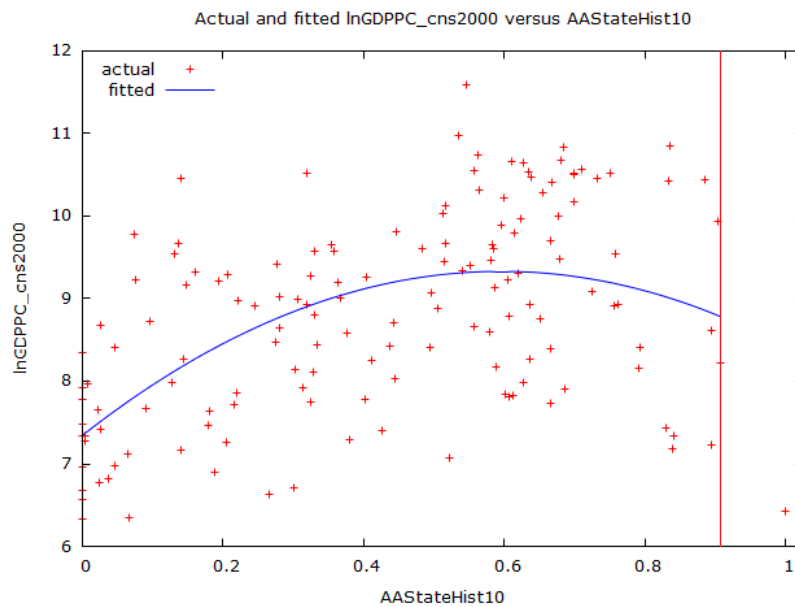
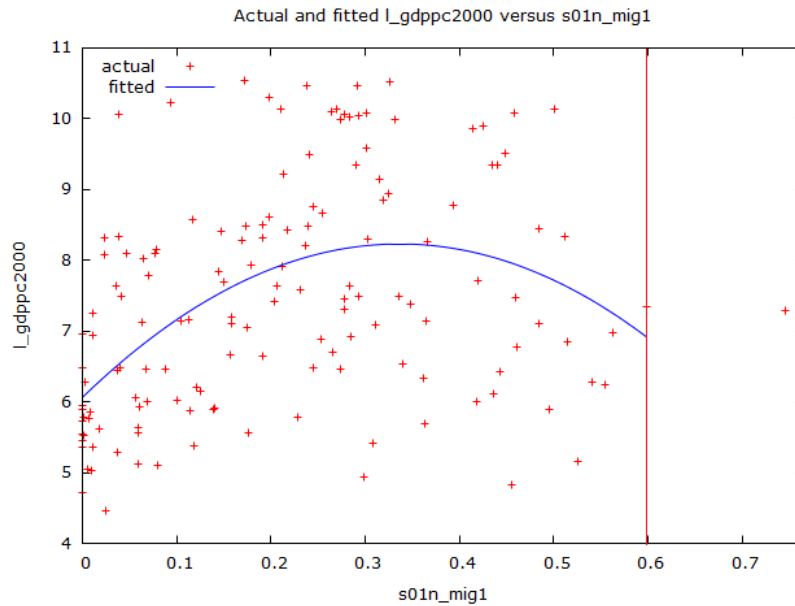
Notes: Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

TABLE 8. OLS results for a 'linear segment check of inverse-u relationship'¹⁵ with corrected and updated data

Dependent variable: ln(GDP per capita 2000)		
	Ancestry-adjusted state history <i>below</i> optimal value for ln(GDP per capita 2000)	Ancestry-adjusted state history <i>above</i> optimal value for ln(GDP per capita 2000)
	(1)	(2)
Ancestry-adjusted state history	3.36*** (0.47)	-3.39* (1.91)
constant	7.59*** (0.16)	11.7*** (1.3)
N	98	50
R squared	0.331	0.082

Notes: Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

15. The comparable check by BOP 2018 is their Table D8 (page 21 of their online appendix).



Figures 3 and 4. BOP’s fitted quadratic regression curve (top) and mine (bottom), showing year-2000 GDP per capita’s ‘hump-shaped’ relationship with ancestry-adjusted state history 3500 BCE–1500 CE and 1 CE–1500 CE, respectively. I have taken the liberty of erasing the portion of the curves to the right of the second-highest value of ancestry-adjusted state history to prevent visual distraction by the disparate behavior of the two curves over values that are all but out of sample.

TABLE 9. WLS results for quadratic relationship with corrected and updated data, dropping cases with missing ancestry-adjusted state history

Dependent variable: ln(GDP per capita 2000)			
Weighted by arable land area 2000			
	(1)	(2)	(3)
Ancestry-adjusted state history		8.08*** (1.29)	6.70*** (0.79)
Ancestry-adjusted state history squared		-8.51*** (1.32)	-6.55*** (0.82)
Absolute latitude	0.0397*** (0.0048)		0.0418*** (0.0042)
Landlocked	-0.695** (0.237)		-0.417** (0.198)
Eurasia	-1.20*** (0.13)		-0.972*** (0.118)
Climate	0.373*** (0.102)		0.329*** (0.087)
constant	7.91*** (0.17)	7.71*** (0.29)	6.46*** (0.23)
N	146	146	146
R squared	0.614	0.227	0.745

Notes: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

TABLE 10. Effect size at different values of x when $y = B_0 + 8.08x - 8.51x^2$

Value of x	Effect on y
0	0.00
0.2	1.28
0.4	1.87
0.6	1.78
0.8	1.02
1	-0.43

What findings are surprising?

Putterman and Weil say that “the ability of these historical measures to predict income today is surprisingly high” (2010, 1651). They therefore suggest we should be surprised by the strength of the correlation between their ancestry-adjusted state history variable, which combines information on year-1500 governance and year-2000 ethnicities, and year-2000 GDP per capita.

I show in Table 11 that year-1500 population¹⁶ and year-2000 population together have an effect on year-2000 GDP per capita—which, personally, I do not find to be surprising—and that the linear relationship between ancestry-adjusted state history and year-2000 GDP per capita is relatively *weaker*. Ancestry-adjusted state history adds 0.03 to the R-squared achieved solely by PW’s controls, but year-1500 population and year-2000 population add 0.06.

TABLE 11. OLS results with corrected and updated data, dropping cases with missing ancestry-adjusted state history

	Dependent variable: ln(GDP per capita 2000)			
	(1)	(2)	(3)	(4)
Ancestry-adjusted state history			1.01*** (0.37)	1.02*** (0.34)
ln(Population 2000)		0.135 (0.084)		0.0576 (0.0861)
ln(Population 1500)		-0.226*** (0.068)		-0.187*** (0.069)
Absolute latitude	0.0304*** (0.0064)	0.0278*** (0.0057)	0.0318*** (0.0060)	0.0290*** (0.0055)
Landlocked	-0.815*** (0.197)	-0.718*** (0.198)	-0.750*** (0.196)	-0.692*** (0.194)
Eurasia	0.107 (0.240)	0.351 (0.228)	-0.169 (0.218)	0.0654 (0.2145)
Climate	0.231** (0.111)	0.274** (0.108)	0.191* (0.113)	0.231** (0.108)
constant	7.74*** (0.19)	8.36*** (0.95)	7.47*** (0.20)	8.86*** (0.95)
N	148	148	148	148
R squared	0.463	0.523	0.496	0.553

Notes: Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

16. Putterman and Weil (2010, 1637) say that in their Table I they “confine [their] analysis to looking at 11 large regions” because “population data for 1500 are very noisy, particularly at the country level.” However, Putterman and Weil actually don’t use the region-level data from their source, a book by Colin McEvedy and Richard Jones (1978). What Putterman and Weil use as the year-1500 populations for the regions is McEvedy and Jones’s data for individual countries, summed up—but Putterman and Weil’s data set doesn’t include every country. For example, McEvedy and Jones’s year-1500 estimate for the entire Caribbean region is 0.3m (1978, 299), but PW don’t use that figure; instead PW sum up the populations of the seven Caribbean countries in their data set, arriving at 0.186m (though even that is incorrect, as it happens). Like Putterman and Weil, then, I use McEvedy and Jones’s country-level estimates for 1500, and I see no reason to judge that data any more noisy than the data in Putterman and Weil’s World Migration Matrix, which after all is also country-level data on where people were living in the year 1500.

Conclusion

The notion of a strong relationship between year-2000 GDP per capita and ancestry-adjusted state history is not robust to weighting by population or arable land area—such weights being sensible methods to address the historical contingency of today’s country borders—but the strong relationships between year-2000 GDP per capita and distance from the equator, being landlocked, and climate *are* robust to those weights. Furthermore, sub-Saharan Africa drives the relationship observed in the PW (2010) regressions, yet the effect of ancestry-adjusted state history among sub-Saharan African countries is zero. And taking into account state history between 3500 BCE and 1 BCE does not transform whatever relationship may appear to exist between early development and current incomes.

Reflect for a moment on the historical contingency of today’s country borders. Why does the PW (2010) finding fail the robustness checks? It’s not only because the robustness checks address spatial autocorrelation *a la* Kelly (2019), but also because the checks treat recent historical accidents such as the partition of Africa as accidents, while in PW (2010) those accidents drive the results. If there were a few countries in sub-Saharan Africa rather than a few dozen, PW’s regressions would have found nothing—that huge t-ratio on the simple regression crumbling down to where p isn’t even below .05. And the number of countries in Africa is hardly a result of anything that happened before 1880, let alone before 1500 or 1 CE.

Further reason to refrain from accepting the conclusions of PW (2010) is the nature of the “World Migration Matrix”—the table used by PW (2010) to adjust state history data for population ancestry. Its builders’ frequent assumptions that year-2000 ethnicities reveal the year-1500 location of ancestors are sometimes unwarranted and are not justified in principle.¹⁷ It neglects, and as a result often

17. Present ethnic categorizations are at best only weakly suggestive regarding migrations of 80, 150, or 400 years ago. “The nations in which we claim citizenship are no more than two hundred years old, and the ethnic groups with which we identify, while sometimes older, have been remarkably changeable,” notes the scholar of migration Patrick Manning (2005, 4). PW (2010) purport to overcome the weak suggestiveness of ethnicity by making “heroic assumptions” (Putterman 2016). By diving into the PW (2010) appendices the immense role of such assumptions can be appreciated. The “World Migration Matrix” often assumes that if a year-2000 person can be classified into an ethnic identity, then there was a corresponding social group in the year 1500 that contained 100 percent of that person’s ancestors. Only somewhat less often, it incorporates the assumption that that year-1500 social group lived entirely within the year-2000 borders of a single country. In combination these assumptions lead PW into some silly statements, e.g., “The majority of Austria’s population (95.2%) is ethnic Austrian” (Appendix-Europe, page 1), by which they mean, incredibly: Almost everyone living in Austria today is descended entirely from ‘Austrians,’ a people who in the year 1500 lived entirely within the borders of present-day Austria. And according to the World Migration Matrix the only other countries where descendants of Austria’s year-1500 population live today

conflicts with, other public data sets about migration.¹⁸ Its sources, primarily American and British encyclopedias, exhibit incompleteness regarding past migrations across present-day country borders, an incompleteness that is surely systematically biased.¹⁹ And as its data is specifically about the locations of people in the year 1500, it would seem that justification is needed for any direct application to data from other time periods, and yet without supplying such justification PW applied the Matrix not to a variable for state development as of the year 1500 but to

are Australia, Canada, Italy, Slovenia, Switzerland, and the United States. Yes, it says *Germany* has no such persons, and further that *none* of the year-1500 ancestors of today's population of Germany lived in the territory of *any* country that borders Germany. It also says that none of the year-1500 ancestors of today's population of Hungary lived in Austria. It says that South Koreans have zero North Korean ancestry, and vice versa! Such absurdities are all over the Matrix.

18. World Bank data on foreign-born populations ([link](#)) shows that 45 of 165 countries had more foreign-born population in the year 2000 than, per the World Migration Matrix, their year-2000 population had 'foreign' year-1500 ancestors (see Box A3). Needless to say, that likely should not be the case for *any* country. Notably, that same World Bank data set was used by Chanda, Cook, and Putterman (2014), who treated its differences between year-2000 and year-1960 foreign-born populations as a measurement of migration and then applied that measurement to the "World Migration Matrix" to produce a "1500–1960 Migration Matrix." I was unable to successfully replicate what Chanda, Cook, and Putterman did to produce that 1500–1960 Migration Matrix (which I nonetheless use in Section 4 above), and the process they used to produce it seems to me flawed. One example of such a flaw is that their process assumes, without acknowledgment, zero differences in natural rate of population change 1960–2000 among persons of different ancestry. A more dramatic example is that in many cases the World Bank data indicated an *increase* in foreign-born population between 1960 and 2000 in a country where the World Migration Matrix had said there was *zero* 'foreign' ancestry in 2000, meaning that any straightforward calculation by Chanda, Cook, and Putterman should have led them in many cases to an absurd conclusion that there was negative foreign ancestry in 1960, but of course the 1500–1960 Matrix contains no negative values.

19. Dramatic mass movements may be recorded in an encyclopedia paragraph or two, but encyclopedias do not recount, in any detail or with any precision, regional patterns of reproduction and relocation over the several centuries after 1500 during which the borders of today's countries were often nonexistent, not meaningful, or both. Consider that of the 138 countries in the World Migration Matrix that have two or more "immediate neighbors" (those sharing a land boundary or separated by less than 24 miles of water), the Matrix says 91 of them have *zero* ancestry from *over half* of their immediate neighbors. And of the 155 countries with at least one immediate neighbor, the Matrix says 61 have zero ancestry from *all* of their immediate neighbors. Early in their paper, PW (2010, 1634–1637) devote a few pages to examining the data as it would be if countries that are immediate neighbors were lumped together. There they write that lumping of immediate neighbors is needed because while "[t]he principal diagonal of the matrix provides a quick indication of differences in the degree to which countries are now populated by the ancestors of their historical populations...in some cases, the diagonal entry may give a misleading impression" because, they say, migration that is from neighboring countries should be considered at least "near-indigenous" (PW 2010, 1635). However, subsequent to those few pages, PW do *not* lump immediate neighbors in any of their analyses. And it is easy to find other researchers who have used the World Migration Matrix but not bothered to do any such lumping: Shriram et al. (2018, 45) create a "historical heterogeneity index" variable that simply "counted the number of source countries" and take that variable to be measuring "the historical likelihood of encountering unfamiliar cultural outgroups." Klasing (2013, 453) constructs a variable using "the share of its current population originating from another country." Spolaore and Wacziarg (2013, 333 n.11) "define whether a country's population today is composed of more than 50 percent of descendants of its 1500 population."

a variable for state history over the entire period 1 CE to 1500.²⁰

BOP (2018) purport to offer and confirm a hypothesis that early development has a bit of a “hump-shaped” relationship with year-2000 GDP per capita: too much early development, like in eastern Asia, is bad; too little, like in sub-Saharan Africa, is also bad; northwestern Europe had just the right amount. Since 2000, though, the countries of eastern Asia and sub-Saharan Africa have seen faster economic growth than have the countries of northwestern Europe.

We should now have a new understanding with regard to the data provided in PW (2010) and BOP (2018). The message of a strong linear relationship, conveyed in PW (2010), is wrong; the message conveyed in BOP (2018), that there is a hump-shaped relationship *because of pre-Common Era history*, is also wrong. So what is the effect of history, ‘deep’ or otherwise, as far as the evidence here goes? One could point to Table 9 and say: If you weight countries by arable land area and regress year-2000 GDP per capita on basic geographic controls you’ll get an R-squared of 0.61. By adding information on world history up through the year 1500 plus information about presumed connections between present-day ethnicities and year-1500 locations of ancestors, and allowing that information to have nonlinear effects, your R-squared increases to 0.74.²¹ If you were to add further information on history more recent than the year 1500, obviously your R-squared would go far higher.

20. The PW (2010) empirical strategy considers people alive in the year 1500 to inherit or exhibit their land’s pre-1500 history, as if (among other presumptions) there was zero migration across year-2000 borders prior to 1500. They offer nothing like the reservations that Comin, Easterly, and Gong (2010, 83) express about the application of the World Migration Matrix to pre-1500 data: “This [applying the Matrix] is straightforward for the 1500 AD technology measure. It is more problematic for the 1000 BC and 0 AD exercise, since we have no data on migrations before 1500 AD. It still seems of interest to correct the 1000 BC and 0 AD measures by the post-1500 migration matrix to test a peoples-rather-than-places technology persistence view. The post-1500 migrations are arguably the most consequential, since the discovery of the New World and the technological advances in oceangoing transport made wholesale replacement of low-technology people by high-technology people more likely than in earlier eras. We could assume either that pre-1500 migrations were random and orthogonal to the error term, or that they also tended to direct high-technology peoples to low-technology places (because of the ease of conquest and the high returns from applying more advanced technology to a previously underdeveloped land area). In the first case, the coefficient on 0 AD and 1000 BC would be unbiased. In the second case, the coefficient would be biased downward, making persistence look lower than it really was.”

21. For comparison’s sake, by adding variables for year-1500 population and year-2000 population to the basic geographic controls, the R-squared increases from 0.61 to 0.70.

Appendix

Box A1. Countries excluded from Putterman and Weil's key regression (their Table IV Panel A col. (6)) because of mistakenly missing values, in descending order by population

- | | | |
|----------------|--------------------------|---------------------|
| • Russia | • Tajikistan | • North Macedonia |
| • Vietnam | • Libya | • Slovenia |
| • Ukraine | • Kyrgyzstan | • Namibia |
| • Myanmar | • Turkmenistan | • Estonia |
| • Tanzania | • Croatia | • Eswatini |
| • Uzbekistan | • Bosnia and Herzegovina | • Cyprus |
| • Iraq | • Moldova | • Fiji |
| • Saudi Arabia | • Lithuania | • Guyana |
| • Kazakhstan | • Eritrea | • Equatorial Guinea |
| • Cambodia | • Lebanon | • Qatar |
| • Belarus | • Albania | • Comoros |
| • Serbia | • Armenia | • Cabo Verde |
| • Azerbaijan | • Liberia | |

Box A2. Countries excluded from all of Putterman and Weil's regressions, or even from their data set entirely, but whose year-2000 population is greater than that of Cabo Verde (the smallest country in their data set for which there should not have been any missing values), in descending order by population

- | | | |
|---------------|------------------------|--------------|
| • North Korea | • United Arab Emirates | • Djibouti |
| • Afghanistan | • Oman | • Bahrain |
| • Cuba | • Kuwait | • Bhutan |
| • Somalia | • Guinea-Bissau | • Suriname |
| • Puerto Rico | • Timor-Leste | • Luxembourg |

Box A3. Countries that, per the World Bank, had a percentage of foreign-born population in the year 2000 that was higher than, per the World Migration Matrix, the percentage of their year-2000 population's *year-1500 ancestors* that were foreign, in descending order by the product of the percentage-point difference and the year-2000 population

- | | | |
|----------------|------------------------|---------------|
| • Germany | • United Arab Emirates | • Spain |
| • Russia | • Japan | • Greece |
| • Saudi Arabia | • United Kingdom | • Netherlands |

- Austria
- Nepal
- Burkina Faso
- South Korea
- Switzerland
- Ethiopia
- Belgium
- Sweden
- Uganda
- Portugal
- Mozambique
- Ireland
- Philippines
- Norway
- Poland
- Tanzania
- Denmark
- Armenia
- China
- Gambia
- Bahrain
- Finland
- Tajikistan
- Luxembourg
- Algeria
- Mauritania
- Slovenia
- Bosnia and Herzegovina
- Czech Republic
- Rwanda
- Tunisia
- Senegal
- Congo, Rep.
- Papua New Guinea
- Lesotho
- Oman

Box A4. Countries that have at least one immediate neighbor and, per the World Migration Matrix, have *zero* year-1500 ancestry from *all* of their immediate neighbors, in descending order by number of immediate neighbors

- China
- Germany
- Brazil
- Tanzania
- Algeria
- Burkina Faso
- Mozambique
- South Africa
- Argentina
- Bolivia
- Colombia
- Congo, Rep.
- Peru
- Senegal
- Uganda
- Uzbekistan
- Angola
- Ethiopia
- Finland
- Greece
- Guatemala
- Indonesia
- Mauritania
- Nigeria
- Tajikistan
- Thailand
- Turkmenistan
- Venezuela
- Chile
- El Salvador
- Eritrea
- Ghana
- Honduras
- Malawi
- Mexico
- Nicaragua
- Norway
- United Arab Emirates
- Bhutan
- Bosnia and Herzegovina
- Costa Rica
- Ecuador
- Guyana
- Nepal
- Netherlands
- Panama
- Papua New Guinea
- Somalia
- Tunisia
- Uruguay
- Australia
- Canada
- Dominican Republic
- Gambia
- Haiti
- Japan
- South Korea
- Lesotho
- Philippines
- Portugal
- Trinidad and Tobago

Data and code

Data and code used in this research is available from the journal website ([link](#)).

References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2002. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics* 117(4): 1231–1294.
- Borcan, Oana, Ola Olsson, and Louis Putterman (BOP).** 2018. State History and Economic Development: Evidence from Six Millennia. *Journal of Economic Growth* 23: 1–40.
- Caplan, Bryan.** 2016a. Cross-Country Regressions and Population Weighting. *EconLog, Library of Economics and Liberty*, June 20. [Link](#)
- Caplan, Bryan.** 2016b. Ancestry, Long-Run Growth, and Population-Weighting. *EconLog, Library of Economics and Liberty*, June 23. [Link](#)
- Chanda, Areendam, C. Justin Cook, and Louis Putterman.** 2014. Persistence of Fortune: Accounting for Population Movements, There Was No Post-Columbian Reversal. *American Economic Journal: Macroeconomics* 6(3): 1–28.
- Cinyabuguma, Matthias M., and Louis Putterman.** 2011. Sub-Saharan Growth Surprises: Being Heterogeneous, Inland and Close to the Equator Does Not Slow Growth Within Africa. *Journal of African Economies* 20(2): 217–262.
- Comin, Diego, William Easterly, and Erick Gong.** 2010. Was the Wealth of Nations Determined in 1000 BC? *American Economic Journal: Macroeconomics* 2(July): 65–97.
- Kelly, Morgan.** 2019. The Standard Errors of Persistence. CEPR Discussion Paper 13783. Centre for Economic Policy Research (London). [Link](#)
- Kelly, Morgan.** 2020. Understanding Persistence. CEPR Discussion Paper 15246. Centre for Economic Policy Research (London). [Link](#)
- Klasing, Mariko J.** 2013. Cultural Dimensions, Collective Values and Their Importance for Institutions. *Journal of Comparative Economics* 41: 447–467.
- Manning, Patrick.** 2005. *Migration in World History*. New York: Routledge.
- McEvedy, Colin, and Richard Jones.** 1978. *Atlas of World Population History*. Harmondsworth, UK: Penguin.
- Putterman, Louis.** 2000. Can an Evolutionary Approach to Development Predict Post-War Economic Growth? *Journal of Development Studies* 36(3): 1–30.
- Putterman, Louis.** 2016. Correspondence with Bryan Caplan. *EconLog, Library of Economics and Liberty*, February 2. [Link](#)
- Putterman, Louis.** 2017. State Antiquity Index (“Statehist”) Version 3.1 and Updates. November 10. [Link](#)
- Putterman, Louis, and David N. Weil (PW).** 2010. Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality. *Quarterly Journal of Economics* 125(4): 1627–1682.

BRIGGEMAN

Shrira, Ilan, Arnaud Wisman, and Kenji Noguchi. 2018. Diversity of Historical Ancestry and Personality Traits Across 56 Cultures. *Personality and Individual Differences* 128: 44–48.

Spolaore, Enrico, and Romain Wacziarg. 2013. How Deep Are the Roots of Economic Development? *Journal of Economic Literature* 51(2): 325–369.

About the Author



Jason Briggeman is the managing editor of *Econ Journal Watch*, an adjunct associate professor of economics at Austin Community College, an associate editor for *Liberal Currents*, and an adjunct fellow of the Niskanen Center. His email address is jason@briggeman.org.

[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6047>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 109–123

From Hume to Smith on the Common Law and English Liberty: A Comment on Paul Sagar

Jacob R. Hall¹

[LINK TO ABSTRACT](#)

In a *Political Theory* article, “On the Liberty of the English: Adam Smith’s Reply to Montesquieu and Hume,” Paul Sagar (2021) treats Smith in relation to Montesquieu and David Hume on the origins, development, and robustness of English liberty. In doing so he fashions something of a divide between Smith and Hume with respect to their views on the development of liberty in England. Sagar suggests a difference between Smith and Hume on the importance of the common law, that “Smith took more seriously than Hume the idea that liberty required not just an appropriate constitution but quotidian security as realized via law” (2021, 18). Sagar goes further to say that Hume would underestimate, or miss entirely, the idea “that liberty must be understood not just in terms of the form of constitution and wider political order, but also regarding the security of citizens as achieved via the legal system, and especially the operation of fair trials” (ibid., 14).

Upon reading and reflecting on Sagar’s paper, I felt that Sagar had failed to do justice to Hume. I drafted a comment along the lines of the present article and submitted it to *Political Theory*, where Sagar’s paper appears, but it was turned down without explanation.

1. Graduate student, George Mason University, Fairfax, VA 22030. I am grateful for feedback from Erik Matson, two anonymous referees, and the attendants of the Invisible Hand Seminar at George Mason University.

Sagar on Smith and Hume

According to Sagar, Montesquieu, in a manner like Tacitus, traces English liberty back to the ancient customs of the Germans, rooted in bottom-up way in local practice and judgment. But in Montesquieu's estimation, English liberty was fragile. In Sagar's article, Hume is cast as counterpart to the Frenchman. Hume, according to Sagar, dated the origins of English liberty to the Glorious Revolution when Parliament secured its supremacy over the king. In Hume's estimation, English liberty was robust.

Sagar does not define what he means by liberty. Nor does he wade into interpreting how Montesquieu, Hume, and Smith use the word themselves. I do not fault him for that. In this paper I myself will avoid wading into debating the many meanings of liberty. But it is worth remembering that there are many meanings of liberty. Hume talks of liberty with many different modifiers. At times, Hume seems to see liberty as synonymous with general rules that are consistently and equitably applied. But other times, Hume's liberty is more akin to a particular constitutional order, particularly the post-revolution British constitution. Daniel Klein and Erik Matson (2020) argue that a central meaning of liberty in Hume is "others not messing with one stuff," which they dub "mere-liberty." Keeping these different senses of liberty in mind, even if in the background, is important.

Sagar then turns to Smith, arguing that we should situate the Scotsman "as intervening in the debate between Montesquieu and Hume on the origins, age, and robustness of English liberty" (2021, 2). He argues that Smith agreed with Hume on the importance of the Glorious Revolution for securing English liberty, but unlike Hume, Smith recognized that the reforms of the Revolution were "grafted onto, and...greatly enhanced by, a wider preexisting legal framework," namely the common law (ibid., 15). For Sagar's Smith, the common law, particularly the legal and administrative reforms of king Edward I (r. 1272–1307), constituted a major development in the story of English liberty, a development that Sagar supposes Hume underappreciated or missed entirely. In the course of making his argument about the importance of the common law in Smith's historical narrative in *Lectures on Jurisprudence*, Sagar overlooks Hume's discussions of the common law in *The History of England*.

Sagar (2021, 14) argues that Hume was "comparably inattentive," relative to Smith, on the importance of English common law for the development of liberty. *Comparably* inattentive—perhaps. But throughout his article Sagar regularly portrays Hume as *highly* inattentive to such historical developments. According to Sagar, Hume's narrative is "focused on the high politics of court and parliament and not the day-to-day affairs of legal administration," and, as a result, Hume

supposedly misses the importance of the common law as an element of English liberty (*ibid.*). Instead, Hume’s theory of English liberty supposedly turns merely on the serendipities of 1688. Sagar faults Hume for his supposed oversights: “Core aspects of English liberty long predated the Glorious Revolution in ways Hume had not appreciated” (*ibid.*, 2).

Sagar is right to highlight Hume’s emphasis on the Glorious Revolution for the development of English liberty. But he is wrong to conclude that Hume failed to recognize earlier elements of liberty prior to 1688. In volume 2 of the *History*, Hume discusses arbitrary measures Edward I took to secure funds for his war with France in 1296. The English barons mount a dissent and subsequently forced Edward I to cease and desist and to renew his oath to uphold Magna Carta. They even empowered knights in each county to investigate and punish royal officials for violating Magna Carta. Hume, reacting to the baronial response, writes:

A precaution, which, though it was soon disused, as encroaching too much on royal prerogative, proves the attachment, which the English in that age bore to liberty, and their well-grounded jealousy of the arbitrary disposition of Edward. (*H*, 2:119–121)²

Nearly 400 years before the Glorious Revolution, Hume seems to believe the English bore an attachment to liberty. Liberty may not have been solidified, and the medieval institutions often worked against it, but its spirit inhabited the island.

Sagar (2021, 8) quotes volume six of the *History* in support of his claim that Hume felt it was “only in 1688 that English liberty was finally established and secured”:

The revolution alone...put an end to all these disputes: By means of it, a more uniform edifice was at last erected: The monstrous incoherence, so visible between the ancient Gothic parts of the fabric and the recent plans of liberty, was fully corrected: And to their mutual felicity, king and people were finally taught to know their proper bounds (*H*, 6:475–476; quoted in Sagar 2021, 8).

Sagar follows Hume’s quotation by saying: “Against Montesquieu, Hume’s verdict was that English liberty was not old but very new indeed” (2021, 8). But Hume is not saying here that only in 1688 did liberty spring forth. Hume’s use of “fully corrected” does not preclude the development of elements prior to the revolution and partial corrections along the way. Earlier in the volume 5, Hume again speaks

2. The notation “*H*, 3:76” means page 76 in the third volume of the Liberty Fund edition of Hume’s *History of England*. I adopt such notation throughout. For Adam Smith’s *Wealth of Nations* I adopt the abbreviation *WN* and for his *Lectures on Jurisprudence* I use *LJ*.

of elements of liberty that existed well before 1688:

The grievances, under which the English laboured, when considered in themselves, without regard to the constitution, scarcely deserve the name; nor were they either burthensome on the people's properties, or anywise shocking to the natural humanity of mankind... [Under Charles I's administration] all ecclesiastical affairs were settled by law and uninterrupted precedent; and the church was become a considerable barrier to the power, both legal and illegal, of the crown. Peace too, industry, commerce, opulence; nay, even justice and lenity of administration, notwithstanding some very few exceptions: All these were enjoyed by the people; and every other blessing of government, except liberty, *or rather the present exercise of liberty*, and its proper security. (*H*, 5:249–250, my emphasis)

In Hume's estimation, and, as Sagar argues, Smith's, something meaningful did happen in 1688, but that did not mean that elements of liberty did not exist or were not upheld prior to 1688.

Sagar admits that Hume discusses habeas corpus a few times in his *History*, but correctly points out that Hume "tended to do so only incidentally" (2021, 14). Sagar then concludes that "there is no sustained effort by Hume to tell a story in which the liberty he believed was only secured by the 1688 revolution was itself augmented and buttressed by a wider legal culture that was by that point centuries old." He goes on to detail Smith's account of the English common law and the legal reforms of Edward I. But Sagar overlooks Volume 1 and 2 of Hume's *History* where his discussions of the common law can be found, which dovetail with Smith's narrative.

As I see it, the divide between Smith and Hume is less than Sagar suggests, if it exists at all. Hume did in fact recognize the English common law as an element of English liberty. Rather than replying to Hume, Smith should be seen as developing on Hume's ideas and disseminating the results.

Sagar provides a fine summary of Smith's story given in the *Wealth of Nations* and the *Lectures on Jurisprudence* concerning the medieval barons' embrace of luxury goods and commerce over retainers, retinues, and other instruments of war and plunder. Smith's narrative serves as a great example of his tendency to take the ideas of his friend, Hume, and refine and improve them. Hume's historical account of the civilizing of the barons occurs over two and a half volumes in a disjointed fashion. Hume only once comes out and identifies the phenomenon clearly (*H*, 3:76). Smith's *WN* account is a concise and straightforward 44 paragraphs. Notice how Smith condenses time in his famous "diamond buckles" passage:

But what all the violence of the feudal institutions could never have effected,

the silent and insensible operation of foreign commerce and manufactures gradually brought about. ... For a pair of diamond buckles perhaps, or for something as frivolous and useless, they [the barons] exchanged the maintenance, or what is the same thing, the price of the maintenance of a thousand men for a year, and with it the whole weight and authority which it could give them. (*WN*, 418–419.10)

We might ask, “How gradually did the civilizing process unfold?” Hume would say 450 years, give or take, and it was by no means monotonic. Smith was synthesizing and reformulating the arguments of Hume to make it coherent and understandable to his audience. By drawing out one distinct process taken up in Hume’s *History*, Smith made the civilizing process of commerce more concrete and graspable.

But we should not say that Hume did not understand the importance of that civilizing process for English political development. He did. He even discusses mechanisms unmentioned in Smith’s account, such as Henry VII’s prohibition on baronial military retainers (*H*, 3:75) and his allowance of the barons to alienate their estates (*H*, 3:77). Nor would we say that Smith disagreed with Hume on the importance of that civilizing process for English liberty. Smith cites Hume favorably as “the only writer who, so far as I know, has hitherto taken notice of it” (*WN*, 412.4).

Sagar does not paint a divide between the two Scots on the dynamic by which the barons became gentlemen. He rightfully notes that Smith built upon Hume (Sagar 2021, 12). But on the importance of the common law, and England’s legal system more generally, he does make a separation between Smith and Hume. That separation strikes me as inappropriate. Hume does not have a chapter on the common law in the *History*, nor is it discussed at length in an appendix. Also, there is no “Of the Common Law” in Hume’s *Essays*. We are impoverished because of that. Hume does, however, take up the common law at many key moments in its development in his *History*. The common law is one example of his larger theme of jural integration and its importance for the development of English liberty.

Just as he builds upon, clarifies, and refines Hume’s account of the barons becoming gentlemen, Smith takes up the common law in the *Lectures on Jurisprudence*. Hume spends two volumes giving his readers a rich display of examples of just how violent and licentious the barons were and how the common law (among other things), backed by the increasing authority of the king, suppressed their local authority. Smith distills Hume and combines him with the work of other scholars and his own knowledge to tell a coherent story about the development of the common law and its importance for English liberty. We will never know the extent to which Hume grasped the importance of the common law for development of English liberty, an importance that Smith teaches in his lectures. Hume had more to do than time allowed for, and he understood more than he passed down to us in

essays. Hume certainly had much more than an inkling of the narrative that Smith would come to flesh out with greater detail.

Jural integration in England

Hume's *History of England* is situated between two grand arrivals to Britain: those of Julius Caesar in 55 BCE and William of Orange in 1688. Between those two men, Hume spends six volumes detailing England's high politics and the slow development of the English constitution. And like another great six-part epic saga that began 200 years after his death, Hume produced the *History* out of order. Nowhere in it does Hume have a sweeping analysis of the English common law. In fact, searching the text for "common law" yields only trivialities.

But that does not mean Hume neglects the development of England's legal system. In a sense, that development is a major theme of Hume's *History*. From Arthur to Henry VIII, England was ruled by multiple, competing powers. In volumes 1 and 2 Hume shows us, often in bloody detail, that the quantum of authority and legitimacy of the average medieval king was built on shifting sands (*H*, 2:283–284). In addition to the king, England was home to a number of powerful actors such as the towns, independent barons, and the Roman Church and its affiliate ecclesiastical bodies, not to mention the potential influence of the Welsh, Scottish, and French aristocracies. Each separate authority had its own jurisdiction, source of power and influence, and instruments for making their voice heard.

The existence of multiple powers capable of violence made medieval England a dangerous place to live. Hume characterizes the ancient constitution as one of minimal economic growth plagued by violence and political instability:

The towns were situated either within the demesnes of the king, or the lands of the great barons, and were almost entirely subjected to the absolute will of their master. The languishing state of commerce kept the inhabitants poor and contemptible; and the political institutions were calculated to render that poverty perpetual. The barons and gentry, living in rustic plenty and hospitality, gave no encouragement to the arts, and had no demand for any of the more elaborate manufactures: Every profession was held in contempt but that of arms: And if any merchant or manufacturer rose by industry and frugality to a degree of opulence, he found himself but the more exposed to injuries, from the envy and avidity of the military nobles. (*H*, 1:463)

In addition to plundering their societal inferiors, the barons engaged in private warfare against each other, leaving the countryside in a continual state of chaos

and lawlessness (*H*, 1:231, 237, 250, 284, 288, 350–351, 371–372, 400, 463, 2:11, 143, 189, 279). Even during periods of relative peace between the bigger political players, “men were never secure in their houses” and bands of robbers, often supported by encastellated barons, were known to plunder entire villages (*H*, 1:69, 288).

A main theme of Hume’s *History* is the integration of the separate powers into a single unified government (Forbes 1975, 263; Whelan 2004, 256; Sabl 2012, 65). Dan Klein and I have produced a lengthy compendium that gives 142 quotations from Hume’s *History* touching upon jural pluralism or jural integration (Hall and Klein 2020). Barry Weingast (2015; 2016; 2017) identifies similar themes in Smith’s *Wealth of Nations* and *Lectures on Jurisprudence*. Integration went hand-in-hand with the increasing power and authority of the king. It was a slow process. Two steps forward, one step back. The personal characters of kings were of great importance, as evidenced by both Hume’s historical narrative and his lengthy character portraits. The strongest and most respectable of kings always seemed to sire the weakest and least respectable heirs. Henry II begets John, and Edward I begets Edward II.

But through the centuries, the authority of the king increased in scale and in scope. Individual kings may have been weak, but the crown was growing stronger. The medieval era eventually gave way to what is now called the early modern period. Dating the achievement of establishing an integrated nation-state is not an exact science. If we wish to be poetic, perhaps that transition occurred at Bosworth when Henry Tudor crushed Richard III and his Yorkist supporters. Henry VII and his successor Henry VIII crafted reforms that led to demilitarization of the English aristocracy on a grand scale (*H*, 3:75, 77). By the time Hume discusses the reign of Elizabeth I in volume 4 the competing powers of the medieval era have fallen away and he speaks of *the* government. The Tudors and Stuarts still faced various political challenges, but the section header “Discontent of the Barons,” used five times by Hume in the first two volumes, is absent from the early modern volumes, that is, volumes 3, 4, 5, and 6.

Hume on the common law

Hume hailed from a family of common-law lawyers.³ Both his father and his two grandfathers were trained formerly in the law and were practiced barristers. He attended school in Edinburgh at an early age, which he abandoned at 14 without receiving his degree. But upon returning Hume took up studying the law, guided

3. I thank an anonymous referee for bringing these biographical facts about Hume to my attention.

partly by Henry Home, Lord Kames (Mossner 2001, 53–54). Ernest Campbell Mossner has praised Hume’s legal knowledge:

In the end, Hume’s legal knowledge, both theoretical and practical, was not inconsiderable. The theoretical, forming an integral part of moral philosophy, appears so frequently in his published works as to require no special comment. Of theoretical jurisprudence, Hume was a master. (Mossner 2001, 55)

Hume also would have had the entirety of the Advocates’ Library collection at his disposal while working on the *History*. Post-Norman England experienced an explosion of administrative documentation, and it is reflected in Hume’s citations in the early volumes.

In medieval England, the rules enforced in the King’s court were in competition with the other courts of the realm. An aggrieved man could seek justice in the county courts which administered local customary law or he could go to the ecclesiastical courts which administered canon law. A town merchant could take his case to a borough court to be judged by the rules of the *Lex mercatoria*. One might even go to one’s feudal lord to make his case under the rules of feudal custom. The royal law, however, was common throughout the realm and was the origination point of what has come to be known as the common law (Hogue 1986/1966).

Arthur Hogue defined common law as “the body of rules prescribing social conduct and justiciable in the royal courts of England” (1986/1966, 5), and I use the phrase *common law* along those lines. At a more abstract level, common law is simply law held in common throughout the polity. A third sense of common law is law worked out through precedent. The three senses of common law are all rooted in the historical development of English common law. The royal law applied to all Englishmen, no matter where the crime was committed or who the perpetrator may have been. As royal judges travelled the kingdom hearing cases they learned and refined their legal judgments. By travelling, they made the royal law common throughout the realm. Through their travels they amassed a bank of precedent that they could call upon in subsequent cases.

Over time, the common law, enforced by the royal courts, subsumed or marginalized its competitors. As the authority and power of the king grew the “justice done in the king’s name by men who [were] the king’s servants became the most important kind of justice” (Pollock and Maitland 2010/1895, 91). The success of the common law went hand in hand with the centralization of power around the king. Hume said as much:

It [the people’s freedom] required the authority almost absolute of the sovereigns, which took place in the subsequent period, to pull down those

disorderly and licentious tyrants, who were equally averse from peace and from freedom, and to establish that regular execution of the laws which, in a following age, enabled the people to erect a regular and equitable plan of liberty. (*H*, 2:525)

As the king's power and authority grew, so did the impact of his laws. Frederick Pollock and Frederic Maitland (1895), Arthur Hogue (1966), Harold Berman (1983), and John Baker (1995), all scholars of the common law, attest to that fact.

Another reason for the rise of the common law cited by Hume was the rediscovery of Justinian's Pandects, a compendium of juristic writings on Roman law (*H*, 2:520). For Hume, the rediscovery in 1130 of Justinian's Pandects was a glimmer of light from a more civilized era that would begin to illuminate a dark world. No other event "tended further to the improvement of the age" (*ibid.*).

With Justinian's Pandects in their hands, the clergy took up legal studies with great zeal. Less than ten years later, according to Hume, lectures in civil law were being given in Oxford. Although Roman civil law never rose to the same level of prominence in England as it did on the continent due to "the jealousy [of] the laity," and perhaps England's island geography, it left a permanent mark on English law (*H*, 2:520–521). The English jurists imitated their civil law equivalents, "rais[ing] their own law from its original state of rudeness and imperfection" (*H*, 2:521). Here Hume complements Larry Siedentop's *Inventing the Individual* (2014) on the importance of Christianity and the Catholic Church for the development of western liberalism. In chapter 16 of his book, Siedentop discusses the re-discovery of Justinian's Pandects, tracing its effects on both ecclesiastical and lay law over the subsequent chapters.

To find Hume's discussions of concrete legal development, we need to look at his coverage of the strongest medieval English kings: Henry II and Edward I. Only these men were able to extend their authority and carry out reform without having their political coalitions turn on them.

The reign of King Stephen (r. 1135–1154) was marked by "The Anarchy," a succession crisis that led to the complete breakdown of civil order in England (*H*, 1:279–295). Henry II (r. 1154–1189), upon winning the war and ascending to the throne, was tasked with cleaning up the mess and restoring order and justice to the kingdom. Hume depicts Henry II as a good and strong king who led England with a steady hand and an "equitable administration" (*H*, 1:359, 301, 370). He was a politically savvy man, as shown by his swift actions to demolish the castles illegally built by the local barons during The Anarchy (*H*, 1:360). He had his share of dark days (*H*, 1:310–338, 348–358), as all medieval kings did, but he was responsible for increasing the power of the monarchy over the licentious barons and executing long-lasting reforms to England's legal system. In 1176, Henry II partitioned

England into four divisions and appointed itinerant justices to travel along a circuit to hear and decide on the cases brought before them in the counties (*H*, 1:359–360). The general eyre, as the law circuit was called, furthered the mission of making the king’s law common throughout the realm. The eyre increased the geographical influence of the king’s laws and regularized Englishmen to its enforcement. It protected the lower gentry and the peasants from the arbitrary violence and corruption of the barons, and, albeit slowly, acted to curb baronial power (*H*, 1:360). This is all Hume’s narrative, not mine, and it is a narrative of the common law.

But the expansion of royal justice and the common law was not a matter of force. Royal justice passed the market test and came to be the preferred court of law because it administered better justice. The justices in eyre were notable men of honor, in contrast to the local courts, thus the respectability of the common law was bolstered by their character (*H*, 1:360). They were also better trained and less corrupt than their local counterparts. Additionally, the common law possessed what no other court in England did—the weight of the monarchy.

After looking at the common law reforms under Henry II, Hume drops common law until his discussion of Edward I (r. 1272–1307). That is understandable. What occupies Hume during the reigns of Richard I (r. 1189–1199), John (r. 1199–1215), and Henry III (r. 1216–1272) is tracing out the events that led to John’s capitulation at Runnymede and the solidification of Magna Carta into the English political ethos. Like Henry II, Edward I inherited a mess. Edward’s father, Henry III, was a relatively weak king who bumbled his way into a civil war against a group of discontented barons led by Simon de Montfort. Upon inheriting the crown, Edward “immediately applied himself to the re-establishment of his kingdom, and to the correcting of those disorders” introduced by Henry III’s weak administration (*H*, 2:75).

Edward I was a strong king. A weak king would never have earned the nickname “Hammer of the Scots,” which is reminiscent of the great Frankish king Charles Martel (688–741 CE), as *martel* in old French means ‘hammer.’ As recognized by Smith and discussed by Sagar, Edward I was a great legal reformer. Smith puts Edward I alongside Henry II as one of the greats in terms of his legislative capacity (*LJ(A)*, v.34).

Hume’s account of Edward I’s legal reforms is similar to Smith’s as recounted by Sagar, but not once does Sagar’s article cite volumes 1 or 2 of the *History*. To diminish the power of the great barons, Edward offered his protection to the gentry, merchants, and serfs by instituting “an exact distribution of justice” and by “a rigid execution of the laws” (*H*, 2:75). He did so by insisting that, as he obeyed Magna Carta with respect to the barons, they too should extend and uphold Magna Carta with respect to their own vassals. He replaced judges that had grown

corrupt under the former administration, and he provided the justice system as a whole with force sufficient to execute the law properly (*H*, 2:75). Hume says that by Edward's actions "the face of the kingdom was soon changed; and order and justice took place of violence and oppression" (*H*, 2:75–76). In fact, Hume argues that Edward's legal reforms were the chief advantage which the English attained from his reign—and even more importantly, that Englishmen "still continue to reap" the benefits of Edward's vigor in Hume's day (*H*, 2:141).

From Hume to Smith on English history

Samuel Pufendorf wrote in the preface to *An Introduction to the History of the Principal Kingdoms of States of Europe*:

I hope therefor, that the Discreet Reader will look favourably upon this Work, not as a Piece design'd for Men of great Learning, but adapted to the Apprehensions and Capacities of young Men, whom I was willing to shew the Way, and, as it were, to give them a tast[e], whereby they might be encouraged to make a further search into this Study. (Pufendorf 2013/1695, 7)

Smith's *Lectures on Jurisprudence* can be read as engaging in the same mission, to impart the history and practice of jurisprudence to his young students. In crafting each lecture, Smith would have distilled the works of a number of writers, to deliver a coherent presentation of the subject matter. Reviewing the course material and preparing a lecture forces the mind to explore multiple potential angles, to see connections between seemingly disjointed phenomena, and to make concrete the abstract. As a result, Smith improved and refined the arguments of a number of great writers, not least Hume.

If a student approached Smith after his lecture to inquire about further reading on feudalism and medieval history, Smith may have pointed him towards the first two volumes of Hume's *History*. Smith cribs from Hume on multiple occasions over the course of his lectures, synthesizing Hume and folding in other bits of information from other sources to round out the lecture's topic. The influence of Hume and his *History* on Smith is clear. The editors of the Liberty Fund edition of the *Lectures* write in their introduction that "Smith's use of Montesquieu is clear from LJ(B), his dependence on Hume's *History* and *Essays* is more pronounced in LJ(A)" (*LJ*, 32). Simply searching on "Hume" in the body of the text of *LJ* yields 53 results, mostly contained in footnotes by the editors showing likeness between what Smith says and things Hume had said previously.

Smith's long discussions of feudal history and law place him in the camp

of Hume, Henry Spelman, Matthew Hale, and Robert Brady in their efforts to debunk the idea of the ancient constitution of Edward Coke and the common-law lawyers. By teaching his students about the origins of feudalism and its implications for English legal developments, Smith educated his students against Coke's ideas. Simply by situating the origins of the common law with the Plantagenets kings, he stakes out a position, alongside Hume, against the idea of the ancient constitution.

As Sagar (2021, 17) discusses, Smith gives some importance to juries in his story of English liberty. Hume had a similar appreciation for the jury. He writes in volume I of his *History* that the jury is the best institution “calculated for the preservation of liberty and the administration of justice” (*H*, 1:77). He attributes the innovative legal practice to Alfred the Great's (r. 886–889) “popular and liberal plan” of the administration of justice (*H*, 1:77).

We moderns think of juries as an obvious form of adjudication. But the main competitor of the jury, trial by ordeal, or trial by combat if the dispute involved multiple parties, was a hardy weed. Hume saw trials by ordeal or combat as a mark of barbarism and backwards superstition (*H*, 1:181, 359, 486). Throughout the *History*, Hume uses the existence of ordeals and trial by combat as a way to gauge England's degree of civilization. When trial by ordeal was outlawed in England during the reign of Henry III, Hume wrote that it was “a faint mark of improvement in the age” (*H*, 2:72). Trial by combat would not ever be officially outlawed, but the legal reforms of Henry II brought about its gradual decline (*H*, 1:486). Hume clearly appreciated the importance of trial by jury and its role in developing a legal system that respected liberty.

Another common theme shared by the Smith and Hume is the importance of island geography for the development of British politics. In Hume's *History*, Britain's island geography protects it from would-be military invaders and provides a buffer from continental politics and foreign influence (*H*, 1:11–12, 207, 229, 299–300, 2:362, 522, 3:51, 88, 146, 348, 4:55). Smith extends Hume by explaining how Britain's island geography explains Britain's long tradition of not having a standing army, which Smith argues was the reason Britain did not fall prey to an authoritarian monarch like the French or the Spanish (*LJ(A)* iv.168–169).

Securing liberty or elements of a system of liberty?

I argue above that Hume recognized common law as an important element of English liberty. Elements come together to form compounds, and compounds are in a sense more than the sum of their parts. That brings me to a final point about

Sagar's article: I find his grouping of Montesquieu, Hume, and Smith into a two-by-two matrix unsatisfactory. Sagar's Montesquieu thinks English liberty is old and fragile. Sagar's Hume thinks English liberty is new and robust. But after learning of Smith's discussions of the common law, how does Sagar want us to think about Smith?

At points in his article, Sagar seems to suggest that Smith fits into the two-by-two matrix. After discussing Smith's Humean affirmation of the Glorious Reformation, Sagar writes: "But this did not mean that English liberty was only as old as 1688, as Hume concluded" (2021, 15). After his account of Smith on the common law, Sagar suggests that Smith thought liberty to be new, but not as new as Hume: "But as a result, English liberty was much older in its core elements than the reforms effected by the Glorious Revolution alone" (*ibid.*, 19). Does Sagar mean to suggest that Smith thought English liberty was mostly secured in 1307?

At other points, Sagar seems to suggest that Smith does not fit neatly into one of the cells of the two-by-two matrix. Instead, Sagar suggests that Smith saw the common law as an element of English liberty, much as I say that Hume did. Sagar (2021, 15) writes that although Smith understands the importance of the Glorious Revolution, Smith believes the 17th-century reforms stuck because they were "grafted onto, and [were] greatly enhanced by, a wider preexisting legal framework."

In the conclusion of his paper, Sagar poses a puzzle to the reader:

[T]here remains obscurity in Smith's account as to how exactly his Humean story of England's constitutional liberty being secured in 1688 meshes with his jurisprudential account of liberty in terms of the security delivered by the common law as already being largely in place before the late seventeenth century, and which (if either) he considers most important. (Sagar 2021, 19)

Reading Smith and Hume as articulating important elements of English liberty helps to resolve Sagar's puzzle. The common law did not secure liberty in England; rather it was a single element in liberty's development.

After recognizing that Smith and Hume thought of the common law as an important element, it does not necessarily follow that they felt English liberty was robust. If it was so robust, why would Smith need to mount an extensive case for liberty in *The Wealth of Nations*? Why outline and argue for his vision of a liberal England if liberty were robustly secured by the Glorious Revolution and its aftermath?

Conclusion

On a careful reading of Hume's *History*, and from an awareness of Smith's pervasive cribbing from Hume, Sagar (2021) may wish to reconsider the gap he sees between Smith and Hume on the importance of English common law for the development of liberty. Smith should be seen as distilling, developing, and even disseminating Hume's interpretations. Smith's treatment of the Edwardian common-law reforms is similar to Hume's in terms of both the factual account and its importance for the English constitution. The development of the English common law can be read in Hume as a byproduct of the increasing power of the crown at the expense of the violent and licentious barons. With the local authority of the barons suppressed, and finally put to rest sometime between the Battle of Bosworth and the reign of Elizabeth I, England's stability was enhanced, allowing a more regular plan of liberty to take root and grow into a salient characteristic feature of the British state.

References

- Baker, John H.** 1995. Personal Liberty Under the Common Law of England, 1200–1600. In *The Origins of Modern Freedom in the West*, ed. R. W. Davis, 178–202. Stanford, Cal.: Stanford University Press.
- Berman, Harold J.** 1983. *The Formation of the Western Legal Tradition*. Cambridge, Mass.: Harvard University Press.
- Forbes, Duncan.** 1985. *Hume's Philosophical Politics*. London: Cambridge University Press.
- Hall, Jacob R., and Daniel B. Klein.** 2020. Jural Pluralism and Jural Integration in David Hume's *History of England*: A Compendium of 142 Quotes. *GMU Working Paper in Economics* (George Mason University, Fairfax, Va.) 20-36. [Link](#)
- Hogue, Arthur.** 1986 [1966]. *Origins of the Common Law*. Indianapolis: Liberty Fund.
- Hume, David.** 1983 (H). *The History of England from the Invasion of Julius Caesar to the Revolution in 1688*, ed. W. B. Todd, 6 vols. Indianapolis: Liberty Fund.
- Hume, David.** 1994. *Essays, Moral, Political, and Literary*, ed. Eugene F. Miller. Indianapolis: Liberty Fund.
- Klein, Daniel B., and Erik W. Matson.** 2020. Mere-Liberty in David Hume. In *A Companion to David Hume*, ed. Moris Polanco, 125–160. Guatemala City: Universidad Francisco Marroquin.
- Mossner, Ernest C.** 2001. *The Life of David Hume*. Oxford, UK: Oxford University Press.
- Pollock, Frederick, and Frederic Maitland.** 2010 [1895]. *The History of English Law before the Time of Edward I*, 2 vols. Indianapolis: Liberty Fund.
- Pufendorf, Samuel.** 2013 [1695]. *An Introduction to the History of the Principal Kingdoms and States of Europe*, ed. Michael J. Seidler, trans. Jodocus Crull. Indianapolis: Liberty Fund.
- Sabl, Andrew.** 2012. *Hume's Politics*. Princeton, N.J.: Princeton University Press.

- Sagar, Paul.** 2021 (forthcoming). On the Liberty of the English: Adam Smith's Reply to Montesquieu and Hume. *Political Theory*. [Link](#)
- Siedentop, Larry.** 2014. *Inventing the Individual*. Cambridge, Mass.: Harvard University Press.
- Smith, Adam.** 1976 [1776] (WN). *The Wealth of Nations*, ed. R. H. Campbell and A. S. Skinner, 2 vols. Oxford, UK: Oxford University Press.
- Smith, Adam.** 1982 (LJ). *Lectures on Jurisprudence*, ed. R. L. Meek, D. D. Raphael, and P. G. Stein. Oxford, UK: Oxford University Press.
- Weingast, Barry R.** 2015. Adam Smith's Industrial Organization of Religion: Explaining the Medieval Church's Monopoly and Its Breakdown in the Reformation. Working paper, October 18. [Link](#)
- Weingast, Barry R.** 2016. The Medieval Expansion of Long-Distance Trade: Adam Smith on the Town's Escape from the Violent and Low-Growth Feudal Equilibrium. *Stanford Law and Economics Olin Working Paper* (Stanford University, Stanford, Cal.) 492. [Link](#)
- Weingast, Barry R.** 2017. Adam Smith's Theory of Violence and the Political Economics of Development. In *Organizations, Civil Society, and the Roots of Development*, ed. Naomi R. Lamoreaux and John Joseph Wallis, 51–82. Chicago: University of Chicago Press.
- Whelan, Frederick G.** 2004. *Hume and Machiavelli: Political Realism and Liberal Thought*. Lanham, Md.: Lexington Books.

About the Author



Jacob Hall is an economics Ph.D. student at George Mason University. He is a Ph.D. fellow at the Mercatus Center and a fellow of the Adam Smith Program at George Mason University. His academic research is in economic history. He also does work on the writings of David Hume, especially the *History of England*. His email address is jhall26@gmu.edu.

[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at JournalTalk:
<https://journaltalk.net/articles/6048>



Comment on Measuring the Size of the Shadow Economy Using a Dynamic General Equilibrium Model with Trends

Manuel A. Gómez¹ and Adrián Ríos-Blanco²

[LINK TO ABSTRACT](#)

Various approaches have been proposed in the literature to estimate the size of the informal or shadow economy. Following Ceyhun Elgin (2020), the approaches can be classified as direct, indirect, and model approaches. Direct approaches provide estimates using direct methods as surveys, interviews, and questionnaires on households or firms. Indirect approaches infer the size of the shadow economy using data on economic indicators, such as currency demand or electricity consumption, or on the discrepancy between actual and registered labor force or between national income and expenditure statistics. Model approaches generally rely on the use of a theoretical model, such as the dynamic general equilibrium (DGE) model and the multiple indicators multiple causes (MIMIC) model, which is based on the use of a specific structural equation model.

As Elgin et al. (2021) argue, the DGE method has several limitations, including the need of a base-year estimate of the size of shadow production from an external source and its reliance on strong and somewhat arbitrary assumptions on the relationship between the productivities in the shadow and formal sectors. Mario Solis-Garcia and Yingtong Xie (2018) propose a DGE method for measur-

1. Department of Economics, Universidade da Coruña, 15071 A Coruña, Spain. We thank two anonymous referees for their useful comments. In particular, the suggestions of one referee have helped to improve the presentation of the Appendix.

2. Graduate student, Universidade da Coruña, 15071 A Coruña, Spain.

ing the size and dynamics of the shadow economy. They apply the method to a set of Latin American and Asian countries. They set up a two-sector dynamic deterministic general equilibrium model with four different exogenous trends, and use the restrictions imposed by the model’s balanced growth to endogenize the shadow productivity trend. This method would avoid imposing an arbitrary assumption about productivities in the shadow and formal sectors.

Unfortunately, the method proposed by Solis-Garcia and Xie (2018) does not serve its purpose of computing the ratio of shadow to formal production. The reason is that the balanced-growth assumption entails that formal and shadow output grow at the same rate, and hence their ratio must be constant at all times and equal to its base-year value, which is taken as given (from Schneider et al. 2010). The problem should have become apparent in the numerical simulations made by the authors. They did not notice the problem because of an error in the formula for computing the growth rate of the productivity shock in the shadow economy. This error propagated to other variables of the model and, eventually, led to a non-constant ratio of shadow to formal output, which gave a sense of plausibility to the simulation results.

The model

The method proposed by Solis-Garcia and Xie (2018) relies on a two-sector (formal and shadow) deterministic DGE model.

Setup

The economy is inhabited by a representative infinitely lived household who solves the utility maximization problem:

$$\begin{aligned} & \max_{\{C_t, N_t, N_{F,t}, N_{S,t}, K_{t+1}, X_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{\phi \Gamma_{H,t} N_t^{1+\chi}}{1+\chi} \right) \\ & \text{subject to :} \quad C_t + \Gamma_{A,t} X_t = (1 - \tau_t) Y_{F,t} + (1 - \rho \hat{s} \tau_t) Y_{S,t} \text{ ,} \\ & \quad \quad \quad K_{t+1} = (1 - \delta) K_t + X_t \text{ ,} \\ & \quad \quad \quad N_t = N_{F,t} + N_{S,t} \text{ ,} \end{aligned}$$

where formal output, $Y_{F,t}$, and shadow output, $Y_{S,t}$, are given by

$$Y_{F,t} = K_t^\alpha (\Gamma_{F,t} N_{F,t})^{1-\alpha}, \tag{1}$$

$$Y_{S,t} = (\Gamma_{S,t} N_{S,t})^\eta. \tag{2}$$

In the maximization problem, C_t is consumption, N_t is total hours worked, X_t is investment, $\tau_t \in [0,1)$ is the income tax rate, K_t is the stock of capital, $N_{F,t}$ and $N_{S,t}$ are hours worked in formal and shadow production, and total output is $Y_t = Y_{F,t} + Y_{S,t}$. The model includes four exogenous permanent productivity shocks: in the household's choice of hours worked, $\Gamma_{H,t}$; the production of investment goods, $\Gamma_{A,t}$; the hours worked in formal production, $\Gamma_{F,t}$; and the hours worked in shadow production, $\Gamma_{S,t}$. The parameter $\beta \in (0,1)$ is the discount factor, $\sigma > 0$ is the inverse of the intertemporal elasticity of substitution, $\phi > 0$ is the disutility of labor, $\chi \geq 0$ is the inverse of the Frisch elasticity of labor supply, $\delta \in (0,1)$ is the depreciation rate of capital, $\rho \in [0,1)$ is the probability of a tax audit, $\hat{s} > 1$ is a tax surcharge, $\alpha \in (0,1)$ is the capital share in formal production, and $\eta > 0$ is the elasticity of work time in shadow production.

The government taxes income at a rate τ_t and uses tax revenues to fund a stream of non-productive expenditure G_t . The informal sector avoids taxation unless caught by a tax audit, which occurs with probability ρ , in which case the government imposes a tax surcharge \hat{s} . Thus the government's budget constraint is

$$G_t = \tau_t Y_{F,t} + \rho \hat{s} \tau_t Y_{S,t}$$

Balanced growth path

Let $g_{Z,t} = Z_t / Z_{t-1}$ denote the (gross) growth rate of a variable Z , so that $Z_t = Z_0 \prod_{j=1}^t g_{Z,j}$ ($t > 0$), and let $g_{i,t} = \Gamma_{i,t} / \Gamma_{i,t-1}$ denote the growth rate of the exogenous productivity shock $i \in \{H, A, F, S\}$, so that $\Gamma_{i,t} = \Gamma_{i,0} \prod_{j=1}^t g_{i,j}$ ($t > 0$).

Solis-Garcia and Xie (2018) use their equilibrium conditions (expressions 3.5 through 3.15 in their article) to calculate the balanced-growth rates, which are derived from their conditions (expressions 4.1–4.10 in their article), which we reproduce for reference:

$$g_C = g_A g_X = g_G = g_Y \tag{3}$$

$$g_K = g_X \tag{4}$$

$$g_N = g_{N_F} = g_{N_S} \quad (5)$$

$$g_A = g_{Y_F} g_K^{-1} \quad (6)$$

$$g_H g_N^\chi = g_{Y_F}^{1-\sigma} g_{N_F}^{-1} \quad (7)$$

$$g_{Y_S} g_{N_S}^{-1} = g_{Y_F} g_{N_F}^{-1} \quad (8)$$

$$g_{Y_F} = g_K^\alpha (g_F g_{N_F})^{1-\alpha} \quad (9)$$

$$g_{Y_S} = (g_S g_{N_S})^\eta \quad (10)$$

$$g_Y = g_{Y_F} = g_{Y_S} \quad (11)$$

$$g_A = g_{P_X} \quad (12)$$

Here, P_X is the decentralized price of investment goods, and income tax rates are assumed to be stationary, $g_\tau = 1$.

There are 14 equalities in the system (3)–(12).³ As Eqs. (6) and (8) can be derived from (3), (4), (5) and (11), we can delete them. The resulting system has 12 equalities, so 12 of the 15 growth rates can be solved for as functions of the three remaining ones. This means that of the growth rates of the exogenous productivity shocks, one can be expressed as a function of the remaining three, namely, g_S is chosen to be a function of g_H , g_F and g_A .

Computation of the ratio of shadow to formal output

Using their balanced-growth conditions, Solis-Garcia and Xie (2018) compute the ratio of shadow to formal production, $Y_{[S/F],t} = Y_{S,t} / Y_{F,t}$ by following these steps:

Step 1. Obtain data for the stock of capital, K_t , formal work time, $N_{F,t}$, and formal output, $Y_{F,t}$. Use them to calculate the time series of the observed growth rates

3. Taking logarithms in (3)–(12), the resulting system is linear in the logarithms of the growth rates so it can be readily discussed and solved.

$\hat{g}_{K,t}$, $\hat{g}_{N_{F,t}}$, and $\hat{g}_{Y_{F,t}}$.

Step 2. Select a base year t_0 and the values of the parameters α , σ , χ , ρ , and \hat{s} .

Step 3. Calibrate the value of the elasticity of shadow work, η .

Step 4. Calculate the base-year value of shadow work time, N_{S,t_0} , taking as given the base-year formal production, Y_{F,t_0} , from actual data, and the base-year ratio of shadow to formal production, $Y_{[S/F],t_0}$, from Schneider et al. (2010). Compute the growth rates of the productivity shock in shadow production, $g_{S,t}$ from the observed growth rates and then the time series of the productivity shock, $\Gamma_{S,t}$.

Step 5. Compute the time series of work time in shadow production $N_{S,t}$ using that $g_{N_{S,t}} = g_{N_{F,t}} = \hat{g}_{N_{F,t}}$. This equation follows from the balanced-growth condition (5) and the assumption that the growth rate of formal work time, $g_{N_{F,t}}$, is equal to its observed value, $\hat{g}_{N_{F,t}}$.

Step 6. Compute the time series of shadow output $Y_{S,t} = (\Gamma_{S,t} N_{S,t})^\eta$ and then the time series of the ratio of shadow to formal output $Y_{[S/F],t} = Y_{S,t} / Y_{F,t}$.

Why does this method fail?

Solis-Garcia and Xie (2018) take as given the base-year value of formal production, Y_{F,t_0} , from actual data, and the base-year value of the ratio of shadow to formal production, $Y_{[S/F],t_0}$, from Friedrich Schneider, Andreas Buehn, and Claudio Montenegro (2010). Hence, the initial value of shadow output is determined by $Y_{S,t_0} = Y_{[S/F],t_0} Y_{F,t_0}$. The balanced-growth condition (11)—Eq. (4.9) in Solis-Garcia and Xie (2018)—says the growth rate of shadow output is equal to the growth rate of formal (and of total) output, which is computed from actual data, $g_{Y_{S,t}} = g_{Y_{F,t}} = \hat{g}_{Y_{F,t}}$. The immediate consequence is that the ratio of shadow to formal output must be constant and equal to its base-year value at every time,

$$Y_{[S/F],t} = \frac{Y_{S,t}}{Y_{F,t}} = \frac{Y_{S,t_0} \prod_{j=t_0+1}^t g_{Y_{S,j}}}{Y_{F,t_0} \prod_{j=t_0+1}^t g_{Y_{F,j}}} = \frac{Y_{S,t_0}}{Y_{F,t_0}} = Y_{[S/F],t_0}, \text{ if } t > t_0,$$

or

$$Y_{[S/F],t} = \frac{Y_{S,t}}{Y_{F,t}} = \frac{Y_{S,t_0} \prod_{j=t+1}^{t_0} g_{Y_{Sj}}}{Y_{F,t_0} \prod_{j=t+1}^{t_0} g_{Y_{Fj}}} = \frac{Y_{S,t_0}}{Y_{F,t_0}} = Y_{[S/F],t_0}, \text{ if } t_0 > t.$$

The method, therefore, generates a constant ratio of shadow to formal production and, therefore, is useless to compute the time series of this ratio. Despite its obviousness, the problem with this method has so far gone unnoticed, probably because of the relative complexity of its implementation.

The question that arises now is why the simulation results in Solis-Garcia and Xie (2018) do not display a constant ratio of shadow to formal production as they should. The reason is that there is an error in their expression (4.14) for g_S which causes the simulated ratio to be variable. Appendix A shows that the true expression for g_S is

$$g_S = g_H^{-\frac{1-\eta}{(\sigma+\chi)\eta}} g_A^{-\frac{\alpha[(1+\chi)-(1-\sigma)\eta]}{(1-\alpha)(\sigma+\chi)\eta}} g_F^{\frac{(1+\chi)-(1-\sigma)\eta}{(\sigma+\chi)\eta}}. \quad (13)$$

In contrast, the expression (4.14) in Solis-Garcia and Xie (2018) is

$$g_S = g_H^{-\frac{1+\eta}{(\sigma+\chi)\eta}} g_A^{-\frac{\alpha[(1+\chi)+(1-\sigma)\eta]}{(1-\alpha)(\sigma+\chi)\eta}} g_F^{\frac{1+\chi+(1-\sigma)\eta}{(\sigma+\chi)\eta}},$$

where the differences with (13) are marked in **red**.

The error in the computation of the growth rate $g_{S,t}$ translates to the productivity shock in shadow production $\Gamma_{S,t} = \Gamma_{S,t_0} \prod_{j=t_0+1}^t g_{S,j}$ if $t > t_0$ (or $\Gamma_{S,t} = \Gamma_{S,t_0} \prod_{j=t+1}^{t_0} g_{S,j}$ if $t_0 > t$), which in turn translates to shadow production $Y_{S,t} = (\Gamma_{S,t} N_{S,t})^\eta$ and, finally, to the ratio of shadow to formal output $Y_{[S/F],t} = Y_{S,t} / Y_{F,t}$. As a result, the ratio of shadow to formal output is variable in their simulation results, which gives the proposed method an appearance of validity. Using the correct formula for g_S in the Matlab scripts made available by the authors, the simulated ratio $Y_{[S/F],t}$ is constant, as expected.⁴

4. The original MATLAB files are available at Solis-Garcia's website ([link](#)). The files, with the error corrected, that reproduce the simulated shadow and formal output and its ratio are available from the journal website ([link](#)).

Conclusions

The DGE method proposed in Solis-Garcia and Xie (2018) was intended to overcome some of the limitations of previous approaches. Unfortunately, it does not serve its purpose of estimating the ratio of shadow to formal production. The goal is valuable, but more research will be needed to achieve it.

Appendix A

Derivation of g_S

In this Appendix we show that the expression (4.14) for the growth rate of the productivity shock in shadow production, g_S , in Solis-Garcia and Xie (2018) is wrong, and we compute its true value. To show the scope and the source of the error, we provide a detailed derivation starting from the restrictions imposed by the model's balanced-growth conditions (3)–(12), which are expressions 4.1–4.10 in Solis-Garcia and Xie (2018).

From (3) and (4) we get

$$g_Y = g_A g_K. \quad (14)$$

Using (11) and (14), we have

$$g_Y = g_{Y_F} = g_{Y_S} = g_A g_K, \quad (15)$$

whereas (5) states that $g_{N_F} = g_{N_S} = g_N$. Using (15) and (5), we substitute g_{Y_F} and g_{Y_S} with $g_A g_K$, and substitute g_{N_F} and g_{N_S} with g_N , into Eqs. (7), (9) and (10), and get

$$g_H g_N^{1+\chi} = g_A^{1-\sigma} g_K^{1-\sigma}, \quad (16)$$

$$g_A g_K^{1-\alpha} = g_N^{1-\alpha} g_F^{1-\alpha}, \quad (17)$$

$$g_A g_K = g_S^\eta g_N^\eta \quad (18)$$

Solving for g_N in (17), we get that

$$g_N = g_A^{\frac{1}{1-\alpha}} g_K g_F^{-1}. \quad (19)$$

Plugging this expression into (16), we have

$$g_A^{1-\sigma} g_K^{1-\sigma} = g_H \left(g_A^{\frac{1}{1-\alpha}} g_K g_F^{-1} \right)^{1+\chi} = g_H g_A^{\frac{1+\chi}{1-\alpha}} g_K^{1+\chi} g_F^{-(1+\chi)}.$$

Solving for g_K , we obtain

$$g_K = g_H^{-\frac{1}{\sigma+\chi}} g_A^{-\frac{\alpha(1-\sigma)+\sigma+\chi}{(1-\alpha)(\sigma+\chi)}} g_F^{\frac{1+\chi}{\sigma+\chi}}, \quad (20)$$

which coincides with expression (4.11) in Solis-Garcia and Xie (2018). Plugging (20) into (19), after simplification we get

$$g_N = g_H^{-\frac{1}{\sigma+\chi}} g_A^{-\frac{\alpha(1-\sigma)}{(1-\alpha)(\sigma+\chi)}} g_F^{\frac{1-\sigma}{\sigma+\chi}}, \quad (21)$$

which coincides with expression (4.12) in Solis-Garcia and Xie (2018).

Solving for g_S in (18) we have

$$g_S = g_A^{\frac{1}{\eta}} g_K^{\frac{1}{\eta}} g_N^{-1}. \quad (22)$$

Plugging the expression (20) for g_K and the expression (21) for g_N into the former equation (22), we obtain

$$\begin{aligned} g_S &= g_A^{\frac{1}{\eta}} \left(g_H^{-\frac{1}{\sigma+\chi}} g_A^{-\frac{\alpha(1-\sigma)+\sigma+\chi}{(1-\alpha)(\sigma+\chi)}} g_F^{\frac{1+\chi}{\sigma+\chi}} \right)^{\frac{1}{\eta}} \left(g_H^{-\frac{1}{\sigma+\chi}} g_A^{-\frac{\alpha(1-\sigma)}{(1-\alpha)(\sigma+\chi)}} g_F^{\frac{1-\sigma}{\sigma+\chi}} \right)^{-1} \\ &= g_A^{\frac{1}{\eta}} g_H^{-\frac{1}{\eta(\sigma+\chi)}} g_A^{-\frac{\alpha(1-\sigma)+\sigma+\chi}{\eta(1-\alpha)(\sigma+\chi)}} g_F^{\frac{1+\chi}{\eta(\sigma+\chi)}} g_H^{\frac{1}{\sigma+\chi}} g_A^{-\frac{\alpha(1-\sigma)}{(1-\alpha)(\sigma+\chi)}} g_F^{-\frac{1-\sigma}{\sigma+\chi}} \\ &= g_H^{-\frac{1-\eta}{(\sigma+\chi)\eta}} g_A^{-\frac{\alpha[(1+\chi)-\eta(1-\sigma)]}{\eta(1-\alpha)(\sigma+\chi)}} g_F^{\frac{(1+\chi)-(1-\sigma)\eta}{(\sigma+\chi)\eta}}. \end{aligned}$$

Thus, we get the true expression (13) for g_S , instead of the wrong expression (4.14) reported in Solis-Garcia and Xie (2018). The error in Solis-Garcia and Xie

(2018) seems to come from inserting (20) and (21) into the erroneous formula $g_S = g_A^{1/\eta} g_K^{1/\eta} g_N$ rather than using the correct equation (22), $g_S = g_A^{1/\eta} g_K^{1/\eta} g_N^{-1}$.

References

- Elgin, Ceyhun.** 2020. *The Informal Economy: Measures, Causes, and Consequences*. New York: Routledge.
- Elgin, Ceyhun, M. Ayhan Kose, Franziska Ohnsorge, and Shu Yu.** 2021. Understanding the Informal Economy: Concepts and Trends. In *The Long Shadow of Informality: Challenges and Policies*, Advance ed., ed. Franziska Ohnsorge and Shu Yu, 35–91. Washington, D.C.: World Bank.
- Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro.** 2010. Shadow Economies All Over the World: New Estimates for 162 Countries from 1999 to 2007. *Policy Research Working Paper* 5356. World Bank (Washington, D.C.). [Link](#)
- Solis-Garcia, Mario, and Yingtong Xie.** 2018. Measuring the Size of the Shadow Economy Using a Dynamic General Equilibrium Model with Trends. *Journal of Macroeconomics* 56(1): 258–275.

About the Authors



Manuel A. Gómez is a Full Professor of Mathematical Methods for Economics and Business at the Department of Economics of the University of A Coruña (Spain). He received his B.S. in Economics from the University of Santiago de Compostela and his Ph.D. from the University of A Coruña. His areas of research are economic growth, dynamic macroeconomics, and public economics. He has published over 50 articles in academic journals. His email is manuel.gomez@udc.es.



Adrián Ríos-Blanco is a Ph.D. candidate in the Department of Economics at the University of A Coruña (Spain). He received both his B.S. in Business Administration and M.S. in Economics from the University of A Coruña. He is interested in public economics, dynamic economics, and shadow economy measurement and its relation with economic growth. His email address is adrian.rios@udc.es.

[Mario Solis-Garcia and Yingtong Xie's reply to this article](#)
[Go to archive of Comments section](#)
[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6049>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 133–141

Response to “Comment on Measuring the Size of the Shadow Economy Using a Dynamic General Equilibrium Model with Trends”

Mario Solis-Garcia¹ and Yingtong Xie²

[LINK TO ABSTRACT](#)

In Solis-Garcia and Xie (2018), our aim was to propose a working method for measuring the size and properties of the shadow economy based on a dynamic deterministic general equilibrium (DGE) model. The approach relied on the dynamics of observed trends along the balanced growth path to account for both the size and the cyclicity of the shadow economy, as the trends impose a set of equilibrium restrictions over the growth rates of the model variables—including shadow sector output.

In a recent contribution, Manuel Gómez and Adrián Ríos-Blanco (2022) argue that the method proposed by our paper fails to accomplish its goal. In particular, their argument is twofold: First, the assumption of a balanced growth path implies a constant shadow-to-formal output ratio. Second, an error on our end hides this inconsistency and gives rise to an unlucky coincidence, in that it generated results that looked sensible but concealed the original flaw in the method. We acknowledge the mathematical error and concede that using the properties of the balanced growth path indeed results in the consequences they

1. Macalester College, Saint Paul, MN 55105.

2. Carleton College, Northfield, MN 55057.

point out. We sincerely regret this error.

That said, we believe that the claim of Gómez and Ríos-Blanco (2022) about our paper failing to accomplish its goals to measure the shadow economy is unfounded. More precisely, we can attain these goals when we relax the balanced growth path assumption; this change allows us to obtain precise estimates of the size of the shadow economy using minimal data, even compared to Solis-Garcia and Xie (2018). Our revised procedure retains the original model but streamlines the calculations, dropping the balanced growth path assumption altogether and relying instead on a subset of the equilibrium conditions to pin down the dynamics of the informal sector. Put differently, the procedure below achieves the goals we were set to deliver in our earlier contribution.

Revisiting Solis-Garcia and Xie (2018)

Again, we still believe that the idea underlying the paper is sound. In a nutshell, the restrictions imposed by a DGE model require shadow output to behave in a certain way—it cannot move around in a random way.

Moreover, if shadow sector producers want to remain underground—which, by definition, is what the informal producers desire!—then relying on theory to back out what by definition doesn't want to be measured is a good idea.

In this sense, our DGE model and approach with trend still delivers what it was set to do: derive precise estimates of the size of the shadow economy using minimal data. The only update we made to our procedure is to drop the balanced growth path assumption. This means our updated method is more general than the original version and calls for less restrictions. We now describe how the updated procedure is able to pin down the dynamics of the informal sector using a subset of original equilibrium conditions.

Equilibrium conditions

The model remains the same but for a minor modification: We set the hours trend Γ_{Ht} to unity. Note that this is without loss of generality, as we still allow shadow- and formal-sector hours to trend up or down. In this sense, the set of equilibrium conditions is virtually identical, and is reproduced below for convenience:

$$C_t + \Gamma_{At}X_t + G_t = K_t^\alpha (\Gamma_{Ft}N_{Ft})^{1-\alpha} + (\Gamma_{St}N_{St})^\eta \quad (1)$$

$$K_{t+1} = (1 - \delta)K_t + X_t \quad (2)$$

$$N_t = N_{Ft} + N_{St} \quad (3)$$

$$\Gamma_{At} C_t^{-\sigma} = \alpha \beta C_{t+1}^{-\sigma} (1 - \tau_{t+1}) K_{t+1}^{\alpha-1} (\Gamma_{F,t+1} N_{F,t+1})^{1-\alpha} + \beta (1 - \delta) \Gamma_{A,t+1} C_{t+1}^{-\sigma} \quad (4)$$

$$\phi N_t^{\chi} = (1 - \alpha) C_t^{-\sigma} (1 - \tau_t) K_t^{\alpha} \Gamma_{Ft}^{1-\alpha} N_{Ft}^{-\alpha} \quad (5)$$

$$\eta (1 - \rho \hat{s} \tau_t) \Gamma_{St}^{\eta} N_{St}^{\eta-1} = (1 - \alpha) (1 - \tau_t) K_t^{\alpha} \Gamma_{Ft}^{1-\alpha} N_{Ft}^{-\alpha} \quad (6)$$

$$G_t = \tau_t K_t^{\alpha} (\Gamma_{Ft} N_{Ft})^{1-\alpha} + \rho \hat{s} \tau_t (\Gamma_{St} N_{St})^{\eta} \quad (7)$$

$$Y_{Ft} = K_t^{\alpha} (\Gamma_{Ft} N_{Ft})^{1-\alpha} \quad (8)$$

$$Y_{St} = (\Gamma_{St} N_{St})^{\eta} \quad (9)$$

$$Y_t = Y_{Ft} + Y_{St} \quad (10)$$

$$P_{Xt} = \Gamma_{At} \quad (11)$$

Note that our choice to set $\Gamma_{Ht} = 1$ only affects condition (5). Also, and consistent with the parametrization choice in Solis-Garcia and Xie (2018), we will set the probability of an audit (ρ) to zero, which modifies (6) and (7) slightly.

Equilibrium growth rates

Gómez and Ríos-Blanco (2022) show that our original choice to use the balanced growth path conditions (namely, $g_{Y_S} = g_{Y_F}$) results in a constant shadow-to-formal output ratio. Our turnaround is to discard the assumption of balanced growth and focus on period-by-period growth rates instead; as we show below, we have enough observed growth rates to back out the size and dynamics of the shadow economy—though we still need a base year value $Y_{[S/F], t_0}$.

To make this operational, note that conditions (5), (6), and (9) imply

$$g_{Nt}^{\chi} = g_{Ct}^{-\sigma} g_{Y_{Ft}} g_{N_{Ft}}^{-1} \quad (12)$$

$$g_{Y_{St}} g_{N_{St}}^{-1} = g_{Y_{Ft}} g_{N_{Ft}}^{-1} \quad (13)$$

$$g_{Y_{St}} = \left(g_{St} g_{N_{St}} \right)^\eta. \quad (14)$$

To reiterate, equations (12)–(14) are not assumed to hold along the balanced growth path, as the time subscripts suggest.

From (12), we get a measure for the total labor growth rate (g_{N_t}) as a function of the growth rates of consumption (g_{C_t}) and formal output and hours ($g_{Y_{Ft}}$ and $g_{N_{Ft}}$, respectively):

$$g_{N_t} = \left(g_{C_t}^{-\sigma} g_{Y_{Ft}} g_{N_{Ft}}^{-1} \right)^{1/\chi}. \quad (15)$$

Similarly, substituting (14) into (13) gives

$$\left(g_{St} g_{N_{St}} \right)^\eta g_{N_{St}}^{-1} = g_{Y_{Ft}} g_{N_{Ft}}^{-1},$$

which allows us to express the shadow productivity growth rate (g_{St}) as a function of shadow and formal labor growth rates ($g_{N_{St}}$ and $g_{N_{Ft}}$, respectively), the growth rate of formal output ($g_{Y_{Ft}}$), and parameter η . A bit of algebra results in

$$g_{St} = \left(g_{Y_{Ft}} g_{N_{Ft}}^{-1} g_{N_{St}}^{1-\eta} \right)^{1/\eta}. \quad (16)$$

Equations (15) and (16) are all we need to back out the dynamic properties of the shadow economy.

Uncovering the shadow economy

Finding shadow labor for the base period

By construction, the shadow-to-formal output ratio in base year t_0 equals

$$Y_{[S/F],t_0} \equiv \frac{Y_{S,t_0}}{Y_{F,t_0}}. \quad (17)$$

Substituting the expression for shadow output, equation (9), we get

$$Y_{[S/F],t_0} = \frac{\left(\Gamma_{S,t_0} N_{S,t_0}\right)^\eta}{Y_{F,t_0}}.$$

From here we can solve for shadow labor in the base period

$$N_{S,t_0} = \frac{\left(Y_{[S/F],t_0} Y_{F,t_0}\right)^{1/\eta}}{\Gamma_{S,t_0}}$$

and, without loss of generality, note we can normalize Γ_{S,t_0} to unity. Thus, base-period shadow labor is simply

$$N_{S,t_0} = \left(Y_{[S/F],t_0} Y_{F,t_0}\right)^{1/\eta}. \quad (18)$$

By inspection, all the terms in the right-hand side of the expression above are observable, except for the value of η . We now explain how to identify this parameter.

Pinning down the value of η

We use the equilibrium condition (6) to solve for η ; as condition (6) holds for every period, it should hold for t_0 as well. Thus,

$$\eta = \frac{(1-\alpha)\left(1-\tau_{t_0}\right) K_{t_0}^\alpha \Gamma_{F,t_0}^{1-\alpha} N_{F,t_0}^{-\alpha}}{\Gamma_{S,t_0}^\eta N_{S,t_0}^{\eta-1}} = \frac{(1-\alpha)\left(1-\tau_{t_0}\right) Y_{F,t_0} N_{F,t_0}^{-1}}{Y_{S,t_0} N_{S,t_0}^{-1}},$$

where the second equality uses the definitions of formal and shadow output, conditions (8) and (9). Rearranging the negative exponents gives

$$\eta = \frac{(1 - \alpha)(1 - \tau_{t_0})Y_{F,t_0}N_{S,t_0}}{Y_{S,t_0}N_{F,t_0}}.$$

Now, equation (17) implies that $Y_{S,t_0} = Y_{[S/F],t_0} Y_{F,t_0}$; using this expression in the denominator of the expression above—and (18) in the numerator—we get

$$\eta = \frac{(1 - \alpha)(1 - \tau_{t_0})Y_{F,t_0} \left(Y_{[S/F],t_0} Y_{F,t_0} \right)^{1/\eta}}{Y_{[S/F],t_0} Y_{F,t_0} N_{F,t_0}}$$

or, simplifying the numerator,

$$\eta = \frac{(1 - \alpha)(1 - \tau_{t_0})Y_{[S/F],t_0}^{(1-\eta)/\eta} Y_{F,t_0}^{1/\eta}}{N_{F,t_0}}. \tag{19}$$

Note that equation (19) is one equation in η . As in the original paper, we use a fixed-point algorithm to find the value of η that solves the equation:

Algorithm 1 (Fixed point calculation of η)

1. Pick a base year t_0 and obtain data for the tax rate τ_{t_0} , the shadow-to-formal output ratio $Y_{[S/F],t_0}$, formal output Y_{F,t_0} , and formal labor N_{F,t_0} .³
2. Pick values for $\{\alpha, \sigma, \chi\}$. Set a value $M \gg 0$ and build a grid with M elements over the interval $[\eta_L, \eta_H]$, where $\eta_L \geq 1$ (see Remark 5.4 in Solis-Garcia and Xie 2018); call this set N .

3. From condition (7) and our assumption of period-by-period government budget balance, we get

$$\tau_{t_0} = \frac{G_{t_0}}{K_{t_0}^\alpha (\Gamma_{F,t_0} N_{F,t_0})^{1-\alpha}} = \frac{G_{t_0}}{Y_{F,t_0}}.$$

(Note that we have already imposed $\rho = 0$.) These terms can be easily backed out from real-world data (recall that G_t is endogenous in our model).

3. For every $\eta_m \in \mathbf{N}$, calculate η following equation (19).
4. Find the entry in \mathbf{N} where $|\eta_m - \eta|$ is minimized; call this value η^* .

Given η^* , our calculated value of shadow labor in period t_0 (hereafter, N_{S,t_0}^*) follows equation (18):

$$N_{S,t_0}^* = \left(Y_{[S/F],t_0} Y_{F,t_0} \right)^{1/\eta^*}.$$

Finding total and shadow labor for the full sample

From condition (3), it's clear that total labor for the base year is

$$N_{t_0} = N_{F,t_0} + N_{S,t_0}.$$

Given N_{S,t_0}^* and the observed value for N_{F,t_0} , this value is easily obtained. Now recall equation (15):

$$g_{Nt} = \left(g_{Ct}^{-\sigma} g_{Y_{Ft}} g_{N_{Ft}}^{-1} \right)^{1/\chi}.$$

All the growth rates on the right-hand side of the expression are observable; from here we can calculate the trend Γ_{Nt}^* (normalized so that $\Gamma_{N,t_0}^* = 1$) and then obtain the full series of total labor, $\{N_t^*\}$, as

$$N_t^* = \Gamma_{Nt}^* N_{t_0}^*.$$

Of course, we can use condition (3) again to obtain the values for shadow labor, $\{N_{St}^*\}$, using that

$$N_{St}^* = N_t^* - N_{Ft}.$$

Finding shadow productivity and output for the full sample

Recall equation (16):

$$g_{St} = \left(g_{Y_{Ft}} g_{N_{Ft}}^{-1} g_{N_{St}}^{1-\eta} \right)^{1/\eta}.$$

Given the set of shadow labor values $\{N_{St}^*\}$, all the values in the right-hand side of the equation are available to us. This allows us to calculate the sequence of shadow productivity rates $\{g_{St}^*\}$ and, in a second step, the trend $\{\Gamma_{St}^*\}$ (normalized so that $\Gamma_{S,t_0}^* = 1$). With these values at hand, we obtain the full series of shadow output $\{Y_{St}^*\}$ following the definition of shadow sector output, equation (9):

$$Y_{St}^* = \left(\Gamma_{St}^* N_{St}^* \right)^\eta.$$

Concluding remarks

In their comment, Gómez and Ríos-Blanco (2022, abs.) argue that “the [Solis-Garcia and Xie 2018] method does not serve its purpose because once correctly implemented it generates a constant ratio of shadow to formal production, which is equal to a base-year estimate taken from an external source.”

In this response, we prove that the authors’ assessment is unfounded. While we agree that our 2018 paper is affected by the issues detailed in their comment, the purpose of our approach—to provide a measure of the shadow output—remains intact and is still based on the cross-equation restrictions of a DGE model. Our updated procedure is an improvement of the 2018 given the same model and trend, but asks for less restrictions.⁴

References

- Gómez, Manuel A., and Adrián Ríos-Blanco.** 2022. Comment on Measuring the Size of the Shadow Economy Using a Dynamic General Equilibrium Model with Trends. *Econ Journal Watch* 19(1): 124–132. [Link](#)
- Solis-Garcia, Mario, and Yingtong Xie.** 2018. Measuring the Size of the Shadow Economy Using a Dynamic General Equilibrium Model with Trends. *Journal of Macroeconomics* 56(1): 258–275.

4. A detailed version of this response will be available shortly at the corresponding author’s website.

About the Authors

Mario Solis-Garcia earned his Ph.D. in economics at University of Minnesota in 2011 and is now an associate professor of economics at Macalester College. His current research interests concern macroeconomics, in particular, the effects of government policy over the business cycle and in the determinants of the shadow economy. His email address is msolisga@macalester.edu.

Yingtong Xie is an assistant professor of economics at Carleton College. Her email address is yxie@carleton.edu.

[Go to archive of Comments section](#)

[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6050>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 142–165

Liberalism in Colombia

Sebastián Rodríguez¹ and Gilberto Ramírez²

[LINK TO ABSTRACT](#)

From at least the lead-up to independence from Spain in 1819 and the subsequent organization of Colombia as a republic, liberalism has been alive in many Colombian thinkers and leaders. Colombia today has a political party that was formed in 1848 and that still bears the name “liberal.” The Colombian Liberal Party began with significant liberal bearings but moved leftward during the twentieth century. As one of the largest parties in Congress, it is generally inclined toward the governmentalization of social affairs, contrary to its original tendencies. Notwithstanding the Liberal Party’s shift, liberal ideas remain influential in the public debate.

This article focuses primarily on the economic aspects of liberal thought in Colombia. Notably, that includes the economic debates that have taken place since the beginning of the nineteenth century on the topics of global and national trade integration, a modern monetary and banking system, and government intervention in the economy. The Colombian story bears likeness to those of other Latin America countries, some of which are told in the present series: Venezuela (Faria and Filardo 2015), Guatemala (Marroquín and Thomas 2015), Mexico (Kuchař 2016), Ecuador (Romero, Hodgson, and Gómez 2018), Peru (Saenz-Armstrong 2018), and Brazil (Berlanza 2020).

1. Graduate student, George Mason University, Fairfax, VA 22030. The authors are very grateful to Fergus Hodgson for editing and feedback, to two anonymous referees, and to participants of the Invisible Hand Seminar at George Mason University for the valuable comments and discussion.

2. National University of Colombia, Bogotá 111321, Colombia.

trade balance.

At the beginning of the nineteenth century, Colombia consolidated under the government of the viceroyalty of New Granada.³ José Ignacio de Pombo (1761–1812) and Antonio de Narváez (1733–1812) stated their concerns about the fiscal policies of the Spanish crown. Their critique focused on the tendency of the crown, engaged in a war against Great Britain, to demand more resources and impede local trade. De Pombo and de Narváez both resided in the port city of Cartagena, the main commercial port of the viceroyalty of New Granada. Cartagena was one of the most important ports of the Spanish crown in the Caribbean Sea and one of the most affected by the crown reforms regarding international commerce.

The selection of Cartagena as headquarters for the Consulate of Commerce confirmed its commercial importance in 1795. Cartagena merchants defended their economic interests via the Consulate, which also addressed concerns of Spanish crown officials such as combating smuggling and fraud to the treasury. The Consulate expounded on the advantages of having a trading system not limited to Spaniards. In reports put before the Consulate, a series of viceroys of New Granada serving between 1773 and 1810—including Manuel Guirior (1708–1788), Manuel Antonio Flórez (1722–1799), Antonio Caballero y Góngora (1723–1796), José Manuel de Ezpeleta (1741–1823), Pedro de Mendinueta (1736–1825), and Antonio José Amar y Borbón (1742–1819)—implied that the expansion of trade would only bring advantages.

However, the crown refrained from taking measures favoring free trade for fear that such measures would favor the colonies' demands for autonomy. In general, crown officials followed the rule implemented under King Carlos III of allowing inhabitants of his domains to trade freely with each other but not with nonsubjects.

The political crisis of 1808 confirmed the crown's fears about its colonies, which led to the abdication of the House of Bourbon dynasty in favor of José Bonaparte, Napoleon's brother. The compromised loyalty of subjects in Europe and the Americas gave rise to divergent positions between the monarch and his subjects.

In the viceroyalty of New Granada, several of the Spanish king's subjects acted to resolve the dilemmas of to whom they owed loyalty. They appointed provisional authorities whose autonomy in some cases quickly evolved—as in the city of Cartagena in November 1811—to a declaration of independence and the establishment of a new government. Cartagena would make its own decisions in

3. The viceroyalty of New Granada was created under the reign of Felipe V of Spain in the year 1739 and covered the territories of the current republics of Panama, Colombia, Ecuador, and Venezuela.

fiscal, monetary, and commercial matters.

The rest of the viceroyalty followed Cartagena's example. However, many of the first economic measures of the new authorities were limited to gathering sufficient resources to finance militias that sought to secure the newly achieved independence.

Armed confrontations between the Spanish army and the revolutionary troops who sought independence broke out between 1812 and 1824 in the New Granada territories. The independence wars and conflicts consumed considerable wealth and human lives and ended with victory for the revolutionary troops. The new Republic of Gran Colombia was established in 1819. It sought to rule over the entire jurisdiction of the old viceroyalty of New Granada.

Impelled by weak finances, the new government resorted to loans from abroad and forced contributions from its citizens. The debt, duties, and challenges of achieving a stable polity shaped many of the coming discussions and debates in the new republic.

One of the first finance ministers of the new republic was José María del Castillo y Rada (1776–1833) from Cartagena. He struggled with the paradox of how to pay off the republic's financial obligations. To do so would require that it maintain old taxes, add new ones, and tax more efficiently. Private enterprises could then perform without significant obstacles. An example of this paradox was the decision to keep intact the *estanco*, a colonial tax set by the government for sales of products for which the government also set the price, such as tobacco, aguardiente (a popular liquor), and salt.

In the debates and discussions that gave rise to the constitution approved in the city of Cúcuta (which governed between 1821 and 1830), some considered the possibility of taxing people's income under a direct and progressive tax, as well as the creation of a national bank that would allow the consolidation of public debt. Neither measure was implemented, however. Meanwhile, many old tributes from the colonial era remained unchanged in the national economy.

The vast territorial jurisdiction now known as Gran Colombia disintegrated and gave rise to the republics of Venezuela (1829), Ecuador (1830), and New Granada (1832). The external debt contracted by the old republic was distributed among the three countries, with half assigned to New Granada. Between 1830 and 1845, the New Granada governments limited themselves to stabilizing their finances while under a debt that far exceeded the annual budget, which meant maintaining a good part of the old Spanish fiscal regime.

Between 1825 and 1829 the former republic of Gran Colombia signed mutual-recognition treaties that favored free trade with the governments of Great Britain, the United States, and the Netherlands. New Granada continued to import more than it exported.

Not until President Tomás Cipriano de Mosquera (1798–1878), who presided over New Granada between 1845 and 1849, did a more significant commercial opening begin. The tobacco *estanco* was eliminated, allowing free cultivation and producing an export boom. One of the leading promoters of the measure was Florentino González (1805–1874), who served as secretary of the treasury to Mosquera and stood out as one of the most tenacious defenders of economic liberalism in New Granada. Free cultivation served a robust European demand for tobacco. Several entrepreneurs took the risk of acquiring land for tobacco cultivation and participated in this bonanza, which continued for more than two decades.

Along with the commercial boom in the making, a new generation of young politicians began to stand out in public opinion with proposals for change. For this generation, the promises of the independence movement were still unfulfilled, including the dismantling of lingering protectionist policies. The measures advanced by González began to excite the new generation, which had been influenced by the lessons of political economy, now taught in the classrooms and the press. Ezequiel Rojas (1804–1873) was one of the leading expositors of economic thought.

Along with González, Rojas was accused of being part of the assassination attempt on Simón Bolívar in September 1828. They were acting on the perception that the recently proclaimed Bolívar dictatorship was the establishment of a tyranny. Both González and Rojas, then, acquired reputations as conspirators.

With the end of the Bolívar government, Rojas entered a teaching job that, for more than three decades, would place him as the dean of political economy in New Granada, a job that he mixed with his political activism. Using the work especially of Jean-Baptiste Say (1767—1832) and Jeremy Bentham (1748—1832), Rojas structured a series of lessons later published under the title of *Moral Philosophy*. In that text, he favored liberal notions of liberty as consistent with reason and human nature (Rojas 1868). In July 1848, Rojas authored a politically influential article in the newspaper *El Aviso* entitled “The Reason for My Vote.” That title refers to his support for José Hilario López’s candidacy and his support for what would become the Liberal Party, which initially proposed a program based on classical liberalism.



Ezequiel Rojas

TABLE 1. Name and territorial composition of Colombia since 1819 to present

Year	Name	Present-day countries
1819–1830	Gran Colombia	Colombia, Ecuador, Panama, Venezuela, and portions of Brazil and Peru
1830–1858	New Granada	Colombia, Panama, and portions of Brazil, Ecuador, and Peru
1858–1863	Confederation of Granada	Colombia, Panama, and portions of Brazil, Ecuador, and Peru
1863–1886	United States of Colombia	Colombia, Panama, and portions of Brazil, Ecuador, and Peru
1886–1903	Republic of Colombia	Colombia and Panama
1903–present	Republic of Colombia	Colombia

The liberal period

During the second half of the nineteenth century, when liberal thought reached a more tangible form, many Latin American countries experimented with liberal reforms, and liberal-leaning political movements gained power. The liberal reforms or aspirations during this period were often reflected in formal constitutions and laws. In the Mexican case, for example, liberal aspirations were reflected in the 1857 federal constitution (Kuchař 2016).

The liberal period in Colombia was more extensive than in most other Latin American countries in the nineteenth century. The period includes, among other things, some years of free banking, the abolition of slavery, partial reductions in tariffs, federalism, religious freedom, and the separation of church and state.

From 1850, though, political ideas constantly clashed, and different groups fought one another with words and weapons alike. There were six civil wars between 1850 and 1886. As in other Latin American countries, Colombia's liberal period was mixed, troubled, and short-lived. The most liberal constitution of 1863 was replaced 23 years later, in 1886.

The liberal period started formally with José Hilario López winning the presidential election of 1848. López's program was characterized by proposals that emphasized individual liberties, the rule of law, and meritocracy in public office and employment. During López's presidential term, many young liberals started to participate in the political process with the firm conviction that the country needed to leave the colonial era behind. These young liberals were primarily influenced by political ideologies in France, but the influences of England and the United States were also important (Safford and Palacios 2002). At the same time, intellectual works like those by Smith, Bentham, Say, David Hume, John Locke, and John Mill began to permeate the minds of some of the influential members of society and helped to configure a narrative around the concepts of economic and individual

freedom.

Special mention must go to Frédéric Bastiat (1808–1850), the famous disseminator of liberal ideas in France that impacted many countries in Europe and the Americas. One of his works, *Economic Sophisms* (Bastiat 1848), was republished in New Granada. Another was translated and expanded as *What Is Seen and What Is Not Seen: Political Economy in a Lesson, Given to His Countrymen in France by Frédéric Bastiat and Interpreted and Offered to His People in Zipaquirá by Eustacio Santamaría* (Bastiat 1853). As in other Latin American countries, the liberal spirit imbued a new generation eager to change colonial traditions and, in their view, bad administrations. Compared to prior generations, this new generation was more educated and more in touch with international associates and events. Many had gone to schools organized in the republican era of the country, following a non-colonial curriculum (Bushnell 1996).

Ezequiel Rojas's endorsement of José Hilario López, expressed in "The Reason for My Vote," is usually seen as the starting point of the Liberal Party in politics. In the article, Rojas summarizes the policy demanded by a party that calls itself liberal as follows:

In short, the Liberal Party wants a government organized to benefit the governed. It wants a republic, a truly representative system, an independent congress, an executive branch with powers limited by law and responsible for an independent judiciary, good laws, and an unmistakably national and American executive power with impartial justice for all, which by its actions takes into account nothing but the public benefit. It wants all this so the obedient will not be slaves to those who govern; so there can be true liberty; so we can liberate ourselves from theocratic government; so those of Granada can truly ensure ownership of themselves and their properties; so these guarantees will not be broken promises. (Llano Isaza 2009, translation by Fergus Hodgson)

Colombia's liberal period gained that name not from a description of the general ideas in the country but from the ruling party. At its start, the Liberal Party united figures who endorsed a liberal tradition and were influenced by Bastiat and other French liberals. In the period of 1844–1880 or so, characterized by the rule of the Liberal Party, we can see that the policies were focused on two major topics, namely the relationship between the church and the state and the country's economic policies. Some of the policies followed a true liberal spirit, in the Smithian sense of "allowing every man to pursue his own interest his own way," within the bounds of commutative justice. However, others fit more into unruly polity struggles over power, influence, and stability, with interest groups playing major roles.

The church and the state

For some Liberal Party members, the Catholic Church represented the old colonial tradition. The church was seen as an institution opposed to the aims and values of the Liberal Party, specifically freedom of religion and freedom of speech. The church was seen, or at least portrayed, as supportive of the Kingdom of Spain. Among the new government's first objectives was reforming the relationship between the church and the state.

The Jesuit order, which had been expelled from the Spanish colonies by King Carlos III, had returned to the country in 1844. The Jesuits were a permanent target of the Liberal Party. One of its first actions was the expulsion of the Jesuits from the country. Frank Safford and Marco Palacios (2002) mention the expulsion as evidence of French influence among the Liberal Party partisans. Eugene Sue's novel *The Wandering Jew*, openly anti-Jesuit, was well-known among liberals.

In 1851 the Liberal Party gained control of the legislative power and passed a law considered radical by a large part of the population. The new law ended the *fuero*, which had accorded to priests the privilege of being judged in ecclesiastical courts, so that priests would be judged in secular court. The law gave the members of the city councils a role in the selection of assigned priests. And in what amounted to a confiscation of church loans, it allowed the owners of property indebted to the church to free themselves of the obligation by paying half of the value to the government instead of to the church (Safford and Palacios 2002). This last concerned the *censo*, a form of mortgage with the church as lender. It allowed the colonial church to become an essential source of financing when lenders were limited or non-existent. The *censo* helped the church to become one of the biggest lenders in the region, as well as one of the biggest landlords. It also helped many merchants to finance their operations (Mena García 1996).

In the 1859 revolution, the church openly favored the Conservative Party. In the 1860s, when the more radically anti-clerical wing of the Liberal Party gained office, their measures were extended to include the confiscation of church properties. Under Tomás Cipriano de Mosquera measures were especially severe and took on a punitive character. In 1861 land ownership by the church (derisively described by detractors as mortmain) was abolished, and a redistribution of the church's land took place. Supposedly the idea was that the confiscation would lead ultimately to better use of the land, and with that in view, sales revenue for the confiscator and greater social product. To many people, the confiscation measures were retaliation against the church for having backed the Conservative Party.

The Liberal Party government's assault on the church was a major misstep for two reasons. First, the radical posture helped the conservatives construct a

narrative of liberal values being unsuited to a Catholic country. Second, in some cases it showed a replacement of liberal ideals for political convenience. Sadly, this departure eventually led to the complete separation of the Liberal Party from liberty itself. This clash shows some of the dilemmas and paradoxes in the union between liberal ideals and real-world politics.

The expulsion of the Jesuits obviously tramples basic freedom, and in particular assaults freedom of sect and religion, a cornerstone of the arc of liberalism. Also, some of the confiscations against the church did not even live up to fiscal promises (Jaramillo and Meisel 2008). When executing the liquidation of confiscated church properties, the amounts collected were significantly lower than expected, though still important for the government. At the same time, most of Mosquera's aggressive reforms faced internal opposition from other Liberal Party members, like the faction called Radicals, which favored *laissez-faire* with regard to religion, in the manner of the United States, and opposed Mosquera's confiscation process.

Economic reforms

The liberal period in Colombia was marked by two economic reforms: territorial political integration and the encouragement of exports.

First, the government supported the political integration of lands around the Magdalena River, the main river in Colombia and the most direct outlet to the sea and therefore to international markets. The process of expanding political integration of territories caught on quickly (García de Kausel 2013).

Second, the government promoted an economic model focused on exports. The primary sources of exports continued to be agricultural, and to a large extent local growth did occur. It is important to remember that many parts of the country were unavailable and inaccessible because of geographical barriers. The political integration of specific regions of the country created new sorts of groupings of the populace and a distinctive working class.

At first, many of the young members of the Liberal Party were defenders of the free-market system. Although many of the export policies were oriented under that ideal, as time under the Liberal Party administrations passed, the idea of the market over the government began to fade. The Conservatives and Liberals at that time had the same ideas for how to incentivize the country's economy. Irrespective of party, the majority believed in dismantling the colonial tax system centered on monopolies (tobacco and *aguardiente*), which constituted a significant obstacle to private activity. Both parties also endorsed the idea that economic development would depend on the growth of international trade (Melo 2015).

Laissez-faire policies and the endorsement of free commerce were present in the popular mind and were not exclusive to members of the Liberal Party.

Since 1845, public opinion had become more market-oriented. Newspaper topics seemed to change from exclusively political news to incorporate economics and business (Safford and Palacios 2002). The focus on economic issues perhaps helped the Liberal Party governments in office, but some of the policies were not significantly implemented in reality. Tariffs took many years to disappear, for example, as the government had no other significant sources of revenue.

Partly from the policies of openness and colonization, New Granada began growing exports. The main export products were tobacco and cinchona bark. The regional entrepreneurs never stuck with one product and were more focused on the international price changes between products, moving between products with high prices. Despite the reduction in taxes, the central government found an increase in customs revenue, stimulated by the export boom of 1850–1875, such that in the 1860s collection levels had returned to those of the 1840s (Melo 2015). Export growth was present between 1875 and 1898, with a recession between 1888 and 1891 (Ocampo 2013). The climate of economic strength during a big part of the period was favorable with the acceptance of market-oriented policies and the laissez-faire vision of many Liberal and Conservative elites.

The 1863 constitution

The liberal spirit of the ruling Liberal Party peaked with the new constitution of 1863, known as the Rionegro Constitution. The document ruled the newly established federal country of the United States of Colombia between 1863 to 1886. It was a formal body of laws that consolidated and amplified some of the previously mentioned reforms and was libertarian in spirit. It established a federalist government with highly independent states and individual liberty as a fundamental goal. The role of the central government was limited to a few functions, and individualism was highlighted over collectivism. The constitution declared in its article 15, section three, that:

Individual freedom is guaranteed, with no limits other than the freedom of another individual; that is, the power to do or omit everything whose execution or omission does not injure another individual in the community. (National Constituent Assembly 1863, our translation)

The constitution also incorporated notable elements such as freedom of thought, press, and religion, the freedom to travel without a passport through the

territory and leave it without restrictions, and the freedom to carry and trade arms. It was inspired by the constitution of the United States of America. In practice, however, the implementation of the constitution faced problems, and the political struggle was more substantial than the liberal ideals.



United States of Colombia (Erhard Schièble 1883)

Colombia's federalization process was framed in a period of continuous war and revolution. The federal government used its power to punish the states that were opposed to the central government. For example, after the Conservative government of Ospina Rodríguez, which had ended in a civil war won by Liberals, the new government removed all governors of Conservative affiliation and installed Liberal ones. The action produced unrest in states with a conservative majority such as Antioquia. The tensions created a continuous animosity between states and the central government. The problems between minority and majority parties were widespread because, as some scholars have shown, electoral fraud was widespread, as regions controlled by one party often showed suspiciously high vote counts for that party (Bushnell 1996).

Many critics started a campaign to support a single strong command, centralizing power. Thus dissatisfaction with the central government actually tended to erode the institution of federalism. To make the situation worse, internal

conflicts of the Liberal Party led to division between factions. Political booty began to appear as the central factor in electoral contests. Corruption and irregularities in government enterprises increasingly appeared in the period. Liberal spirits were often broken by participation in governing and electoral politics, leading to compromise. For example Manuel Murillo Toro, who in 1857 had argued that the state should not take responsibility for road-building, supported government participation in public works during his presidency and became a full advocate of it by the 1870s (Safford 1988).

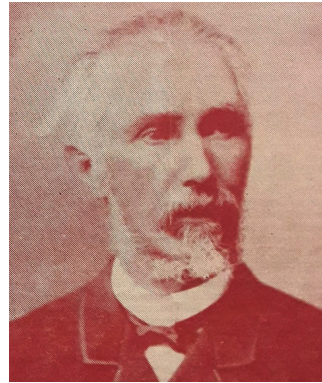
After the so-called liberal period, the Conservatives gained office in 1880, led by Rafael Núñez, and started a process known as “la Regeneración.” Many of the liberal reforms were overturned, including the 1863 constitution, producing a drastic change in the structure of the state. The Catholic Church also benefited from the change in government and started a recovery process. In 1888, the Conservative Party government signed a concordat with the Vatican that exemplifies the conservative vision and reflected the Colombian people’s relationship with the church. The first lines are as follows:

The Catholic, Apostolic, and Roman religion is the religion of Colombia; the public powers recognize it as an essential element of social order, and they are obliged to protect it and make it respected, the same as the ministers, preserving it, at the same time in full enjoyment of rights and prerogatives. (*Concordato Celebrado Entre La Santa Sede Apostolica y El Gobierno de La República de Colombia, y Otros Documentos Eclesiásticos y Civiles, Helguera, J. León Collection 1895, our translation*)

The Regeneración period was not exempt from economic debate. On the one hand, most of the reforms proposed by Núñez were seen as statist and interventionist, in contrast to liberal policies that were often of a laissez-faire nature. One of the significant debates was about the banking system. As of 1880, all Colombian banks were private, and their notes were widely accepted, including in the payment of taxes. Nevertheless, Núñez wanted to establish a national bank early in his presidency and founded the bank in 1881 by using a New York loan backed with the revenues of the Panama railroad. Private banks were obligated to accept national banknotes, and the government eventually prohibited the circulation of private banknotes. The national bank also got involved in a great controversy about printing money to finance the government’s military and bureaucratic apparatus.

The existence of the free banking system under the liberal constitution shows us the broad vision of liberal concepts of the time and the importance of the economic reforms for the liberals. The reforms of the 1860s allowed the evolution

of the credit system that the Roman Catholic church previously controlled. The Colombian system was stable, promoted competition, and eliminated barriers to the entry of new banks, and although it did not last long enough to reach considerable maturity there were 42 banks in the country by 1882. Probably the most significant period of instability in the free banking period was related to the civil war of 1876, caused by political unrest rather than the operability of the banks (Meisel 1992). The creation of a national bank was not prompted by failings in the system of private banks but rather corresponded more to the need of the government to finance its debts



Miguel Samper

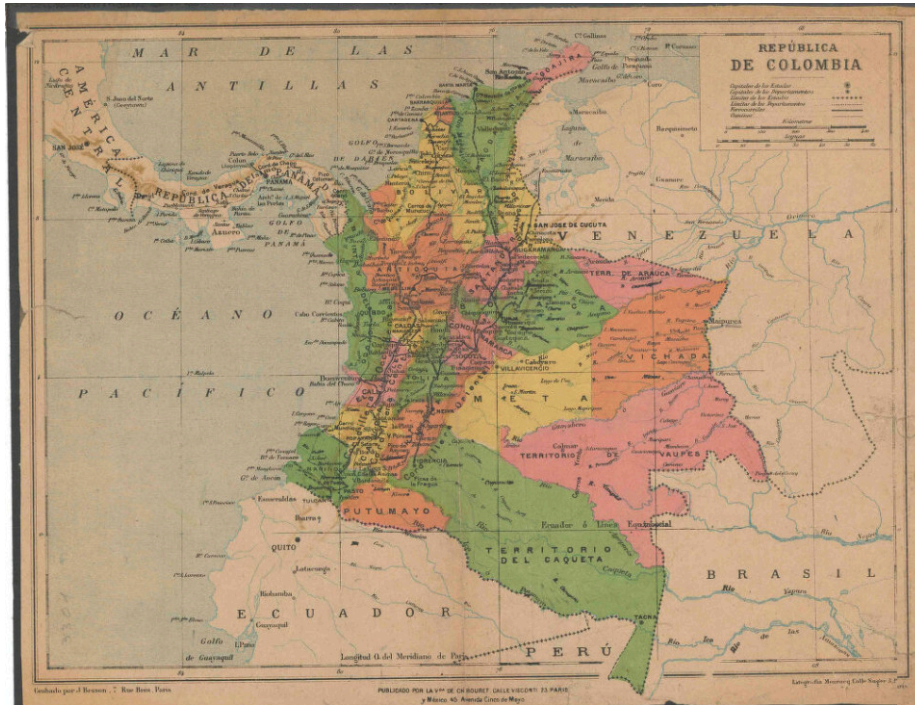
(Echeverri 1994). Two notable liberals, Miguel Samper and Salvador Camacho Roldán, stood out as defenders of free banking and opponents of the national bank, participating in an ongoing debate with President Núñez and Miguel Antonio Caro. Samper and Camacho both argued that paper money would end up being a source of government financing and that the government would eventually be granted monopoly control over emission. Time has proved those liberals right. In 1886, an unprecedented monetary crisis led to an inflationary spiral and eventually hyperinflation, problems attributed to the national bank (Echeverri 1994). The free banking experience in the country seems to have been positive, without a major crisis or instability of the system. Several proposals were discussed in Congress before 1920 to implement the free-banking system again, but none of them became law (Meisel 1992).

The twentieth century and the rise of interventionism

The worst civil war in Colombia's history, the War of a Thousand Days, began in 1899 and ended in 1903. It left between 60,000 and 100,000 dead in a country of about 4 million (Luzardo-Luna 2019). The conflict was a battle between the Liberal and Conservative parties.

The abrupt end of the liberal reforms with the *Regeneración* in 1885 had worsened the relationship of the two dominant political parties and fueled the animosity leading to the War of a Thousand Days. After the war ended, the political discussion was exhausted. Both parties recognized the catastrophe and the futility of the violence. After three years of turmoil, the country was submerged in a fiscal

crisis and faced disruption in the labor system and production. Postwar politics was also affected by the U.S.-backed separation of Panama in 1903 and subsequent efforts by Colombians to recover the lost territory (Bergquist 1986).



Republic of Colombia, 1920

After the war, the Liberal Party would not have a president elected until 1930, but from 1910 the nature of the government was bipartisan in composition. The debate had calmed down, and topics like the relationship of church and state became less pertinent. The climate of economic opinion continued to support the import-export model, but the growing popularity of coffee dynamized the exports cycle and contributed to a rise in economic growth as well as an increase in government revenue. From 1905 to 1925, exports of 60-kilo sacks of coffee rose from 500,000 to 2 million. At the same time, the price of coffee grew on international markets, guided by an increase in worldwide demand (Beyer 1947).

The role of coffee is fundamental to understanding the history and development of Colombia. However, it is also important to understand the generally positive attitude of the people toward external markets and trade. The coffee boom fostered entrepreneurial activity in the areas where coffee was cultivated. Unlike other countries such as Brazil, where cultivation was concentrated in the hands of big landlords, Colombia sprouted small- and medium-sized coffee farms that

contributed to economic dynamism. In the 1932 census, more than 150,000 coffee farms were registered in the country (Bergquist 1986).

While coffee farms grew and international prices rose, the country's government also got bigger. The war was fresh in people's minds, and radical liberal reforms were seen as something of another time. Many of the Liberal Party members were in tune with what was happening in other parts of the world and began to replace classically liberal perspectives with socialist and progressive ones.

Other figures, like Carlos Eugenio Restrepo, who won the 1910 presidential election, searched for moderation in government interventionism. Restrepo, who previously was influential in the configuration of the Republican Union—a political party that championed the ideals of republican government, religious tolerance, and laissez-faire economics—was previously a member of the Conservative Party. However, his vision of the country was compatible with classical liberal principles. His influence was significant during this period. He was also a minister in the administration of Enrique Olaya Herrera in 1930, the first Liberal Party president after the war.

A mission, led by the economist Edwin Kemmerer, reformed the banking, monetary, and fiscal systems of Colombia, Chile, Ecuador, Bolivia, and Peru and helped establish new central banks in those countries. The Kemmerer mission implemented Colombia's new central bank in 1923, but there had been debate about creating a central bank for more than a decade prior (García de Kausel 2013). The central-bank project was amply criticized by Restrepo and some of his former ministers, such as Tomas Eastman, based on arguments similar to those of Samper some decades earlier. Restrepo spoke of “a large group of monopolies in favor of the Central Bank” and that it would “prohibit other entities from any legal business in the same monopolized industries or branches” (García de Kausel 2013). Eastman (1923) commented: “If this project is realized, the existing banks will have to be liquidated today since they will not be able to enter as shareholders of the new bank, nor will they be able to subsist independently of it” (our translation). Despite these liberal arguments against the project, the central bank was approved.



Carlos Eugenio Restrepo

Postwar years

After World War II, and true to the spirit of the times worldwide, came years of state intervention in the Colombian economy. Colombia had already been imposing exchange controls and development credits to support an industrialization model along the lines of ‘import substitution,’ the policy in vogue across Latin America.

One of the Colombians decisively supporting the rise of state intervention in the economy was Carlos Lleras Restrepo (1908–1994)—no relation to Carlos Eugenio Restrepo. An intellectual member of the Colombian Liberal Party, Lleras Restrepo was head of the Ministry of Finance between 1938 and 1942 under the government of Eduardo Santos, delegate to the Bretton Woods monetary conference in 1944, delegate to the commercial conference in Havana in 1947, and finally president of the country from 1966 to 1970. An influential disciple of the economist and several times finance minister Esteban Jaramillo, Lleras Restrepo believed that state intervention was vital both in extraordinary moments and in general economic development.

Two of Lleras Restrepo’s main economic policies were the Exchange Statute of March 1967 and the constitutional reform of 1968. The Exchange Statute, also known as Decree-Law 444, unified various exchange rates and established import controls and export-promotion measures. It replaced the multiple exchange rates with a single fluctuating exchange rate to implement small, gradual ‘crawling peg’ devaluations. The plan sought to keep exports favorable compared to imports without suffering external shocks that would trigger inflation. Inflation increased, nonetheless, through gradual devaluation and growing public spending.

The constitutional reform of 1968 aimed to strengthen state planning mechanisms with so-called indicative planning schemes. This included the preparation, design, and execution of a National Development Plan to be implemented for four years, the duration of a presidential term. Such a measure continues to operate today in Colombia and consists of a national budget execution plan focused on ‘priority’ sectors. The interventionist spirit of the reform is summarized in article 32 of the then-operative Political Constitution of Colombia:

Free enterprise and private initiative are guaranteed within the limits of the common good, but the general direction of the economy will be in the care of the government. By mandate of the law, it will intervene in the production, distribution, use, and consumption of goods and in public and private services to rationalize and plan the economy to achieve integral development.

The state will also intervene, by mandate of the law, to give full employment to human and natural resources, within an income and salary

policy, according to which economic development has as its primary objective social justice and harmonious improvement of the community, and the proletarian classes in particular. (Legislative act 1 of 1968, our translation)

The rise of interventionism

The postwar climate of opinion is well represented by the businessman and intellectual Hernan Echavarría Olozaga (1911–2006), who had studied economics at the London School of Economics. Echavarría published a text in 1948 entitled *Full Employment and Other Issues*, applying John Maynard Keynes's thought to Colombia (Echavarría Olózaga 1948). Echavarría would add more far-reaching reflections in his best-known work, *Common Sense in the Economy* (1958). In that book he explains the importance of considering multiple effects that occur in the economy when proposing a particular economic policy.

Echavarría claims to defend both private companies and the democratic nature of taxes to achieve investments of general utility. He endorses precise state intervention while respecting private property and free enterprise (Echavarría Olózaga 2003/1958, 154, 214–215, 331–340). A rejection of socialism is explicit in Echavarría, as socialism employs means unsuited to the goal of social betterment and violates individual liberties (*ibid.*, 306–330).

Three other figures representative of the time, Enrique Caballero Escovar (1910–1994), Hernando Agudelo Villa (1923–2010), and Carlos Sanz de Santamaría (1905–1992) offered defenses of freedom supposedly compatible with gradual and moderate state interventionism led by competent and well-trained personnel. The books *Capitalization and Employment* (Caballero Escovar 1959), *Towards a Modern Liberalism* (Agudelo Villa 1968), and *Silent Revolution* (Sanz De Santamaría 1971) each summarize its author's thinking after his experiences at the forefront of economic policies led by the government or from a business union or official multilateral entity. Caballero served as an advisor to the National Association of Industrialists (ANDI), Agudelo was president of the National Federation of Merchants (FENALCO) and minister of finance in the government of Alberto Lleras Camargo between 1958 and 1961, and Sanz de Santamaría was minister of finance under the government of Guillermo León Valencia between 1962 and 1964 and president of the inter-American committee of Alliance for Progress, the Latin American development aid program promoted by U.S. president John F. Kennedy from 1961.

The climate of interventionist opinion became the rule of the central economic policies to be followed by the governments of the so-called National

Front, a governing coalition of the Liberal and Conservative political parties that between 1958 and 1974 sought to guarantee economic development as a primary obstacle against the revolutionary theses of socialism. Caballero affirmed ‘capitalization’—that is, capital-intensive investments—as a fundamental requirement for employing a growing, economically active population. Unemployment would spell political instability and social discontent, a situation favorable to leftist revolutionary impulses (Caballero Escovar 1959, 19–34). Agudelo (1968, 39–61, 121–132) bet on technocratic economic guidance to supposedly lead to egalitarian transformations, thus disrupting radical socialist or communist ideas. Finally, Sanz de Santamaría (1971, 239–246) maintained that the true revolution of transforming people’s quality of life in Colombia and Latin America was a silent one: a revolution not dependent on inciting violence.

From the mid-1970s, the climate of opinion gradually changed when recessionary conditions arose under Keynesian-type economic policies and the criticisms of figures such as Friedrich Hayek and Milton Friedman acquired force. In Colombia, a reintroduction of classical-liberal economic thought was aided especially by two think tanks, the Fundación de Estudios para la Libertad, led by Alberto del Corral in Bogotá, and the Instituto de Estudios Sociales y Económicos, led by Juan Delgado Padilla in Cali. Those two men exemplified the businessperson who was also an intellectual. Under the auspices of Del Corral and his think tank, *Ten Lessons in Economics* by Faustino Ballve was reissued in 1963. Ballve was a Spaniard living in Mexico and a friend of Ludwig von Mises, whose text (Ballve Pallicer 2012) would become one of the best introductions to the Austrian School of Economics. Delgado Padilla’s group distributed several brochures and a bulletin between 1962 and 1982 that disseminated ideas of liberal thinkers such as Hayek, Friedman, Ayn Rand, Wilhelm Röpke, Ludwig von Mises, Manuel Ayau, Alberto Benegas Lynch, and Gustavo Velazco among others. The scope and impact of Del Corral and Delgado’s work in both think tanks has yet to be studied in any detail.

Another systematic work was developed by Pablo Eduardo Victoria Wilches (b. 1942), who by 1976 was already forming a proposal for economic liberalism around his text *Towards a Liberal Economy: A Comprehensive Economic Program for Colombia*. Victoria (1976) criticized the conventional countercyclical policies of Keynesianism and their application in Colombia, relying on Friedman and Hayek. Later he deepened these theses in his text *Analytical Macroeconomics*, where he developed a proposal to privatize the issuance of currency under free banking (Victoria 1993, 329–338). Victoria served as a congressman between 1990 and 1998.

In some way, the previous Keynesian consensus came to be tempered and complemented by a recovery of economic wisdom that, according to Hayek in an interview with the Colombian Diego Pizano Salazar given in 1979, seeks to

explain how it is possible to adapt to the unknown (Hayek 2008). The trend toward socialism was not reversed, but there was more awareness of incentives, knowledge limitations, and the role of the price mechanism.

A critique of state intervention in Colombia's economy was fashioned by Jorge Felipe Ospina Sardi (b. 1947), in *Towards a New State Intervention Scheme* (Ospina Sardi 1986). As director of the National Planning Department (DNP), the main entity in charge of nationwide planning, Ospina called for the state to moderate its active role in the economy and make its interventions less damaging. In this way, the state would be:

possibly smaller in size but with a more significant presence in those areas where its action is essential—for example in justice, security, primary health care, drinking water, and minimal infrastructure in regions that do not have it. Its presence would be less in the direct management of commercial and industrial companies, especially in high-risk and complex administration activities. On the other hand, it would be a state more fitting with the community's requirements and more inclined to utilize the private sector to provide services such as cleaning, higher education, and recreation. Finally, a state thus conceived would provide a more stable and predictable frame of reference for developing private initiative, an irreplaceable source of economic growth and social advancement. (Ospina Sardi 1986, our translation)

Another notable contribution came from Enrique Peñalosa Londoño (b. 1954) in his text *Capitalism: The Best Option?*, which is a transcript of a television series with the same name that was broadcast in October 1989. Peñalosa (1989) explains how capital accumulation advances the wealth and well-being of countries and the consolidation of democratic governments. He compares the performance of various Asia-Pacific economies that, with fewer natural resources than Colombia, achieved higher exports and per capita income thanks to better incentives for investment and savings (*ibid.*, 13–32, 79–100). Peñalosa served as mayor of Bogotá from 1997 to 2000 and again from 2015 to 2019.

Since 1987, promotion of liberal public policies has come from the Institute of Political Science (ICP), a think tank that serves as the main affiliate in Colombia of the Atlas Network. The ICP was mainly promoted by Hernan Echavarría Olozaga, the intellectual entrepreneur mentioned above. A member of the ICP, Carlos Lemos Simmonds, was responsible for a publication, *The Thief State*, which incorporated his reflections on the abuses of state power. He suggested that the government preys on the citizens it claims to govern (Lemos Simmonds 1991). Simmonds became a member of the Constituent Assembly that in 1991 gave rise to the constitutional charter currently governing Colombia.

It is important to mention the role of Cesar Gaviria's presidential period,

1990 to 1994. Gaviria, of the Liberal Party, implemented a series of policies oriented toward opening the economy to international markets. The process sought to relax many of the protectionist policies and to induce economic actors to compete globally. Such market-oriented reforms were common in several Latin American countries during the 1980s and 1990s. As in other countries, Colombia's reforms were political and left much to be desired yet can mostly be counted improvements (Edwards and Steiner 2008). Gaviria's reforms included a gradual elimination of tariffs on imports, a new tributary system, a banking reform that allows banks to diversify and expand their offerings, and incentivizing foreign investments. Gavira's administration also tried to reduce the size of the state by reforming many government institutions. The reforms had a significant impact, measured by the increase in exports and imports the country experienced in the period. Some scholars suggest that the growth from the reforms was not as significant as expected for different reasons, including several sources of sustained violence and the 1998 financial crisis. The liberal reforms of Gaviria mark a point where Colombia became integrated in international markets.

In the last part of the twentieth century, Sergio Arboleda University contributed to the dissemination of economic liberal ideas. The first dean of the School of Economics, Mario Jaramillo Contreras, helped to include Colombia in the first indexes carried out by the Fraser Institute to measure economic freedom. In that index, Colombia has maintained a moderately free position to this day. Sergio Arboleda University accommodated the thesis of one of its founders, Enrique Gómez Hurtado (1927–2019), to promote reform in favor of the decriminalization of the drug trade, such as cocaine, whose prohibition continues to generate disorder in Colombia (Gómez Hurtado 1993, 35–44). Gómez was a congressman between 1991 and 2010 and was a leader of the Colombian Conservative Party.

A singularly important initiative in promoting liberal ideas in Colombia was the role played by the Hayek Foundation, which between 2011 and 2015 carried out three international seminars on the Austrian School of Economics. The seminars involved such liberal figures as Lawrence H. White, Adrian Ravier, Maria Blanco, Butler Shafer, Peter Klein, and Jorg Guido Hulsmann. The Foundation also commissioned a Spanish reprint of the abridged version of Hayek's *The Road to Serfdom*, from the condensed text in *Reader's Digest*.

Also noteworthy are some new outfits. A political organization called the Libertarian Movement was created in 2015 amid Daniel Raisbeck's candidacy for mayor of Bogotá. Since that campaign, Raisbeck has been one of Colombia's most visible representatives of libertarianism. More recently, LiberTank was created in 2019, based in Medellín. LiberTank focuses on disseminating economically liberal ideas on topics like employment, the minimum wage, and public investment.

Conclusions

Ideas in the same vein as Smithian liberalism have played an active part in the history of Colombia, from its constitution as a democratic republic to recent times. Similar to what occurred in the leading countries of the Western Hemisphere, Colombia witnessed a rise in liberalism, reflected in liberal policies adopted in the second half of the nineteenth century in favor of free trade, free banking, and limited government. Liberalism remained relevant throughout the twentieth century but, as in the rest of the world, it ceded ground to ideologies more favorable to the governmentalization of social affairs. The direction of Colombia's political reforms has largely been in step with worldwide trends.

Liberalism nevertheless continues to pulsate in Colombia. Its future is unwritten.

References

- Agudelo Villa, Hernando.** 1968. *Hacia Un Liberalismo Moderno*. Bogotá: Ediciones Tercer Mundo.
- Asamblea Nacional Constituyente.** 1863. Constitución Política de Los Estados Unidos de Colombia de 1863. [Link](#)
- Asamblea Nacional Constituyente.** 1886. Constitución Política de 1886. [Link](#)
- Ballve Pallicer, Faustino.** 2012. *Los Fundamentos de La Ciencia Económica: Diez Lecciones de Economía*. Bogotá: Unión Editorial.
- Bastiat, Frederic.** 1848. *Sofismas Económicos*. Bogotá: Imprenta de Ancizar.
- Bastiat, Frederic.** 1853. *Lo Que Se ve y Lo Que No Se ve o La Economía Política En Una Lección, Dada a Sus Paisanos En Francia Por Frederic Bastiat, e Interpretada y Ofrecida a Los Suyos de Zipaquirá*, ed. and trans. Eustacio Santamaría. New York: U.S. Post Office Printing.
- Bergquist, Charles W.** 1986. *Coffee and Conflict in Colombia, 1886–1910*. Durham, N.C.: Duke University Press.
- Berlanza, Lucas.** 2020. Liberalism in Brazil. *Econ Journal Watch* 17(2): 405–441. [Link](#)
- Beyer, Robert.** 1947. *The Colombian Coffee Industry: Origins and Major Trends 1740–1940*. Ph.D. diss., University of Minnesota.
- Bushnell, David.** 1996. *The Making of Modern Colombia: A Nation in Spite of Itself*. Berkeley, Cal.: University of California Press.
- Cabellor Escovar, Enrique.** 1959. *Capitalización y Empleo*. Bogotá: ANDI.
- Concordato celebrado entre La Santa Sede Apostólica y El Gobierno de La República de Colombia, y Otros Documentos Eclesiásticos y Civiles.** 1895. Bogotá: Imprenta de Antonio M. Silvestre. [Link](#)
- Congreso de Colombia.** 1968. Acto Legislativo 1 de 1968. [Link](#)
- Eastman, Tomas.** 1923. El Proyecto Sobre El Banco de Emisión Tiene Parte Del Trabajo Extranjero y Parte Del Trabajo Indígena. *El Espectador* (Medellín), June 7–8.
- Echavarría Olózaga, Hernan.** 1948. *Pleno Empleo y Otros Temas*. Bogotá: Minerva.

- Echavarría Olózaga, Hernan.** 2003 [1958]. *El Sentido Común En La Economía*. Bogotá: Alfaomega.
- Echeverri, Lina.** 1994. Banca Libre: La Experiencia Colombiana. In *Ensayos de Historia Monetaria y Bancaria de Colombia*, ed. Fabio Sánchez Torres, 305–331. Bogotá: TM Editores.
- Edwards, Sebastian, and Roberto Steiner.** 2008. *La Revolución Incompleta: Las Reformas de Gaviria*. Bogotá: Editorial Norma.
- Erhard Schièble, Georges.** 1883. Antiguo Virreinato de La Nueva Granada, Hoy Estados Unidos de Colombia y República Del Ecuador. In *Les Etats Unies de Colombie: Précis d'Histoire et de Géographie Physique, Politique et Commerciale*. Paris: C. Marpon and E. Flammarion.
- Faria, Hugo J., and Leonor Filardo.** 2015. Venezuela: Without Liberals, There Is No Liberalism. *Econ Journal Watch* 12(3): 375–399. [Link](#)
- García de Kausel, Gloria Patricia.** 2013. *La Bandera Sigue al Mercado. El Paso Del Banco de Emisión al Banco Central y Su Incidencia En La Soberanía de Las Naciones En Colombia y América Latina: 1914–1945*. Tesis de posgrado, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata (La Plata, Argentina).
- Gómez Hurtado, Enrique.** 1993. *La Tragedia de La Droga: Perspectiva de Su Solución*. Bogotá: Universidad Sergio Arboleda.
- Hayek, Friedrich A.** 2008. Un Dialogo Con El Profesor Friedrich von Hayek [interview by Diego Pizano Salazar]. *Apuntes Del Cenes* 27(46): 287–300.
- Jaramillo, Roberto, and Adolfo Meisel.** 2008. Más Allá de La Retórica de La Reacción: Análisis Económico de La Desamortización En Colombia 1861–1888. *Cuadernos de Historia Económica y Empresarial* 22. Banco de la República (Cartagena). [Link](#)
- Kuchař, Pavel.** 2016. Liberalism in Mexican Economic Thought, Past and Present. *Econ Journal Watch* 13(1): 129–167. [Link](#)
- Lemos Simmonds, Carlos.** 1991. *El Estado Ladrón*. Bogotá: Fundación Nueva República.
- Llano Isaza, Rodrigo.** 2009. *Historia Resumida del Partido Liberal Colombiano*. Bogotá: Partido Liberal Colombiano.
- Luzardo-Luna, Ivan.** 2019. *Colombia's Slow Economic Growth: From the Nineteenth to the Twenty-First Century*. Cham, Switzerland: Palgrave Macmillan.
- Marroquín, Andrés, and Fritz Thomas.** 2015. Classical Liberalism in Guatemala. *Econ Journal Watch* 12(3): 460–478. [Link](#)
- Meisel, Adolfo.** 1992. Free Banking in Colombia. In *The Experience of Free Banking*, ed. Kevin Dowd, 93–102. London: Routledge.
- Melo, José Orlando.** 2015. Las Vicisitudes Del Modelo Liberal 1850–1899. In *Historia Económica de Colombia*, ed. José Antonio Ocampo and Germán Colmenares. Bogotá: Fondo de Cultura Económica.
- Mena García, María del Carmen.** 1996. Censos Eclesiásticos y Propiedad Urbana. *Caravelle: Cahiers Du Monde Hispanique et Luso-Brésilien* 66: 5–26. [Link](#)
- Mosquera, Tomás Cipriano de.** 1852. Carta de La República de Nueva Granada, Conforme a Su Última División Política. In *Memoria Sobre la Geografía Física y Política de la Nueva Granada*. New York: Ferdinand Mayer.
- Ocampo, José Antonio.** 2013. *Colombia y La Economía Mundial, 1830–1910*, 2nd ed. Bogotá:

Universidad de los Andes.

Ospina Sardi, Jorge. 1986. *Hacia Un Nuevo Esquema de Intervención Del Estado*. Washington, D.C.: Centro Internacional Para la Empresa Privada.

Peñalosa Londoño, Enrique. 1989. *Capitalismo, La Mejor Opción?* Bogotá: Fundación Hacia el Desarrollo.

Rojas, Ezequiel. 1868. *Filosofía Moral*. Bogotá: Imprenta de la Nación.

Romero, Pedro, Fergus Hodgson, and María Paz Gómez. 2018. Liberalism in Ecuador. *Econ Journal Watch* 15(1): 105–128. [Link](#)

Saenz-Armstrong, Patricia. 2018. Economic Liberalism in Peru. *Econ Journal Watch* 15(2): 179–200. [Link](#)

Safford, Frank. 1988. The Emergence of Economic Liberalism in Colombia. In *Guiding the Invisible Hand: Economic Liberalism and The State in Latin American History*, ed. Joseph L. Love and Nils Jacobsen. New York: Praeger.

Safford, Frank, and Marco Palacios. 2002. *Colombia: Fragmented Land, Divided Society*. New York: Oxford University Press.

Sanz De Santamaría, Carlos. 1971. *Revolución Silenciosa*. Mexico City: Fondo De Cultura Económica.

Victoria Wilches, Pablo. 1976. *Hacia Una Economía Liberal Un Programa Económico Integral Para Colombia*. Bogotá: Ediciones Tercer Mundo.

About the Authors



Sebastián Rodríguez Duran is an Adam Smith Program fellow and third-year Ph.D. student in the Department of Economics at George Mason University. He earned an M.A. in economics from Francisco Marroquín University in Guatemala and a B.S. in business from EAFIT University in Colombia. His research interests include development economics, economics and religion, and political economy. His email address is srodri5@gmu.edu.



Gilberto Ramírez Espinosa is a historian with the National University of Colombia (Bogotá), with a master's in geography from the University of the Andes. He is a history professor at Sergio Arboleda University and the editorial director of Unión Editorial Colombia. He was regional director in Colombia of Students for Liberty between 2016 and 2018 and coordinator of events for the Bastiat Society chapter in Bogotá between 2018 and 2020. He was part of the think-tank team at the Center for Innovation and Leadership at the University of Gran Colombia

(Bogotá) in 2019. His email address is gilberto.ramirez@unioneditorial.net.

[Go to archive of Character Issues section](#)

[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6051>



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 19(1)
March 2022: 166–175

The Market for Goods and the Market for Ideas¹

Ronald H. Coase

[LINK TO ABSTRACT](#)

The normal treatment of governmental regulation of markets makes a sharp distinction between the ordinary market for goods and services and the activities covered by the First Amendment—speech, writing, and the exercise of religious beliefs—which I call, for brevity, “the market for ideas.” The phrase, “the market for ideas,” does not describe the boundaries of the area to which the First Amendment has been applied very exactly. Indeed, these boundaries do not seem to have been very clearly drawn. But there can be little doubt that the market for ideas, the expression of opinion in speech and writing and similar activities, is at the center of the activities protected by the First Amendment, and it is with these activities that discussion of the First Amendment has been largely concerned.

The arguments that I will be considering long antedate the passage of the First Amendment (which obviously incorporated views already held) and there is some danger for economists, although not necessarily for American lawyers, in confining our discussion to the First Amendment rather than considering the general problem of which it is a part. The danger is that our discussion will tend to concentrate on American court opinions, and particularly those of the Supreme Court, and that, as a result, we will be led to adopt the approach to the regulation of markets found congenial by the courts rather than one developed by economists,

1. Copyright American Economic Association; reproduced with permission of the *American Economic Review*. The piece was originally published in the *American Economic Review* Vol. 64, No. 2, Papers and Proceedings of the Eighty-sixth Annual Meeting of the American Economic Association (May, 1974), pp. 384–391; it was reproduced verbatim (though block-indenting for long quotations) in Coase’s *Essays on Economics and Economists* (University of Chicago Press, 1994), pp. 64–74.

a procedure which already has gone a long way to ruin public utility economics and has done much harm to economic discussion of monopoly problems generally. This approach is confining in another way, since, by concentrating on issues within the context of the American Constitution, it is made more difficult to draw on the experience and thought of the rest of the world.

What is the general view that I will be examining? It is that, in the market for goods, government regulation is desirable whereas, in the market for ideas, government regulation is undesirable and should be strictly limited. In the market for goods, the government is commonly regarded as competent to regulate and properly motivated. Consumers lack the ability to make the appropriate choices. Producers often exercise monopolistic power and, in any case, without some form of government intervention, would not act in a way which promotes the public interest. In the market for ideas, the position is very different. The government, if it attempted to regulate, would be inefficient and its motives would, in general, be bad, so that, even if it were successful in achieving what it wanted to accomplish, the results would be undesirable. Consumers, on the other hand, if left free, exercise a fine discrimination in choosing between the alternative views placed before them, while producers, whether economically powerful or weak, who are found to be so unscrupulous in their behavior in other markets, can be trusted to act in the public interest, whether they publish or work for the *New York Times*, the *Chicago Tribune* or the Columbia Broadcasting System. Politicians, whose actions sometimes pain us, are in their utterances beyond reproach. It is an odd feature of this attitude that commercial advertising, which is often merely an expression of opinion and might, therefore, be thought to be protected by the First Amendment, is considered to be part of the market for goods. The result is that government action is regarded as desirable to regulate (or even suppress) the expression of an opinion in an advertisement which, if expressed in a book or article, would be completely beyond the reach of government regulation.

This ambivalence toward the role of government in the market for goods and the market for ideas has not usually been attacked except by those on the extreme right or left, that is, by fascists or communists. The Western world, by and large, accepts the distinction and the policy recommendations that go with it. The peculiarity of the situation has not, however, gone unnoticed, and I would like to draw your attention to a powerful article by Aaron Director. Director quotes a very strong statement by Justice William O. Douglas in a Supreme Court opinion, a statement which is no doubt intended as an interpretation of the First Amendment, but which obviously embodies a point of view not dependent on constitutional considerations. Justice Douglas said: "free speech, free press, free exercise of religion are placed separate and apart; they are above and beyond the police power; they are not subject to regulation in the manner of factories, slums, apartment

houses, production of oil and the like” (*Beaubarnis v. Illinois* 1952). Director remarks of the attachment to free speech that it is “the only area where *laissez-faire* is still respectable” (Director 1964, 5).

Why should this be so? In part, this may be due to the fact that belief in a free market in ideas does not have the same roots as belief in the value of free trade in goods. To quote Director again: “The free market as a desirable method of organizing the intellectual life of the community was urged long before it was advocated as a desirable method of organizing its economic life. The advantage of free exchange of ideas was recognized before that of the voluntary exchange of goods and services in competitive markets” (1964, 3). In recent years, particularly, I think in America (that is, North America), this view of the peculiar status of the market for ideas has been nourished by a commitment to democracy as exemplified in the political institutions of the United States, for whose efficient working a market in ideas not subject to government regulation is considered essential. This opens a large subject on which I will avoid comment. Suffice it to say that, in practice, the results actually achieved by this particular political system suggest that there is a good deal of “market failure.”

Because of the view that a free market in ideas is necessary to the maintenance of democratic institutions and, I believe, for other reasons also, intellectuals have shown a tendency to exalt the market for ideas and to depreciate the market for goods. Such an attitude seems to me unjustified. As Director said: “the bulk of mankind will for the foreseeable future have to devote a considerable fraction of their active lives to economic activity. For these people, freedom of choice as owners of resources in choosing within available and continually changing opportunities, areas of employment, investment, and consumption is fully as important as freedom of discussion and participation in government” (1964, 6). I have no doubt that this is right. For most people in most countries (and perhaps in all countries), the provision of food, clothing, and shelter is a good deal more important than the provision of the “right ideas,” even if it is assumed that we know what they are.

But leave aside the question of the relative importance of the two markets; the difference in view about the role of government in these two markets is really quite extraordinary and demands an explanation. It is not enough merely to say that the government should be excluded from a sphere of activity because it is vital to the functioning of our society. Even in markets which are mainly of concern to the lower orders, it would not seem desirable to reduce the efficiency with which they work. The paradox is that government intervention which is so harmful in the one sphere becomes beneficial in the other. The paradox is made even more striking when we note that at the present time it is usually those who press most strongly for an extension of government regulation in other markets who are most anxious

for a vigorous enforcement of the First Amendment prohibitions on government regulation in the market for ideas.

What is the explanation for the paradox? Director's gentle nature does not allow him to do more than hint at it:

A superficial explanation for the preference for free speech among intellectuals runs in terms of vertical interests. Everyone tends to magnify the importance of his own occupation and to minimize that of his neighbor. Intellectuals are engaged in the pursuit of truth, while others are merely engaged in earning a livelihood. One follows a profession, usually a learned one, while the other follows a trade or a business. (Director 1964, 6)

I would put the point more bluntly. The market for ideas is the market in which the intellectual conducts his trade. The explanation of the paradox is self-interest and self-esteem. Self-esteem leads the intellectuals to magnify the importance of their own market. That others should be regulated seems natural, particularly as many of the intellectuals see themselves as doing the regulating. But self-interest combines with self-esteem to ensure that, while others are regulated, regulation should not apply to them. And so it is possible to live with these contradictory views about the role of government in these two markets. It is the conclusion that matters. It may not be a nice explanation, but I can think of no other for this strange situation.

That this is the main explanation for the dominance of the view that the market for ideas is sacrosanct is certainly supported if we examine the actions of the press. The press is, of course, the most stalwart defender of the doctrine of freedom of the press, an act of public service to the performance of which it has been led, as it were, by an invisible hand. If we examine the actions and views of the press, they are consistent in only one respect: they are always consistent with the self-interest of the press. Consider their argument that the press should not be forced to reveal the sources of its published material. This is termed a defense of the public's right to know—which is interpreted to mean that the public has no right to know the source of material published by the press. To desire to know the source of a story is not idle curiosity. It is difficult to know how much credence to give to information or to check on its accuracy if one is ignorant of the source. The academic tradition, in which one discloses to the greatest extent possible the sources on which one relies and thus exposes them to the scrutiny of one's colleagues, seems to me to be sound and an essential element in the search for truth. Of course, the counterargument of the press is not without validity. It is argued that some people would not express their opinions honestly if it became known that they really held these opinions. But this argument applies equally to all expressions of views, whether in government, business, or private life, where confidentiality is necessary for frankness. However, this consideration has

commonly not deterred the press from revealing such confidences when it was in their interest to do so. Of course, it would also impede the flow of information to reveal the sources of the material published in cases in which the transmission of the information involved a breach of trust or even the stealing of documents. To accept material in such circumstances is not consistent with the high moral standards and scrupulous observance of the law which the press expects of others. It is hard for me to believe that the main thing wrong with the Watergate affair was that it was not organized by the *New York Times*. I would not wish to argue that there are not conflicting considerations in all these cases which are difficult to evaluate. My point is that the press does not find them difficult to evaluate.

Consider another example which is in many ways more striking: the attitude of the press to government regulation of broadcasting. Broadcasting is an important source of news and information; it comes within the purview of the First Amendment. Yet the program content of a broadcasting station is subject to government regulation. One might have thought that the press, devoted to the strict enforcement of the First Amendment, would have been constantly attacking this abridgment of the right of free speech and expression. But, in fact, they have not. In the forty-five years which have passed since the formation of the Federal Radio Commission (now transformed into the Federal Communications Commission), very few doubts about the policy have been expressed in the press. The press, which is so anxious to remain unshackled by government regulation, has never exerted itself to secure a similar freedom for the broadcasting industry.

Lest you think that I manifest a hostility to the American press, I would like to point out that the British press has acted in a similar fashion. In this case the contrast between actions and proclaimed beliefs is even stronger since what was established in Britain was a government-controlled monopoly of a source of news and information. It might have been thought that this affront to the doctrine of freedom of the press would have appalled the British press. It did not. They supported the broadcasting monopoly, mainly, as far as I can see, because they saw the alternative to the British Broadcasting Corporation (*BBC*) as commercial broadcasting and, therefore, as involving increased competition for advertising revenue. But if the press did not want competition for advertising revenue, they also did not want increased competition in the supply of news. And so they did their best to throttle the *BBC*, at least as a purveyor of news and information. When the monopoly was originally established (when it was still the *British Broadcasting Company*), the *BBC* was prohibited from broadcasting news and information unless obtained from certain named news agencies. No news could be broadcast before 7 p.m. and broadcasts likely to affect adversely the sale of newspapers faced other restrictions as well. Gradually, over the years, these restrictions were relaxed as a result of negotiations between the press and the *BBC*. But it was not until after the

outbreak of World War II that the *BBC* broadcast a regular news bulletin before 6 p.m.²

But, it may be argued, the fact that businessmen are mainly influenced by pecuniary considerations is no great discovery. What else would one expect from the money-grubbers of the newspaper world? Furthermore, it may be objected, because a doctrine is propagated by those who benefit from it does not mean that the doctrine is unsound. After all, have not free speech and a free press also been advocated by high-minded scholars whose beliefs are determined by what is true rather than by more sordid considerations? There has surely never been a more high-minded scholar than John Milton. As his *Areopagitica* “for the liberty of unlicensed printing” is probably the most celebrated defense of the doctrine of freedom of the press ever written, it seemed to me that it would be worthwhile to examine the nature of his argument for a free press. Milton’s work has another advantage for my purpose. Written in 1644, that is, long before 1776, we can see the character of the argument before there was any general understanding of how competitive markets worked and before the emergence of modern views on democracy.

It would be idle for me to pretend that I could act as a guide to Milton’s thought. I know too little of seventeenth century England and there is much in Milton’s pamphlet the meaning of which I cannot discern. Yet, there are passages which leap across the centuries and for whose interpretation no scholarship is needed.

As one would expect, Milton asserts the primacy of the market for ideas: “Give me the liberty to know, to utter, and to argue freely according to conscience, above all liberties” (1959/1644, 44). It is different from the market for goods and should not be treated in the same way: “Truth and understanding are not such wares as to be monopolised and traded in by tickets and statutes and standards. We must not think to make a staple commodity of all the knowledge in the land, to mark and license it like our broadcloth and our wool-packs” (*ibid.*, 29). The licensing of printed material is an affront to learned men and to learning:

When a man writes to the world, he summons up all his reason and deliberation to assist him; he searches, mediates, is industrious, and likely consults and confers with his judicious friends; after all which done he takes himself to be informed in what he writes, as well as any that writ before him. If in this the most consummate act of his fidelity and ripeness no years, no industry, no former proof of his abilities can bring him to that state of maturity as not to be still mistrusted and suspected, unless he carry his considerate

2. For a discussion of the attitude of the press to the monopoly of British broadcasting, see Coase 1950, 103–110, 192–193.

diligence, all his midnight watchings...to the hasty view of an un leisured licenser, perhaps much his younger, perhaps far his inferior in judgment, perhaps one who never knew the labour of book-writing, and, if he be not repulsed or slighted, must appear in print like a puny with his guardian and his censor's hand on the back of his title to be his bail and surety, that he is no idiot or seducer, it cannot be but a dishonour and derogation to the author, to the book, to the privilege and dignity of learning. (Milton 1959/1644, 27)

Licensing is also an affront to the common people:

Nor is it to the common people less than a reproach; for if we be so jealous over them, as that we dare not trust them with an English pamphlet, what do we but censure them for a giddy, vicious, and ungrounded people, in such a sick and weak state of faith and discretion, as to be able to take nothing down but through the pipe of a licenser. (Milton 1959/1644, 30)

In the market for ideas, the right choices are made: "Let [truth] and falsehood grapple; who ever knew Truth put to the worse in a free and open encounter" (Milton 1959/1644, 45). Those who undertake the job of licensing will be incompetent. A licenser should be, according to Milton, "studious, learned, and judicious." But this is not what we are likely to get: "we may easily foresee what kind of licensers we are to expect hereafter: either ignorant, imperious, and remiss, or basely pecuniary" (ibid., 25). The licensers are more likely to suppress truth than falsehood: "if it come to prohibiting, there is aught more likely to be prohibited than truth itself; whose first appearance to our eyes bleared and dimmed with prejudice and custom is more unsightly and unplausible than many errors..." (47). Nor does Milton fail to tell us that the licensing scheme against which he was writing came about as a result of industry pressure: "And how it got the upper hand...there was in it the fraud of some old patentees and monopolisers in the trade of bookselling" (50).

In the formation of Milton's views, self-interest may perhaps have played a part, but there can be little doubt that his argument embodies a good deal of intellectual pride of the kind to which Director refers. The writer is a learned man, diligent and trustworthy. The licenser would be ignorant, incompetent, and basely motivated, perhaps "younger" and "inferior in judgment." The common man always chooses truth as against falsehood. The picture is a little too one-sided to be wholly convincing. And if it has been convincing to the intellectual community (and apparently it often has), it is surely because people are easily persuaded that what is good for them is good for the country.

I do not believe that this distinction between the market for goods and the market for ideas is valid. There is no fundamental difference between these two

markets and, in deciding on public policy with regard to them, we need to take into account the same considerations. In all markets, producers have some reasons for being honest and some for being dishonest; consumers have some information but are not fully informed or even able to digest the information they have; regulators commonly wish to do a good job, and though often incompetent and subject to the influence of special interests, they act like this because, like all of us, they are human beings whose strongest motives are not the highest.

When I say that the same considerations should be taken into account, I do not mean that public policy should be the same in all markets. The special characteristics of each market lead to the same factors having different weights, and the appropriate social arrangements will vary accordingly. It may not be sensible to have the same legal arrangements governing the supply of soap, housing, automobiles, oil, and books. My argument is that we should use the same *approach* for all markets when deciding on public policy. In fact, if we do this and use for the market for ideas the same approach which has commended itself to economists for the market for goods, it is apparent that the case for government intervention in the market for ideas is much stronger than it is, in general, in the market for goods. For example, economists usually call for government intervention, which may include direct government regulation, when the market does not operate properly—when, that is, there exist what are commonly referred to as neighborhood or spillover effects, or, to use that unfortunate word, “externalities.” If we try to imagine the property rights system that would be required and the transactions that would have to be carried out to assure that anyone who propagated an idea or a proposal for reform received the value of the good it produced or had to pay compensation for the harm that resulted, it is easy to see that in practice there is likely to be a good deal of “market failure.” Situations of this kind usually lead economists to call for extensive government intervention.

Or consider the question of consumer ignorance which is commonly thought to be a justification for government intervention. It is hard to believe that the general public is in a better position to evaluate competing views on economic and social policy than to choose between different kinds of food. Yet there is support for regulation in the one case but not in the other. Or consider the question of preventing fraud, for which government intervention is commonly advocated. It would be difficult to deny that newspaper articles and the speeches of politicians contain a large number of false and misleading statements—indeed, sometimes they seem to consist of little else. Government action to control false and misleading advertising is considered highly desirable. Yet a proposal to set up a Federal Press Commission or a Federal Political Commission modeled on the Federal Trade Commission would be dismissed out of hand.

The strong support enjoyed by the First Amendment should not hide from

us that there is, in fact, a good deal of government intervention in the market for ideas. I have mentioned broadcasting. But there is also the case of education, which, although it plays a crucial role in the market for ideas, is subject to considerable regulation. One might have thought that those who were so anxious to obstruct government regulation of books and other printed material would also find such regulation in the field of education obnoxious. But, of course, there is a difference. Government regulation of education commonly accompanies government financing and other measures (such as compulsory school attendance) which increase the demand for the services of intellectuals and, therefore, their incomes (see E. G. West 1967, 101). So self-interest, which, in general, would lead to support for a free market in ideas, suggests a different attitude in education.

Nor do I doubt that detailed study would reveal other cases in which groups of practitioners in the market for ideas have supported government regulation and the restriction of competition when it would increase their incomes, just as we find similar behavior in the market for goods. But interest in monopolizing is likely to be less in the market for ideas. A general policy of regulation, by restricting the market, would have the effect of reducing the demand for the services of intellectuals. But more important, perhaps, is that the public is commonly more interested in the struggle between truth and falsehood than it is in the truth itself. Demand for the services of the writer and speechmaker depends, to a considerable extent, on the existence of controversy—and for controversy to exist, it is necessary that truth should not stand triumphant and alone.

Whatever one may think of the motives which have led to the general acceptance of the present position, there remains the question of which policies would be, in fact, the most appropriate. This requires us to come to some conclusion about how the government will perform whatever functions are assigned to it. I do not believe that we will be able to form a judgment in which we can have any confidence unless we abandon the present ambivalence about the performance of government in the two markets and adopt a more consistent view. We have to decide whether the government is as incompetent as is generally assumed in the market for ideas, in which case we would want to decrease government intervention in the market for goods, or whether it is as efficient as it is generally assumed to be in the market for goods, in which case we would want to increase government regulation in the market for ideas. Of course, one could adopt an intermediate position—a government neither as incompetent and base as assumed in the one market nor as efficient and virtuous as assumed in the other. In this case, we ought to reduce the amount of government regulation in the market for goods and might want to increase government intervention in the market for ideas. I look forward to learning which of these alternative views will be espoused by my colleagues in the economics profession.

References

- Coase, Ronald H.** 1950. *British Broadcasting, A Study in Monopoly*. London: London School of Economics and Political Science.
- Director, Aaron.** 1964. The Parity of the Economic Market Place. *Journal of Law and Economics* 7(Oct.): 1–10.
- Milton, John.** 1959 [1644]. *Areopagitica, A Speech for the Liberty of Unlicensed Printing*. New York: St. Martin's Press.
- West, E. G.** 1967. The Political Economy of American Public School Legislation. *Journal of Law and Economics* 10(Oct.): 101–128.

Cases Cited

Beauharnis v. Illinois, 343 U.S. 250, 286, 1952.

About the Author



Ronald Coase (1910–2013) was a British economist who became Clifton R. Musser Professor of Economics at the University of Chicago Law School. He received the Nobel Memorial Prize in Economic Sciences in 1991. He beseeched us to take responsibility for the total effect of our action, which, if spelling the governmentalization of social affairs, meant facing up to the downsides of governmentalization. He wrote two essays on Adam Smith, one focused on *The Wealth of Nations*, the other *The Theory of Moral Sentiments*. The latter ends with these words: “Adam Smith would not have thought it sensible to treat man as a rational utility-maximiser. He thinks of man as he actually is—dominated, it is true, by self-love but not without some concern for others, able to reason but not necessarily in such a way as to reach the right conclusion, seeing the outcomes of his actions but through a veil of self-delusion. No doubt modern psychologists have added a great deal, some of it correct, to this eighteenth century view of human nature. But if one is willing to accept Adam Smith’s view of man as containing, if not the whole truth, at least a large part of it, realisation that his thought has a much broader foundation than is commonly assumed makes his argument for economic freedom more powerful and his conclusions more persuasive.”

[Go to archive of Watchpad section](#)
[Go to March 2022 issue](#)



Discuss this article at Journaltalk:
<https://journaltalk.net/articles/6052>