American Sociological Review
1–7

DOI: 10.1177/0003122417691952
journals.sagepub.com/home/asr

Culture as a Random Treatment: A Reply to Chou

Javier G. Polavieja[a]

Abstract

In a 2015 *ASR* article, I introduced SISTER, a new method to estimate the causal effects of culture using migrant populations. Chou raises significant concerns about SISTER and concludes that the method is flawed. I contend that this conclusion is incorrect because it is based on a mischaracterization of the method's identification assumptions. Specifically, Chou disregards that SISTER exploits cultural variation across multiple countries of origin/ancestry as the main source of identification. I argue that SISTER can comply with the IV assumptions precisely because it is a multiple-origin method that conceptualizes culture of birth as a random treatment. I discuss potential threats to the exogeneity condition and offer several recommendations for future applications of the method.

Keywords

culture, migration, quantitative methods, instrumental variables

In a 2015 *ASR* article, I proposed a new method to tackle the problem of endogenous preferences in the estimation of cultural effects (Polavieja 2015). The SISTER method combines imputation regression, instrumental-variable (IV) estimation, and cross-national sampling using migrant populations as a means to capture the exogenous influence of culture on people's behavior. In his comment, Chou raises significant concerns about SISTER and concludes that the method "is an unnecessary and obscurant alternative to standard IV methods" and hence "should not be used." In this reply, I provide a careful examination of Chou's arguments. I contend that Chou's main conclusion that SISTER is flawed by design is incorrect as a general statement, because it is based on a mischaracterization of the method's identification assumptions. Specifically, Chou disregards that SISTER exploits cultural variation across multiple countries of origin/ancestry as the main source of identification. I argue that SISTER can comply with the IV assumptions precisely because it is a multiple-origin method that conceptualizes culture of birth as a random treatment. Chou is right, however, that there are a number of important potential threats to the exogeneity condition in SISTER. While I believe such potential threats could indeed invalidate particular applications of the method, they do not invalidate the whole approach. Hence the question, in my view, is not whether SISTER is right or wrong, sound or flawed, in general, but under which specific conditions particular applications of the

[a]Universidad Carlos III de Madrid

Corresponding Author:
Javier Polavieja, Department of Social Sciences, Universidad Carlos III de Madrid, C/ Madrid, 135 28903 Getafe, Madrid, Spain
E-mail: javier.polavieja@uc3m.es

method are likely to fail. I welcome the opportunity to clarify the assumptions behind my own application of the method to the case of female labor force participation (FLFP) and to address Chou's legitimate concerns about omitted variables and selection bias.

For ease of exposition, I first clarify the mechanics of SISTER and then take up Chou's concerns, starting with his most fundamental challenges to the method.

SISTER AS AN EXTENSION OF IV METHODS

What is the impact of traditionalism on women's labor supply? The general causal model for women's participation decision, Y, is given by the following equation:

$$Y = \alpha + \tau T + \mathbf{G}\gamma + \varepsilon \tag{1}$$

where T is women's traditionalism, $\mathbf{G}$ is a vector of control variables, and ε is an error term that summarizes all other causes of FLFP.

The problem of endogenous preferences arises because people's preferences and beliefs are influenced by a whole set of environmental factors that also influence their actions. These environmental factors act as a big omitted confounder generating a correlation between T and ε. This means we cannot identify the causal effect of T on Y by regression (in other words, the OLS estimate of τ is bound to be biased). The method of IV allows researchers to consistently estimate τ if they observe an instrument Z that complies with two conditions. The first condition is the *relevance* condition, which requires that the instrument is correlated with T net of the control variables, $\mathbf{G}$; formally: $Cov(Z,T|\mathbf{G}) \neq 0$. The second condition is the *exogeneity* condition, which requires that the instrument is uncorrelated with the error term; formally: $Cov(Z,\varepsilon) = 0$. The exogeneity condition also implies that Z must only affect Y through its effect on T, an implication known as the *exclusion restriction*.

The trouble is when people are observed within the bounds of a single society, it is extremely hard for researchers to find an instrument that is not itself affected by the same environmental factors that simultaneously influence T and Y. This poses a formidable methodological challenge for the estimation of cultural effects. To tackle this problem, the SISTER method exploits migration across different social environments as a key source of identification.

SISTER generates an instrument, $\hat{Z}_i$, for immigrants' trait of interest, T_i, by imputing from non-migrating equivalents j, observed at different countries of origin/ancestry c, using a set of variables, $\boldsymbol{X}_i$, as imputation predictors.[1] This is done in two steps. First, we model non-migrants' traditionalism ($\boldsymbol{T}_j$) at country of origin using a set of multiple regressions of the form:

$$\hat{T}_j = \hat{\gamma}_{0,c_j} + \sum\nolimits_{s=1}^{k} \hat{\gamma}_{s,c_j} X_{s,j} \tag{2}$$

where c_j is the country of non-migrating individual j; $\hat{\gamma}_{0,c_j}$ is the intercept for c_j; and $\hat{\gamma}_{s,c_j}$ is a vector of multiple regression coefficients.

Second, we estimate the imputed value of traditionalism for migrants ($\hat{Z}_i$) by multiplying their own values of $\boldsymbol{X}$ by the coefficients estimated at their respective countries of origin/ancestry (c):

$$\hat{Z}_i = \hat{\gamma}_{0,c_i} + \sum\nolimits_{s=1}^{k} \hat{\gamma}_{s,c_i} X_{s,i} \tag{3}$$

It is key to understand that this method exploits variation across *multiple* cultures of origin/ancestry as the main source of identification. The intuition is simple: migration can help us identify the exogenous component of culture because it removes people from their original embedding social environments. We can call this idea the *epidemiological principle* (Fernandez 2011). In other words, it is only because immigrant women's degree of traditionalism varies with country of origin/ancestry that we can identify the "cultural component" of this trait. Hence, for SISTER to work, there must be variation in origin/ancestry.

The SISTER instrument $\hat{Z}_i$ (called *synthetic traits*) thus contains information not only of individual covariates ($\boldsymbol{X}_i$), but also, and crucially so, from the country of origin/ancestry (c_i). In the 2015 article, I fitted 23 different imputation regressions for immigrants from 23 different origin/ancestry countries using information from 23 different country samples of non-migrating donors.[2] Each of these imputation regressions included an intercept, $\hat{\gamma}_{0,c_j}$, which can be understood as the net average value that traditionalism takes in each country of origin/ancestry.

The relevance condition for a valid instrument in SISTER can thus be formally expressed as follows:

$$Cov(\hat{\gamma}_{0,c_i} + \ldots + \hat{\gamma}_{k,c_i} X_{k_i}, T_i \mid \mathbf{G}_i) \neq 0 \qquad (4)$$

and the exogeneity condition is:

$$Cov(\hat{\gamma}_{0,c_i} + \ldots + \hat{\gamma}_{k,c_i} X_{k_i}, \varepsilon_i) = 0 \qquad (5)$$

Refuting Chou's Claim That SISTER Is Unnecessary

Chou claims that the imputation step in SISTER in unnecessary because the imputation regression can only yield valid instruments $\hat{Z}_i$ if it includes (at least one) valid non-imputed regressor(s) X_i. Hence, his argument goes, researchers should ignore the imputation step entirely and use such valid non-imputed regressor(s) directly as instrument(s) in standard IV regression. Thus, for Chou, there is never a reason to use SISTER.

Chou is right that the exogeneity of $\hat{Z}_i$ implies the exogeneity of at least one imputation parameter. Yet he disregards that SISTER is a multiple origin method, which means $\hat{\gamma}_{0,c_i}$ is also a parameter of $\hat{Z}_i$. This omission has crucial implications for the validity of his argument.

To be clear, Chou's critique would be impeccable if we observed women coming from only one single country of origin/ancestry (i.e., if $c_i = 1$ for all i); $\hat{Z}_i$ would then be collinear with variables already included in the structural model (Equation 1), unless the researcher had an extra instrument for identification (i.e., an exogenous variable different from ($\mathbf{G}_i$). In this case, *but only in this case*, SISTER would indeed be superfluous, because researchers could (and for clarity should) simply use that exogenous variable directly as the instrument, as Chou argues. Yet it is easy to see that when multiple origins are considered, $\hat{\gamma}_{0,c_i}$ is no longer a constant across individuals, as Chou assumes,[3] but varies with origin (because different countries have different net average values of traditionalism). SISTER uses this variation for identification. Conceptually, this amounts to considering culture of origin/ancestry as a random treatment. SISTER exploits the "lottery of birth" and asks: how would a woman's employment propensity change had her culture of origin/ancestry been different? Intuitively, to answer this question we need variation in the treatment (i.e., we need multiple origins). When multiple origins are considered, SISTER can provide identification even if $\boldsymbol{X}_i$ consists only of (a subset of) the variables included in the structural model, $\mathbf{G}_i$, provided that the imputed instrument is still relevant.[4]

Variation in the intercepts estimated at country of origin/ancestry thus acts as the main source of identification in SISTER. Yet it is not the only source. The use of (any subset of) $\mathbf{G}_i$ as imputation predictors allows us to improve identification by also exploiting variation in the imputation coefficients across different origins (e.g., differences in the way age and schooling correlate with traditionalism in each country of origin/ancestry).[5] Conceptually, this amounts to using variation in the degree of cultural homogeneity at the country of origin/ancestry as an additional source of identification.[6] Because $\mathbf{G}_i$ covariates are controls in the structural equation and thus orthogonal to the error term, using them as imputation predictors requires making no additional exogeneity assumptions. If, by contrast, the researcher wishes to impute on covariates not included in the structural model, she will have to justify theoretically why they are assumed exogenous.

A key advantage of the imputation step in SISTER is that it provides a parsimonious way of modelling the first stage in IV estimation by taking a large number of potential instruments and condensing them in a scalar using external data. To see this, note that if researchers were to follow Chou's advice to skip this step, they would have to introduce a large number of instruments (at the very least 1–*c* country of origin/ancestry dummies) and this would inevitably reduce instrument strength. The imputation step can thus refine the instrument by (1) providing more precise estimates of culture of origin variation based on non-migrant samples; and (2) addressing the "many instruments" problem, which is one of the most important sources of estimation bias in standard IV methods (see, e.g., Andrews and Stock 2005).[7]

Refuting Chou's Claim That SISTER Threatens Transparency and Validity

Chou further claims that the imputation step "hides the dependence of SISTER on *all* non-imputed regressors" and notes that this "can also diminish the validity of empirical estimates." SISTER, he concludes, is thus both obscurant and harmful. Again, this conclusion seems based on an inexact characterization of the method. To be clear, Chou would be absolutely right *if and only if* imputation predictors potentially affecting FLFP were excluded from the structural equation. For example, if we were to use women's schooling to impute synthetic traditionalism but did not control for schooling in the structural equation, then the effect of schooling on FLFP would be hidden in the effect of the imputed instrument, $\hat{Z}_i$. In this example, the exclusion restriction would be violated because schooling directly affects FLFP (as well as traditionalism), making $\hat{Z}_i$ invalid. This is why all $\boldsymbol{X}$s that are potential predictors of FLFP should always be introduced also as controls in the structural equation $\mathbf{G}_i$, as I stated in the 2015 article. The inclusion of $\mathbf{G}_i$ covariates in the imputation step can never harm the validity of $\hat{Z}_i$ simply because they are predictors of FLFP in the structural equation (and hence uncorrelated with the error term).[8]

In summary, SISTER is not flawed by construction: by exploiting cultural variation across multiple countries of origin/ancestry (i.e., variation in both the intercepts and the slopes of the imputation regression) synthetic traits can *in theory* provide identification. Now the question that needs to be addressed is under which specific conditions particular applications of the SISTER method are likely to fail. As with any other IV method, this implies discussing potential threats to the exogeneity condition.

OMITTED VARIABLE BIAS (OVB) AS A THREAT TO EXOGENEITY IN SISTER

Clearly, as Chou asserts, the SISTER method will likely fail if immigrants bring unobserved attributes with them that are jointly correlated with their cultural traits and economic outcomes. Because variation in origin/ancestry is the main instrument in SISTER, it is crucially important to consider potential country-level effects on FLFP not properly measured in the structural equation. This is a major challenge for SISTER. A case in point is unobserved ability. Chou notes that unobserved differences in ability (linked to origin) can pose a fundamental threat to the method's validity, because ability might affect FLFP directly.[9] For example, if the quality of educational systems differed by country of origin/ancestry in a way that is correlated with levels of traditionalism and women's participation decisions, then respondents' schooling would not capture all the variation in relevant skills linked to FLFP.

To address concerns about OVB, I carried out numerous robustness tests in an online supplement to the 2015 article (henceforward the 2015 supplement). These tests suggest that potential differences in unobserved skills are probably not a major source of bias in the European Social Survey. This is actually unsurprising, since most countries of origin/ancestry in the dataset are advanced European economies with highly developed educational systems (which is one of the key advantages of exploiting intra-European migration).[10] Many other robustness tests for omitted skill bias are of course possible and indeed advisable in future applications of the method. If data are available, researchers should seek to directly control for individual measures of skills/ability as well as for parental education and skills. Alternatively, they could test the sensitivity of results to the choice of individual-level controls potentially linked to ability (Luttmer and Singhal 2011). Other ways of testing for potential unobserved skill bias include controlling for macro-economic conditions and the quality of the educational system at country of origin/ancestry (Fernandez and Fogli 2009); removing migrants from countries with low-quality education from the sample (see the 2015 supplement); using standard Mincer earning equations to capture differences in productivity by origin/ancestry (Fernandez 2011); estimating sibling fixed-effects models (Finseraas and Kotsadam 2015); or introducing measures of linguistic distance between country of origin/ancestry and country of destination/birth. Many of these tests exploit origin variation to address potential OVB. It is obvious, however, that not all forms of OVB can be tested empirically. This means assumptions about unobserved heterogeneity must be made, as with any other IV.

Is the Exclusion Restriction Implausible in SISTER?

Chou argues that the exclusion restriction in SISTER is "implausible" because traditionalism will likely influence other behaviors (e.g., childbearing and educational attainment) that affect FLFP. Note that Chou is assuming these variables are unobserved. In fact, in the 2015 article both were observed *post migration*. Yet while I controlled for education in all models, I purposely left family composition variables out of the best-specified model (Polavieja 2015:178, Figure 2; 182, Table 4). The reason is simple: traditionalism affects family decisions jointly with labor-market decisions (i.e., traditional women are more likely to marry, have children, and withdraw from the labor market). This implies that the full impact of traditionalism on FLFP can only be captured *before* controlling for family structure variables. I stress this point to illustrate the importance of always inquiring about the substantive (theoretical) meaning of variables, for what might appear as a "lurking" violator of exclusion could, on closer look, turn out to be a crucial channeling variable.

But Chou is of course right that unobserved *pre-migration* characteristics should be a cause for concern. In his discussion of "first-best assumptions," Chou concludes that "unfortunately, there is no fix for this problem." I disagree. One very easy way of addressing this potential problem is to consider second-generation immigrants as the target sample. In the 2015 supplement, I showed that findings were robust to restricting the sample to the second and the 1.5 generations. Obviously, the second generation has no pre-migration characteristics to worry about because they did not migrate.

SELECTION INTO MIGRATION BIAS (SB)

SB also poses a major threat to SISTER. Chou argues that because migration is not a random phenomenon, applying SISTER could even invalidate a valid instrument. Chou draws on both simple fictional illustrations and complex simulation experiments to prove his point (the latter presented in his online supplement [http://asr.sagepub.com/supplemental]). Chou's illustrations assume traditionalism is exogenous in the population. This amounts to assuming away the very problem we want to solve. For space, I address his most complex simulations in the online supplement. There I show that when both endogenous preferences and selective migration are formalized jointly, Monte Carlo simulations can yield very different results to those reported by Chou. Does this mean we should be unconcerned with potential SB? Absolutely not: it only means that fictional illustrations and simulation experiments cannot shed definitive light on this problem.

Potential sources of SB can, however, be investigated *empirically*. There are a host of robustness tests one can do, including testing again for the sensitivity of results to individual-level controls potentially linked to the migration decision; exploring the correlation between recent migrants' values and those of their non-migrant equivalents to test for potential selection on the trait; controlling for macro-economic conditions at both origin and destination to test for selection on the outcome; and testing the sensitivity of results to the choice of origin sample (by removing migrants from particular sending countries where selection effects seem most plausible). If data permits, researchers could also investigate more directly the reasons behind migration decisions. The epidemiological literature in economics provides numerous examples of robustness tests for SB (for a discussion, see Fernandez 2011). Of course, not all forms of SB can be fully addressed empirically and again, as with any other method, assumptions must be made. SISTER users should always be aware of these assumptions and perform as many robustness tests for potential SB as appropriate to their specific research questions and data (see, e.g., the 2015 supplement). Finally, note that using second-generation migrants as the target sample will logically diminish the potential biasing impact of selective migration.

CONCLUSIONS

Epidemiological approaches challenge the view that "moral systems" are not randomly assigned. The lottery of birth assigns people to different cultures. If people never moved out of these cultures, the problem of endogenous preferences would seem unsolvable. But people move across cultures and very often they do so for reasons unrelated to either cultural traits or the outcomes of interest. This provides a unique opportunity to identify the causal effect of underlying cultural values without sacrificing classical theory's great ambition to understand the motivation for human action. What Chou calls the "unmoved" aspects of culture can be identified because they are *portable*.

I conclude this reply by making six simple recommendations for future applications of SISTER. First, if data permits, apply SISTER to the second generation. Second, use observations for as many countries of ancestry as possible. Third, use a very simple imputation regression. Fourth, make sure all imputation regressors are theoretically defensible. Fifth, check that all imputation regressors that have the potential to directly affect the outcome are included in the structural equation. And, sixth, always test thoroughly for potential sources of OVB and SB.

Tackling the problem of endogenous preferences has crucial implications for theory and research. I invite scholars interested in the study of cultural effects to take this problem seriously and to join us in the ongoing search for new and imaginative solutions.

Funding

This study is part of the project *New Approaches to Immigration Research, NewAir*, funded by the Spanish Ministry of Economy and Competitiveness (CSO2016-78452).

Acknowledgments

I wish to thank Michael Biggs, Robert Fishman, Lucinda Platt, and the *ASR* reviewers for their helpful comments. I also wish to thank very specially Sophocles Mavroeidis for his generous input and his invaluable help in programming the simulation experiments presented in the online supplement to this reply. All errors are my own.

Notes

1. Imputation predictors "should seek to capture relevant sources of socialization (in the trait of interest) that operate within nations of origin" (Polavieja 2015:175).
2. In total, I used more than 40,000 donors to impute synthetic traditionalism for roughly 3,000 migrant women.
3. Note that Equations 4 to 7 in Chou's comment disregard variation in country of origin/ancestry.
4. To see this formally, note that $Cov\left(\hat{Z}_i, T_i|\mathbf{G}_i\right) = Cov(\hat{\gamma}_{0,c_i} + \mathbf{G}_i^T \hat{\gamma}_{G,c_i}, T_i|\mathbf{G}_i) \neq 0$ precisely because of the variation in $\hat{\gamma}_{0,c_i}$, where $\hat{\gamma}_{G,c_i}$ is a vector of the coefficients of $\mathbf{G}_i$ in country of origin c_i. This means that, contrary to what I hesitantly suggested in the 2015 article, imputing on exogenous parameters not included in the structural equation is not necessary for SISTER to provide identification. All we need is origin variation under the sole assumption that culture of birth is exogenous. This is an important addition to Polavieja (2015).

5. I provide both formal and empirical proof of this statement in the online supplement (see Parts 1 and 2 [http://asr.sagepub.com/supplemental]).
6. Within countries, cultural traits vary across regions, generations, and social groups (Polavieja 2015:170–72), but the extent of this variation can itself be country-specific (i.e., some countries are more culturally homogeneous than others).
7. For an empirical proof, see the online supplement (Part 3, Table S2).
8. For a formal proof, see Part 1 of the online supplement.
9. Unobserved ability might also affect the probability of migration, and hence it is a potential source of selection bias.
10. In my view, the idea that differences in unobserved ability can affect intra-European migrants' FLFP seems a bit far-fetched. Unobserved skills are probably more likely to influence earnings, occupational attainment, or career progression.

References

Andrews, Donald W. K., and James H. Stock 2005. "Inference with Weak Instruments." NBER Technical Working Paper No. 313, August (http://www.nber.org/papers/t0313.pdf).

Fernandez, Raquel. 2011. "Does Culture Matter?" Pp. 481–510 in *Handbook of Social Economics*, edited by J. Benhabib, M. O. Jackson, and A. Bisin. North-Holland: Elsevier.

Fernandez, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1(1):146–77.

Finseraas, Henning, and Andreas Kotsadam. 2015. "Ancestry Culture, Local Institutions, and Situated Agency: A Study of Female Labor Supply." Unpublished manuscript, Oslo University (https://andreaskotsadam.files.wordpress.com/2012/02/culture-and-female-labor-supply.pdf).

Luttmer, Erzo F. P., and Monica Singhal. 2011. "Culture, Context and Taste for Redistribution." *American Economic Journal: Economic Policy* 3(1):157–79.

Polavieja, Javier G. 2015. "Capturing Culture: A New Method to Estimate Exogenous Cultural Effects using Migrant Populations." *American Sociological Review* 80(1):166–91.

Javier G. Polavieja is Banco Santander Professor of Sociology at Universidad Carlos III de Madrid. His main fields of research are labor-market stratification, economic sociology, political sociology, and immigration research. His work has appeared in the *American Sociological Review, American Journal of Sociology, European Sociological Review, Social Forces, Socio-Economic Review, Labour Economics*, and other outlets. In 2015 he was appointed Fellow of the European Academy of Sociology.