Broken Instruments

Trevor Gallen * Ben Raymond

October 2020

Abstract

Repeated use of the same potentially related instrumental variables by a literature can "collectively invalidate" these instruments. This paper examines two ways in which this can happen. First, when instruments sharing significant sources of variation are used to instrument multiple distinct covariates, it is increasingly likely the exclusion restriction was not satisfied in any individual specification from the outset. Second, when a variable is documented to affect many outcomes that are likely to be highly or even mildly persistent, using lagged values of that variable as an instrument is likely to violate the exclusion condition. This paper produces a dataset of approximately 960 instrumental variables papers from 1995-2019 in highly-ranked economics general interest and field journals. We find six groups of commonly-used instruments whose literatures, taken together, suggest they are likely to fail the strict exogeneity condition: (i) elevation and bodies of water (ii) sibling structure (iii) ethnicity/ethnolinguistic fractionalization (iv) religion (v) weather and (vi) immigrant enclaves. Taken together, these potentially related instruments have been used in 86 “top five" publications and 317 well-ranked field or general interest journals, with 189 total uses cataloged from 2011 onwards. We propose a Hausman-like test for suspect regressions and discuss its asymptotic properties. We then apply it to two IV papers, finding little reason to be concerned about one, and tentative evidence to be concerned about the other.

JEL Classification: C13, C26, C36
Keywords: Instrumental variables, commonly-used instruments, exclusion restriction

---

*The views expressed herein are those of the authors and do not necessarily reflect the views of the Bureau of Labor Statistics. The authors would like to thank Natalie Bau, Tim Bond, Chad Syverson, Devin Pope, Ian Fillmore, Dominic Smith, Jake Schild, Eliza Forscythe, Yana Gallen, Josh Gottleib, Scott Imberman, Mohitosh Kejriwal, Tim Moore, Casey Mulligan, Jonathan Skinner, and Steven Levitt for helpful comments and dicussions. Correspondence: Purdue University Krannert School of Business, West Lafayette, IN 47906. Tel.: (765) 496-2458. Email: tgallen@purdue.edu. Web: www.tgallen.com

I Introduction

A strength of empirical economic research is its focus on causality and mechanisms that underlie economic phenomena. A crucial tool in determining causality has been instrumental variables regression, typically used to isolate a source of variation that identifies a single causal channel. Unfortunately, finding a good instrument is difficult. Consequently, a strong instrument for an important phenomenon accepted by the literature often becomes pervasively used.

In general, a strong instrument with the capacity to be used for many purposes is more likely to be invalid. In Morck and Yeung (2011)'s words, "...each successful use of an instrument creates an additional latent variable problem for all other uses of that instrument."[1] Of course, reality does not change when a paper is published, but the relative probability a researcher assigns to potential exogeneity violations should increase the more the instrument is shown to have strong effects on other, seemingly unrelated variables. To better understand this phenomenon, we document the use of instrumental variables in well-regarded economics journals from 1995-2019. Our analysis of these instruments uncovered six groups of potentially related instruments used 86 times in "top five" journals and 317 times in well-ranked field or general interest journals including the top five.

We focus on two types of collective invalidation, giving examples from the literature. The first comes from repeated use of the same or highly-related instrument for multiple distinct covariates. Consider an instrument that affects both segregation in a city (and thereby education, income, and single motherhood by race), as well as the exposure of that city to national housing price changes (via housing supply elasticity), and thereby affects recovery from recession. If either segregation or disparate racial educational and work outcomes caused by segregation affect the rate at which a city recovers from a recession, the second paper is threatened. If housing price changes affect education and income by race, the first paper is threatened. Both seem likely, ex ante, but are less likely to be noticed or considered when one study concerns urban economics while the other is macroeconomic. The second issue comes when lagged values of a covariate are used as an instrument for current values of the same covariate, and that covariate is established to affect dozens of plausibly durable outcomes. For instance, if immigration affects housing investment, human capital investment, native migration, and business investment, four extremely durable outcomes, then it appears plausible that past immigrant flows would have affects on outcomes-of-interest measured in decades, again violating the exclusion restriction.

[1] Bazzi and Clemens (2013) also enunciate this issue, suggesting that literatures may "collectively invalidate" the use of an instrument.

Two solutions to this problem may seem simple: first, include all the predictors used in other papers as controls. We show that this does not always solve the problem. Instead, including other endogenous regressors as controls may bias an otherwise well-identified regression.[2] We discuss such cases to better understand when this might be true, and show when it is a problem in our Monte Carlo exercises: typically generated when there is simultaneity between endogenous regressors.

We create a database of approximately 960 instrumental variables papers from well-regarded journals, and identify six strains of literature in which the exclusion restriction should attract particular attention.[3] First, "elevation and bodies of water," used to isolate exogenous components of housing supply, segregation, governance structure, dam location, and infrastructure cost, and broadband provision among other variables. Second, "sibling structure," used to isolate exogenous components of family size and fertility, father presence, parental wages, child schooling, age at grandparenthood, welfare receipt and geographical mobility, among other uses. Third, ethnicity/ethnolinguistic fractionalization, which is used to instrument for rule of law, corruption, democracy, income and investment, social trust, institutions, creditor protections, and welfare-state generosity. Fourth, religion, which is used to instrument for land regulation, social trust, national uncertainty aversion, the free press, private school share, bank regulation, and work ethic. Fifth, weather, which is used to instrument for agricultural and fishing productivity, economic growth, energy prices, commodity prices, pollution, population, migration, water quality, political changes, and managerial moods.

Our sixth instrument, past immigration, while only instrumenting for a single covariate (current immigrant share or flow) affects dozens of outcomes such as education, housing supply, political equilibria, and firm capital investment, which are likely to be highly persistent, particularly when taken jointly, leading to a violation of the exclusion restriction. Second, we find instruments that are potentially subject to contamination caused by the persistence of outcome variables. For example, in the case of preexisting immigrant enclaves as an instrument for flow immigration, the persistence of latent variables affected by past immigration is likely to increase nonlinearly as the number of outcomes affected by immigration increases, violating the exclusion condition.

Figure 1 depicts the uses of these six instruments in all surveyed journals individually and jointly over

[2]If our equation is over-identified with multiple instruments, we could correctly include these suspected endogenous regressors as controls. In the cases we study, this over-identified case is rare.

[3]Many of these are not identical, but are typically highly correlated, which is why we adopt the phrase "potentially related" rather than "identical." Average hours of sunshine in an MSA and average cloud cover are highly related, as is average hours of sunshine and average temperature. Similarly, ethnic, language, and ethnolinguistic fractionalization are highly related (the lowest pairwise correlation is 0.70), and are each significantly correlated with religious fractionalization (Alesina et al., 2003).

time. While the use of one of these instruments is mostly limited to between two and five papers a year, the cumulative use of these instruments in economics journals reaches 317 papers by 2019.[4] Moreover, the total use of these instruments has been relatively steady since 2006 at approximately 13 papers per year and has not significantly declined for any of them over time. If anything, the data show the use of these IVs is at best leveling off after increasing precipitously from 2002-2013.[5]

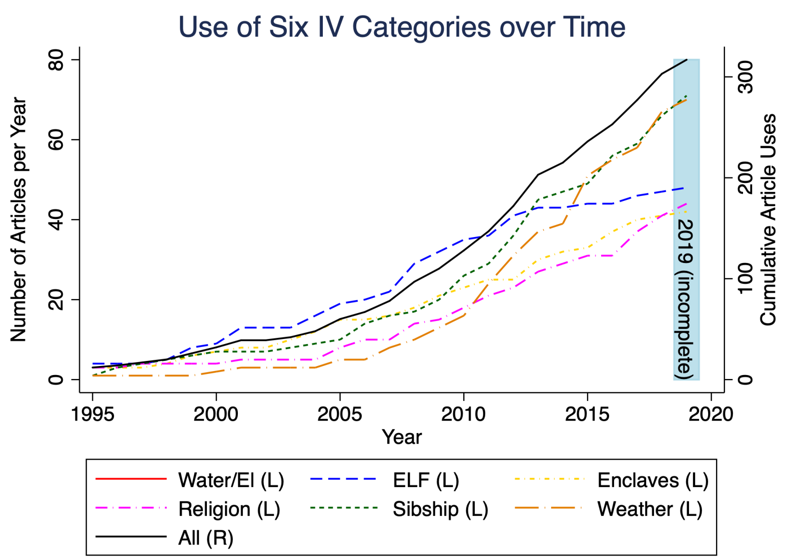

Figure 1: This figure depicts the use of six groups of potentially related instruments in well-ranked economics journals from 1995-2019: (i) elevation and bodies of water (ii) sibling structure (sibship) (iii) ethnicity/ethnic fractionalization (ELF) (iv) religion (v) weather and (vi) immigrant enclaves. Individual yearly uses for each variable and their sum are given by the left axis (L). Cumulative uses of all instruments are given by the right axis (R).

To better grapple with this problem, we produce a new statistical test to shed light on when a researcher might be concerned about using common and potentially related instruments. Our test is comparable to the Hausman test: we run a single-paper regression ignoring other potential IV uses, and run the same regression including other potential IV uses as exogenous controls.[6] The sources of bias in the two regressions is generated by different issues. Consequently, a failure of coefficients to be different suggests either an unlikely coincidence of biases or that both biases are small. We differ from the Hausman test

[4]We include available categorized articles through September 2019.

[5]We are aware of a long literature on instrumental variables preexisting 1990. However, it is only in the 1990's that mechanical statements of simultaneous equations with excluded variables makes way for clear treatment and committed defense of instrument exogeneity in the papers we surveyed.

[6]We emphasize that this approach is a test, and does not advocate automatically including "bad controls" to estimate the point coefficient.

in that we do not assume efficiency, and thereby allow for nonzero covariance of estimated coefficients. In both Monte Carlo tests and our applications, allowing for covariance between estimated coefficients is important. To understand this test, we provide Monte Carlo evidence on various versions of the IV estimator. We find significant change in the coefficient of interest when adding in other potential endogenous variables as controls is a useful filter to reject results. Inversely, when the coefficient of interest does not change even when other potential endogenous variables are included, the regressor typically has good mean square error properties.

Our paper touches on several other papers criticizing some of these same instruments. Bazzi and Clemens (2013) discusses the issue of collective invalidation of instruments in the context of growth regressions and to our knowledge is the first to mathematically set down the idea of a literature's collective invalidation. Sarsons (2015) discusses why rainfall likely does not affect conflict solely through income shocks, and may be a bad instrument for conflict. We complement her work by noting more than fifteen other endogenous variables the instrument is applied to understand more than thirty other outcomes. Dell, Jones, and Olken (2014) review the literature on weather shocks, but avoid the concerns we raise focusing on the effect of weather on variables, rather than narrow causal chains that weather as an instrument is claimed to uncover. Jaeger, Ruist, and Stuhler (2018) argue that because general equilibrium adjustments take time, and because immigration shocks are correlated over time, short- and long-run effects of immigration may be misstated. Our contribution is to argue the many documented outcomes increases the potential persistence of the system: if outcomes influence one another even weakly over time, persistence of a shock grows nonlinearly with the number of outcomes. Heath et al. (2019) make an argument related to this paper's focus: the repeated use of natural experiments increases the likelihood of false discoveries. Heath et al. (2019) examine two in particular: the Regulation SHO pilot and business combination laws, which we do not touch on in this paper. Kolesár et al. (2015) note that if the direct effects of our instrument on our variable of interest is orthogonal to the direct effects of our endogenous regressor, then we can design an estimator that recovers the true effect of our endogenous regressor on the outcome of interest. As we discuss, because of the nature of these six instruments, we do not believe this is a reasonable assumption in the large majority of cases we document. Finally, Young (2019) notes that a number of top instrumental variables papers rely on outliers, which plays a significant role in our small-sample Monte Carlo findings.

Section 2 discusses a framework for thinking about multiple-paper exogeneity violations. Section 3 discusses our data collection and criteria for journal inclusion. Section 4 discusses in detail each of the six categories of potentially problematic instruments. Second 5 produces Monte Carlo evidence on IV estimator performance in a multiple-paper setting. Section 6 applies this test to Rupert and Zanella (2018) and Mian and Sufi (2014). Section 7 concludes.

II Framework

In this section, we outline potential pitfalls stemming from repeated use of the same instrument or use lagged values of a variable as an instrument for a covariate that affects many outcomes. The first is the obvious "direct" violation of exogeneity described by Morck and Yeung (2011). When two separate papers use the same variable Z as an instrument for both X_1 and X_2's relationship with outcomes Y_1 and Y_2 respectively, we must be concerned with whether X_1 affects Y_2 or X_2 affects Y_1, a standard violation of exogeneity.[7] However, it is also possible that controlling for X_2 or Y_2 induces a violation of the exclusion restriction where none would have occurred in the absence of controls. The third occurs when covariates are used as proxies for a broader concept (e.g. using education as a proxy for human capital), as the significance of other outcomes from other papers may contradict the causal story. The last occurs when a variable affects many outcomes that are potentially persistent. Even if the persistence of each individual outcome variable is relatively weak, the variables can generate substantial persistence jointly. This strong joint persistence can induce significant serial correlation, invalidating the instrument.

Figure 2 demonstrates the first two "direct" and "indirect" violations using directed acyclic graphs (DAGs). The lefthand DAG ("Possibility 1") shows two papers, Paper 1 and Paper 2. Paper 1 is interested in the relationship between X_1 and Y_1. Paper 2 is interested in the relationship between X_2 and Y_2. These relationships are confounded by unobserved variables ξ and ν, respectively, and require an instrumental variable Z, which is the same for both papers. The implicitly-assumed DAG for Paper 1 is circled and given by the solid red lines, while the implicitly-assumed DAG for Paper 2 is circled and given by dashed blue lines. While both papers establish the solid red and dashed blue lines as important, their joint publication gives rise to concerns depicted in the dotted black lines (among others).

In Possibility 1, X_2 may affect X_1 or Y_1, and Y_2 may affect Y_1. This might be the case, for instance, if a city's elevation gradient and the presence of bodies of water affected both local changes in housing prices during a national housing market downturn, which affects local change in employment (Paper 1) as well as segregation and local school choice, which affects local educational outcomes (Paper 2). Consistent

[7]This concern remains if two papers use highly related instruments, such as

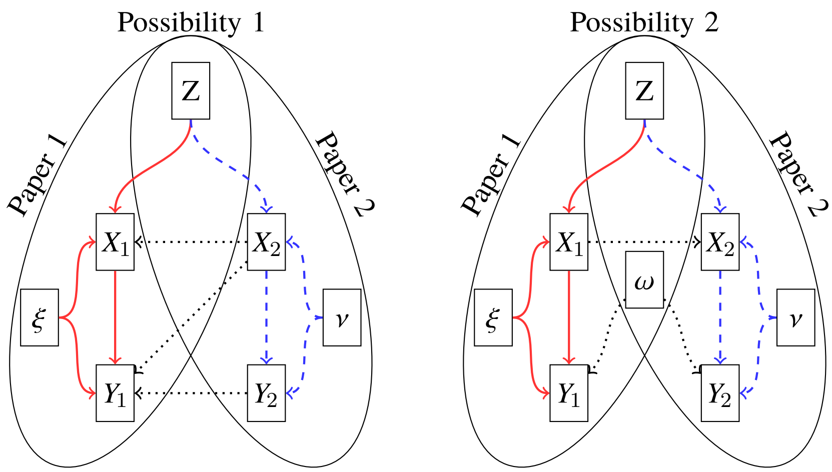

Figure 2: This figure uses a DAG to depict two different possibilities when the same instrument is used in two different papers. In each example, the implicit DAG for the two papers are circled and labelled. In both examples, the first paper's DAG is depicted using a solid red line, and estimates X_1's relationship to Y_1 using Z as an instrument, removing the confounder ξ. Similarly, the second paper's DAG is depicted using a dashed blue line, and estimates X_2's relationship to Y_2 using Z as an instrument, removing the confounder ν. The difference between the two examples is the dashed black lines, which are unmodelled in either paper. As discussed in the text, for Paper 1 to accurately estimate X_1's effect on Y_1, in the first example, it must control for both X_2 and Y_2. However, in the second example, it must *not* control for both X_2 and Y_2, as this would induce a violation of strict exogeneity, even though no such violation is present in either of the two papers separately.

with Figure 2, it is likely that segregation and local school choice might affect both the change in housing prices and cyclical changes in local employment and that educational outcomes may also cause cyclical changes in employment. The importance of Paper 2 raises concerns about Paper 1's causal channel and identification. Fortunately, in this situation, because Paper 2's outcomes are known, controlling for X_2 and Y_2 is enough to properly identify the affect of X_1 on Y_1 in Paper 1.[8] A paper analyzing the effect of housing prices at the business cycle need only to control for segregation measures in this case.

The same cannot be said in Possibility 2. Note in Possibility 2, the two outcomes are completely causally unrelated in regards to Z and X's and are only affected by the same variable ω. In Possibility 2, Paper 1's *not* controlling for X_2 and Y_2 is sufficient for identification.[9] However, controlling for X_2 and Y_2, induces spurious correlation where there was none via residual regression and Berkson's paradox between

[8]It is easy to see that the effect of X_1 on Y_1 will be consistently estimated by instrument X_1 with Z and controlling for X_2 and Y_2, because they partial out any confounding effects, closing backdoor paths and do not conditioning on descendants, in the terminology of Pearl (2009).

[9]This can be seen from the fact that ω is a confounder, but only has the opportunity to bias X_1's estimates via the inclusion of X_2 and Y_2. Without X_2 and Y_2, the instrument's exogeneity condition trivially holds.

X_1 and Y_1. In this way, in Possibility 2, Paper 1's estimates of the effect of X_1 and Y_1 are contaminated even though there is no casual chain. For a concrete example, consider the case in which sibling composition instruments for family size, which affects both child mortality (as an adult) and child geographic mobility. It is possible these two causal chains are largely unrelated and each paper may independently use the instrument correctly. However, if a child's IQ affects both their adult mortality and their adult geographic mobility, controlling for child geographic mobility induces a correlation via IQ, violating the exclusion restriction.

While repeated use of an instrument for multiple combination of endogenous variables and outcomes is cause for concern, Figure 2 makes it clear there is no panacea. In the first case, controlling for the relevant endogenous covariates and outcomes is enough to produce a valid estimator, while in the second, controlling for the same endogenous covariates and outcomes is enough to invalidate an otherwise valid estimator by inducing spurious correlations.

The third problem deals with the causal scrum that occurs when each new outcome is found to be significant. First, as the number of outcomes proliferates, the likelihood that some seemingly innocuous control is in fact a mediator through an unthought-of channel increases, and controlling for it would invalidate the estimate of the total effect of X_1 on Y_1. Second, it raises the possibility that significance on Y_2 suggests the "proxy" story behind the relationship between X_1 and Y_1 is incorrect.[10] In this case, every additional outcome opens up an additional potential causal channel to exogeneity violation. The idea behind this concern is depicted with a concrete example in Figure 3. Brückner and Ciccone (2011) use rainfall to instrument for transitory negative income shocks, which are positively related to democratic regime change. The paper argues this is consistent with the theory in Acemoglu and Robinson (2001), that democratic improvements may occur during recessions, when income is low and so the opportunity cost of unrest is low. The proxy for the cost of fighting a regime is income. However, rainfall's effect on income is also found to affect tax revenue, capital investment, risk preferences, population size, school attendance, and urbanization, which may muddle estimates of the causal effect of lower income. One example of how these might muddle causal effects: while the opportunity cost of potential revolutionaries is lower, rainfall is also found to affect government revenues (Brückner, 2012). If the African countries studied are less able to smooth such transitory shocks, state capacity may fall, allowing both increased dissatisfaction and a lower ability of the government to resist change, imposed internally or externally via conditional loans. While the established relationship with income is correct, income as a clean proxy for the opportunity cost

[10] This point is also made in Heath et al. (2019), which studies two natural experiments: state business combination laws and Regulation SHO pilot that have been used for more than 120 academic papers.

of revolution is threatened. The more causal chains, the less likely a proxy variable is a clean proxy.

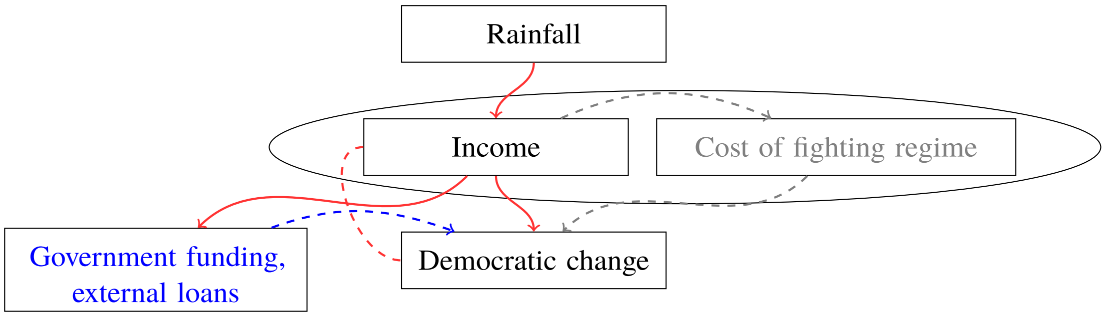

Figure 3: This figure shows an example of how proxy variables are often used. Solid black text and red lines are actual variables. In gray is the relationship discussed in the paper, using income as a proxy for opportunity cost. In blue is a potential second paper's threat to using income solely as a proxy for opportunity cost. In the first paper, income is used as a proxy for the cost of fighting a regime, which leads to democratic change. However, another paper establishing that income also affects government revenues raises concerns about using income as a proxy for opportunity cost.

Figure 4 depicts the fourth and final way in which multiple uses of an instrument may induce exogeneity violations. As in Card (2001), suppose we wish to estimate the effect of immigrants on local education or local land values. However, identification is threatened by the tendency of immigrants to immigrate to places where wages are high, housing prices are low, and education is high quality per unit cost. This produces confounders that threaten identification, depicted as dotted arrows connecting ΔX_t with Y_t^1 and Y_t^2. To obtain accurate estimates, it is thus necessary to find an instrument for current immigration flows that does not affect local education or local land values.

A traditional solution to this particular problem is to use past immigration as an instrument for current immigration. Bartel (1989) finds immigrants tend to immigrate to places where there are already many immigrants. He argues this is due to "supply side" factors such as culture, linguistic familiarity, or "weak ties" that may be unrelated to "demand side" factors, such as city-level productivity. This reasoning implies current immigrant levels determine a city's exposure to national immigration trends. Thus, if immigration from Mexico increases by 10%, it is plausible that cities with preexisting Mexican immigrant populations will absorb proportionally more of that immigration flow than states with no preexisting population, producing a valid instrument that increases immigration for reasons other than productivity.

However, the proposed relationship between current immigration flows and past immigration flows is subject to criticism. If we believe that ΔX_t affects Y_t^1 and Y_t^2, by assumption it is also the case that ΔX_{t-1} affected Y_{t-1}^1 and Y_{t-1}^2. This yields three important avenues by which X_{t-1} may affect Y_t^1 and Y_t^2 other

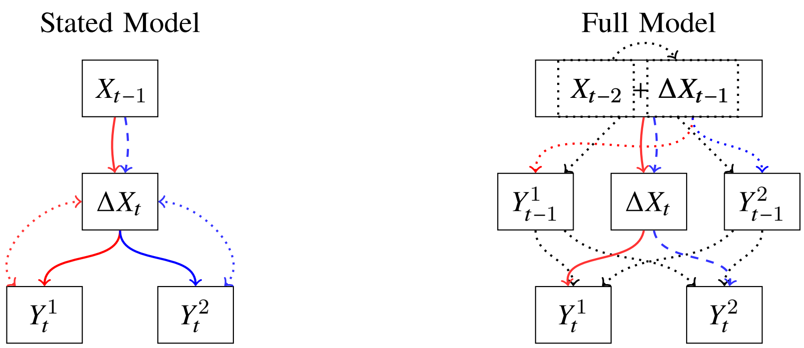

Figure 4: This figure generates a DAG for the "immigrant enclaves" and similar instruments. In the stated model, previous immigration (X_{t-1}) influences immigration flows ΔX_t. In its most "Bartik"-style form, local immigration levels determine exposure to national changes in immigration, which then affect a variety of outcomes, such as education and land values (Y_t^1 and Y_t^2). The "full model" recognizes that if each of the levels of Y_t^1 and Y_t^2 is affected by ΔX_t, then Y_{t-1}^1, and Y_{t-1}^2, and other (unmodelled) covariates are certain to be affected by ΔX_{t-1}, and potentially even X_{t-2}. For instance, we might expect past education and past wages both to affect current education and current wages. These unmodelled effects, that threaten identification and whose likelihood is revealed by multiple outcomes being established, are depicted as black dotted lines. For convenience, the explicit error terms have been omitted from the diagram.

than through ΔX_t. First, it is likely that Y_{t-1}^1 (and similarly Y_{t-1}^2) affects either or both Y_t^1 and Y_t^2.[11] Since X_{t-1} affects Y_{t-1}^1, this establishes another channel through which past immigration can affect outcomes. In the given example, it is quite likely that past education and land values are durable goods. Both are connected to human or structural capital and both have low depreciation rates. Consequently, lagged values of immigration affect current levels of education and capital via persistence of these variables, not simply through immigration today, posing a problem for identification.

Second, it is plausible past change in immigration ΔX_{t-1} not only affects Y_{t-1}^1 and Y_{t-1}^2, but also the level of immigration directly. For instance, the immigrant stock (X_{t-1}, rather than ΔX_{t-1}) likely directly affects land values and education levels, rather than simply through immigrant flows (see, for instance, our discussion of the many long-run effects of ethnolinguistic fractionalization in this paper). Third, while the stock of past immigrants X_{t-2} may affect Y_{t-1}^1 and Y_{t-1}^2, and therefore other variables, it is also guaranteed to indirectly affect them through its effect on ΔX_{t-1}. This is a standard problem with dynamic instruments. The past immigrant stock X_{t-1} is the sum of the past immigrant X_{t-2} and ΔX_{t-1}. However, by assumption, X_{t-2} affects ΔX_{t-1}.

How important is this problem of multiple outcomes? Consider the following vector autoregression for

[11] A similar argument is made in Jaeger, Ruist, and Stuhler (2018). Our contribution is to document the sheer weight of papers finding potential long-run effects

k outcomes, in which all other past outcomes lagged one period weakly affect every other past outcome:

$$\begin{aligned} Y_{1,t} &= \sum_{i=1}^{k} \beta_{1i} Y_{i,t-1} + \epsilon_{1,t} \\ Y_{2,t} &= \sum_{i=1}^{k} \beta_{2i} Y_{i,t-1} + \epsilon_{2,t} \\ &\vdots \qquad \vdots \\ Y_{k,t} &= \sum_{i=1}^{k} \beta_{ki} Y_{i,t-1} + \epsilon_{k,t} \end{aligned}$$

or, organizing this into a standard VAR matrix:

$$Y = \beta X + \epsilon$$

Consider the case in which β's diagonals are significantly greater than zero (own autocorrelation is significant at one-year lags), but less than one in absolute value. Even if the off-diagonals are only slightly greater than zero, the persistence of a shock increases geometrically as a function of the number of outcomes k and the size of the off-diagonal relations $\beta_{k,j}$, $k \neq j$. Consider the case in which Y_1 (for instance, immigration) is shocked, via $\epsilon_{1,t}$. A number of studies have shown that immigration affects future immigration (via $\beta_{1,1}$), but also native public/private educational decisions (Y_2), maternal labor supply (Y_3), wages (Y_4), voting behavior (Y_5), prices of service goods (Y_6), native health (Y_7), supply of nurses (Y_8), native task intensities (Y_9) and so on. These each plausibly affect one another over time. For instance, we would expect educational decisions, maternal labor supply, and child-rearing choices to potentially affect long-run prices of service goods and/or wages. Indeed, in the case of education and child-rearing decisions, we may expect a delayed response not captured by the 1-period VAR above.

What is the effect of adding more potential outcomes? Table 1 depicts the results of increasing k, the magnitude of the off-diagonals $\beta_{k,j}$, $j \neq k$, and the diagonal element $\beta_{k,k}$ in a simulated computational exercise. Table 1 displays two important facts. The relationship between $Y_{1,t+20}$ and $Y_{k\neq 1,t+20}$ is strongly convex in k and $\beta_{k,j}$ when $k > 1$ and when $\beta_{k,j} > 0$.[12] Indeed, for reasonable levels of autocorrelation, such as that caused by $\beta_{k,k} = 0.95$, when $\beta_{k,j} = 0$ and $k = 8$, there is no effect on other variables, and a unit shock to $Y_{1,t}$ increases $Y_{1,t+20}$ by "only" 0.36, and $Y_{2,t}$ by zero. However, if the off diagonal elements

[12]The geometric relationship between k and $\beta_{k,j}$ can also be shown succinctly the fact that the maximum eigenvalue of β increases linearly as a function of $\beta_{k,j}$ and k.

Table 1: Effects on $Y_{x,t+20}$ from a shock to $\epsilon_{1,t}$

	Effects on $Y_{1,t+20}$ from a shock to $\epsilon_{1,t}$					
	$\beta_{k,k} = 0.8$			$\beta_{k,k} = 0.95$		
k	$\beta_{k,j} = 0$	$\beta_{k,j} = 0.02$	$\beta_{k,j} = 0.04$	$\beta_{k,j} = 0$	$\beta_{k,j} = 0.02$	$\beta_{k,j} = 0.04$
1	0.012	0.012	0.012	0.36	0.36	0.36
2	0.012	0.013	0.017	0.36	0.39	0.49
3	0.012	0.015	0.029	0.36	0.43	0.70
4	0.012	0.017	0.050	0.36	0.48	1.08
5	0.012	0.021	0.092	0.36	0.55	1.73
6	0.012	0.026	0.17	0.36	0.64	2.85
7	0.012	0.033	0.32	0.36	0.75	4.76
8	0.012	0.042	0.59	0.36	0.91	7.99
9	0.012	0.055	1.08	0.36	1.10	13.37
10	0.012	0.073	1.95	0.36	1.36	22.29
	Effect on $Y_{j\neq1,t+20}$ from a shock to $\epsilon_{1,t}$					
	$\beta_{k,k} = 0.8$			$\beta_{k,k} = 0.95$		
k	$\beta_{k,j} = 0$	$\beta_{k,j} = 0.02$	$\beta_{k,j} = 0.04$	$\beta_{k,j} = 0$	$\beta_{k,j} = 0.02$	$\beta_{k,j} = 0.04$
1	0	0	0	0	0	0
2	0	0.006	0.013	0	0.16	0.33
3	0	0.008	0.024	0	0.20	0.55
4	0	0.011	0.046	0	0.25	0.93
5	0	0.014	0.088	0	0.31	1.58
6	0	0.019	0.17	0	0.40	2.70
7	0	0.026	0.31	0	0.52	4.61
8	0	0.035	0.58	0	0.67	7.83
9	0	0.048	1.07	0	0.87	13.22
10	0	0.066	1.95	0	1.13	22.14

Table 1: This table displays the results of the simulated computational example of the effect on $Y_{1,t+20}$ after introducing a unit shock to $\epsilon_{1,t}$ twenty periods earlier, depending on the own autocorrelative term ($\beta_{k,k}$, the off-diagonal autocorrelative term ($\beta_{k,j}$) and the number of linked outcomes (k). The top panel displays effects on Y_1, while the bottom panel displays effects on $Y_{j\neq1}$.

are only 0.02, which would appear feasible when discussing relationships like education, wages, labor supply, and prices, the effect of a unit shock to $Y_{1,t}$ on $Y_{1,t+20}$ and $Y_{2,t+20}$ are 0.91 and 0.87, dramatically higher in both cases. Indeed, adding a ninth potential variable causes the effect on all variables to be increasing over time, meaning cities diverge in the relevant economic covariates in the long run.[13] As additional, important, and durable economic outcomes such as those in the immigration literature are found to be affected by immigration, it makes it more likely that the immigration stock far in the past may have ongoing effects today, violating the exclusion restriction. As we have stated, while reality does not

[13]For a related discussion of how the instrument may muddle short- and long-run effects of immigration see Jaeger, Ruist, and Stuhler (2018).

change when an additional paper is published, these additional papers may cause a researcher to update their priors about whether or not the state of the world is well-described by the first row of Table 1.

III Data

Our data collection process contained two parts. First, from 1995-2019, all uses of articles with the phrase “instrumental variable" in the American Economic Review, the Journal of Political Economy, the Quarterly Journal of Economics, Econometrica, and the Review of Economic Studies were catalogued. This resulted in approximately one thousand IV uses being examined. Many of these papers included common instruments, such as (1) dynamic panel instruments outlined in Arellano and Bond (1991) or Blundell and Bond (1998) (2) competing product characteristics or counts as in Berry, Levinsohn, and Pakes (1995) (3) Bartik (1991) instruments, and (4) Frankel and Romer (1999) “gravity" based instruments. This paper does not discuss any of these extremely common instruments, except in the case of “immigrant enclaves."

We then examined this list and found the six potentially concerning instruments discussed above. Following this, we searched for uses of these instruments on economic journal-hosting websites, Elsiever, Online Wiley, JSTOR, and some journal-specific websites. If an article cited use of an instrument or a related instrument, we pursued that lead, yielding some journals not in our original search criteria but relevant to and cited in the literature. Our target journal was typically in the top sixty journals in RePeC’s journal rankings. This garnered additional 231 papers relevant to these six instruments. Importantly, the text below cannot possibly touch on all the uses of these instruments, and consequently we attempt to discuss only the most highly relevant papers, though we include other papers in our tables.

IV Six Categories of Related Instruments

We discus each of the six instrumental variable groups separately, highlighting the common variation and correlations shared by the instruments in each group that pose a threat to identification.[14]

[14]This common variation is the real source of concern for empirical work. Even if the IVs in a group are not identical, the fact that they are significantly correlated suggests previous assertions of the validity of these instruments are likely misguided.

Elevation and bodies of water

Changes in elevation and the presence of bodies of water (including rivers and streams) are used as an instrument 21 times in “top five" papers, and 23 times in top field or general interest journals. They are used, either implicitly or explicitly, to instrument for approximately fifteen outcomes: (1) change in housing prices (2) change in housing stock (3) city density (4) farming (including income) outcomes (5) enterprise (6) segregation (7) school governance structure (8) number of county governments (9) presence of dams (10) cost of highways (11) broadband provision (12) share of developed land (13) access to international markets (14) access to domestic market center and (15) presence of piped water. These fifteen instrumented covariates are then used to estimate more than thirty-five unique outcomes, from household health, longevity, and fertility, to firm earnings, worker wages, household education decisions, and investment outcomes, to air quality, industrial composition, trade openness, and educational efficiency.

Perhaps the most important recent paper in this literature is Saiz (2010), which produces city-specific estimates of the elasticity of housing supply. Because housing prices and quantities are determined by supply and demand, and because residents in more inelastic cities are more able to increase prices through more regulation, the paper must instrument for both the demand-side of housing and for regulation, which is coincident with the elasticity itself. To deal with these issues, the paper first produces novel measures of developable land by calculating the amount of MSA area either lost to internal bodies of water (including rivers, lakes, wetlands, and other water features), lost to oceans and the Great Lakes, or lost because the soil gradient is 15% or more. To instrument for both demand and regulation, Saiz (2010) uses the interaction of land unavailability with a shift-share industrial composition instrument, immigration shocks, and average January hours of sun, along with their levels. In analyzing the simultaneous role of regulation, the paper instruments for local housing regulation using the share of local expenditure spent on protective inspections and the nontraditional Christian share in 1970. With instrumented regulation and quantities, Saiz (2010) is able to estimate housing supply elasticities. Importantly for this paper, Saiz (2010) uses immigration shocks, average city weather 1941-1970, and elevation and bodies of water together as instruments. While this paper flags these instruments as concerning, Saiz reports that the results are similar if each instrument, including additional instruments of climate or immigration, is used *separately* as well. Moreover, the author also notes that his first stage correlations hold even conditional on coastal status. These implicit overidentification-style tests may help alleviate some concerns about using these instruments.

Saiz’s measure has been influential: the paper has more than 1200 citations in Google Scholar as of 2019. A number of top journal articles have been based around using the housing supply elasticity measures of Saiz (2010) along with the change in national housing prices, in a Bartik-like framework. Mian and Sufi (2009) and Glaeser, Gyourko, and Saiz (2008) both establish strong relationships between Saiz's measure and housing price changes, but do not use it as an instrument. However, Mian and Sufi (2011), Mian, Rao, and Sufi (2013), and Mian and Sufi (2014) use Saiz's elasticities combined with national housing prices to measure the exposure of MSA's to exogenous housing price declines. This is used to instrument the actual housing price decline, which then affects debt growth, employment growth, and the change in consumption for each paper respectively. It is also used as an instrument for the same variable to understand the effect of housing downturn exposure on the purchase of automobiles (Berger and Vavra, 2015), corporate investment (Chaney, Sraer, and Thesmar, 2012), and industrial composition (Beaudry, Green, and Sand, 2012). This multiple use of the same instrument on the same covariate for different outcomes may not ordinarily be concerning.[15] However, religion has been used to establish housing supply elasticity, and local religiosity (instrumented for with area ancestry) may causally affect education, marriage and divorce rates, and welfare disability receipt (Gruber, 2005). Religion is also related to social trust, bank regulation, beliefs about redistribution, and private school presence (Zak and Knack, 2001; Barth et al., 2013; Guiso, Sapienza, and Zingales, 2006; West and Woessmann, 2010). It is plausible the latent variables affecting education, marriage, social trust, beliefs about redistribution, and actual welfare receipt affect business cycle dynamics, corporate investment, and household portfolio choice.

Elevation and the presence of water around a city are also important factors in dam construction (Duflo and Pande, 2007). Duflo and Pande use river gradients as an instrument for dam construction in India, finding that "...A gentle river gradient (1.5-3 percent) increases the number of dams, while a steep gradient reduces it. However, a very steep river gradient (more than 6 percent) increases dam construction." Dams are not unimportant: as the authors report, 19% of the world's electricity supply, and 30% of irrigated land is generated via dam. Lipscomb, Mobarak, and Barilam (2013) find the presence of dams (instrumented using river gradient) to be causally related to local longevity, income, education, infant mortality, poverty rates, urbanization, illiteracy, human capital measures, and years of school. Taken together, these papers suggest many economic variables may be affected by rivers, bodies of water, and gradient changes other than via the elasticity of housing supply.[16] Finally, dams are found to mitigate the effects of weather

[15]Saiz's values for land elasticity, not used in conjunction with national housing prices, has also been found to be linked to the long-run income and population convergence patterns of cities (Ganong and Shoag, 2017), suggesting that high-income cities with relatively less elastic housing may have experienced atypically low income convergence compared to historical norms.

[16]It is also worth noting that because dams are important sources of electricity, and their power generation is subject to weather conditions, weather interacted with dams (discussed below) has been used as an instrument for electricity production as well

shocks (discussed in another section) on agricultural production (Sarsons, 2015).

Rivers and bodies of water instrument not only housing availability and dam presence, but also segregation (Cutler and Glaeser, 1997), school competition (Hoxby, 2000)[17], and the number of county governments (Alesina, Baqir, and Hoxby, 2004; Hatfield and Kosec, 2013, 2019). While these are separate endogenous covariates, they are linked to rivers because rivers "divide MSAs into natural subunits." Instrumented-for segregation is found to differentially affect a number of important economic outcomes, such as education, income, and single motherhood, all by race (Cutler and Glaeser, 1997). Indeed, the fact that bodies of water are linked to differential labor market outcomes by race suggests that other outcomes, such as consumer spending over the business cycle, may differ in these cities for reasons other than housing prices. Instrumented-for school governance structure is found to improve student achievement and school efficiency. Smaller governments are found to compete less on regulation, and as a consequence affects economic development (Hatfield and Kosec, 2013) and air quality and industrial composition in the form of employment in polluting industries (Hatfield and Kosec, 2019).

Finally, while not used as an instrument per se, Felkner and Townsend (2011) find that "growth in enterprise is more likely in areas of lower, flatter elevation and in areas closer to rivers and waterways." They additionally find being near these naturally enterprise-friendly areas spurs enterprise in surrounding areas, suggesting a strong spatially autocorrelated role for both elevation and rivers. Additionally, the presence of rivers have been used as a source of potential traffic congestion in Winston and Langer (2006).

Figure 5 below summarizes a subset of papers dealing with elevation and water graphically. Blue are the instruments or measurements created from the presence of elevation changes and bodies of water. In pink are the instrumented-for endogenous covariates. In red are outcomes. As can been seen, slopes and bodies of water have the potential to affect an incredible variety of covariates in independent ways. Because most of these covariates of interest are important, such as a city's segregation, or educational efficiency, housing elasticity, or presence of a dam, they are likely to affect other economic outcomes of interest, which makes causal inference using the instrument difficult.

It is important to note that Kolesár et al. (2015) provide a novel potential solution to our identification problem. Specifically, they note that, in the language of Figure 2, if the direct effects of Z on Y_1 (say through the alternative channel of X_2) are orthogonal to the effects of X_1 on Y, then we can design a

(Allcott, Collard-Wexler, and O'Connell, 2016), leading to potentially differential effects of weather and climate for these cities. The authors also note that this may lead to different industrial composition in the long run.

[17]Hoxby (2000)'s definition of streams included inlets, lakes, ponds, marshes and swamps if they are "roughly curvilinear in form."

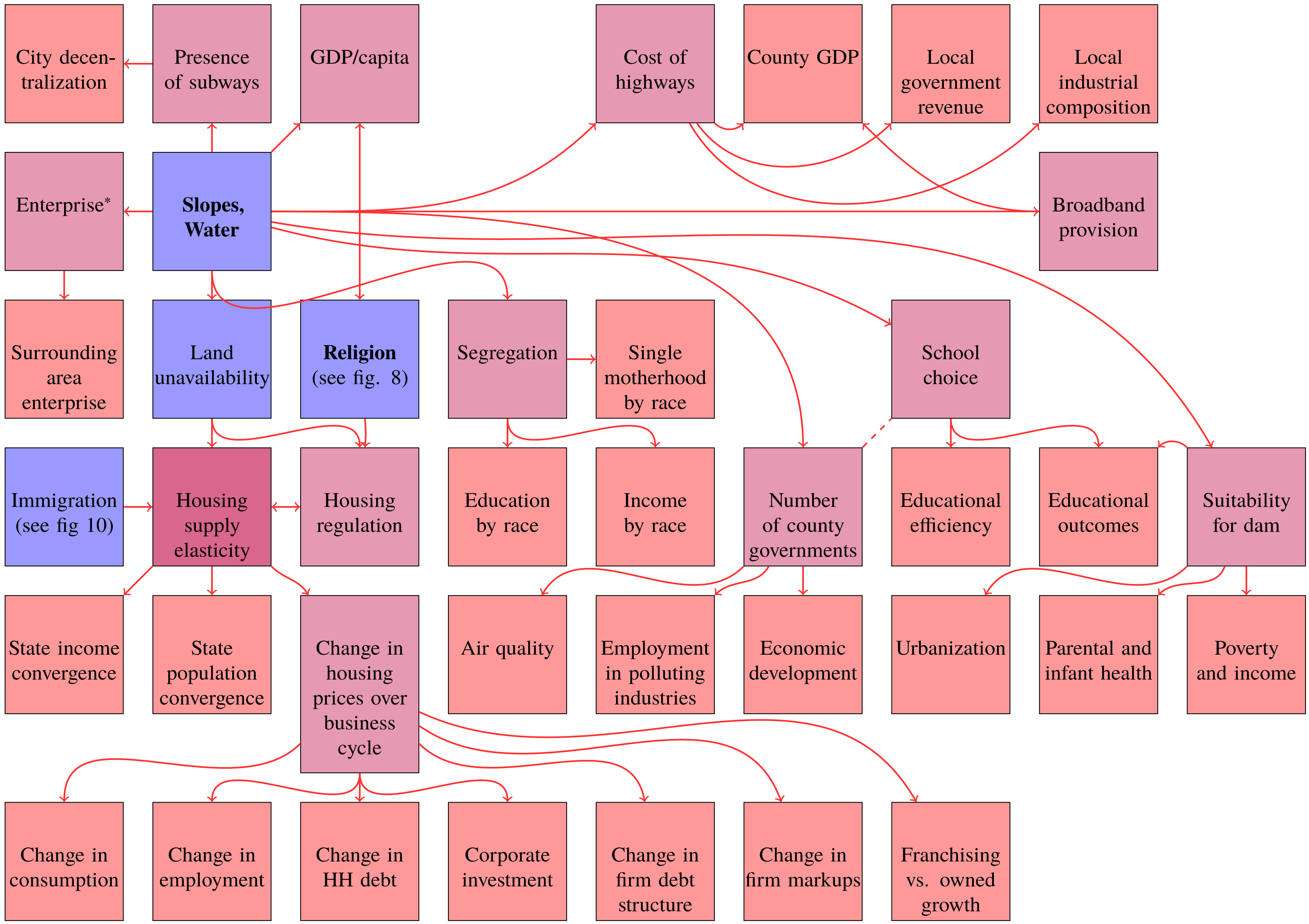

Figure 5: This figure summarizes selected research using elevation and bodies of water as an instrument. Blue are the instruments or measurements created from the presence of elevation changes and bodies of water. In pink are the instrumented-for endogenous covariates. In red are outcomes. Enterprise* is not directly treated as an instrument. A more complete list of 55 paper-instrument-outcomes can be found in Appendix Table B: IV Tables. Some variables, such as “city decentralization" and “urbanization" or “economic development" "GDP/capita" and “poverty and income" are listed in multiple places for legibility.

bias-corrected two-stage-least-squares estimator that is consistent. However, an inspection of Figure 5 suggests the assumption that X_2's effect (or Z's direct effect) on Y_1 are orthogonal to the effects of X_1 is an uphill climb, but by no means impossible. For example: consider the concern that slopes/bodies of water affect the ease of segregation, which affects education and income by race. If these affect an MSA's employment during the great recession in a way other than through the change in housing prices, we have provided a cause for concern. However, if the effects of (1) MSA-level segregation, (2) education by race and (3) income by race on employment over the business cycle is orthogonal to the effect of housing prices on employment, then a consistent estimator may still be recovered. We argue that Figure 5's focus on so many important economic outcomes that are likely to be at the center of a complex causal web (income, education, housing prices, urbanization), there remains no fit-all solution to the problem we identify.

Sibling structure

Perhaps because it is available in many datasets, the age/gender composition of an individual's siblings is a popular instrument, and is used 12 times in "top 5" papers and 36 times other well-ranked journals. In one early paper, gender mix is used as an instrument for women's education, which affects earnings (Butcher and Case, 1994). However, Angrist and Evans (1998) argue households desire a mixed sibling sex composition, as households whose first two children are girls are empirically more likely to have a third child. Thus, they use sibling gender mix to instrument for family size, which affects maternal labor supply. Gender mix as an instrument for family size is also used to study child private school attendance and academic performance (Conley and Glauber, 2006), parental marital status and non-traditional family living, income, health insurance, blood pressure and obesity (Cáceres-Delpiano and Simonsen, 2012), as well as child BMI and illness (Palloni, 2017).

However, the gender and age mix of children is also used as an instrument for welfare generosity. According to housing and urban development rules, children of opposite sex above a certain age cannot be required to share a bedroom. Thus, a household with two girls is less likely to receive generous housing benefits. Consequently, child gender mix is also used as an instrument for children living in a public housing project, which affects the neighborhood they live in, their school quality, and whether or not children repeat a grade (Currie and Yelowitz, 2000).

While households are more likely to have a third child if they initially have two girls (as opposed to mixed gender or two boys), they are also more likely to get divorced if their first child is a girl. Lundberg and Rose (2003) show a link between first-born child sex and female marital status, finding first-born sons are more likely to cause a mother to transition into marriage. Bedard and Deschenes (2005) use first child sex as an instrument for marital dissolution, which affects female earnings (differentially by income category). Ananat and Michaels (2008) also find that having a female first-born child increases marital dissolution, showing it causes dispersion in earnings, increasing both very high and very low income households. Dahl and Moretti (2008) find households with a firstborn son are more likely to remain intact, which they argue is due to the desire of fathers for sons. Instrumenting for absentee fathers with the firstborn child's gender, they estimate an effect on family income of $128/year (intent to treat), or $18,000/year for the households whose father leaves due to child gender. These papers suggest child gender composition and order strongly influence family dissolution in addition to family size, and may affect many of the same variables of interest through a distinct causal channel.

Because women become mothers at an earlier age than men become fathers, first-born child sex is also used as an instrument for the age at which a parent becomes a grandparent. (Rupert and Zanella, 2018) find that a woman whose first child was a girl becomes a grandmother 2.5 years earlier than if her first child was a boy. The instrumented effect of being a grandparent is to reduce the total annual hours of women aged less than 80 who have children above the age of 14 an economically meaningful amount: by 32% (from 1600 hours per year). Dynamic optimization suggests a mother with knowledge of this is likely to change her education, marriage, and labor force decisions earlier in life for reasons other than simple family size, offering a third causal channel.

The age and gender mix of siblings is also used as an instrument for sibling contribution to parental care, which affects a child's own contribution to parental care (Antman, 2012), parental nursing home use (Houtven and Norton, 2004), child depressive symptoms (Coe and Van Houtven, 2009), and child employment and hours worked (Bolin, Lindgren, and Lundborg, 2008). It is also used as an instrument for migration, which affects earnings (Abramitzky, Boustan, and Eriksson, 2012) and parental outcomes (Antman, 2010).

Number of siblings itself is used as an instrument for a variety of outcomes. For instance, Levin and Plug (1999); Taber (2001); Korpi and Tåhlin (2009) use it as an instrument for schooling, which affects wages. The same method is also used (with other instruments) to show the effect of schooling on a smoker's choice to quitting smoking (Sander, 1995). While Ziliak and Kniesner (1999) allow the number of children to affect labor supply (but not other parameters) in their model, they do not formally use it as an instrument.

Closely related are studies that use twins as an instrument. Typically twins are also used to instrument

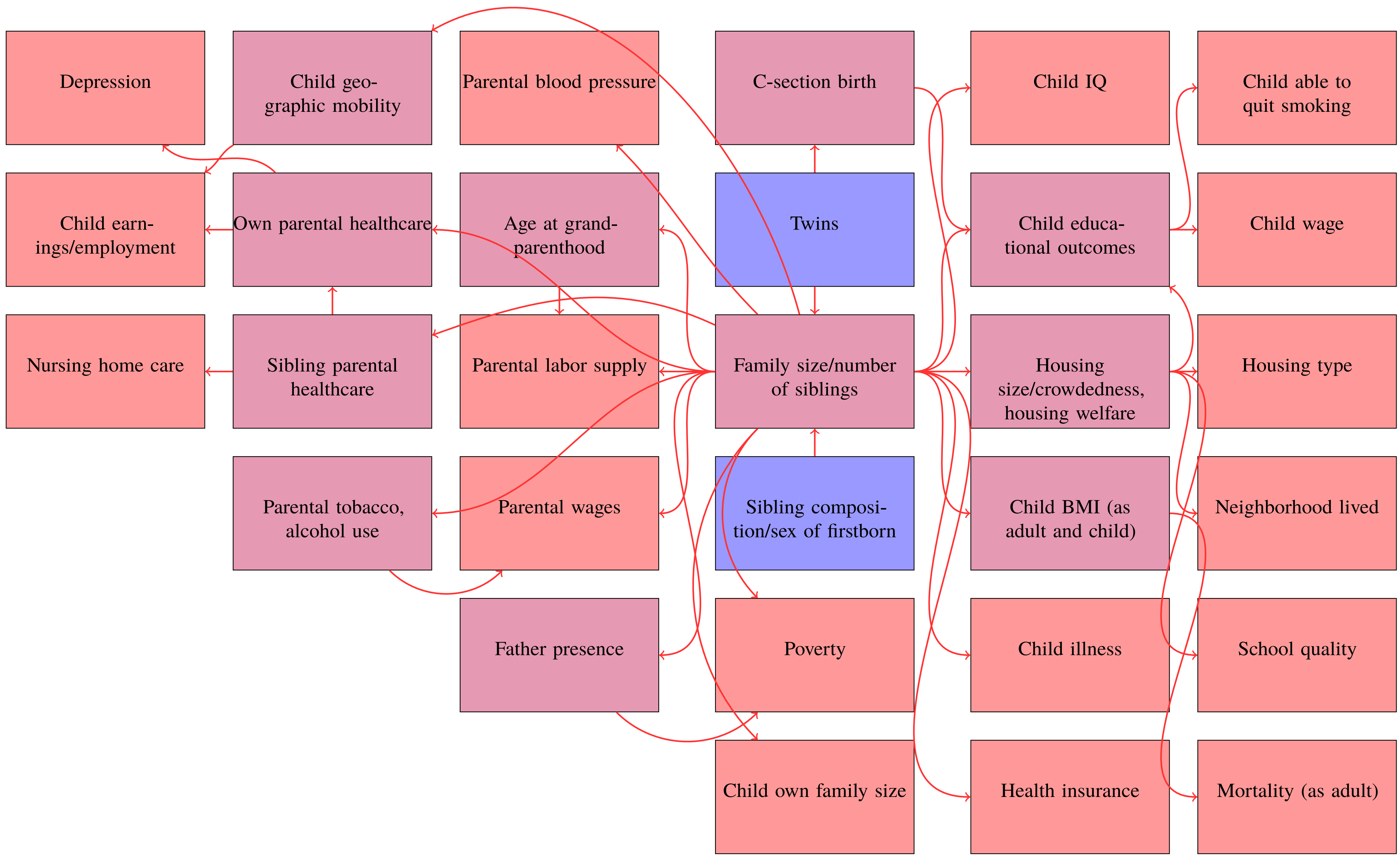

Figure 6: This figure summarizes selected research using sibling structure. The main instruments are in blue, the endogenous covariates are in purple, and the outcomes are in pink. Some papers use family size and sibling structure directly to instrument for other covariates of interest, so it is also linked to endogenous covariates.

for family size, and instrumented family size affects whether or not a mother is ever married, her labor force participation, earnings, family earnings, welfare receipt, and poverty status (Bronars and Grogger, 1994; Angrist and Evans, 1998)[18]. Twins as an instrument for family size also affect divorce rates (Jacobsen, Pearce, and Rosenbloom, 2001), child personality (Fletcher and Kim, 2019), whether or not a child is living without a father (Dahl and Moretti, 2008), child IQ (Black, Devereux, and Salvanes, 2010), child schooling (Åslund and Grönqvist, 2010), children's own family size (Kolk, 2015), and children's geographic distance from their mother (Holmlund, Rainer, and Siedler, 2013). Figure 6 depicts a subset of the highly interlinked set of regressions that use gender/age composition or twins as instruments.

One potential flaw of using twins as an instrument for family size is that it not only affects the age composition of siblings, which as described affects a number of outcomes via a different channel than family size, but also the gender composition. There appears to be an excess of same-sex non-identical twin pairs even beyond that which would be predicted by already-unbalanced non-twin live birth ratios (James (1971)). Moreover, approximately 75% of twin births were delivered via Cesarean section which appears to be related to long-run child cognitive development, potentially via Cesarean-related after affects such as disturbed gut bacteria (Polidano, Zhu, and Bornstein, 2017). It is also probable short-run child development affected by cesarean-section or twin births affects a multitude of parental choices and child outcomes: for instance, Dahl and Moretti (2008) find evidence that fathers are more likely to leave children with health problems. Finally, by compressing birth events, twin births, also affect mother's age at potential sibling's birth, which has itself been used as an instrument for the presence of a younger sibling, and thereby probability of marriage by an older sibling (Vogl, 2013), maternal employment and thereby child adverse health events (Morrill, 2011), and a child's desired number of children (Rasul, 2008). This suggests twin births may have affects on child and parental outcomes not simply via family size per se, but also through age composition, gender composition, and Cesarean section births.

While information on siblings is present in many datasets, it is used for such a large variety of outcomes that its use as an instrument must be limited. Many of these uses, such as the likelihood siblings will be present to take care of an elderly parent, child geographic mobility, and parent's age at grand-parenthood are likely to affect decision-making such as education earlier in the lifecycle via household planning. Worse, while a household member may know which sibling is likely to live near parents when older and provide care, that information is far less likely to be available to researchers.

[18]Bronars and Grogger (1994) only report IV estimates in text, not tables, but produce tables in which IV estimates may be backed out.

Ethnolinguistic fractionalization and language

Ethnolinguistic fractionalization (ELF) and similar concepts, including ethnic fractionalization, linguistic fractionalization, ethnolinguistic polarization, or fraction speaking a European language[19], are frequently used to examine the effect of “institutions" or “governance" on relevant economic outcomes, most importantly growth. However, governance is a broad concept, and these ethno-linguistic measures are used to discuss multiple distinct but interrelated concepts. Usefully, Kaufmann, Kraay, and Zoido-Lobatón (1999) define three distinct components of governance: (1) the rule of law, (2) bureaucratic efficiency/effectiveness, and (3) graft/bribery. As noted in Hall and Jones (1999), a useful economic definition might quantitatively map these to the wedge between private costs and returns, or private returns and social returns. However, in practice these indices are taken as given, assuming the difference in a wedge from increasing a corruption index from a one to a two, and then from a two to a three are the same, though they might not be.[20] There are therefore three immediate problems with how ELF and related concepts are commonly used. First, different papers focus on very different governance concepts, second, the indices are not necessarily well-scaled, and third, it is used to instrument for a variety of endogenous covariates, even beyond governance.

With regards to the first issue, one of the first papers to examine ethnolinguistic fractionalization, Mauro (1995), uses ELF to instrument, in different regressions, for both corruption’s effect on investment and growth and then for bureaucratic efficiency’s effect on investment and growth. Notably, bureaucratic efficiency includes judiciary efficiency, red tape, and corruption. However these three concepts are not the same: for instance, using Mauro’s data, we calculate that even after extracting the portion of “red tape" that covaries with “corruption," there is still a strongly positive correlation between red tape and ELF. While such regressions might plausibly suggest institutions affected by ELF matter, it is not clear which institution actually are. Indeed, after controlling for investment in separate regressions, Mauro (1995) notes corruption’s statistical importance on growth vanishes, while bureaucratic efficiency’s does not. This suggests ELF acts through corruption to reduce investment, but also through other channels related to bureaucratic efficiency (including investment itself).

[19]While fraction speaking English or another European language is potentially different in concept, the correlation between speaking a European language and ethnolinguistic fractionalization is -0.29 (taken from Hall and Jones (1999) and Bazzi and Clemens (2013) data). ELF and related concepts are used 4 times in “top 5" papers and 37 times in other, well-regarded journals. In a regression of fraction speaking a European language on ethnolinguistic fractionalization, for every 1% more fractionalized a society becomes, it is -0.4% less likely to speak a European language.

[20]We could find no evidence that the governance indices commonly used are well-scaled. This scaling problem has the potential to flip regression results, as it does for test scores in Bond and Lang (2013) and happiness rankings in Bond and Lang (2019) .

As an instrument for corruption alone, ELF has been used by Michaelides et al. (2015) and Michaelides, Milidonis, and Nishiotis (2019) to instrument for a country's Transparency International's "Corruptions Perceptions Index," which we denote as TI Score. While the TI score is generated from a range of 13 different sources, the questions primarily appear to focus on the inappropriate use of public funds for private gains and bribery.[21]They find higher corruption is linked to information leakages before government debt downgrades and high volatility of asset prices and exchange rates.

ELF is also used as an instrument for "governance" more broadly, which typically includes not only corruption but also governmental efficiency and the rule of law. For instance, in their large sample regressions, Rodrik, Subramanian, and Trebbi (2004) use fraction speaking english as an instrument, among others, to instrument for both the rule of law and for market integration, finding per-capita GDP is primarily determined by institutions rather than integration. Importantly, the fraction speaking english and other European languages had a significant effect on both concepts. Wang (2013) uses ELF as an instrument on the first principal component of all three types of governance from the Worldwide Governance Index, finding the first principle component of "governance" significantly affects R&D spending. Méon and Sekkat (2008) allow ELF, along with distance to the equator and legal origin, to instrument separately for three different concepts of governance. They find the rule of law may cause a statistically significant increase in non-manufactured goods (but not total trade or manufactured goods), while governmental efficiency (specifically the regulatory framework) affects manufacturing exports (but not total or nonmanufactured exports). Finally Méon and Sekkat also find democracy appears to cause increases in nonmanufactured trade, but not total or manufactured trade. Ades and Glaeser (1995) use ELF, along with other variables, as an instrument for a country being under both dictatorship and its trade policies, which affects main city size.

While noting "social infrastructure" should measure the wedge between social and private returns, Hall and Jones (1999) construct their measure by taking the sample average of measures of (1) law and order, (2) bureaucratic quality (3) corruption (4) expropriation risk (5) government repudiation of contracts and (6) years open to international trade, with the first five given 10% weight each and the last 50% weight. They then instrument this measure with fraction speaking English and fraction speaking a European language, finding it is correlated strongly with total factor productivity. Similarly, Alcala and Ciccone (2004) use the fraction of a country speaking a European language, along with other instrumental variables, to instrument for both institutional quality, (which is a combination of government effectiveness, rule of law, and graft,

[21]Most of the sources are surveys, one of which is Mauro (1995)'s own source Business International.

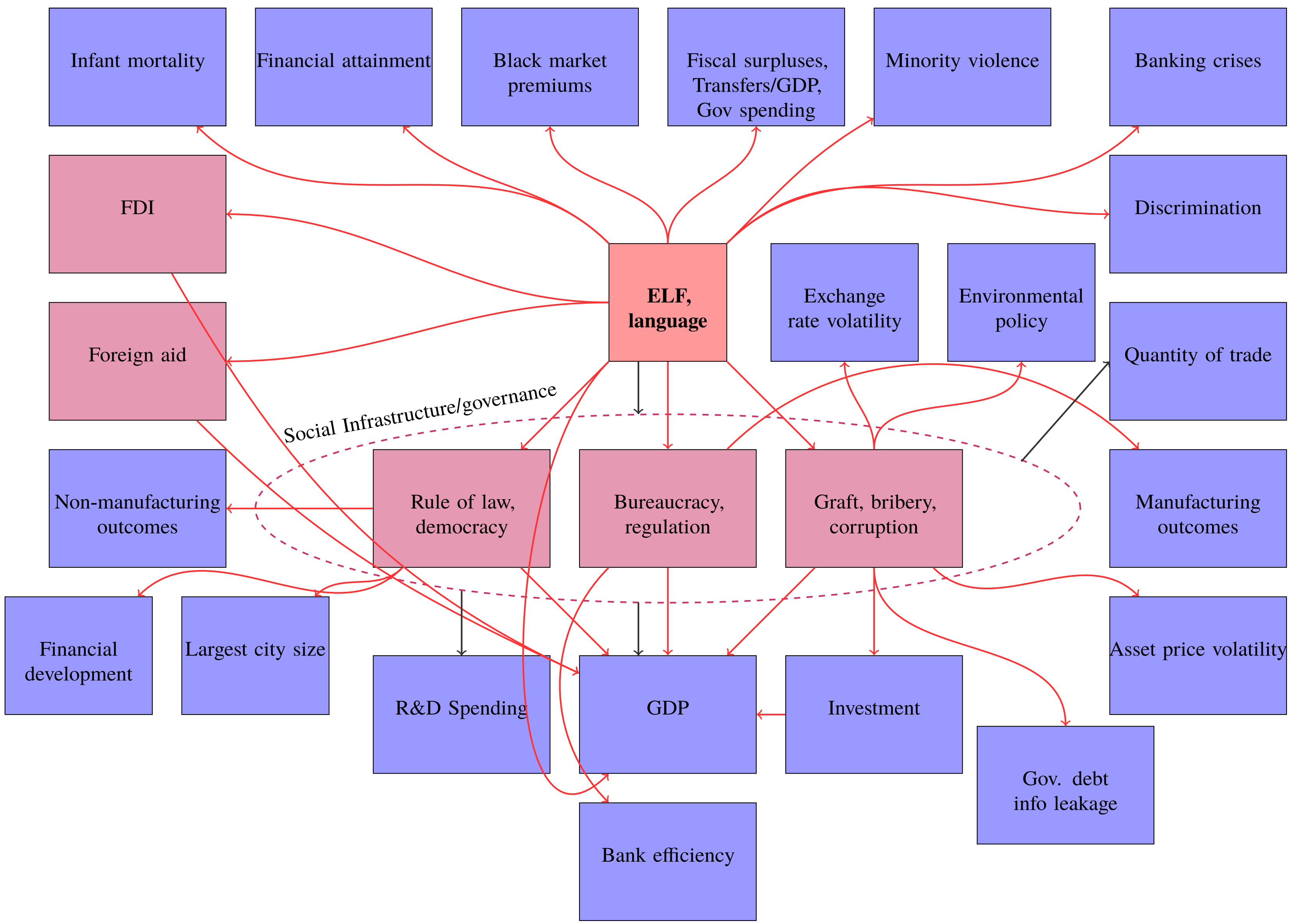

Figure 7: This figure summarizes selected research using ethnolinguistic fractionalization as an instrument. Instrument is in blue, endogeneous covariates in purple, and outcomes in pink. Because ELF is used for both agglomerations of (1) rule of law/democracy, (2) bureaucracy and (3) regulation and graft, bribery and corruption, as well as each individually, we depict with black lines generating from a dashed circle outcomes that use agglomerations (“social infrastructure/governance").

to be comparable with Hall and Jones) and for quantity of trade (real openness). Importantly, European language is significant in both regressions.

While Hall and Jones argued European languages are naturally correlated with the extent of Western European influence on social infrastructure, it also likely plays a different role through aid. However, having a common language with foreign aid givers is also used as an instrument for foreign aid, which causes growth (Rajan and Subramanian, 2008). Relatedly, Kourtellos, Tan, and Zhang (2007) find ethnolinguistic fractionalization interacted with aid had differential effects on growth. Moreover, in a secondary regression Delgado, McCloud, and Kumbhakar (2014) instrument for foreign direct investments using ELF, finding it positively related to growth. However, this potentially undermines the suggested relationship between aid and growth. In addition, language is also used as an instrument for gender discrimination, which affects migration from a country (Ruyssen and Salomone, 2018).

Finally, three papers find reduced-form relationships of ELF with a variety of outcomes, with suggestive causal evidence. LaPorta et al. (1999) find ELF is negatively associated with a wide range of government performance inferiority, from property rights and regulation to corruption, delays, tax compliance, public goods, and government intervention. However, they find this is typically reflected through per-capita income and pushed out by latitude: only public good provision and state ownership of firms remain after controlling for those two. Two less causally-focused papers relate ELF to an enormous slew of variables including school attainment, financial attainment, black market premiums, fiscal surpluses, infrastructure spending, discrimination, and minority violence, (Easterly and Levine, 1997), and also banking crises, property rights, business regulation, transfers and subsidies as a fraction of GDP, democracy and political rights, and infant mortality (Alesina et al., 2003). The sheer volume of ELF's correlation with so many potentially interrelated variables may give a researcher pause in interpreting causal channels. For instance , (1) foreign direct investment, (2) foreign aid, (3) rule of law and democracy (4) bureaucracy and regulation and (5) graft, bribery and corruption all give rise to distinct causal channels.

Religion

Like ethnolinguistic fractionalization, national, local, or personal religion is a frequently-measured and unambiguously important factor for numerous economic outcomes. Because religion shapes culture, it is frequently used as an instrument for many endogenous covariates and as a control, and shows up seven times in "top 5" papers and 37 times in other well-regarded journals. And, like ethnolinguistic fractionalization, it is also occasionally instrumented-for. We discuss both literatures, as the outcomes of some papers naturally interact with the instrumental validity of others.

Historically, religion is discussed by a number of papers that were relatively careful about causality, but noted it was suggestively and intuitively linked to a number of outcomes. Guiso, Sapienza, and Zingales (2003), while "well aware of the difficulty in interpreting the observed correlation as causal effects," find religion is linked to an enormous number of outcomes: attitudes toward a variety of topics, such as cooperation, government, working women, the market economy, and racism, but also legal rules and societal thriftiness. Similarly, LaPorta et al. (1999) link religion to property rights, regulation, tax rates, corruption, bureaucratic delays, the size of government, public rights, infant morality, illiteracy and schooling, and democracy. Stulz and Williamson (2003) find it correlated with investor productions, McCleary and Barro (2006a) with economic growth, and Brainerd and Menon (2014) with infant mortality.

These correlational papers are important when examining instrumental papers, as they highlight the causal structure of religion's effects. We suggest they may help eliminate some stories. For instance, if any of previous variables, such as bureaucratic delay, tax rates, or corruption affect housing prices in ways other than regulation, the very commonly-used measurements of land elasticities from Saiz (2010) may be mismeasured. Similarly, it is difficult to reconcile all these correlations if religion's only effect on entrepreneurship and savings is through trust, as in Guiso, Sapienza, and Zingales (2006).

Other than being used as an instrument for social trust and for land regulation, religion is used in a variety of other situations. Its use in instrumenting for social trust extends to Bjørnskov (2012), which finds it also affects the rule of law and schooling expenditures. It is used as an instrument for societal respect and responsibility in Breuer and McDermott (2013), which affects per-capita GDP, and for democracy, which also affects per-capita GDP (Mobarak, 2005). It is also used as an instrument for national uncertainty aversion, which affects differential industry-level growth in Huang (2008). Hakkala, Norbäck, and Svaleryd (2008) and Arin et al. (2011) use it as an instrument for corruption, which affects FDI as well as the ability of the government to consolidate spending during business cycle contractions. It also is used as an instrument for the free press, which affects the ability of a government to deceive, specifically by stating an exchange rate regime different from the de facto regime Méon and Minne (2014).

At a micro level, religion is more frequently instrumented-for. For instance, Gruber (2005) instruments local religiousity with area ancestry, finding local religiousity to be causally linked to household income, welfare receipt, marriage, and divorce. These present important explanatory hurdles that any paper using religion as an instrument must pass before being given credence. Similarly, distance to Wittenberg is used as an instrument for protestantism in Becker and Woessmann (2009, 2008, 2018), finding that it affects

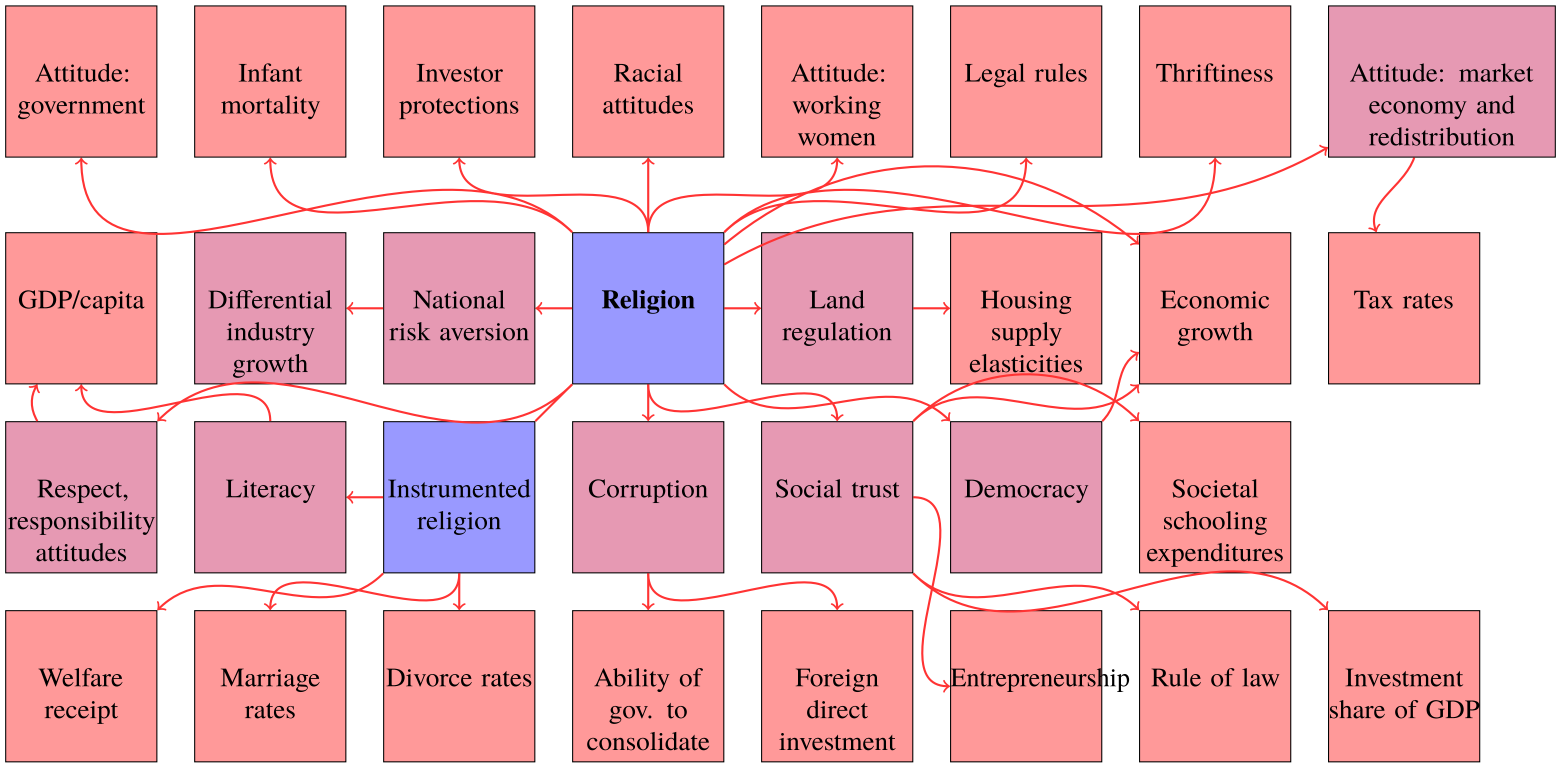

Figure 8: This figure summarizes selected research using religion as an instrument. Instrument is in blue, endogeneous covariates in purple, and outcomes in pink. Many papers link it correlationally to outcomes, which provide potential causal tests of IV papers. Many micro papers instrument for individual religion, and find it correlated with a number of important outcomes. While these papers may not be threatened by other papers in this graph, they may threaten papers and are included for that reason.

historical literacy, the gender education gap, and suicide rates. Another instrument for religious (mission) presence has been the initial missionary treks in Mexico, which are found to affect a slew of extremely long-run primary, secondary, and postsecondary school outcomes as well as catholicism (Waldinger, 2017). Another interesting historical instrument for protestantism is a prince's religion after the Peace of Ausburg, which affects hours worked four centuries later (Spenkuch (2017)). McCleary and Barro (2006b) help cement a concern about religiousity by showing that GDP/capita (instrumented using latitude and land-locked status) affects country-wide religiosity. The same instruments are used to extract exogenous variation in political constraints, which affect the presence of a state religion Barro and McCleary (2005). Ultimately, when religion is instrumented for, important causal channels are opened, and papers using religion as an instrument must grapple with these confirmed causal channels.

To conclude, religion has been shown to be statistically related to an enormous number of important economic variables. Similarly, it is causally linked in a number of papers to many disparate outcomes. The sheer variety of these outcomes suggests a single causal channel is unlikely. The variety of outcomes and instrumented endogenous variables suggest religion affects many important but different aspects of life, including culture, institutions, governments, regulation, personal morals and preferences.

Weather

Weather, as distinct from climate, is frequently used as an instrument for a plethora of endogenous covariates, and is in 30 "top five" publications and 40 other well-regarded publications. Perhaps most prominently, rainfall is used as an instrument for income, which affects the likelihood of conflict (Miguel, Satyanath, and Sergenti, 2004), discussed at length by Sarsons (2015). It is used as an instrument for growth or local income shocks, which affects local witch killings (Miguel, 2005), land invasions (Hidalgo et al., 2010), democratic change (Burke and Leigh, 2010; Brückner and Ciccone, 2011), consumption (Kazianga and Udry, 2006), remittances (Arezki and Brückner, 2012a), sale of durable investment goods (Fafchamps, Udry, and Czukas, 1998), trade balance (Brückner and Gradstein, 2013a), urbanization (**?**), manufacturing output, employment and capital investment (Lee, 2018) and the rate of time preference of households (Tanaka, Camerer, and Nguyen, 2010; Di Falco et al., 2019).[22] Rainfall also instruments for both agricultural and fishing productivity (Angrist, Graddy, and Imbens, 2000). Interestingly, while variation in planted crops' susceptibility to rain has been used as a way to generate differential income

[22]Sarsons (2015) finds rainfall does not affect the agricultural production of districts in India that were downstream of dams, but that conflict in these districts measured via ethnic riots persisted, suggesting conflict was caused by another channel other than income shocks. Sarsons leaves further investigation open for future research.

shocks, Kochar (1999) uses crops planted as a way to better measure weather expectations, suggesting exposure to weather-based income risk shocks are not random.

Rainfall is also used as an instrument for a variety of other outcomes. Rainfall's effects on hydroelectric power and cold/hot days are both used as an instrument for electricity shortages, which affect manufacturing outcomes and industrial output (Allcott, Collard-Wexler, and O'Connell, 2016; Fisher-Vanden, Mansur, and Wang, 2015). Temperature, dew points, and humidity are also used to instrument for energy demand, which affects energy prices (Ito and Reguant, 2016). However, agricultural and energy prices are not the only prices affected: rainfall instruments for international commodity prices, and grain prices specifically, (Brückner, 2012b; Mehlum, Miguel, and Torvik, 2006), which affect government tax revenue and crime rates (both property and violent) respectively. It is also used as an instrument for child wages, which affects school attendance (Jacoby and Skoufias, 1997). Indeed, the effect of last month's rainfall on food-staple prices appears to affect this month's trade policy utilization (Giordani, Rocha, and Ruta, 2016). Moreover, rainfall's lag structure itself is used to identify systems of supply and demand equations for staple commodities. Roberts and Schlenker (2013) note past weather shocks affect the supply of storable commodities (via inventories), while current weather shocks affect demand (again via future inventories). Consequently, past rainfall and current rainfall together instrument for supply and demand of corn, rice, soybeans, and wheat.

One prominent paper (Sarsons, 2015) expresses concern about rainfall as an instrument for conflict. Sarsons notes rainfall does not affect income as severely in areas downstream from dams. However, these places still show a strong relationship between rainfall and conflict, suggesting rainfall influenced conflict in ways other than income. While Sarsons (2015) did not find evidence of migration in India in her examination of rainfall as an instrument, others have. Prominently, Munshi (2003) uses past rainfall in a migrant's home community as an instrument for past migration, which affects the size of a migrant community in the U.S., and therefore an immigrant's occupation.[23] Past rainfall's effect on migration also serves to increase ties to a distant location, which reduces savings of stayers, presumably because it reduces the precautionary motive (Giles and Yoo, 2007). Consistent with rainfall-induced migration playing an important role, Jayachandran (2006) finds rainfall affects agricultural productivity, which affects wages differentially depending on the migration costs a district faces. More extreme weather appears to affect migrant flows as well, which acts as an instrument for hourly earnings, lost weeks of work, access to relief jobs, and going from full- to part-time work (Boustan, Fishback, and Kantor, 2010). Drought is also used

[23]Giulietti, Wahba, and Zenou (2018) also find that weak ties in a distant location caused by past rainfall's effects on migration increases the probability of own migration.

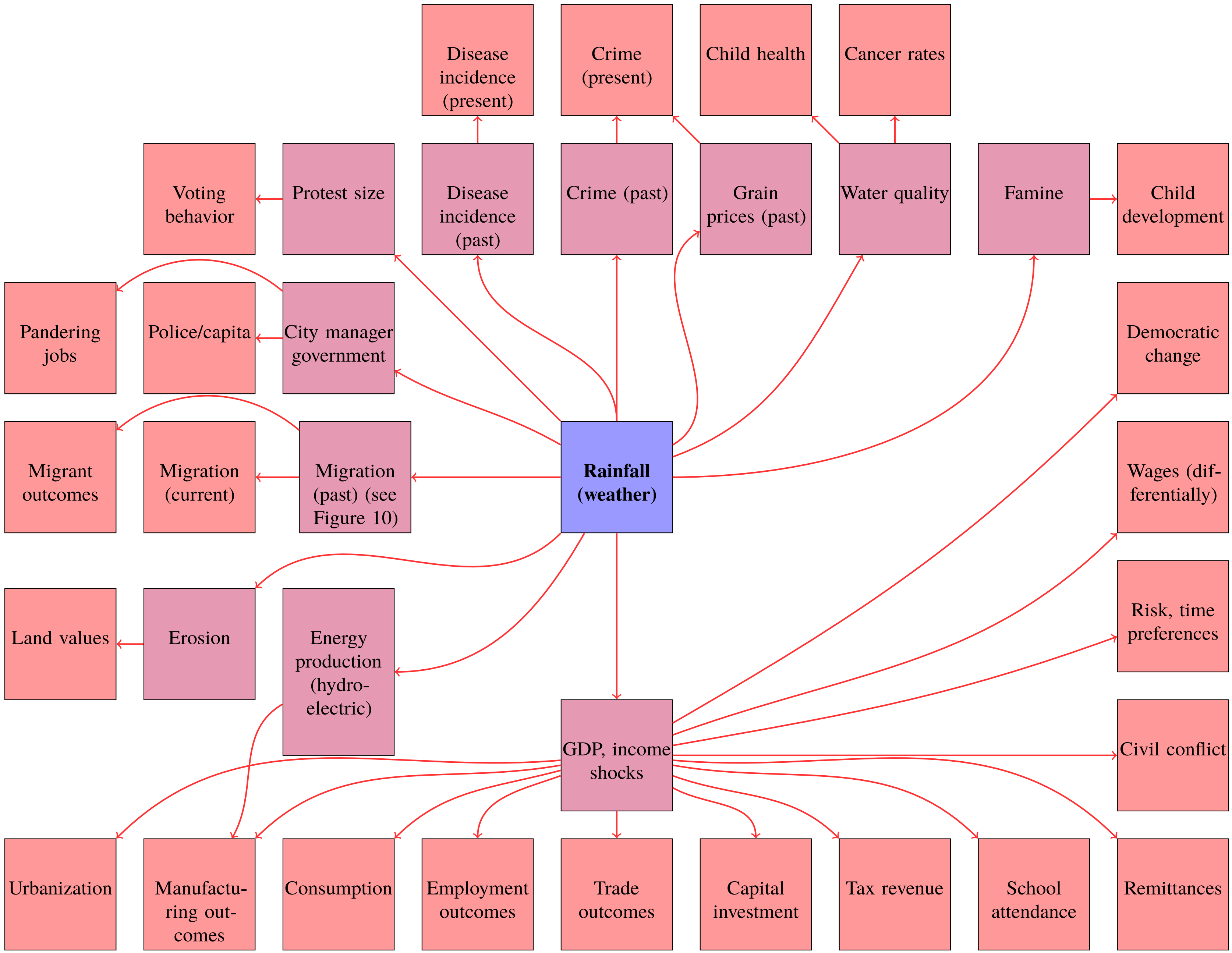

Figure 9: This figure summarizes selected research using weather as an instrument. Instrument is in blue, endogeneous covariates in purple, and outcomes in pink.

as an instrument for population, which affects civil conflict (Brückner, 2010).

Weather has a number of potentially more unusual channels as well. Sky cover has been used as an instrument for managerial expansion beliefs (moods), which affect actual hiring and capital investment (Chhaochharia et al., 2018). It reduces the likelihood of industrial inspection that day, which increases industrial pollution (Lin, 2013). It affects pollution in rivers, which affects cancer rates (Ebenstein, 2012). Because viruses thrive in cold and dry conditions, and affect socialization patterns, lagged weather is used as an instrument for disease transmission in flu, acute diarrhea, and chickenpox (Adda, 2016). Rainfall, combined with proximity to health clinics, instruments for acute illness, which affects sectoral labor supply (Adhvaryu and Nyshadham, 2017). It affects protests and thereby later voting behavior (Madestam et al., 2013), and meetings of loan groups, which affects loan defaults (Feigenberg, Field, and Pande, 2013). Rainfall variance (among other variables) is used as an instrument for land concentration, which affects banks per capita (Rajan and Ramcharan, 2011). Finally, it affects first week movie performances, which affects later sales (Gilchrist and Sands, 2016a).

Historical rainfall shocks have also been important. When the United States was on the gold standard, rainfall's effect on cotton production appears to have significantly affected industrial production via a monetary channel (Davis, Hanes, and Rhode, 2009). Vlaicu and Whalley (2016) document that adoption of a city manager system frequently occurred after severe pre-1936 rainfall shocks. Using historical rainfall events as an instrument for a city manager government (suggesting long-run effects of weather shocks), they find city managers appear to be less prone to political pandering, measured via lower police officer hires and higher officer employment volatility. However, lack of rainfall in the 1930's also instrumentally affected erosion on farmland, and appears to have had long-run effects on agricultural land values for decades afterward (Hornbeck, 2012).[24] Rainfall and rainfall variability is also used as a predictor (but not instrument) for enterprise activity, which affects the enterprise activity of surrounding areas (Felkner and Townsend, 2011).

With so many potential avenues for affecting households and nations, it is perhaps no surprise that rainfall is correlated with long-run effects. For instance, early-life rainfall is correlated with improved health, schooling, and socioeconomic status for women decades later (Maccini and Yang, 2009). Like religion, the sheer variety of endogenous covariates that are instrumented-for, and the number of different outcomes suggests weather, as an instrument, should be looked at carefully. We extend the concern of

[24]Using drought as an IV for erosion is only reported in an appendix. Most results are taken as a direct comparison between high- and low-erosion areas.

Sarsons (2015) about rainfall's use as an instrumental variable for income, which affects conflict, by documenting fifteen other uses of rainfall as an instrument that may help explain her results, as well as noting more than thirty other outcomes that may be affected by rainfall.

Immigrant enclaves

Immigrant enclaves are used as instruments in 12 "top five" papers and 58 other well-regarded articles. Its use as an instrument begins with Bartel (1989)'s documentation that immigrants appear to migrate to locations where other immigrants have already located. Altonji and Card (1991) use historical immigrant share by industry as an instrument for new immigrant shares by industry, which affect wages. The instrument took on its more classical form in Card (2001), which used historical immigration patterns to predict (instrument for) immigrant inflows, which affect native outflows, employment/population ratios for natives and immigrants, and wages. Importantly, the instrument is not valid under serial correlation: if, for instance, a city underwent a permanent and unmitigated productivity shock that brought both past immigrants and current immigrants to a city while simultaneously increasing wages of natives, then the instrumented effect of immigrants on native wages would be misstated. Unlike other instruments in this paper, the past presence of immigrants is typically only used to instrument for the flow (or new stock) of immigrants today. However, it is found to be connected with a number of highly durable outcomes, potentially inducing serial correlation and raising concerns regarding its validity.

As an instrument, enclaves have been used to measure the effect of immigrants on a number of educational outcomes. Insofar as human capital is a durable good, and education has long-lasting effects on a labor market, we expect the instrument to invalidate itself. Immigrants appear to affect high-school dropout rates (Card and Lewis, 2007), the public/private school choice of natives and student/teacher ratio in public school (Farre, Ortega, and Tanaka, 2018), the choice to major in science and engineering (Orrenius and Zavodny, 2015), vocational enrollment and specialization (Røed and Schøne, 2016). Even cyclical patterns in immigrant presence at universities affects domestic student enrollment in college (Shih, 2017).

Using this same instrument, there is also evidence that change in immigrant populations affect the prices of non-traded, immigrant-intensive goods and services, such as housekeeping and gardening (Cortes, 2008). With lower prices for goods and services that are substitutes for household production, women in the top quartile of the wage distribution increase average hours of market work and more frequently work schedules in excess of 50 or 60 hours in response to an immigration shock (Cortés and Tessada, 2011).

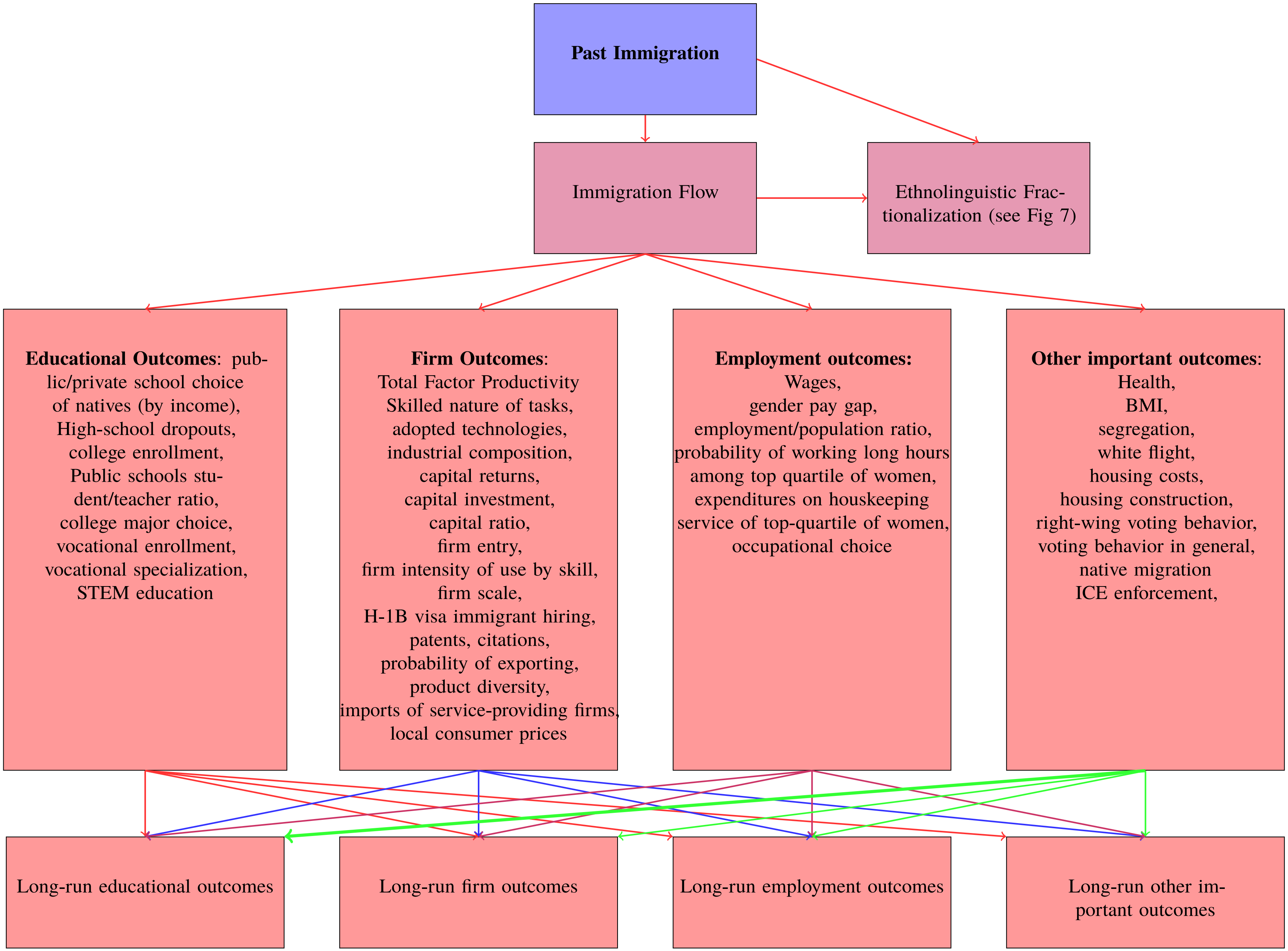

Figure 10: This figure summarizes selected research using immigrant enclaves as an instrument.

If immigrants affect mother's time at work, plausibly affecting both long-run career decsions (see, for instance, Adda, Dustmann, and Stevens (2017)), and long-run child outcomes (see, for instance, Havnes and Mogstad (2015)), then it is possible long-run effects of an immigration shock today are larger than the short-run effects, as affected agents are able to adjust behavior in the first case and affected children grow up and join the labor market in the second.

Another source of potentially long-run effects is through firms. Using the enclave instrument, immigrants are found to affect firm capital intensity, investment, and returns (Baum-Snow, Freedman, and Pavan, 2018; Lafortune, Lewis, and Tessada, 2019). It is also found to affect firm intensity of labor use by skill level, firm scale, and firm entry (Dustmann and Glitz, 2015). High-skilled immigration appears to affect patents and citations (Draca, Machin, and Witt, 2011; Peri, Shih, and Sparber, 2015; Bosetti, Cattaneo, and Verdolini, 2015; Ganguli, 2015). It also affects firm probability of exporting, export sales, and diversity of product (Parrotta, Pozzoli, and Sala, 2016; Ottaviano, Peri, and Wright, 2018). Because there is reason to think exporting causally increases firm productivity (see (Park et al., 2010)), this may yield another channel for persistent effects of immigration. Indeed, (Peri, 2012; Ottaviano, Peri, and Wright, 2018) also confirm productivity effects of immigration. Relatedly, immigration also appears to affect the technologies adopted by the firm, such as computers and automation (Lewis, 2011), and the skilled nature of tasks (Peri and Sparber, 2009; D'Amuri and Peri, 2014; Giuntella et al., 2018).

Seven additional channels might allow for persistence in immigration's effects, with each's relevance established through the enclave instrument. First, immigrants appear to affect both rental and housing prices (Saiz, 2007), suggesting another form of slow-depreciating capital to generate long-run effects. Second, growth in the immigrant share may affect political dynamics: Halla, Wagner, and Zweimüller (2017) find immigration causally affects right-wing vote share. Third, it may change the city's composition via out-migration, as is found in the short and long-run in (Morales, 2018). Fourth, larger migrant networks appear to increase both migrant entrepreneurial capital investment and profits (Woodruff and Zenteno, 2007). Sixth, locations with more unemployment see changes in local immigration enforcement (Makowsky and Stratmann, 2014), and immigration enforcement causally affects immigrant poverty rates (Amuedo-Dorantes, Arenas-Arroyo, and Sevilla, 2018). Finally, it's important to note the historical growth instrument ethnolinguistic fractionalization is strongly tied to immigrant share. While ethnolinguistic fractionalization is discussed in the text above, the idea that it causally affects corruption, the share of the labor force in small enterprises, immigrant language acquisition and the size of the black market is significant. Just as religion's inclusion in the estimates of MSA-level housing elasticities joins two potentially concerning instruments together, so too is immigration linked to ELF.

All of these effects are separate from the most studied effect of immigrants: on wages. As in Card (2001), a number of studies find negative effects on low skilled wages (and employment). Dustmann, Frattini, and Preston (2013) find immigration decreases the lower tails and increases the upper tails of the wage distribution. Peri, Shih, and Sparber (2015) find that STEM workers via the H-1B visa have a strong positive effect on college-educated natives and a weak positive effect on non-college educated natives, and Gould (2018) finds an expanding effect of immigrants on the 90/10 ratio in manufacturing wages. Dustmann and Glitz (2015) find large negative effects on wages in the non-traded sector, but no effect on the tradable sector in Germany. Similarly, Morales (2018) finds differential effects of the migrant share on women in the short- and long-run, depending on the skill group. Because heterogeneous effects have been estimated by skill, income, education, and gender, and age[25], concerns about local area treatment effects arise (Imbens and Angrist, 1994). Consistent with this concern, if one-third of the increase in wage inequality is due to city size and nonlinear agglomeration economies, the LATE effect becomes stronger (Baum-Snow and Pavan, 2013).

V Testing for Invalid Instruments

Much of our evidence given above is descriptive, and relies on sheer volume of causal connections established in the literature to raise concern about an instrument's use. But some uses may be valid, even if others are not. In this section, we develop a Hausman-like test for instrumental validity and apply it in the next section.

Consider a set of causal relationships between proposed instrument Z, proposed endogenous variables X_1 and X_2, and outcomes of interest Y_1 and Y_2. The relationship of both X_1 to Y_1 and X_2 to Y_2 are of interest to a researcher. However, confounders ξ_1 and ξ_2 confound the relationship between X_1 and Y_1, and X_2 and Y_2 respectively.[26]. Instrument Z is proposed in each paper because the means of $Z'\xi_1$ and $Z'\xi_2$ are both plausibly zero, solving a confounding problem. One possibility is that X_1 may cause X_2, or X_2 may cause X_1. For parsimony and tractability this simultaneity is captured via a correlated error term ω_x, which also allows for shared variation from Z, either directly or because of simulteneity[27]. Finally, ω_y

[25]Kerr, Kerr, and Lincoln (2015) find different effects of skilled immigrants on skilled worker employment by age.

[26]Rather than further introducing complex notation with a scalar weight multiplying the confounder, we equivalently rescale the variance of other error terms in our Monte Carlo simulation, so that the contribution of ξ in a regression varies in the same way as introducing a scalar would provide..

[27]For instance, if X_1 causes X_2, and X_2 causes X_1, then both X_1 and X_2 will contain variation from both ξ_1 and ξ_2, and will

also allows for joint variation between Y_1 and Y_2 not otherwise controlled for and presumed to be iid to other errors. Equation 1-4 describe the proposed system of equations below. The assumed covariance structure is given in 5. Figure 11 displays the causal relationships and covariance structure of this flexible framework graphically below.

$$Y_1 = \beta_{11} X_1 + \beta_{21} X_2 + \xi_1 + \omega_y + \epsilon_{y_1} \tag{1}$$

$$Y_2 = \beta_{12} X_1 + \beta_{22} X_2 + \xi_2 + \omega_y + \epsilon_{y_2} \tag{2}$$

$$X_1 = \gamma_1 Z + \xi_1 + \omega_x + \epsilon_{x_1} \tag{3}$$

$$X_2 = \gamma_2 Z + \xi_2 + \omega_x + \epsilon_{x_2} \tag{4}$$

$$\begin{bmatrix} Z \\ \omega_x \\ \xi_1 \\ \xi_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_z^2 & \sigma_{z,\omega} & 0 & 0 \\ \sigma_{z,\omega} & \sigma_\omega^2 & \sigma_{\omega,\xi_1} & \sigma_{\omega,\xi_2} \\ 0 & \sigma_{\omega,\xi_1} & \sigma_{\xi_1}^2 & 0 \\ 0 & \sigma_{\omega,\xi_2} & 0 & \sigma_{\xi_2}^2 \end{bmatrix} \right) \tag{5}$$

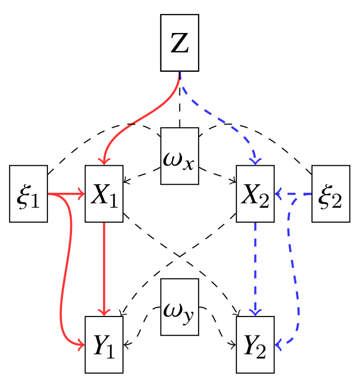

Figure 11: This figure graphically depicts the assumed causal structure given in Equations 1-5, and captures a broad set of possibilities, including those discussed in Figure 2. As a strong instrument, Z affects both X_1 and X_2. Both X_1 and X_2 are potentially correlated, either due to Z or due to simultaneity, captured by ω_x. Both X_1 and X_2 have a confounding issue ξ_1 and ξ_2 with outcomes of interest Y_1 and Y_2, which gave reason for an IV estimation to begin with. Dashed lines with no arrows denote the nonzero off-diagonal covariances of equation 5.

A researcher wishing to estimate β_{11} would ordinarily estimate β_{11} via OLS in Equation 1 which is by the researchers assumption biased, and via instrumental variables using Z to instrument for X_1 , which is

contain common variation from Z, captured by a correlation of Z with ω_x.

Table 2 Asymptotic Distributions of Four Estimators

Estimator	Notation	Asymptotic Mean	Asymptotic Variance
OLS, no controls	$\widehat{\beta_{11}}^{OLS,NC}$	$\beta_{11} + \beta_{21}\frac{Cov(X_1,X_2)}{Var(X_1)} + \frac{Cov(X_1,\xi_1)}{Var(X_1)}$	$\frac{\sigma^2_{Y\|X_1}}{Var(X_1)}$
OLS, controls	$\widehat{\beta_{11}}^{OLS,C}$	$\beta_{11} + \frac{Cov(X_1^*,\xi_1^*)}{Var(X_1^*)}$	$\frac{\sigma^2_{Y\|X_1,X_2}}{Var(X)}$
IV, no controls	$\widehat{\beta_{11}}^{IV,NC}$	$\beta_{11} + \beta_{21}\frac{Cov(Z,X_2)}{Cov(Z,X_1)}$	$\sigma^2_{Y\|X_1,Z}\frac{Var(Z)}{Cov(Z,X_1)^2}$
IV, controls	$\widehat{\beta_{11}}^{IV,C}$	$\beta_{11} + \frac{Cov(Z,\xi_1^*)}{Cov(Z^*,X_1^*)}$	$\sigma^2_{Y\|X_1,X_2,Z}\frac{Var(Z^*)}{Cov(Z^*,X_1^*)^2}$

Table 2: Table 2 displays the asymptotic means and variances of four estimators for β_{11} in the system of equations given by Equations 1-5 and displayed graphically in Figure 2. "OLS" denotes simple ordinary least squares regression, and the OLS moments of the asymptotic distributions are included for comparison purposes. "IV" denotes the use of Z as an instrument for X_1. "Controls" denotes inclusion of X_2 as an exogenous control. X_1^* and ξ_1^* denote the residual of X_1 and ξ_1 after being regressed on X_2.

by the researcher's assumption unbiased, (or less biased). We examine the asymptotic distributions of four possible estimators: OLS and IV of X_1 and Y_1 using Z as an instrument for X, either including X_2 or excluding X_2 as an exogenous control (denoted with a superscript C or NC respectively).[28]

Letting starred letters (for instance, X_1^*) denote the value of those covariates residual of a regression on X_2, the asymptotic distribution for each of the four estimators in terms of variables and primitives from Equations 1-5 are given in Table 2. One of the core insights in this paper is that when Paper 2 establishes the first stage of Z on X_2, they document $Cov(Z, X_2) \neq 0$, opening room for bias in the third row of Table 2: the stronger the first stage in Paper 2, the more concerned we should be for Paper 1's use of the instrument, all else equal.

Notice that $Cov(Z, \xi_1^*)$ introduces bias via controlling for X_2 when instrumenting X_1 with Z. This different potential source of bias is important for our test. Through what mechanism can $Cov(Z, \xi_1^*) \neq 0$ given that $Cov(Z, \xi_1) = 0$ and $Cov(\xi_1, \xi_2) = 0$? This is possible even in our simple example above because of the presence of ω_x. If there is any simultaneity or correlation between X_1 and X_2, so that the covariance of X_2 and ξ_1 is not zero, then controlling for X_2 re-introduces the confounder into X_1 in the first stage. That is, $Cov(X_2, \xi_1) = Cov(\omega_x, \xi_1)$ in our example, which is not assumed to be zero and indeed will not be zero in the presence of simulteneity between X_1 and X_2, even if $\beta_{21} = 0$. The potential biases in controlling for endogenous variables may vary significantly from the biases present when not controlling for the variable.

To better understand our test, one should compare the two potentially distinct sources of bias in using

[28]In what follows, we assume that the ordinary conditions for asymptotic covariances such as the Grenader conditions and relevance hold for both the data, instruments, and particularly the instruments and data after being conditioned on other data. In particular, we assume that $\underset{n\to\infty}{plim}\frac{Z'X_1^*}{n} = Q_{ZX^*}$ is a positive definite matrix, where $X_1^* = (I - X_2(X)2'X)2)^{-1}X)2')X_2$, i.e. X_1 multiplied by X_2'S annihilator matrix. This would not hold, for instance, if all of $X_2 = Z$.

instrumental variables with and without controls. For these two estimators to be equal asymptotically would require a significant “coincidence" of biases between the covariance of Z and X_2 and Z and the confounder of the first paper residual of X_2.

Test

To diagnose this issue, we propose a test similar to a Durbin-Wu-Hausman test (Hausman, 1978), testing whether or not the coefficients of two regressions are the same. The idea behind our test is intuitive. As seen in the fourth column of Table 2, the causes of asymptotic bias between IV with and without controls are generated by different covariances. The first is caused by common variation in X_2 and X_1's confounder, as is the case when our X's are simultaneous or caused by the same unobserved factor that also directly affects Y_1. The second is caused by an entirely different issue: X_2 is affecting Y_1 directly. We suggest that the issue of simulteneity between X_1 and X_2 and X_2 directly affecting Y_1 are relatively independent effects. Consequently it would be surprising if the two biases turned out to be the same asymptotically. However, when both biases are small, the two estimators will be the same. If one took two estimators whose asymptotic bias was drawn randomly from a distribution, then holding their standard errors constant, finding the two estimators are indistinguishable suggests the drawn bias term is small, because the two draws are drawn by two relatively independent sources, as suggested by Table 2. If the two estimators are statistically different, it suggests that at least one of the biases is large, and potentially both, suggesting caution when adopting the estimated coefficients.

Consequently, we propose running the standard instrumental variables regression and the same regression with other papers endogenous variables as exogenous controls. Conditional on those two estimators being statistically indistinguishable, we conclude the main specification is unlikely to have large bias.[29] While our estimator takes the form of a Hausman test, we cannot adopt the assumption that either of our estimators is efficient, which is needed to show that the covariance between estimators is zero in the Hausman test. Instead, we derive the covariance of estimators in our framework and find that it is “likely" to be nonzero, given reasonable concerns about small exogeneity violation and simultaneity.[30] However, because small-sample biases between two IV estimators are very likely correlated, we propose

[29] Oster (2019) argues "coefficient stability is informative only if authors also consider the importance of the controls in explaining the variance of the outcome." Our test, though framed differently, is consistent with the spirit of Oster's critique. If the R^2 doesn't increase much when X_2 is added to the specification, then our new coefficient isn't related to our outcome of interest. Moreover, if cov(X_1,X_2)=0, then there is no need to worry about bias from excluding X_2. This is reflected in the fact that the derived distribution of the test statistic depends critically on the covariance structure of the presented framework.

[30] We also find the estimated covariance between the two estimators is a significant source of variance in the difference between estimators in both our empirical examples.

estimating the covariance via bootstrapping.

The covariance of $\widehat{\beta}^{IV,NC}$ and $\widehat{\beta}^{IV,C}$ is, asymptotically[31]:

$$Cov(\widehat{\beta}^{IV,NC},\widehat{\beta}^{IV,C}) = \beta_{21}^2 Cov\left(\frac{(Z'Z)^{-1}Z'\omega_x}{(Z'Z)^{-1}Z'X_1},\frac{(Z'Z)^{-1}Z'\xi_1}{(Z'Z)^{-1}Z'X_1}\right) \tag{6}$$

Noting that we recover the Hausman result when $\beta_{21} = 0$, i.e. when one of our estimators is efficient.

With this covariance in hand, we can write the distribution of the test statistic:

$$\left(\widehat{\beta}^{IV,NC} - \widehat{\beta}^{IV,C}\right) \sim \mathcal{N}\left(\beta_{21}\frac{Cov(Z,X_2)}{Cov(Z,X_1)} - \frac{Cov(Z,\xi_1^*)}{Cov(Z^*,X_1^*)}, Var(\widehat{\beta}^{IV,NC}) + Var(\widehat{\beta}^{IV,C}) - 2Cov(\widehat{\beta}^{IV,NC},\widehat{\beta}^{IV,C})\right) \tag{7}$$

Where $Var(\widehat{\beta}^{IV,NC})$ and $Var(\widehat{\beta}^{IV,C})$ are given in Table 2 and the covariance is given in Equation 6. Equation 7 highlights that in order for the mean of the difference between the estimators to be zero, there would have to be a significant coincidence of biases between covariance of Z with X_2 on one hand, and Z and the residual variation of the first paper's confounder that is orthogonal to X_2 in the other. This, combined with Equation 6 also highlights our difference with a standard Hausman test–there is typically a significant positive covariance between the two estimators, which the Hausman test assumes is zero due to the assumption of efficiency.

For the reasons discussed in Hausman (1978) (pp. 1255), allowing for a nontrivial covariance between the two estimators leads to a situation in which power functions cannot easily be calculated in closed form. The strongest practical limitations to our test are likely to come from regressions with large standard errors relative to effect size. Young (2019) documents that, in top published IV work, IV standard errors are typically nearly five times larger than those of OLS, and partially as a consequence, the 95 percent confidence intervals of IV include the OLS point estimate roughly eighty percent of the time. Partially mitigating this concern, we provide Monte Carlo evidence that a larger number of potential IVs improves the power performance of our test, as the likelihood that a large deviation of an estimator from the truth is due to the issues of multiple IV rather than outliers increases.

One criticism of our test is that it does not provide a solution if an instrument fails. We liken a rejection of the null of two IV estimates being the same in our test to the failure of a robustness check in standard OLS: sometimes the instrument is not appropriate, and uncovering causal mechanisms is difficult. At

[31]In this case, small samples are likely to tend to have positive covariances because the sample errors are correlated. For instance, $Cov((Z'X_1)^{-1}Z'\epsilon_{y_1},(Z^{*'}X_1^*)^{-1}Z^{*'}\epsilon_{y_1})$ will be tend to be positive even when $E(Z'\epsilon_{y_1}) = 0$ purely due to noise. To see why, consider the case in which the correlation of $Z^{*'}$ and Z approaches one, so that the regression coefficients covariance tends toward $Var(Z)$ variance even as their mean tends toward zero.

minimum, a paper using an instrument that fails this test has a clear need to choose and defend a preferred specification, given that results are dependent on specification.

V.1 Monte Carlo Exercise

To understand this problem, we conduct Monte Carlo tests of the multiple-use instrument problem. Consider a single instrument z, that instruments n_x potentially causally-related variables x^j, where j can be considered to be a separate paper's use of an instrument and unique covariate. Each covariate x^j is determined by z, a random shock ϵ, a confounding shock η (which may be correlated between j's) and, after those inputs are realized (to avoid iterating to a fixed point), is affected by and affects other x's. We therefore have a pre-simultaneity $x^j_{pre,i}$ that is a function only of z, the confounder of x with y, and its own idiosyncratic variation, and a post-simultaneity x^j_i which is a function of $x^j_{pre,i}$ for all possible values of j:

$$x^j_{pre,i} = \gamma^j z_i + \eta^j_i + \epsilon^j_i \tag{8}$$

$$x^j_i = x^j_{pre,i} + \sum_{k \neq i} \delta^{k,j} x^k_{pre,i} \tag{9}$$

These x's potentially each affect both their "own" y^j but also those of other j's, and are confounded without instrumentation by some $x - y$ paper-specific confounder η^j:

$$y^j_i = \sum_{k=1}^{n_x} \beta^{k,j} x^{k,j}_i + \eta^j_i + \omega^j_i \tag{10}$$

To connect our simulated system to our framework, Equations 8-10 are depicted graphically in Figure 12. The idea behind one of our main exercises, increasing n_x, is suggested by the grayed-out portions of the figure. As we increase the number of potentially-related endogenous variables, we fill in the missing causal lines.

Our assumed coefficients are given in Table 3. Our choices of parameterization are driven by the assumption that (1) the main instrument is strong and likely to be significantly above a first-stage excluded instrument F-statistic of ten (2) the effect of Paper i's covariates of interest ($i \neq 1$), is significantly below the direct effect of X^1_1 on Y_1, which is always unity. In expectation the effect of other paper's covariates of interest is one-tenth the magnitude (3) the effect of the same Paper i's covariates on X is typically small–in expectation making up less than one twentieth the variance of X. We consider this parameterization to be one in which a researcher might reasonably think that IV would reduce bias because all other X's effects

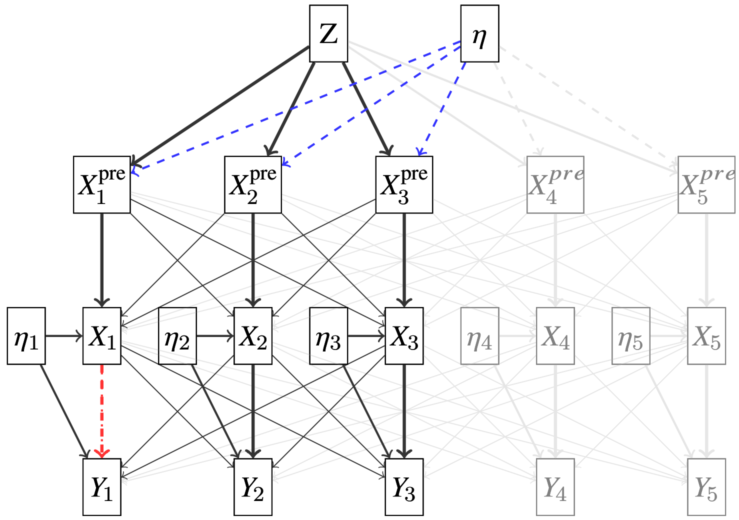

Figure 12: This figure displays the DAG illustrating our Monte Carlo exercises. We are interested in the estimated coefficient between X_1 and Y_1, the dotted red line (set to be unity in all exercises). Thick black lines, such as between Z and X^{Pre}'s, and X_i and the corresponding Y_i, are larger coefficients, in expectation. Thin lines, such as between X_1^{pre} and X_2, denote small average coefficients. When we change the correlation in our Monte Carlo exercises, we are adjusting the strength of η on each of X's relative to Z depicted as the dashed blue line, effectively setting the common component of Z in all X's. When we consider a new "potential IV" (increase n_x) we add an extra set of variables X_4, X_5, Y_4, and Y_5 (greyed out) to this DAG. The consequences of the existence of new potential IVs is the focus of this paper.

individually are small compared to bias, though collectively may not be.

We run three potential regressions for various levels of n_x: (1) a "standard" OLS regression of y^j on x^j, (2) the "standard" IV regression, which examines the effect of x^j instrumented by z on y^j, and (3) "augmented" IV regression, which uses the x^j's of other papers as exogenous controls in IV regression (2). We choose our main instrument to be normal-inverse-gamma distributed, so that within a Monte-Carlo iteration, the variance of the instrument is drawn from an inverse gamma distribution. Conditional on that variance, the instrument is distributed normally with mean zero. For convenience and interpretability, we denote this normal-inverse-gamma distribution as $\mathcal{N} - \Gamma^{-1}(\mu, mean(\sigma), stdev(\sigma))$ where μ denotes the mean of the normally distributed variable (conditional on σ), and $mean(\sigma)$ and $stdev(\sigma)$ denote the required α and β in a gamma distribution to generate an expected standard deviation σ $mean(\sigma)$ with standard deviation $stdev(\sigma)$.[32] Thus when z is denoted $\mathcal{N} - \Gamma^{-1}(0, 0.1, 0.1)$ it states that across simulations, z has a mean of zero, and an expected standard deviation of 0.1. However, the third parameter introduces (gamma-distributed) variance to the simulation's drawn standard deviation of z 0.01 (standard deviation

[32]This is equivalent to choosing $\alpha = \frac{(mean(\sigma))^2 + 2*(stdev(\sigma))^2}{(stdev(\sigma))^2}$ and $\beta = \frac{mean(\sigma)*((\mu^\sigma)^2 + (stdev(\sigma))^2)}{(stdev(\sigma))^2}$ in an Inverse-Gamma distribution.

0.1), so that the strength of our instrument z varies across Monte Carlo simulations, with sometimes low draws for the variance of z within a "paper" (simulation). The shocks for x's, y's and confounders (ϵ, ω, and η) are distributed i.i.d.

Table 3: Distribution of Monte Carlo Parameters

Description	Variable	Distribution
Main instrument strength	γ^j	$\mathcal{N}(3, 0.09)$
Other regression instrument sign	p^j	$2 * (\mathcal{B}(0.5) - 0.5)$
Effect of X's on one another	$\delta^{k,j}$	$\mathcal{N}(0, 0.01)$
Main effect of interest, x^1's on y^1	$\beta^{1,1}$	1
Effect of x^j's on own y^j	$\beta^{j,j}$, $j \neq 1$	$\mathcal{N}(1, 0.04)$
Effect of x^j's on other y^j	$\beta^{k,j}$, $k \neq j$	$\mathcal{N}(0.2, 0.09)$
Distribution of instrument	z	$\mathcal{N} - \Gamma^{-1}\,(0, 0.1, 0.1)$
Distribution of confounder	η^j	$\mathcal{N} - \Gamma^{-1}\,(0, 0.1, 0.1)$
Distribution of noise in x	ϵ^j	$\mathcal{N} - \Gamma^{-1}\,(0, 0.1, 0.1)$
Distribution of noise in y	ω^j	$\mathcal{N} - \Gamma^{-1}\,(0, 0.1, 0.1)$

Table 3: This tables displays the parameters governing the Monte Carlo distribution summarized in Equations 8-10 and graphically in Figure 12. The $\mathcal{N} - \Gamma^{-1}$ is used to model shocks that are mean zero with an expected standard deviation of 0.1 and a standard deviation of that standard deviation of 0.1 (distributed Gamma). $\mathcal{B}$ denotes the Bernoulli distribution, and is used to flip coefficient signs so that 50% contributed positively to some outcome, and 50% contribute negatively.

In what follows, we follow common practice in instrumental variables papers and include only estimators for which the F-statistic in the main IV regression is above ten (Stock and Yogo, 2005)[33], and the estimated coefficients are between -8 and 10 (the true coefficient is one), simulating an estimation selection process in which a researcher discards an IV if it generates nonsensical results.[34] One potential filter for estimators is to accept an estimate β^{IV} if and only if the "standard" estimate isn't statistically significantly different from the "augmented" results, via our Hausman-like test. We term the IV-estimated coefficient that does not differ from the augmented IV coefficient our "conditional" estimator, and find that it has good properties (low mean square error among accepted coefficients). Figure 13 depicts the mean square error of the four estimators for $\widehat{\beta^{1,1}}$ as a function of distinct potential IV paper uses, where the "true" value $\frac{\partial y^1}{\partial x^1}$ is equal to unity and the fourth estimator uses the selected standard IV point estimates described above, which removes approximately 10-20% of estimated coefficients in small samples, and

[33] We are aware of significant criticism of this cutoff. It is chosen to mimic a filter on datasets observed in the literature.

[34] Dropping unusual estimates is certainly done in empirical research. For instance, Cho and Rust (2017) report considering using interest-free loan installment offers as an instrument for interest rates on consumer demand, but rejected the instrument when it yielded an upward-sloping demand curve. Failure to do this immediately yields worse mean squared error for the IV estimator due to rare extreme estimates. For instance, while the lowest 1st percentile of IV regression coefficient is -2.5, the lowest 0.1th percentile is -17, and the 0.01th percentile is -190, with a kurtosis of over 190,000, rather than the 3 we would expect from a standard normally distributed variable

60-80% in large sample, as the power of the test increases. The left two figures depicts “small sample" properties of estimators, with each set of papers generated using 1,000 observations. The right two depict “large sample" properties, with 100,000 observations each. The top two allow for no correlation between X’s, so that controlling for endogenous X’s is the correct procedure. The bottom two allow for correlation between the X’s, so that any given pair of X’s have a correlation of 0.2 in expectation.

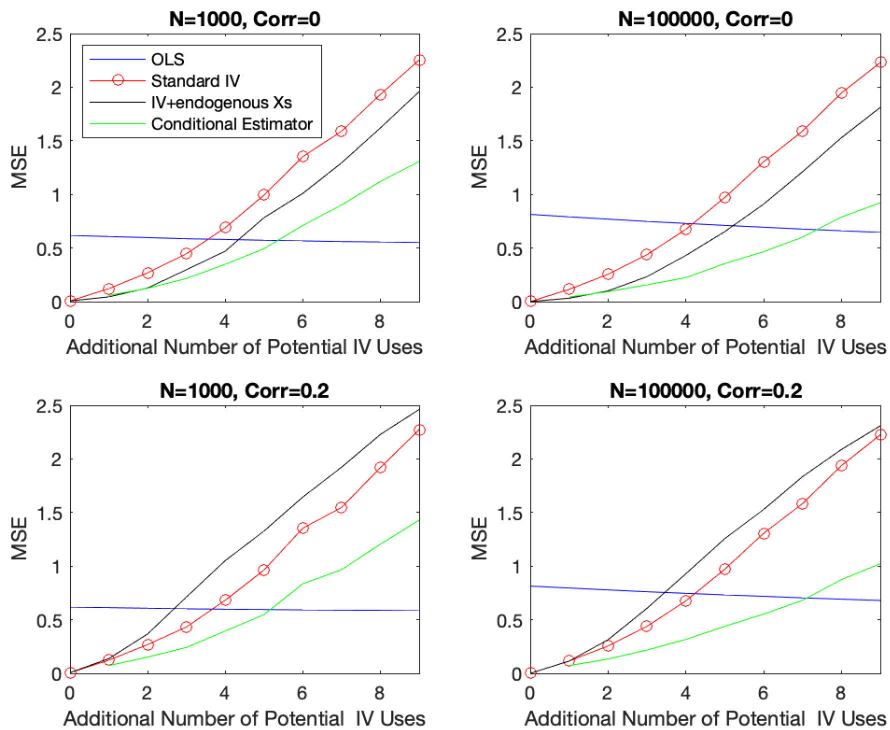

Figure 13: This figure depicts the mean-squared error of four estimators, generated from 31,000 simulations of Equations 8-10 and parameterized by Table 3. The first is an OLS regression of y^1 on x^1. The second is the “standard" IV estimator, which instruments for x^1 using z. The third augments the second by including other paper’s endogeneous variables as controls. The fourth adds other paper’s outcomes as controls. The last uses standard IV estimates if and only if either the second and the third, the second and the fourth, or all three coefficients are not statistically distinguishable at the 10% level and all F-statistics are above the (Stock and Yogo, 2005) threshold of 10.

For low-levels of potential instrumental variable uses, OLS is inferior to IV in both small- and large-sample. Importantly, however, the bias of simple OLS does not greatly differ depending on the number of IV uses, while our other estimators do. Controlling for endogenous X’s significantly improves the estimator (reduces MSE) when correlation between X’s is low, but is not a dependable method for improving the estimator under reasonable correlation (as in the case of simultaneity) between endogenous

X's. In general, our conditional estimator does a very good job, typically dominating all other considered regressions in large samples with large number of IV uses. To put the importance of potential IV uses in perspective, in our setup with many observations and little correlation, by the time the number of independent IV uses grows to about three, standard IV is worse than OLS, but the IV estimator with controls remains a better estimator than OLS until four potential uses are reached. Figure 14 illustrates the power of our test. Because in expectation all estimators are biased in our setup, we define power as the fraction of times our test correctly rejects IV estimation when it is farther from the true causal effect than OLS. For instance, with six potential IV uses and a sample size of 100,000, our filter catches between 78-83% of IV estimators whose point estimate is worse than OLS, depending on the correlation between X's. With small samples, IV in our calibration has reasonably large standard errors, so we detect fewer bad point estimates: between 20-23%.

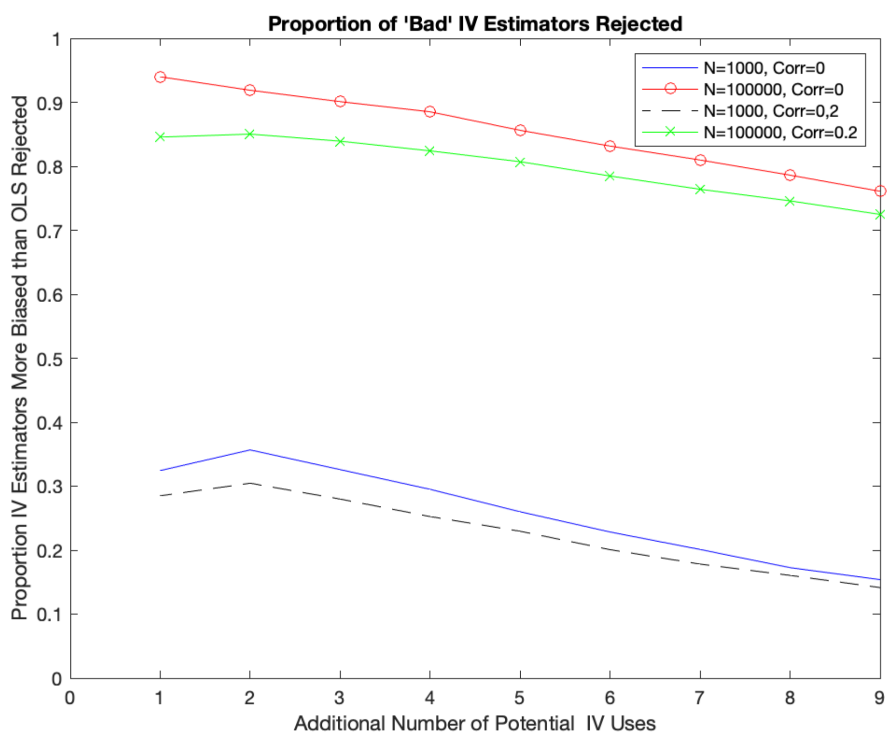

Figure 14: This figure depicts the fraction of "bad" estimates kept by our conditional estimator. We define a "bad" IV estimate as one in which the absolute distance of $\widehat{\beta^{IV,NC}}$ from the true β is greater than $\widehat{\beta^{OLS,NC}}$, so that OLS would be preferred.

We conclude by noting that simply controlling for other endogenous X's plays an important role in testing whether using a common instrument is valid. When the resulting IV coefficients do not differ significantly from the short regression, the estimator has a dramatically lower mean square error. Moreover, when a common IV estimator is present, the likelihood of generating wildly significant coefficients is greatly increased, particularly in small samples for reasons similar to those driving IV's significance in Young (2019).

VI Applications

We apply our proposed tests to two papers in different literatures. First, we examine Rupert and Zanella (2018), which establishes a strong relationship between grandchild birth and grandmother's labor supply. To get an exogenous source of variation for age at grandparenthood, the authors use whether or not a grandmother had a daughter or son first (sibship). The authors explicitly recognize the concerns of multiple-use instruments, focusing on the role of divorce. They argue that because firstborn girls may induce divorce, which increases women's labor supply, they may be finding lower bounds. Second, we examine Mian and Sufi (2014), who use the housing supply elasticities estimated by Saiz (2010) to produce an instrument for housing price changes over the business cycle, which in turn affects employment at the MSA level. Because Saiz (2010) uses elevation and bodies of water (and in the primary regression, religion) as instruments, there may be concern that housing supply elasticities (caused by elevation changes and bodies of water) may also affect segregation, road quality, presence of dams, size of city government, which may in turn affect an MSA's employment drop in 2008-2009.

VI.1 Rupert and Zanella 2018

To apply our Hausman-like procedure to Rupert and Zanella (2018), we gather several additional variables shown to be related to the instrument used to produce their main results concerning grandmother labor supply elasticities. We include (i) housing crowdedness, (ii) child education, (iii) child BMI, (iv) child mobility, (v) parent alcohol use, and (vi) parent tobacco use. For housing crowdedness, we use the person per room (PPR) measure. For child education, we use indicator values denoting the highest level of education achieved by a child. For child BMI, we construct the measure using the standard formula $\text{BMI} = \frac{kg}{m^2}$ and create indicator variables for whether or not a parent ever had an obese or underweight child. For child mobility, we construct an only-child indicator variable.[35] For parent alcohol use, we

[35]Rainer and Siedler (2009) show only children are less likely to move far away from home than individuals with siblings. The authors suggest this result is born out of a sense of duty to care for aging parents. Since individuals with siblings can possibly have a sibling living near their parents, this increases the likelihood they move elsewhere.

create an indicator variable that takes a value of one if an individual ever reports being a frequent drinker, defined as having a drink at least "several times a week." Similarly, tobacco use is controlled for with an indicator variable taking a value of one if an individual ever reports smoking more than nine cigarettes per day. The data used in Rupert and Zanella (2018) as well as the data for our additional controls are all publicly available from the Panel Survey of Income Dynamics (PSID). We use all available data from the PSID core sample to date, covering the years 1968-2017.[36]

Table 4 Rupert and Zanella (2018) Replication

Panel A			
	Log conditional hours		
	(1)	(2)	(3)
Grandparent	-0.366*	-0.661*	-0.549*
	(0.174)	(0.316)	(0.266)
Endogenous Controls	No	No	Yes
Conditional Data	No	Yes	Yes
N	56374	25316	25316
P-value of Hausman-like test			0.127

Standard errors in parentheses, * p<0.05, ** p<0.01

Panel B

Variance-Covariance Matrix			
	β_1	β_2	β_3
β_1	(.)	(.)	(.)
β_2	(.)	0.0707	0.0825
β_3	(.)		0.0997

Table 4: Panel A: Column (1) displays our replication for Rupert and Zanella (2018)'s instrumental variables estimates of the effect of grandchildren on the labor supply of grandmothers. Column (2) shows results using the same specification as (1), but conditioning on the availability of data for additional potentially endogenous covariates. That is, we restrict (2) to only include observations that have data for the additional covariates we intend to use to test the robustness of this IV but use the same specification as in (1). Column (3) shows the results using the conditional data after incorporating additional potentially endogenous controls into the specification. Table 4: Panel B: This panel shows the variance-covariance matrix obtained from our estimation of column (2) and column (3) in Panel A. Note, though the difference between β_2 and β_3 is small relative to each βs variance, the covariance is quite large relative to each βs variance. This effectively makes the standard error of the difference small, leading to a pvalue of 0.127.

In the first column in Panel A of Table 4, we report replicated results for females from Rupert and Zanella (2018)'s Table 7, column 10, which shows the effect of being a grandparent for seniors using whether or not a senior's firstborn child was female as an instrument.[37] In column (2) of Table 4, we run

[36]This differs slightly from Rupert and Zanella (2018) who use data from 1968-2015. Our point estimates are little changed.

[37]We would like to thank Peter Rupert and Giulio Zanella for sharing data and code with us that greatly assisted in the

the same regressions while conditioning on the data being available in both specifications (referred to as conditional data). While the results are quite different, it is worth noting that we lose a substantial portion of the sample due to missing data in our additional exogenous regressors. While column (2) depicts Rupert and Zanella's regressions on conditional data, column (3) adds our potentially endogenous controls. We use the sample covariance of individual moment conditions with and without controls to estimate covariance of estimated coefficients in a joint GMM estimation.

We focus our comparison on the results in columns (2) and (3) of Table 4 Panel A. While the difference between the regression coefficients is likely economically important, a t-test reveals this difference is marginally insignificant (p=.127): our test fails to formally reject Rupert and Zanella (2018). It is important to highlight the role the covariance plays in this result. The covariance is large relative to each βs variance. Thus, even though the difference between the two coefficients is small relative to each βs variance, the relatively large covariance makes the standard error of the difference small, resulting in a small (but marginally insignificant) pvalue. This result certainly does not invalidate Rupert and Zanella (2018), but it is still concerning. It suggests that child composition may (at least partially) threaten the specific results of the effect of grandchild birth on grandparent labor hours.

VI.2 Mian and Sufi 2014

To apply our Hausman-like test to Mian and Sufi (2014), we compare three of their specifications to those same specifications with potentially endogenous variables added as exogenous controls. The authors seek to understand how local housing price changes and the associated decline in household net worth may affect employment during the great recession. In one of their main instrumental variable specifications, the authors regress the change non-tradable employment, defined as either (i) restaurant and retail store employment or (ii) geographically concentrated industries, on the change in housing net worth at the county level. To get an exogenous change in housing prices, the authors instrument the change in housing net worth using the housing supply elasticity measured by Saiz (2010). We focus on Table 3 in their paper. The authors include two important sets of controls, both of which may include variables this paper suggests may be endogenous. The first is industry-level controls, (both housing supply elasticity and slopes/bodies of water may affect industrial composition) and the second is demographic controls such as fraction white, median household income, and the poverty level. We compare these main instrumental

replication.

variables specifications against the same specifications with included endogenous controls.[38]

As additional exogenous controls we include measures by MSA of (i) segregation (ii) density (iii) local government fractionalization (iv) dams (v) roads per capita (vi) broadband provision (vii) rail length and (viii) population growth. For segregation, we include the index of dissimilarity at the tract level within the county and its square. For density we include area and population in levels and logs. For local government fractionalization we proxy using number of independent school authorities divided by the MSA area. For dams, we use counts of the number of major dams in the MSA. To control for roads, we include a measure of miles of roads per person per area. For broadband, we use the average of broadband availability by zip code in the early 2000's. For population growth, we include log population growth from 1990 to 2000. We also include the potential endogenous controls for demographics Mian and Sufi acknowledge in their paper within our set of potentially endogenous variables.

Specifications (1), (3), and (5) in Table 5 replicate Mian and Sufi (2014)'s Table 3 specifications (5), (6), and (7). Column (1) examines rest and retail employment using instrumented housing net worth and using two-digit industries as controls. Column (3) replaces rest and retail employment with geographically concentrated employment. And specification (5) adds county-level demographic controls to the first. Adjacent specifications (2), (4), and (6) add in both our new potentially endogenous covariates as well as Mian and Sufi's demographic controls. While Mian and Sufi report both spatially-adjusted standard errors and clustered standard errors, they note that clustered standard errors are larger. To make our test conservative in terms of rejection, we take clustered standard errors as well. We calculate the covariance between regression coefficients of different regressions by estimating the two jointly and allowing arbitrary correlation within clusters across regressions. This means for instance that the observation-level error between specification (1) for San Bernardino County in California and the error in specification (2) (the same regression but with additional controls) for San Diego County in California can be arbitrarily correlated, because they are in the same cluster.

We apply our test between the three specifications of interest and display the results at the bottom of the first panel. The first two reject cross-specification coefficient equality at the 10% and the 5% level, respectively. The last fails to reject equality of coefficients. As might be expected, observation-level errors display a high positive covariance, driving us to estimate a high covariance between estimators, with similarly high covariances occurring if we instead bootstrapped the covariances. Recall that the variance

[38]Because Broomfield County, Colorado only became an independent county in 2001, but our segregation and population controls use census data before that, we drop Broomfield, resulting in only 539 observations, not 540 as Mian and Sufi have. Their results are little changed by the exclusion.

Table 5 Mian and Sufi (2014) Replication

Panel A						
	(1) Rest. & Retail	(2) Rest. & Retail	(4) Geog. Concen.	(5) Geog. Concen.	(3) Rest. & Retail	(6) Rest. & Retail
Δ Housing Net Worth	0.374**	0.610**	0.208*	0.466**	0.489**	0.610**
	(0.062)	(0.153)	(0.086)	(0.125)	(0.127)	(0.153)
Industry Controls	Yes	Yes	Yes	Yes	Yes	Yes
Possibly Endogenous Dem. Controls	No	Yes	No	Yes	Yes	Yes
Other Endogenous variable as controls	No	Yes	No	Yes	No	Yes
P-value of Hausman-like test		0.070		0.017		0.159
N	539	539	539	539	539	539

Standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01

Panel B						
	Variance-Covariance Matricies					
	$\beta^{IV,NC}$	$\beta^{IV,C}$	$\beta^{IV,NC}$	$\beta^{IV,C}$	$\beta^{IV,NC}$	$\beta^{IV,C}$
$\beta^{IV,NC}$	0.017	0.011	0.007	0.005	0.016	0.016
$\beta^{IV,C}$	(.)	0.023	(.)	0.016	(.)	0.023

Table 5: The top panel of Table 5 displays our replication for Mian and Sufi (2014)'s instrumental variables estimates of the effect of a change in housing net worth on employment at the MSA level. Column (1), (3), and (5) replicate Mian and Sufi's instrumental variables results from their Table 3, columns 5, 6, and 7. Adjacent columns (2), (4), and (5) add in possibly endogenous demographic controls (which are included in Mian and Sufi's Column (5) as well as other possibly endogenous controls outlined above. We report the p-value from our Hausman-like test of the difference between an instrumental variable estimator and the same estimator with potentially endogenous covariates added as controls at the bottom of the first panel. All standard errors are calculated by clustering at the state level. Because these p-values are surprisingly low given the standard errors of point estimates, the second panel reports the estimated variance-covariance matrix of the estimators: as with Rupert and Zanella, the covariance between estimates is high.

of the difference between estimators is the sum of the variance of each estimator minus the product of two and their covariance. For instance, comparing specifications (4) and (5), while the point estimates with and without our controls have variances of 0.007 and 0.016 respectively, their covariance is a comparably large 0.005. As with our Rupert and Zanella (2018) results, this example again highlights the importance of our deviation from the Durbin-Wu-Hausman test.

While our test highlights the need to be clear about the causal model behind their regression, our test does not and cannot conclude that Mian and Sufi (2014)'s regression is invalid. It could be instead that including the endogenous variables as controls renders invalid while their specification is valid. However, following our Monte Carlo tests, it raises the possibility that OLS may be less biased that IV in this case.

VII Conclusion

This paper has discussed some of the issues that arise with “popular" instruments and has discussed six categories of potentially problematic instruments. The use of these potentially troubling instruments is not rare: by our count, which puts only a lower bound on the problem, 317 papers used these instruments, and 86 “top five" papers. Nor has the use of these instruments declined over time.

Importantly, this paper does not condemn instrumental variables as typically practiced. Some of the examples in this paper are surely well-identified. Moreover, the vast majority of IV papers do not use these instruments, but use instruments that are idiosyncratic to their application, or that are less likely to cause concern. We have not focused on these papers.

To better understand when multiple unique uses of an instrument is valid, we propose a new test related to the Hausman test: running a “single paper" IV regression ignoring the other potentially endogenous covariates, and comparing the regression coefficient of interest to an IV regression that includes all those potentially endogenous variables as exogenous controls. We show that because the two potential sources of bias in these two regressions differ, statistical equality between coefficients suggests that either their biases are both small, or they are coincidentally similar. We differ from the Hausman test by estimating the covariance between coefficients of interest.

We then apply our test to two high-quality instrumental variables papers: Rupert and Zanella (2018), which uses firstborn girls as an instrument for age at which one becomes a grandparent, which affects labor supply, and Mian and Sufi 2014, which uses Saiz (2010)'s elevation and bodies of water derived elasticities as an instrument for housing price changes, which affect non-tradable employment. In the first case, our test formally fails to reject differences between estimated coefficients. In the second case, we find tentative reason for concern, depending on whether or not Mian and Sufi are correct to control for demographics.

We provide two clear positive messages going forward: first, more awareness should be paid to the notion that literatures, or sets of literatures, can “collectively invalidate" an instrument. Second, instrumental variables estimates that are robust to the inclusion of other endogenous controls are likely to be good estimators.

We also submit that a surprising number of papers have used the phrase “while this instrument has been used in [another paper], we are the first to use it in this context." While this is typically used to describe a contribution, it should also be a warning. Just because an instrument has “passed what might be called the American Economic Review (AER)-test," in Rodrik, Subramanian, and Trebbi (2004)'s colorful phrasing, does not mean it is a good instrument for a new paper.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5):1832–56.

Acemoglu, Daron, Camilo García-Jimeno, and James A Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *The American Economic Review* 105 (8):2364–2409.

Acemoglu, Daron and James A. Robinson. 2001. "A Theory of Political Transitions." *American Economic Review* 91 (4):938–963.

Adda, Jerome. 2016. "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data." *The Quarterly Journal of Economics* 131 (2):891–941.

Adda, Jerome, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy* 125 (2):293 – 337.

Addario, Sabrina Di and Eleonora Patacchini. 2008. "Wages and the City. Evidence from Italy." *Labour Economics* 15 (5):1040 – 1061.

Ades, Alberto F. and Edward L. Glaeser. 1995. "Trade and Circuses: Explaining Urban Giants." *The Quarterly Journal of Economics* 110 (1):195–227.

Adhvaryu, Achyuta and Anant Nyshadham. 2017. "Health, Enterprise, and Labor Complementarity in the Household." *Journal of Development Economics* 126:91 – 111.

Ager, Philipp and Markus Brückner. 2013. "Cultural diversity and economic growth: Evidence from the US during the age of mass migration." *European Economic Review* 64:76 – 97.

Akay, Alpaslan, Amelie Constant, and Corrado Giulietti. 2014. "The impact of immigration on the well-being of natives." *Journal of Economic Behavior and Organization* 103:72 – 92.

Alcala, Francisco and Antonio Ciccone. 2004. "Trade and Productivity." *The Quarterly Journal of Economics* 119 (2):613–646.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *The Quarterly Journal of Economics* 114 (4):1243–1284.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112 (2):348–396.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8 (2):155–194.

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport. 2016. "Birthplace diversity and economic prosperity." *Journal of Economic Growth* 21 (2):101–138.

Alesina, Alberto and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *The Quarterly Journal of Economics* 115 (3):847–904.

Alexeev, Michael and Robert Conrad. 2009. "The Elusive Curse of Oil." *The Review of Economics and Statistics* 91 (3):586–598.

Algan, Yann and Pierre Cahuc. 2009. "Civic Virtue and Labor Market Institutions." *American Economic Journal: Macroeconomics* 1 (1):111–145.

Ali, Fathi, Norbert Fiess, and Ronald Macdonald. 2011. "Climbing to the top? Foreign direct investment and property rights." *Economic Inquiry* 49 (1):289–302.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India." *The American Economic Review* 106 (3):587–624.

Altonji, Joseph and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives." In *Immigration, Trade, and the Labor Market*. National Bureau of Economic Research, Inc, 201–234.

Ampaabeng, Samuel K. and Chih Ming Tan. 2013. "The long-term cognitive consequences of early childhood malnutrition: The case of famine in Ghana." *Journal of Health Economics* 32 (6):1013 – 1027.

Amuedo-Dorantes, Catalina, Esther Arenas-Arroyo, and Almudena Sevilla. 2018. "Immigration enforcement and economic resources of children with likely unauthorized parents." *Journal of Public Economics* 158:63 – 78.

Ananat, Elizabeth O. and Guy Michaels. 2008. "The Effect of Marital Breakup on the Income Distribution of Women with Children." *The Journal of Human Resources* 43 (3):611–629.

Anderson, Siwan and Mukesh Eswaran. 2009. “What determines female autonomy? Evidence from Bangladesh.” *Journal of Development Economics* 90 (2):179 – 191.

Angrist, Joshua D. and William N. Evans. 1998. “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” *The American Economic Review* 88 (3):450–477.

Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens. 2000. “The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish.” *The Review of Economic Studies* 67 (3):499–527.

Antman, Francisca M. 2010. “Adult Child Migration and the Health of Elderly Parents Left Behind in Mexico.” *The American Economic Review Papers and Proceedings* 100 (2):205–208.

———. 2012. “Elderly Care and Intrafamily Resource Allocation when Children Migrate.” *The Journal of Human Resources* 47 (2):331–363.

Arellano, Manuel and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *Review of Economic Studies* 58 (2):277–297.

Arezki, Rabah and Markus Brückner. 2012a. “Rainfall, financial development, and remittances: Evidence from Sub-Saharan Africa.” *Journal of International Economics* 87 (2):377 – 385.

———. 2012b. “Rainfall, financial development, and remittances: Evidence from Sub-Saharan Africa.” *Journal of International Economics* 87 (2):377 – 385.

Arin, K. Peren, Viera Chmelarova, Eberhard Feess, and Ansgar Wohlschlegel. 2011. “Why are corrupt countries less successful in consolidating their budgets?” *Journal of Public Economics* 95 (7):521 – 530.

Åslund, Olof and Hans Grönqvist. 2010. “Family size and child outcomes: Is there really no trade-off?” *Labour Economics* 17 (1):130 – 139.

Aydemir, Abdurrahman B. and Murat G. Kirdar. 2017. “Quasi-experimental impact estimates of immigrant labor supply shocks: The role of treatment and comparison group matching and relative skill composition.” *European Economic Review* 98:282 – 315.

Baker, Scott R. 2018. “Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data.” *Journal of Political Economy* 126 (4):1504–1557.

Balan, David J. and Stephen Knack. 2012. "The correlation between human capital and morality and its effect on economic performance: Theory and evidence." *Journal of Comparative Economics* 40 (3):457 – 475.

Barone, Guglielmo, Alessio D'Ignazio, Guido de Blasio, and Paolo Naticchioni. 2016. "Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior." *Journal of Public Economics* 136:1 – 13.

Barone, Guglielmo and Sauro Mocetti. 2011. "With a little help from abroad: The effect of low-skilled immigration on the female labour supply." *Labour Economics* 18 (5):664 – 675.

Barone, Guglielmo and Gaia Narciso. 2015. "Organized crime and business subsidies: Where does the money go?" *Journal of Urban Economics* 86:98 – 110.

Barro, Robert J. and Rachel M. McCleary. 2005. "Which Countries Have State Religions?" *The Quarterly Journal of Economics* 120 (4):1331–1370.

Bartel, Ann P. 1989. "Where Do the New U.S. Immigrants Live?" *Journal of Labor Economics* 7 (4):371–391.

Barth, James R., Chen Lin, Ping Lin, and Frank M. Song. 2009. "Corruption in bank lending to firms: Cross-country micro evidence on the beneficial role of competition and information sharing." *Journal of Financial Economics* 91 (3):361–388.

Barth, James R., Chen Lin, Yue Ma, Jesús Seade, and Frank M. Song. 2013. "Do bank regulation, supervision and monitoring enhance or impede bank efficiency?" *Journal of Banking and Finance* 37 (8):2879 – 2892.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* No. wbsle in Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.

Baten, Joerg and Jan Luiten van Zanden. 2008. "Book production and the onset of modern economic growth." *Journal of Economic Growth* 13 (3):217–235.

Batista, Catia and Pedro C. Vicente. 2011. "Do Migrants Improve Governance at Home? Evidence from a Voting Experiment." *The World Bank Economic Review* 25 (1):77–104.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan. 2018. "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics* 10 (4):1–42.

Baum-Snow, Nathaniel and Ronni Pavan. 2013. "Inequality and City Size." *The Review of Economics and Statistics* 95 (5):1535–1548.

Bazzi, Samuel and Michael A. Clemens. 2013. "Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth." *American Economic Journal: Macroeconomics* 5 (2):152–186.

Beaudry, Paul, David A. Green, and Benjamin Sand. 2012. "Does Industrial Composition Matter for Wages? A Test of Search and Bargaining Theory." *Econometrica* 80 (3):1063–1104.

Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2003. "Law, endowments, and finance." *Journal of Financial Economics* 70 (2):137 – 181.

———. 2005. "SMEs, Growth, and Poverty: Cross-Country Evidence." *Journal of Economic Growth* 10 (3):199–229.

———. 2006. "Bank concentration, competition, and crises: First results." *Journal of Banking and Finance* 30 (5):1581 – 1603.

Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2010. "The trade-off between fertility and education: evidence from before the demographic transition." *Journal of Economic Growth* 15 (3):177–204.

Becker, Sascha O., Markus Nagler, and Ludger Woessmann. 2017. "Education and religious participation: city-level evidence from Germany's secularization period 1890–1930." *Journal of Economic Growth* 22 (3):273–311.

Becker, Sascha O. and Ludger Woessmann. 2008. "Luther and the Girls: Religious Denomination and the Female Education Gap in Nineteenth-Century Prussia." *The Scandinavian Journal of Economics* 110 (4):777–805.

———. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *The Quarterly Journal of Economics* 124 (2):531–596.

———. 2018. “Social Cohesion, Religious Beliefs, and the Effect of Protestantism on Suicide.” *The Review of Economics and Statistics* 100 (3):377–391.

Bedard, Kelly and Olivier Deschenes. 2005. “Sex Preferences, Marital Dissolution, and the Economic Status of Women.” *Journal of Human Resources* 40 (2).

Bell, Brian, Francesco Fasani, and Stephen Machin. 2013. “Crime and Immigration: Evidence from large immigrant waves.” *The Review of Economics and Statistics* 95 (4):1278–1290.

Bennedsen, Morten, Hans Christian Kongsted, and Kasper Meisner Nielsen. 2008. “The causal effect of board size in the performance of small and medium-sized firms.” *Journal of Banking and Finance* 32 (6):1098 – 1109.

Bennedsen, Morten, Kasper Meisner Nielsen, Francisco Pérez-González, and Daniel Wolfenzon. 2007. “Inside the Family Firm: The Role of Families in Succession Decisions and Performance.” *The Quarterly Journal of Economics* 122 (2):647–691.

Bennett, Daniel. 2012. “Does Clean Water Make You Dirty? Water Supply and Sanitation in the Philippines.” *The Journal of Human Resources* 47 (1):146–173.

Berger, David and Joseph Vavra. 2015. “Consumption Dynamics During Recessions.” *Econometrica* 83 (1):101–154.

Bernheim, B. Douglas, Andrei Shleifer, and Lawrence H. Summers. 1985. “The Strategic Bequest Motive.” *Journal of Political Economy* 93 (6):1045–1076.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile Prices in Market Equilibrium.” *Econometrica* 63 (4):841–890.

Bianchi, Milo, Paolo Buonanno, and Paolo Pinotti. 2012. “Do immigrants cause crime?” *Journal of the European Economic Association* 10 (6):1318–1347.

Biavaschi, Costanza. 2013. “The labor demand was downward sloping: Disentangling migrants inflows and outflows, 1929-1957.” *Economics Letters* 118 (3):531 – 534.

Bjerk, David. 2010. “Thieves, thugs, and neighborhood poverty.” *Journal of Urban Economics* 68 (3):231 – 246.

Bjørnskov, Christian. 2012. "How Does Social Trust Affect Economic Growth?" *Southern Economic Journal* 78 (4):1346–1368.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2010. "Small Family, Smart Family? Family Size and the IQ Scores of Young Men." *Journal of Human Resources* 45 (1).

Bleakley, Hoyt and Aimee Chin. 2004. "Language Skills and Earnings: Evidence from Childhood Immigrants." *The Review of Economics and Statistics* 86 (2):481–496.

Blundell, Richard and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87 (1):115–143.

Böhme, Marcus H, Ruth Persian, and Tobias Stöhr. 2015. "Alone but better off? Adult child migration and health of elderly parents in Moldova." *Journal of Health Economics* 39:211 – 227.

Bolin, K., B. Lindgren, and P. Lundborg. 2008. "Your next of kin or your own career?: Caring and working among the 50+ of Europe." *Journal of Health Economics* 27 (3):718 – 738.

Bond, Timothy N. and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results." *The Review of Economics and Statistics* 95 (5):1468–1479.

Bond, Timothy N and Kevin Lang. 2019. "The Sad Truth About Happiness Scales." *Journal of Political Economy* 127 (4).

Bosetti, Valentina, Cristina Cattaneo, and Elena Verdolini. 2015. "Migration of skilled workers and innovation: A European Perspective." *Journal of International Economics* 96 (2):311 – 322.

Boustan, Leah Platt. 2010. "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration." *The Quarterly Journal of Economics* 125 (1):417–443.

Boustan, Leah Platt, Price V. Fishback, and Shawn Kantor. 2010. "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression." *Journal of Labor Economics* 28 (4):719–746.

Bove, Vincenzo and Leandro Elia. 2017. "Migration, Diversity, and Economic Growth." *World Development* 89:227 – 239.

Brainerd, Elizabeth and Nidhiya Menon. 2014. “Seasonal effects of water quality: The hidden costs of the Green Revolution to infant and child health in India.” *Journal of Development Economics* 107 (C):49–64.

Breuer, Janice Boucher and John McDermott. 2013. “Respect, responsibility, and development.” *Journal of Development Economics* 105:36 – 47.

Bronars, Stephen and Jeffrey Grogger. 1994. “The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment.” *American Economic Review* 84 (5):1141–56.

Brückner, Markus. 2010. “Population Size and Civil Conflict Risk: Is there a Causal Link?” *The Economic Journal* 120 (544):535–550.

———. 2012a. “Economic growth, size of the agricultural sector, and urbanization in Africa.” *Journal of Urban Economics* 71 (1):26 – 36.

———. 2012b. “An instrumental variables approach to estimating tax revenue elasticities: Evidence from Sub-Saharan Africa.” *Journal of Development Economics* 98 (2):220 – 227.

Brückner, Markus and Antonio Ciccone. 2011. “Rain and the Democratic Window of Opportunity.” *Econometrica* 79 (3):923–947.

Brückner, Markus and Mark Gradstein. 2013a. “Effects of transitory shocks to aggregate output on consumption in poor countries.” *Journal of International Economics* 91 (2):343 – 357.

———. 2013b. “Exogenous volatility and the size of government in developing countries.” *Journal of Development Economics* 105:254 – 266.

Brunnschweiler, Christa N. 2008. “Cursing the Blessings? Natural Resource Abundance, Institutions, and Economic Growth.” *World Development* 36 (3):399 – 419.

Brückner, Markus. 2012. “An instrumental variables approach to estimating tax revenue elasticities: Evidence from Sub-Saharan Africa.” *Journal of Development Economics* 98 (2):220 – 227.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. “Causes of Sprawl: A Portrait from Space.” *The Quarterly Journal of Economics* 121 (2):587–633.

Burke, Paul J. and Andrew Leigh. 2010. “Do Output Contractions Trigger Democratic Change?” *American Economic Journal: Macroeconomics* 2 (4):124–157.

Butcher, Kristin F. and Anne Case. 1994. "The Effect of Sibling Sex Composition on Women's Education and Earnings." *The Quarterly Journal of Economics* 109 (3):531–563.

Cáceres-Delpiano, Julio and Marianne Simonsen. 2012. "The toll of fertility on mothers' wellbeing." *Journal of Health Economics* 31 (5):752 – 766.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19 (1):22–64.

———. 2009. "Immigration and Inequality." *American Economic Review* 99 (2):1–21.

Card, David and John DiNardo. 2000. "Do Immigrant Inflows Lead to Native Outflows?" *American Economic Review* 90 (2):360–367.

Card, David and Ethan G. Lewis. 2007. "The Diffusion of Mexican Immigrants During the 1990s: Explanations and Impacts." In *Mexican Immigration to the United States*, NBER Chapters. National Bureau of Economic Research, Inc, 193–228.

Carranza, Eliana. 2014. "Soil Endowments, Female Labor Force Participation, and the Demographic Deficit of Women in India." *American Economic Journal: Applied Economics* 6 (4):197–225.

Casabonne, Ursula and Charles Kenny. 2012. "The Best Things in Life are (Nearly) Free: Technology, Knowledge, and Global Health." *World Development* 40 (1):21 – 35.

Cashin, Paul, Kamiar Mohaddes, and Mehdi Raissi. 2017. "Fair weather or foul? The macroeconomic effects of El Niño." *Journal of International Economics* 106:37 – 54.

Castells-Quintana, David. 2017. "Malthus living in a slum: Urban concentration, infrastructure and economic growth." *Journal of Urban Economics* 98:158 – 173. Urbanization in Developing Countries: Past and Present.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics* 95 (5):1549–1561.

Cawley, John. 2004. "The Impact of Obesity on Wages." *The Journal of Human Resources* 39 (2):451–474.

Cawley, John and Chad Meyerhoefer. 2012. "The medical care costs of obesity: An instrumental variables approach." *Journal of Health Economics* 31 (1):219 – 230.

Cesur, Resul and Naci Mocan. 2018. "Education, religion, and voter preference in a Muslim country." *Journal of Population Economics* 31 (1):1–44.

Chadi, Adrian and Matthias Krapf. 2017. "The protestant fiscal ethic: Religious confession and euro skepticism in Germany." *Economic Inquiry* 55 (4):1813–1832.

Chaijaroen, Pasita. 2019. "Long-lasting income shocks and adaptations: Evidence from coral bleaching in Indonesia." *Journal of Development Economics* 136:119 – 136.

Chaney, Eric. 2013. "Revolt on the Nile: Economic shocks, religion, and political power." *Econometrica* 81 (5):2033–2053.

Chaney, Thomas, David Sraer, and David Thesmar. 2012. "The Collateral Channel: How Real Estate Shocks Affect Corporate Investment." *American Economic Review* 102 (6):2381–2409.

Charles, Kerwin Kofi and Purvi Sevak. 2005. "Can family caregiving substitute for nursing home care?" *Journal of Health Economics* 24 (6):1174 – 1190.

Chen, Natalie and Luciana Juvenal. 2016. "Quality, trade, and exchange rate pass-through." *Journal of International Economics* 100:61 – 80.

Chetty, Raj, László Sándor, and Adam Szeidl. 2017. "The Effect of Housing on Portfolio Choice." *The Journal of Finance* 72 (3):1171–1212.

Chhaochharia, Vidhi, Dasol Kim, George M. Korniotis, and Alok Kumar. 2018. "Mood, firm behavior, and aggregate economic outcomes." *Journal of Financial Economics* .

Chiswick, Barry R. and Paul W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13 (2):246–288.

Cho, SungJin and John Rust. 2017. "Precommitments for Financial Self-Control? Micro Evidence from the 2003 Korean Credit Crisis." *Journal of Political Economy* 125 (5):1413–1464.

Churchill, Sefa Awaworyi and Russell Smyth. 2017. "Ethnic Diversity and Poverty." *World Development* 95:285 – 302.

Coe, Norma B. and Courtney Harold Van Houtven. 2009. "Caring for mom and neglecting yourself? The health effects of caring for an elderly parent." *Health Economics* 18 (9):991–1010.

Cole, Matthew A., Robert J.R. Elliott, Giovanni Occhiali, and Eric Strobl. 2018. "Power outages and firm performance in Sub-Saharan Africa." *Journal of Development Economics* 134:150 – 159.

Combes, Pierre-Philippe, Bruno Decreuse, Morgane Laouénan, and Alain Trannoy. 2016. "Customer Discrimination and Employment Outcomes: Theory and Evidence from the French Labor Market." *Journal of Labor Economics* 34 (1):107–160.

Conley, Dalton and Rebecca Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *The Journal of Human Resources* 41 (4):722–737.

Cortes, Patricia. 2008. "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data." *Journal of Political Economy* 116 (3):381–422.

Cortés, Patricia and Jessica Pan. 2014. "Foreign nurse importation and the supply of native nurses." *Journal of Health Economics* 37 (C):164–180.

———. 2019. "When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap." *Journal of Labor Economics* 37 (2):351–398.

Cortés, Patricia and Jose Tessada. 2011. "Low-Skilled Immigration and the Labor Supply of Highly Skilled Women." *American Economic Journal: Applied Economics* 3 (3):88–123.

Cruces, Guillermo and Sebastian Galiani. 2007. "Fertility and female labor supply in Latin America: New causal evidence." *Labour Economics* 14 (3):565 – 573.

Currie, Janet, Matthew Neidell, and Johannes F. Schmieder. 2009. "Air pollution and infant health: Lessons from New Jersey." *Journal of Health Economics* 28 (3):688 – 703.

Currie, Janet and Aaron Yelowitz. 2000. "Are public housing projects good for kids?" *Journal of Public Economics* 75 (1):99 – 124.

Cutler, David M. and Edward L. Glaeser. 1997. "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics* 112 (3):827–872.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation." *The Review of Economics and Statistics* 90 (3):478–497.

Dahl, Gordon B. and Enrico Moretti. 2008. “The Demand for Sons.” *The Review of Economic Studies* 75 (4):1085–1120.

D‘Amuri, Francesco, Gianmarco IP Ottaviano, and Giovanni Peri. 2010. “The labor market impact of immigration in Western Germany in the 1990s.” *European Economic Review* 54 (4):550–570.

D’Amuri, Francesco and Giovanni Peri. 2014. “Immigration, jobs, and employment protection: evidence from Europe before and during the Great Recession.” *Journal of the European Economic Association* 12 (2):432–464.

Davey Smith, George, Jonathan AC Sterne, Abigail Fraser, Per Tynelius, Debbie A Lawlor, and Finn Rasmussen. 2009. “The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study.” *BMJ* 339.

Davis, James C. and J.Vernon Henderson. 2003. “Evidence on the political economy of the urbanization process.” *Journal of Urban Economics* 53 (1):98 – 125.

Davis, Joseph H., Christopher Hanes, and Paul W. Rhode. 2009. “Harvests and Business Cycles in Nineteenth-Century America.” *The Quarterly Journal of Economics* 124 (4):1675–1727.

Delgado, Michael S., Nadine McCloud, and Subal C. Kumbhakar. 2014. “A generalized empirical model of corruption, foreign direct investment, and growth.” *Journal of Macroeconomics* 42:298 – 316.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2014. “What Do We Learn from the Weather? The New Climate-Economy Literature.” *Journal of Economic Literature* 52 (3):740–798.

Dercon, Stefan. 2004. “Growth and shocks: evidence from rural Ethiopia.” *Journal of Development Economics* 74 (2):309 – 329.

Dettling, Lisa J. and Melissa S. Kearney. 2014. “House prices and birth rates: The impact of the real estate market on the decision to have a baby.” *Journal of Public Economics* 110:82 – 100.

Di Falco, Salvatore, Peter Berck, Mintewab Bezabih, and Gunnar Köhlin. 2019. “Rain and impatience: Evidence from rural Ethiopia.” *Journal of Economic Behavior and Organization* 160:40 – 51.

Dittmar, Jeremiah E. 2011. “Information technology and economic change: The impact of the printing press.” *The Quarterly Journal of Economics* 126 (3):1133–1172.

Draca, Mirko, Stephen Machin, and Robert Witt. 2011. "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks." *The American Economic Review* 101 (5):2157–2181.

Duflo, Esther and Rohini Pande. 2007. "Dams." *The Quarterly Journal of Economics* 122 (2):601–646.

Duranton, Gilles and Matthew A. Turner. 2018. "Urban form and driving: Evidence from US cities." *Journal of Urban Economics* 108:170 – 191.

Dustmann, Christian, Francesca Fabbri, and Ian Preston. 2005. "The Impact of Immigration on the British Labour Market." *The Economic Journal* 115 (507):F324–F341.

———. 2011. "Racial Harassment, Ethnic Concentration, and Economic Conditions." *The Scandinavian Journal of Economics* 113 (3):689–711.

Dustmann, Christian, Tommaso Frattini, and Ian P. Preston. 2013. "The Effect of Immigration along the Distribution of Wages." *The Review of Economic Studies* 80 (1):145–173.

Dustmann, Christian and Albrecht Glitz. 2015. "How Do Industries and Firms Respond to Changes in Local Labor Supply?" *Journal of Labor Economics* 33 (3):711–750.

Easterly, William and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics* 112 (4):1203–1250.

Ebenstein, Avraham. 2012. "The Consequences of Industrialization: Evidence from Water Pollution and Digestive Cancers in China." *The Review of Economics and Statistics* 94 (1):186–201.

Emran, M. Shahe and Zhaoyang Hou. 2013. "Access to markets and rural poverty: Evidence from household consumption in China." *The Review of Economics and Statistics* 95 (2):682–697.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller. 2017. "Culture, Work Attitudes, and Job Search: Evidence from the Swiss Language Border." *Journal of the European Economic Association* 15 (5):1056–1100.

Evans, William N. and Robert M. Schwab. 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *The Quarterly Journal of Economics* 110 (4):941–974.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *The Review of Economic Studies* 81 (3 (288)):1046–1070.

Faberman, R. Jason and Matthew Freedman. 2016. “The urban density premium across establishments.” *Journal of Urban Economics* 93:71 – 84.

Fafchamps, Marcel. 1993. “Sequential Labor Decisions Under Uncertainty: An Estimable Household Model of West-African Farmers.” *Econometrica* 61 (5):1173–1197.

Fafchamps, Marcel, Christopher Udry, and Katherine Czukas. 1998. “Drought and saving in West Africa: are livestock a buffer stock?” *Journal of Development Economics* 55 (2):273 – 305.

Fajnzylber, Pablo, Daniel Lederman, and Norman Loayza. 2002. “Inequality and Violent Crime.” *The Journal of Law and Economics* 45 (1):1–39.

Fan, Ying, Kai-Uwe Kühn, and Francine Lafontaine. 2017. “Financial Constraints and Moral Hazard: The Case of Franchising.” *Journal of Political Economy* 125 (6):2082–2125.

Farré, Lídia, Libertad González, and Francesc Ortega. 2011. “Immigration, Family Responsibilities and the Labor Supply of Skilled Native Women.” *The B.E. Journal of Economic Analysis & Policy* 11 (1):1–48.

Farre, Lidia, Francesc Ortega, and Ryuichi Tanaka. 2018. “Immigration and the public-private school choice.” *Labour Economics* 51:184 – 201.

Favara, Giovanni and Jean Imbs. 2015. “Credit Supply and the Price of Housing.” *American Economic Review* 105 (3):958–92.

Feigenberg, Benjamin, Erica Field, and Rohini Pande. 2013. “The Economic Returns to Social Interaction: Experimental Evidence from Microfinance.” *The Review of Economic Studies* 80 (4):1459–1483.

Felbermayr, Gabriel and Jasmin Gröschl. 2013. “Natural disasters and the effect of trade on income: A new panel IV approach.” *European Economic Review* 58:18 – 30.

Felkner, John S. and Robert M. Townsend. 2011. “The Geographic Concentration of Enterprise in Developing Countries.” *The Quarterly Journal of Economics* 126 (4):2005–2061.

Fenske, James. 2015. “African polygamy: Past and present.” *Journal of Development Economics* 117:58 – 73.

Fisher-Vanden, Karen, Erin T. Mansur, and Qiong (Juliana) Wang. 2015. “Electricity shortages and firm productivity: Evidence from China’s industrial firms.” *Journal of Development Economics* 114:172 – 188.

Fletcher, Jason M. and Jinho Kim. 2019. “The effect of sibship size on non-cognitive Skills: Evidence from natural experiments.” *Labour Economics* 56:36 – 43.

Forlani, Emanuele, Elisabetta Lodigiani, and Concetta Mendolicchio. 2015. “Impact of Low-Skilled Immigration on Female Labour Supply.” *The Scandinavian Journal of Economics* 117 (2):452–492.

Frankel, Jeffrey A. and David H. Romer. 1999. “Does Trade Cause Growth?” *American Economic Review* 89 (3):379–399.

Fredriksson, Per G. and Jim R. Wollscheid. 2010. “Party Discipline and Environmental Policy: The Role of "Smoke-filled Back Rooms".” *The Scandinavian Journal of Economics* 112 (3):489–513.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton. 2000. “Dodging the grabbing hand: the determinants of unofficial activity in 69 countries.” *Journal of Public Economics* 76 (3):459 – 493.

Furtado, Delia and Heinrich Hock. 2010. “Low Skilled Immigration and Work-Fertility Tradeoffs among High Skilled US Natives.” *American Economic Review* 100 (2):224–28.

Ganguli, Ina. 2015. “Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?” *Journal of Labor Economics* 33 (S1):S257–S288.

Ganong, Peter and Daniel Shoag. 2017. “Why has regional income convergence in the U.S. declined?” *Journal of Urban Economics* 102:76 – 90.

Gergaud, Olivier, Andrew J. Plantinga, and Aurelie Ringeval-Deluze. 2017. “Anchored in the past: Persistent price effects of obsolete vineyard ratings in France.” *Journal of Economic Behavior and Organization* 133:39 – 51.

Gilchrist, Duncan Sheppard and Emily Glassberg Sands. 2016a. “Something to Talk About: Social Spillovers in Movie Consumption.” *Journal of Political Economy* 124 (5):1339–1382.

———. 2016b. “Something to Talk About: Social Spillovers in Movie Consumption.” *Journal of Political Economy* 124 (5):1339–1382.

Giles, John. 2006. "Is life more risky in the open? Household risk-coping and the opening of China's labor markets." *Journal of Development Economics* 81 (1):25 – 60.

Giles, John and Kyeongwon Yoo. 2007. "Precautionary Behavior, Migrant Networks, and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China." *The Review of Economics and Statistics* 89 (3):534–551.

Giordani, Paolo E., Nadia Rocha, and Michele Ruta. 2016. "Food prices and the multiplier effect of trade policy." *Journal of International Economics* 101:102 – 122.

Giulietti, Corrado, Jackline Wahba, and Yves Zenou. 2018. "Strong versus weak ties in migration." *European Economic Review* 104:111 – 137.

Giuntella, Osea and Fabrizio Mazzonna. 2015. "Do immigrants improve the health of natives?" *Journal of Health Economics* 43:140 – 153.

Giuntella, Osea, Fabrizio Mazzonna, Catia Nicodemo, and Carlos Vargas-Silva. 2018. "Immigration and the reallocation of work health risks." *Journal of Population Economics* .

Glaeser, Edward L., Joseph Gyourko, and Albert Saiz. 2008. "Housing supply and housing bubbles." *Journal of Urban Economics* 64 (2):198 – 217.

Glewwe, Paul. 1991. "Investigating the determinants of household welfare in Côte d'Ivoire." *Journal of Development Economics* 35 (2):307 – 337.

Glewwe, Paul, Hanan G Jacoby, and Elizabeth M King. 2001. "Early childhood nutrition and academic achievement: a longitudinal analysis." *Journal of Public Economics* 81 (3):345 – 368.

González, Libertad and Francesc Ortega. 2011. "How do very open economies adjust to large immigration flows? Evidence from Spanish regions." *Labour Economics* 18 (1):57 – 70.

Gonzalez-Navarro, Marco and Matthew A. Turner. 2018. "Subways and urban growth: Evidence from earth." *Journal of Urban Economics* 108:85 – 106.

Gould, Eric D. 2018. "Explaining the Unexplained: Residual Wage Inequality, Manufacturing Decline and Low-skilled Immigration." *The Economic Journal* 129 (619):1281–1326.

Goux, Dominique and Eric Maurin. 2005. "The effect of overcrowded housing on children's performance at school." *Journal of Public Economics* 89 (5):797 – 819.

Gruber, Jonathan. 2005. "Religious Market Structure, Religious Participation, and Outcomes: Is Religion Good for You?" *The B.E. Journal of Economic Analysis and Policy* 5 (1).

Gruber, Jonathan and Daniel M. Hungerman. 2008. "The Church Versus the Mall: What Happens When Religion Faces Increased Secular Competition?" *The Quarterly Journal of Economics* 123 (2):831–862.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2003. "People's opium? Religion and economic attitudes." *Journal of Monetary Economics* 50 (1):225–282.

———. 2006. "Does Culture Affect Economic Outcomes?" *The Journal of Economic Perspectives* 20 (2):23–48.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Economic Journal: Economic Policy* 5 (4):167–99.

Hakkala, Katariina Nilsson, Pehr-Johan Norbäck, and Helena Svaleryd. 2008. "Asymmetric Effects of Corruption on FDI: Evidence from Swedish Multinational Firms." *The Review of Economics and Statistics* 90 (4):627–642.

Hall, Robert E. and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *The Quarterly Journal of Economics* 114 (1):83–116.

Halla, Martin, Alexander F Wagner, and Josef Zweimüller. 2017. "Immigration and Voting for the Far Right." *Journal of the European Economic Association* 15 (6):1341–1385.

Hansen, Casper Worm. 2013. "The diffusion of health technologies: Cultural and biological divergence." *European Economic Review* 64:21 – 34.

Harari, Mariaflavia and Eliana La Ferrara. 2018. "Conflict, Climate, and Cells: A Disaggregated Analysis." *The Review of Economics and Statistics* 100 (4):594–608.

Hatfield, John William and Katrina Kosec. 2013. "Federal competition and economic growth." *Journal of Public Economics* 97:144 – 159.

———. 2019. "Local Environmental Quality and Interjurisdictional Spillovers." *Economica* 86 (343):569–606.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6):1251–1271.

Havnes, Tarjei and Magne Mogstad. 2015. "Is universal child care leveling the playing field?" *Journal of Public Economics* 127:100 – 114. The Nordic Model.

Heath, Davidson, Matthew Ringgenberg, Mehrdad Samadi, and Ingrid M. Werner. 2019. "Reusing Natural Experiments." Working Paper 2019-03-021.

Hémet, Camille and Clément Malgouyres. 2018. "Diversity and Employment Prospects: Neighbors Matter!" *Journal of Human Resources* 53 (3):825–858.

Heyes, Anthony and Soodeh Saberian. 2019. "Temperature and Decisions: Evidence from 207,000 Court Cases." *American Economic Journal: Applied Economics* 11 (2):238–65.

Hidalgo, F. Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson. 2010. "Economic determinants of land invasions." *The Review of Economics and Statistics* 92 (3):505–523.

Hilber, Christian A.L. and Frederic Robert-Nicoud. 2013. "On the origins of land use regulations: Theory and evidence from US metro areas." *Journal of Urban Economics* 75:29 – 43.

Hodler, Roland. 2006. "The curse of natural resources in fractionalized countries." *European Economic Review* 50 (6):1367 – 1386.

Holl, Adelheid. 2016. "Highways and productivity in manufacturing firms." *Journal of Urban Economics* 93:131 – 151.

Hollard, Guillaume and Omar Sene. 2016. "Social capital and access to primary health care in developing countries: Evidence from Sub-Saharan Africa." *Journal of Health Economics* 45:1 – 11.

Holmlund, Helena, Helmut Rainer, and Thomas Siedler. 2013. "Meet the Parents? Family Size and the Geographic Proximity Between Adult Children and Older Mothers in Sweden." *Demography* 50 (3):903–931.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review* 102 (4):1477–1507.

Houston, Joel F., Chen Lin, Ping Lin, and Yue Ma. 2010. "Creditor rights, information sharing, and bank risk taking." *Journal of Financial Economics* 96 (3):485 – 512.

Houston, Joel F., Chen Lin, and Yue Ma. 2011. "Media ownership, concentration and corruption in bank lending." *Journal of Financial Economics* 100 (2):326 – 350.

Houtven, Courtney Harold Van and Edward C. Norton. 2004. "Informal care and health care use of older adults." *Journal of Health Economics* 23 (6):1159 – 1180.

———. 2008. "Informal care and Medicare expenditures: Testing for heterogeneous treatment effects." *Journal of Health Economics* 27 (1):134 – 156.

Hoxby, Caroline M. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90 (5):1209–1238.

Hsieh, Chang-Tai and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2):1–39.

Huang, Rocco R. 2008. "Tolerance for uncertainty and the growth of informationally opaque industries." *Journal of Development Economics* 87 (2):333 – 353.

Huber, John D. and Piero Stanig. 2011. "Church-state separation and redistribution." *Journal of Public Economics* 95 (7):828 – 836.

Hull, Peter and Masami Imai. 2013. "Economic shocks and civil conflict: Evidence from foreign interest rate movements." *Journal of Development Economics* 103:77 – 89.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.

Iranzo, Susana and Giovanni Peri. 2009. "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States." *The Review of Economics and Statistics* 91 (2):420–431.

Ito, Koichiro and Mar Reguant. 2016. "Sequential Markets, Market Power, and Arbitrage." *American Economic Review* 106 (7):1921–57.

Ivus, Olena and Matthew Boland. 2015. "The employment and wage impact of broadband deployment in Canada." *Canadian Journal of Economics/Revue canadienne d'économique* 48 (5):1803–1830.

Iyer, Sriya and Anand Shrivastava. 2018. "Religious riots and electoral politics in India." *Journal of Development Economics* 131:104 – 122.

Jack, William and Tavneet Suri. 2014. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review* 104 (1):183–223.

Jacob, Brian, Lars Lefgren, and Enrico Moretti. 2007. "The Dynamics of Criminal Behavior: Evidence from Weather Shocks." *The Journal of Human Resources* 42 (3):489–527.

Jacobsen, Joyce P., James Wishart Pearce, and Joshua L. Rosenbloom. 1999. "The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment." *Journal of Human Resources* 34 (3):449–474.

———. 2001. "The effects of child-bearing on women's marital status: using twin births as a natural experiment." *Economics Letters* 70 (1):133–138.

Jacoby, Hanan G. and Ghazala Mansuri. 2010. "Watta Satta: Bride Exchange and Women's Welfare in Rural Pakistan." *American Economic Review* 100 (4):1804–1825.

Jacoby, Hanan G. and Emmanuel Skoufias. 1997. "Risk, Financial Markets, and Human Capital in a Developing Country." *The Review of Economic Studies* 64 (3):311–335.

Jaeger, David A, Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact of Immigration." Working Paper 24285, National Bureau of Economic Research.

James, William H. 1971. "Excess of Like Sexed Pairs of Dizygotic Twins." *Nature* 232 (5308):277–278.

Jayachandran, Seema. 2006. "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries." *Journal of Political Economy* 114 (3):538–575.

Jiang, Fuxiu, Zhan Jiang, Kenneth A. Kim, and Min Zhang. 2015. "Family-firm risk-taking: Does religion matter?" *Journal of Corporate Finance* 33:260 – 278.

Johar, Meliyanni and Hajime Katayama. 2012. "Quantile regression analysis of body mass and wages." *Health Economics* 21 (5):597–611.

Johnson, Noel D. and Mark Koyama. 2017. "Jewish communities and city growth in preindustrial Europe." *Journal of Development Economics* 127:339 – 354.

Kaestner, Robert. 1991. "The Effect of Illicit Drug Use on the Wages of Young Adults." *Journal of Labor Economics* 9 (4):381–412.

Kaestner, Robert and Neeraj Kaushal. 2012. "Effect of immigrant nurses on labor market outcomes of US nurses." *Journal of Urban Economics* 71 (2):219 – 229.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. "Aggregating governance indicators." Policy Research Working Paper Series 2195, The World Bank.

Kazianga, Harounan and Christopher Udry. 2006. "Consumption smoothing? Livestock, insurance and drought in rural Burkina Faso." *Journal of Development Economics* 79 (2):413 – 446. Special Issue in honor of Pranab Bardhan.

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln. 2015. "Skilled Immigration and the Employment Structures of US Firms." *Journal of Labor Economics* 33 (S1):S147–S186.

Khandker, Shahidur R. 2012. "Seasonality of income and poverty in Bangladesh." *Journal of Development Economics* 97 (2):244 – 256.

Kinnan, Cynthia, Shing-Yi Wang, and Yongxiang Wang. 2018. "Access to Migration for Rural Households." *American Economic Journal: Applied Economics* 10 (4):79–119.

Kirchmaier, Isadora, Jens Prüfer, and Stefan T. Trautmann. 2018. "Religion, moral attitudes and economic behavior." *Journal of Economic Behavior and Organization* 148:282 – 300.

Kline, Brendan and Justin L. Tobias. 2008. "The wages of BMI: Bayesian analysis of a skewed treatment-response model with nonparametric endogeneity." *Journal of Applied Econometrics* 23 (6):767–793.

Klomp, Jeroen and Jakob de Haan. 2013. "Popular protest and political budget cycles: A panel data analysis." *Economics Letters* 120 (3):516 – 520.

Kochar, Anjini. 1999. "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India." *The Review of Economics and Statistics* 81 (1):50–61.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and Inference With Many Invalid Instruments." *Journal of Business & Economic Statistics* 33 (4):474–484.

Kolk, Martin. 2015. "The causal effect of an additional sibling on completed fertility: An estimation of intergenerational fertility correlations by looking at siblings of twins." *Demographic Research* 32 (51):1409–1420.

König, Michael D, Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. 2017. "Networks in Conflict: Theory and Evidence From the Great War of Africa." *Econometrica* 85 (4):1093–1132.

Korpi, Tomas and Michael Tåhlin. 2009. "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974-2000." *Labour Economics* 16 (2):183 – 193.

Kourtellos, Andros, Chih Ming Tan, and Xiaobo Zhang. 2007. "Is the relationship between aid and economic growth nonlinear?" *Journal of Macroeconomics* 29 (3):515 – 540. Special Issue on the Empirics of Growth Nonlinearities.

Kukharskyy, Bohdan. 2016. "Relational contracts and global sourcing." *Journal of International Economics* 101:123 – 147.

Lach, Saul. 2007. "Immigration and Prices." *Journal of Political Economy* 115 (4):548–587.

Lafortune, Jeanne. 2013. "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market." *American Economic Journal: Applied Economics* 5 (2):151–178.

Lafortune, Jeanne, Ethan Lewis, and Jose Tessada. 2019. "People and Machines: A Look at the Evolving Relationship between Capital and Skill in Manufacturing, 1860-1930, Using Immigration Shocks." *The Review of Economics and Statistics* 101 (1):30–43.

Lafortune, Jeanne, José Tessada, and Carolina González-Velosa. 2015. "More hands, more power? Estimating the impact of immigration on output and technology choices using early 20th century US agriculture." *Journal of International Economics* 97 (2):339 – 358.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics and Organization* 15 (1):222–279.

Lassen, David Dreyer. 2007. "Ethnic divisions, trust, and the size of the informal sector." *Journal of Economic Behavior and Organization* 63 (3):423 – 438.

Lee, Iona Hyojung. 2018. "Industrial output fluctuations in developing countries: General equilibrium consequences of agricultural productivity shocks." *European Economic Review* 102:240 – 279.

Levin, Jesse and Erik J.S Plug. 1999. "Instrumenting education and the returns to schooling in the Netherlands." *Labour Economics* 6 (4):521 – 534.

Lewis, Ethan. 2011. "Immigration, skill mix, and capital skill complementarity." *The Quarterly Journal of Economics* 126 (2):1029–1069.

Li, Qiang. 2013. "Language and urban labor market segmentation: Theory and evidence." *Journal of Urban Economics* 74:27 – 46.

Light, Audrey. 2001. "In?School Work Experience and the Returns to Schooling." *Journal of Labor Economics* 19 (1):65–93.

Lin, Liguo. 2013. "Enforcement of pollution levies in China." *Journal of Public Economics* 98:32 – 43.

Lind, Jo Thori. 2007. "Fractionalization and the size of government." *Journal of Public Economics* 91 (1):51 – 76.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barilam. 2013. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." *American Economic Journal: Applied Economics* 5 (2):200–231.

Luke, Nancy and Kaivan Munshi. 2011. "Women as agents of change: Female income and mobility in India." *Journal of Development Economics* 94 (1):1 – 17.

Lundberg, Shelly and Elaina Rose. 2003. "Child gender and the transition to marriage." *Demography* 40 (2):333–349.

Lusardi, Annamaria and Olivia S. Mitchell. 2007. "Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth." *Journal of Monetary Economics* 54 (1):205 – 224. Carnegie-Rochester Conference Series on Public Policy: Economic Consequences of Demographic Change in a Global Economy April 21-22, 2006.

Maccini, Sharon and Dean Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99 (3):1006–26.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do political protests matter? Evidence from the Tea Party Movement." *The Quarterly Journal of Economics* 128 (4):1633–1686.

Makowsky, Michael D. and Thomas Stratmann. 2014. "Politics, unemployment, and the enforcement of immigration law." *Public Choice* 160 (1):131–153.

Mangyo, Eiji. 2008. "The effect of water accessibility on child health in China." *Journal of Health Economics* 27 (5):1343 – 1356.

Marino, Marianna, Pierpaolo Parrotta, and Dario Pozzoli. 2012. "Does labor diversity promote entrepreneurship?" *Economics Letters* 116 (1):15 – 19.

Mauro, Paolo. 1995. "Corruption and Growth." *The Quarterly Journal of Economics* 110 (3):681–712.

———. 1998. "Corruption and the composition of government expenditure." *Journal of Public Economics* 69 (2):263 – 279.

Mazzocco, Maurizio and Shiv Saini. 2012. "Testing Efficient Risk Sharing with Heterogeneous Risk Preferences." *American Economic Review* 102 (1):428–68.

McCaig, Brian and Thanasis Stengos. 2005. "Financial intermediation and growth: Some robustness results." *Economics Letters* 88 (3):306 – 312.

McCleary, Rachel M. and Robert J. Barro. 2006a. "Religion and Economy." *The Journal of Economic Perspectives* 20 (2):49–72.

———. 2006b. "Religion and Economy." *Journal of Economic Perspectives* 20 (2):49–72.

Mehlum, Halvor, Edward Miguel, and Ragnar Torvik. 2006. "Poverty and crime in 19th century Germany." *Journal of Urban Economics* 59 (3):370 – 388.

Méon, Pierre-Guillaume and Geoffrey Minne. 2014. "Mark my words: Information and the fear of declaring an exchange rate regime." *Journal of Development Economics* 107:244 – 261.

Méon, Pierre-Guillaume and Khalid Sekkat. 2008. "Institutional quality and trade: which institutions? Which trade?" *Economic Inquiry* 46 (2):227–240.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *The Quarterly Journal of Economics* 128 (4):1687–1726.

Mian, Atif and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics* 124 (4):1449–1496.

———. 2011. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis." *American Economic Review* 101 (5):2132–56.

———. 2014. "What Explains the 2007-2009 Drop in Employment?" *Econometrica* 82 (6):2197–2223.

Michaelides, Alexander, Andreas Milidonis, and George P. Nishiotis. 2019. “Private information in currency markets.” *Journal of Financial Economics* 131 (3):643 – 665.

Michaelides, Alexander, Andreas Milidonis, George P. Nishiotis, and Panayiotis Papakyriakou. 2015. “The adverse effects of systematic leakage ahead of official sovereign debt rating announcements.” *Journal of Financial Economics* 116 (3):526 – 547.

Miguel, Edward. 2005. “Poverty and Witch Killing.” *The Review of Economic Studies* 72 (4):1153–1172.

Miguel, Edward and Mary Kay Gugerty. 2005. “Ethnic diversity, social sanctions, and public goods in Kenya.” *Journal of Public Economics* 89 (11):2325 – 2368.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. “Economic Shocks and Civil Conflict: An Instrumental Variables Approach.” *Journal of Political Economy* 112 (4):725–753.

Mobarak, Ahmed. 2005. “Democracy, Volatility, and Economic Development.” *The Review of Economics and Statistics* 87 (2):348–361.

Morales, Juan S. 2018. “The impact of internal displacement on destination communities: Evidence from the Colombian conflict.” *Journal of Development Economics* 131:132 – 150.

Morck, Randall and Bernard Yeung. 2011. “Economics, History, and Causation.” *Business History Review* 85 (1):39–63.

Morrill, Melinda Sandler. 2011. “The effects of maternal employment on the health of school-age children.” *Journal of Health Economics* 30 (2):240 – 257.

Munshi, Kaivan. 2003. “Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market.” *The Quarterly Journal of Economics* 118 (2):549–599.

Neal, Derek. 1997. “The Effects of Catholic Secondary Schooling on Educational Achievement.” *Journal of Labor Economics* 15 (1):98–123.

Neidell, Matthew and Janet Currie. 2005. “Air Pollution and Infant Health: What Can We Learn from California’s Recent Experience?” *The Quarterly Journal of Economics* 120 (3):1003–1030.

Orrenius, Pia M. and Madeline Zavodny. 2015. “Does Immigration Affect Whether US Natives Major in Science and Engineering?” *Journal of Labor Economics* 33 (S1):S79–S108.

Ortega, Francesc and Giovanni Peri. 2014. “Openness and income: The roles of trade and migration.” *Journal of International Economics* 92 (2):231 – 251.

Oster, Emily. 2019. “Unobservable selection and coefficient stability: Theory and evidence.” *Journal of Business & Economic Statistics* 37 (2):187–204.

Ottaviano, Gianmarco I. P., Giovanni Peri, and Greg C. Wright. 2013. “Immigration, Offshoring, and American Jobs.” *American Economic Review* 103 (5):1925–59.

Ottaviano, Gianmarco I.P. and Giovanni Peri. 2005. “Cities and cultures.” *Journal of Urban Economics* 58 (2):304 – 337.

Ottaviano, Gianmarco I.P., Giovanni Peri, and Greg C. Wright. 2018. “Immigration, trade and productivity in services: Evidence from U.K. firms.” *Journal of International Economics* 112 (C):88–108.

Palloni, Giordano. 2017. “Childhood health and the wantedness of male and female children.” *Journal of Development Economics* 126:19 – 32.

Park, Albert, Dean Yang, Xinzheng Shi, and Yuan Jiang. 2010. “Exporting and firm performance: Chinese exporters and the Asian Financial Crisis.” *The Review of Economics and Statistics* 92 (4):822–842.

Parrotta, Pierpaolo, Dario Pozzoli, and Davide Sala. 2016. “Ethnic diversity and firms’ export behavior.” *European Economic Review* 89:248 – 263.

Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. New York, NY, USA: Cambridge University Press, 2nd ed.

Peri, Giovanni. 2011. “Rethinking the area approach: Immigrants and the labor market in California.” *Journal of International Economics* 84 (1):1 – 14.

———. 2012. “The effect of immigration on productivity: evidence from U.S. states.” *The Review of Economics and Statistics* 94 (1):348–358.

Peri, Giovanni, Agnese Romiti, and Mariacristina Rossi. 2015. “Immigrants, domestic labor and women’s retirement decisions.” *Labour Economics* 36:18 – 34.

Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. “STEM Workers, H-1B Visas, and Productivity in US Cities.” *Journal of Labor Economics* 33 (S1):S225–S255.

Peri, Giovanni and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1 (3):135–169.

Polidano, Cain, Anna Zhu, and Joel C Bornstein. 2017. "The relation between cesarean birth and child cognitive development." *Scientific reports* 7 (1):11483.

Rainer, Helmut and Thomas Siedler. 2009. "O brother, where art thou? The effects of having a sibling on geographic mobility and labour market outcomes." *Economica* 76 (303):528–556.

Rajan, Raghuram G. and Rodney Ramcharan. 2011. "Land and Credit: A Study of the Political Economy of Banking in the United States in the Early 20th Century." *The Journal of Finance* 66 (6):1895–1931.

Rajan, Raghuram G. and Arvind Subramanian. 2008. "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *The Review of Economics and Statistics* 90 (4):643–665.

Rasul, Imran. 2008. "Household bargaining over fertility: Theory and evidence from Malaysia." *Journal of Development Economics* 86 (2):215 – 241.

Razin, Assaf, Efraim Sadka, and Phillip Swagel. 2002. "Tax burden and migration: a political economy theory and evidence." *Journal of Public Economics* 85 (2):167 – 190.

Rehman, Faiz Ur and Paolo Vanin. 2017. "Terrorism risk and democratic preferences in Pakistan." *Journal of Development Economics* 124:95 – 106.

Riley, Emma. 2018. "Mobile money and risk sharing against village shocks." *Journal of Development Economics* 135:43 – 58.

Robb, Alicia M. and David T. Robinson. 2012. "The Capital Structure Decisions of New Firms." *The Review of Financial Studies* 27 (1):153–179.

Roberts, Michael J. and Wolfram Schlenker. 2013. "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate." *The American Economic Review* 103 (6):2265–2295.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth* 9 (2):131–165.

Røed, Marianne and Pål Schøne. 2016. "Impact of Immigration on Inhabitants' Educational Investments." *The Scandinavian Journal of Economics* 118 (3):433–462.

Rosenthal, Stuart S. and William C. Strange. 2008. "The attenuation of human capital spillovers." *Journal of Urban Economics* 64 (2):373 – 389.

Rothstein, Jesse. 2007. "Does Competition Among Public Schools Benefit Students and Taxpayers? Comment." *American Economic Review* 97 (5):2026–2037.

Rubin, Jared. 2014. "Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation." *The Review of Economics and Statistics* 96 (2):270–286.

Rupert, Peter and Giulio Zanella. 2018. "Grandchildren and their grandparents' labor supply." *Journal of Public Economics* 159:89 – 103.

Rupp, Nicholas G. and George M. Holmes. 2006. "An Investigation into the Determinants of Flight Cancellations." *Economica* 73 (292):749–783.

Ruyssen, Ilse and Sara Salomone. 2018. "Female migration: A way out of discrimination?" *Journal of Development Economics* 130 (C):224–241.

Sá, Filipa. 2015. "Immigration and House Prices in the UK." *Economic Journal* 125 (587):1393–1424.

Saiz, Albert. 2007. "Immigration and housing rents in American cities." *Journal of Urban Economics* 61 (2):345 – 371.

———. 2010. "The Geographic Determinants of Housing Supply." *The Quarterly Journal of Economics* 125 (3):1253–1296.

Sander, William. 1995. "Schooling and smoking." *Economics of Education Review* 14 (1):23 – 33.

Sarsons, Heather. 2015. "Rainfall and Conflict: A Cautionary Tale." *Journal of Development Economics* 115 (July):62–72.

Shi, Wei and Yinuo Tang. 2015. "Cultural similarity as in-group favoritism: The impact of religious and ethnic similarities on alliance formation and announcement returns." *Journal of Corporate Finance* 34:32 – 46.

Shih, Kevin. 2017. “Do international students crowd-out or cross-subsidize Americans in higher education?” *Journal of Public Economics* 156:170 – 184.

Siegel, Jordan I., Amir N. Licht, and Shalom H. Schwartz. 2013. “Egalitarianism, Cultural Distance, and Foreign Direct Investment: A New Approach.” *Organization Science* 24 (4):1174–1194.

sing Kung, James Kai and Chicheng Ma. 2014. “Can cultural norms reduce conflicts? Confucianism and peasant rebellions in Qing China.” *Journal of Development Economics* 111:132 – 149. Special Issue: Imbalances in Economic Development.

Skoufias, Emmanuel. 1996. “Intertemporal substitution in labor supply: Micro evidence from rural India.” *Journal of Development Economics* 51 (2):217 – 237.

Smith, Christopher L. 2012. “The Impact of Low-Skilled Immigration on the Youth Labor Market.” *Journal of Labor Economics* 30 (1):55–89.

Spenkuch, JÃ¶rg L. 2017. “Religion and work: Micro evidence from contemporary Germany.” *Journal of Economic Behavior and Organization* 135:193 – 214.

Stock, James and Motohiro Yogo. 2005. *Testing for Weak Instruments in Linear IV Regression*. New York: Cambridge University Press, 80–108.

Strobl, Eric. 2011. “The Economic Growth Impact of Hurricanes: Evidence from U.S. Coastal Counties.” *The Review of Economics and Statistics* 93 (2):575–589.

Stuen, Eric T., Ahmed Mushfiq Mobarak, and Keith E. Maskus. 2012. “Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US Doctoral Programmes.” *The Economic Journal* 122 (565):1143–1176.

Stulz, Rene M. and Rohan Williamson. 2003. “Culture, openness, and finance.” *Journal of Financial Economics* 70 (3):313 – 349.

Taber, Christopher R. 2001. “The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?” *The Review of Economic Studies* 68 (3):665–691.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen. 2010. “Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam.” *American Economic Review* 100 (1):557–71.

Tavares, José and Romain Wacziarg. 2001. “How democracy affects growth.” *European Economic Review* 45 (8):1341 – 1378.

Team, Cebu Study. 1992. “A child health production function estimated from longitudinal data.” *Journal of Development Economics* 38 (2):323 – 351.

Valentina, Dimitrova-Grajzl, Grajzl Peter, Guse A. Joseph, and Smith J. Taylor. 2016. “Racial Group Affinity and Religious Giving: Evidence from Congregation-Level Panel Data.” *The B.E. Journal of Economic Analysis & Policy* 16 (2):689–725.

van Ours, Jan C. 2004. “A pint a day raises a man’s pay; but smoking blows that gain away.” *Journal of Health Economics* 23 (5):863 – 886.

Vlaicu, Razvan and Alexander Whalley. 2016. “Hierarchical accountability in government.” *Journal of Public Economics* 134:85 – 99.

Vogl, Tom S. 2013. “Marriage Institutions and Sibling Competition: Evidence from South Asia.” *The Quarterly Journal of Economics* 128 (3):1017–1072.

Waldinger, Maria. 2017. “The long-run effects of missionary orders in Mexico.” *Journal of Development Economics* 127:355 – 378.

Wang, Cong. 2013. “Can institutions explain cross country differences in innovative activity?” *Journal of Macroeconomics* 37:128 – 145.

West, Martin R. and Ludger Woessmann. 2010. “"Every Catholic Child in a Catholic School": Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries.” *The Economic Journal* 120 (546):F229–F255.

Winston, Clifford and Ashley Langer. 2006. “The effect of government highway spending on road users’ congestion costs.” *Journal of Urban Economics* 60 (3):463 – 483.

Woodruff, Christopher and Rene Zenteno. 2007. “Migration networks and microenterprises in Mexico.” *Journal of Development Economics* 82 (2):509 – 528.

Wozniak, Abigail and Thomas J. Murray. 2012. “Timing is everything: Short-run population impacts of immigration in US cities.” *Journal of Urban Economics* 72 (1):60 – 78.

Yermack, David. 2014. "Tailspotting: Identifying and profiting from CEO vacation trips." *Journal of Financial Economics* 113 (2):252 – 269.

Young, Alwyn. 2019. "Consistency without Inference: Instrumental Variables in Practical Application." Working papers, London School of Economics.

Zak, Paul J. and Stephen Knack. 2001. "Trust and Growth." *The Economic Journal* 111 (470):295–321.

Ziliak, James P. and Thomas J. Kniesner. 1999. "Estimating Life Cycle labor Supply Tax Effects." *Journal of Political Economy* 107 (2):326–359.

Zolotoy, Leon, Don O'Sullivan, and Yangyang Chen. 2019. "Local religious norms, corporate social responsibility, and firm value." *Journal of Banking and Finance* 100:218 – 233.