HIAS-E-44

Some Remarks on the Causal Inference for Historical Persistence

Katsuo Kogure
Center for Southeast Asian Studies, Kyoto University, 46 Shimoadashi-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan

May 2017
reviced in September 2017

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8604 http://hias.ad.hit-u.ac.jp/

HIAS discussion papers can be downloaded without charge from:
http://hdl.handle.net/10086/27202
https://ideas.repec.org/s/hit/hiasdp.html

Some Remarks on the Causal Inference for Historical Persistence *

Katsuo Kogure†

September 20, 2017

Abstract

A growing body of literature examines the relationships between historical events and contemporary economic outcomes. Recent studies estimate the causal effects using detailed historical data and contemporary microdata of individuals and/or households. In this paper, we discuss conceptual and econometric issues inherent in the causal inference following the potential outcomes framework. Using an empirical example, we also discuss a simple alternative approach to avoid these issues that is coherent with the potential outcomes framework.

JEL Codes: C01, N01
Keywords: history, economic development, causal effect, Rubin causal model

*I am grateful to Keisuke Hirano for his insightful comments on this issue. I am also grateful to Jun Goto, Kosuke Imai, Nobuyoshi Kikuchi, Hisaki Kono, Masahiro Kubo, Fumiharu Mieno, Chiaki Moriguchi, Ryo Nakajima, Fumio Ohtake, Yoshito Takasaki, Kensuke Teshima, Ken Yamada, and seminar participants at the 3rd Hitotsubashi Summer Institute 2017, Kyoto University, Kansai Labor Workshop, and Nihon Fukushi University for helpful discussions and suggestions. In addition, I am indebted to Melissa Dell and *Econometrica* for allowing me to access her data and programs. This research is supported by Grant-in-Aids for Scientific Research No.15K17044, Ministry of Education, Science and Culture, Japan.

†Center for Southeast Asian Studies, Kyoto University, 46 Shimoadachi-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan. Tel: +81-75-753-7302. Email: katsukogure@gmail.com.

Recent years have seen growing interest in the causal relationships between historical events and contemporary economic outcomes (see Nunn (2009, 2014) for reviews). One feature of the more recent studies is the use of detailed historical data and contemporary microdata of individuals and/or households to identify the causal effects of history on outcomes (e.g., Dell (2010), Nunn and Wantchekon (2011)). This paper discusses conceptual and econometric issues inherent in the causal inference following the Rubin causal model, a framework for causal inference based on potential outcomes (Rubin (1974, 1977)).[1]

Although the existing empirical studies do not always explicitly follow the potential outcomes framework, we follow the framework because it enables us to raise most effectively the issues inherent in the causal inference for historical persistence with microdata; many modern empirical works that examine causal questions adopt this framework (see, e.g., Heckman et al. (1999), Angrist and Pischke (2009), Duflo et al. (2008), Imbens and Wooldridge (2009), Imbens and Rubin (2015)). Using an empirical example, we also discuss a simple alternative approach to avoid certain issues that is coherent with the Rubin causal model.

The paper proceeds as follows. Section I provides a brief description of the Rubin causal model, specifying the major premises. Section II considers the causal inference for historical persistence with contemporary microdata and discusses fundamental problems inherent in the causal inference. We also discuss a simple alternative approach to address these problems. Section III provides an empirical example based on the approach and Section IV concludes.

[1]The potential outcomes were originally introduced by Fisher (1935) and Neyman (1923) for randomized experiments and extended by Rubin (1974, 1977) for non-randomized studies. Holland (1986) labeled this framework the Rubin causal model.

I. Overview of Causal Inference Using Potential Outcomes

Essential Elements of the Rubin Causal Model. The Rubin causal model consists of three essential elements (e.g., Holland and Rubin (1988)). The first is a set (population) of units, U, the size of which is denoted by N, indexed by $i = 1, \ldots, N$. Examples of units include individuals, households, firms, counties, states, and countries. The second is a set of treatments, D, with each unit being exposed to one of the treatments. For simplicity, we assume two treatments, $D_i = \{1, 0\}$, where $D_i = 1$ if unit i is exposed to the treatment and $D_i = 0$ if unit i is not. The third is a response variable, Y, that is recorded for each unit after its exposure to either of the treatments.

Causal Inference. The Rubin causal model assumes that each unit i has two potential outcomes, $Y_i(1)$ and $Y_i(0)$, where $Y_i(1)$ is the value of the response that would be observed if unit i received the treatment while $Y_i(0)$ is the value that would be observed if the same unit did not. Let Y_i^{obs} denote the realized and observed outcome: $Y_i^{obs} = Y_i(D_i) = Y_i(1) \cdot D_i + Y_i(0) \cdot (1 - D_i)$. The potential outcomes enable us to define causal effects at three levels: unit level, population level, and subpopulation level (e.g., Holland and Rubin (1988)).

The unit-level causal inference is defined as the difference between the two potential outcomes for the same unit:

$$Y_i(1) - Y_i(0).$$

The population-level causal inference is defined as the expectation of the difference in the unit-level causal effect over population:

$$E[Y_i(1) - Y_i(0)].$$

The subpopulation-level causal inference is defined in many ways. One definition is the expectation of the difference in the unit-level causal effect over the

subpopulation with covariates, X_i:

$$E[Y_i(1) - Y_i(0)|X_i].$$

Fundamental Problem of Causal Inference. For causal inference at any level, we always face the problem that we can never observe both $Y_i(1)$ and $Y_i(0)$ at the same time. We can observe, at most, either $Y_i(1)$ or $Y_i(0)$. Thus, it is impossible to directly observe the causal effects at all three levels ("fundamental problem of causal inference," Holland (1986)).

Assumptions. Causal inference relies on assumptions. Three key assumptions are normally used. The first is the stable unit treatment value assumption (SUTVA), which requires that the potential outcomes of unit i are not affected by the treatments received by any other units and there are no multiple versions of treatments (Rubin (1980, 1986)). The second assumption is unconfoundedness (Rosenbaum and Rubin (1983)),[2]

$$\big(Y_i(1), Y_i(0)\big) \perp D_i|X_i.$$

Under this assumption, the treatment assignment, D_i, is statistically independent of the potential outcomes, $Y_i(1)$ and $Y_i(0)$, given X_i. The third assumption is overlap,

$$0 < Pr(D_i = 1|X_i) < 1.$$

This assumption ensures overlap in the covariate distribution of treatments and controls. The combination of unconfoundedness and overlap is referred to as "strong ignorability" (Rosenbaum and Rubin (1983)).

[2]Unconfoundedness is closely related to the notion of exogeneity in the econometrics literature (Manski et al. (1992)). The term unconfoundedness is also referred to as "selection on observable" (Barnow et al. (1980)) and the "conditional independence assumption" (Lechner (2001), Angrist and Pischke (2009)).

Identification. The three assumptions justify causal inference, as follows. Suppose that we are interested in learning the conditional average treatment effect, $E[Y_i(1)-Y_i(0)|X_i]$. Under the three assumptions, the average treatment effect can be identified by relying only on observed outcomes:

$$\begin{aligned} E[Y_i(1) - Y_i(0)|X_i] &= E[Y_i(1)|X_i] - E[Y_i(0)|X_i] \\ &= E[Y_i(1)|X_i, D_i = 1] - E[Y_i(0)|X_i, D_i = 0] \\ &= E[Y_i^{obs}|X_i, D_i = 1] - E[Y_i^{obs}|X_i, D_i = 0]. \end{aligned}$$

Here, since $E[Y_i(1)|X_i, D_i = 1]$ and $E[Y_i(0)|X_i, D_i = 0]$ do not depend on D_i under the unconfoundedness assumption, the second equality holds. In addition, based on the overlap assumption, we can estimate both $E[Y_i^{obs}|X_i, D_i = 1]$ and $E[Y_i^{obs}|X_i, D_i = 0]$ for a subpopulation with covariates X_i.

Premises of the Rubin Causal Model. The Rubin causal model (summarized above) is based on two premises: (1) Units *exist* within a specific time period and (2) the action of treatments and the measurement of outcomes take place on a *common* unit.[3] Holland (1986) discusses these two premises, emphasizing the significance of the *role of time* in the causal inference versus the associational inference, the standard statistical model that simply relates two variables over population.

Types of Variables. Holland (1986) also discusses types of variables. Since treatments must occur within a specific time period, variables are classified into two types: pre-treatment variables and post-treatment variables (variables determined before and after receiving the treatments, respectively). The latter may be affected by treatments while the former are not. Thus, it

[3]Frameworks have also been developed to address any lack in parts of the units within a specified time frame due to missing outcomes (dropout) following treatment non-compliance and "truncation-by-death" (e.g., Frangakis and Rubin (2002), Zhang and Rubin (2003)).

is possible to identify the causal effects of the treatments by comparing post-treatment variables (outcomes) between treatment and control groups with similar values of covariates (under the assumptions).

II. Causal Inference for Historical Persistence

Following the potential outcomes framework, we now consider the causal inference for historical persistence with contemporary microdata.

A. The Standard Approach

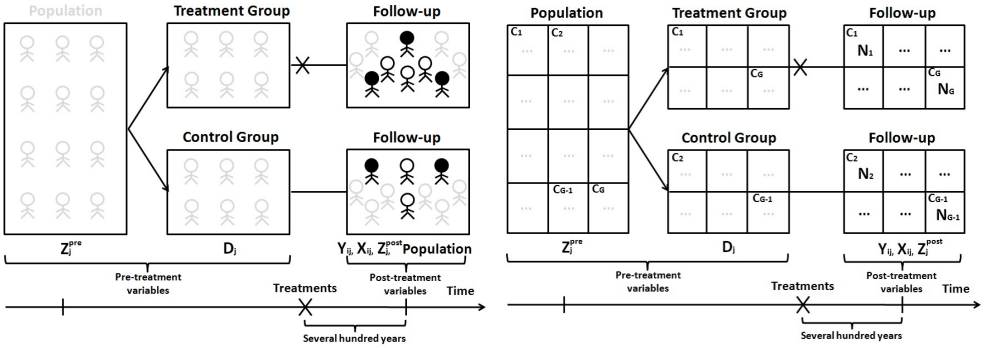
Let us first look at the standard approach used in existing studies. These studies examine historical events such as colonial institutions and Africa's slave trade (e.g., Dell (2010), Nunn and Wantchekon (2011); for reviews, see Nunn (2009, 2014)). Here we assume a historical event (treatment) to be the protection of property rights in a specific country during the colonial era and examine the long-run impacts on contemporary individual income.[4]

To identify the causal effects, researchers may estimate something like the following regression equation under linearity and constant treatment effect assumptions:

$$Y_{ij} = \alpha + \rho D_j + X_{ij}^{'}\beta + Z_j^{'}\gamma + \epsilon_{ij}, \tag{1}$$

where Y_{ij} is the income of individual i in region j; D_j is the treatment variable, which takes the value 1 if individual i lives in region j where property rights were protected during the colonial era and 0 otherwise; X_{ij} is a vector of relevant individual characteristics; Z_j is a vector of relevant regional characteristics; ρ is the parameter of interest.

[4]See, e.g., Acemoglu et al. (2005) for the importance of the protection of property rights for economic development. While some studies with aggregated data use historical events as instruments for the determinants of current domestic institutions (e.g., Acemoglu et al. (2001)), we consider the direct impacts of a historical event on contemporary individual outcomes. For convenience, we assume no missing values on the income and other relevant variables.



A. The Standard Approach B. An Alternative Approach

Figure 1: Two Approaches to Causal Inference for Historical Persistence

One major concern about the identification of ρ is the endogeneity of the protection of property rights, namely, that the exogenous assumption, $\epsilon_{ij} \perp D_j | X_{ij}, Z_j$, may not hold. Existing studies address such endogeneity problems on the variable of interest through quasi-experimental designs (e.g., instrumental variable strategies, regression discontinuity designs) with a limited subsample, which may satisfy the assumption.

Another concern is obtaining a valid inference. Since the variable of interest, D_j, varies only at the regional level, not the individual level, researchers may be concerned that the failure to account for the presence of common group errors generates estimated standard errors dramatically biased downward (Moulton (1986)). To correct the standard errors, researchers may use cluster-robust standard errors for inference, allowing for arbitrary correlation among the errors, ϵ_{ij}, within each region.

B. Remarks on the Standard Approach

The standard approach, however, faces some fundamental problems. We illustrate the problems in Figure 1A, which depicts the time frame for the evaluation. Unlike the evaluation of standard social programs such as job training

(e.g., LaLonde (1986)), the evaluation of historical events involves a long time span, the duration of which often exceeds several hundred years (i.e., easily exceeds a human's life expectancy). This innate distinctive feature can lead to the following conceptual and econometric issues.

Population. First, population is defined for contemporary individual units (not those having been directly exposed to the treatments, unlike the evaluation of standard social programs). In this context, since all individual units (denoted by N) emerge *after* the treatments, *their existence* is potentially affected by the treatments (and subsequent various factors such as intermediate variables or mediators, most of which are unobserved to researchers). Specifically, the distribution of the population (and samples drawn from the population) can be affected by the treatments. Thus, the choice of individual units as causal units generally introduces post-treatment bias (detailed below).

Covariate Selection. Second, covariate selection in empirical works is often guided by relevant economic theories and relevant previous research findings (e.g., Cameron and Trivedi (2005)). Such covariate selection, however, does not always apply in this context. This is because the variables selected for covariates in other studies are often post-treatment variables for which the adjustments generally introduce post-treatment bias (Rosenbaum (1984) or "bad control" (Angrist and Pischke (2009)); historical impacts often go beyond the scope of the relevant economic theories and empirical studies. In equation (1), the variables of individual characteristics, X_{ij}, are post-treatment variables. The variables of regional characteristics, Z_j, may include not only pre-treatment variables, Z_j^{pre}, but also post-treatment variables, Z_j^{post}. The adjustments for X_{ij} and Z_j^{post} in the regression generally introduce post-treatment bias.

SUTVA. Third, since the treatment assignment varies at the cluster (re-

gional) level, we require two additional assumptions regarding SUTVA: (1) no interference between clusters and (2) intact clusters (Hong and Raudenbush (2006)). In this context, the validity of the latter assumption may be of concern. For example, some individuals in a control (treatment) region might actually have been born in a treatment (control) region, spent some time there, and then migrated to the control (treatment) region. In this case, the assumption of no multiple versions of treatment may be violated. Also, potential outcomes of an individual might be affected by others with different versions of treatment through social interactions.

Inference. Fourth, since the existence of contemporary individual units can be affected by the treatments, the nature of dependence within regions (within-cluster correlation) can also be affected by the treatments. Thus, the estimated cluster-robust standard errors can be affected by the treatments.

C. The Identification Problems

This subsection more formally considers the identification problems in using microdata of individuals to identify the causal effects of history on their outcomes. Here we only consider the problems for our choice of individual units as causal units. We do not consider the problems for our adjustments for post-treatment variables. This is because the latter problem has long been recognized and discussed in the existing literature (e.g., Rosenbaum (1984), Wooldridge (2005), Angrist and Pischke (2009), Elwert and Winship (2014), Acharya et al. (2016), Montgomery et al. (2017)) but the former has not. For simplicity, we assume that no pre-treatment variables are recorded.

First, in the causal inference, to avoid having an ill-defined population (discussed above), we specify a larger population consisting of all *potential* individuals (i.e., superpopulation, denoted by M), each of whom can potentially

exist at the time of the "follow-up survey" (depicted in Figure 1A). Its distribution is not affected by the treatments (or the mediators). Thus, it is feasible to apply the potential outcomes framework to those potential individuals.

In this case, an ideal experiment (hypothetical one) would be that all these potential individuals (superpopulation) are randomly assigned to the treatment or control group and that all of them are actually observed at the time of the follow-up survey. We let $Y_{ij}(D_j)$ denote the potential income of potential individual i in region j given D_j. The parameter of our interest is the average treatment effect (ATE) for the superpopulation,

$$\tau = E[Y_{ij}(1) - Y_{ij}(0)] = E[Y_{ij}(1)] - E[Y_{ij}(0)].$$

However, what we can actually observe is quite different, as follows. Let $S_{ij}(D_j)$ denote the potential outcome for the existence of potential individual i in region j given D_j;[5] for example, $S_{ij}(1) = 1$ ($S_{ij}(0) = 1$) implies that potential individual i in region j would actually exist when assigned treatment (control) while $S_{ij}(1) = 0$ ($S_{ij}(0) = 0$) implies that the individual would not. Then, what we actually observe is the average observed difference between individuals who actually existed in the treatment and control groups,

$$\hat{\triangle} = E[Y_{ij}^{obs}|S_{ij}^{obs} = 1, D_j = 1] - E[Y_{ij}^{obs}|S_{ij}^{obs} = 1, D_j = 0].$$

Here again we use the superscript "obs" to distinguish between potential outcomes, which are not always observed, and the observed outcome.

We note that this comparison is problematic if the treatments affect the existence of potential individuals. For example, if a historical event significantly affected the survival of ancestors in such a way that poor, less educated, or

[5]To keep the notation simple, we do not explicitly specify the mediators. Here we assume that SUTVA holds.

low-ability people or people in poor health were more likely to die, then the observed and unobserved characteristics of contemporary individuals (i.e., the descendants of survivors) that may affect the outcomes of interest can systematically differ between the treatment and control groups. Importantly, this *sample selection problem* (e.g., Heckman (1979)) can arise even when the historical event is randomized (at the regional level).

To identify the ATE based on a finite sample of observed individuals, we would require the assumption that the historical event is completely randomized (at the contemporary individual level):

$$(Y_{ij}(1), Y_{ij}(0), S_{ij}(1), S_{ij}(0)) \perp D_j.$$

This assumption implies that the historical event has no systematic effect on the existence of potential individuals. However, in reality, whether or not each individual exists does depend on his or her treatment assignment and/or subsequent various factors; we *cannot* always define $Y_{ij}(1)$ and $Y_{ij}(0)$ for all potential individuals.

To see more details about this, let us follow the idea of the principal stratification approach (Frangakis and Rubin (2002)), which allows us to classify the individuals who actually existed ($S_{ij}^{obs} = 1$) into the following six groups according to the joint values of the two potential existence indicators:

- $EE = \{i : S_{ij}(1) = S_{ij}(0) = 1\}$, those who would exist regardless of their treatment assignment; both $Y_{ij}(1)$ and $Y_{ij}(0)$ are defined in $\mathbb{R}$ (the set of real numbers);
- $EN = \{i : S_{ij}(1) = 1 \text{ and } S_{ij}(0) = 0\}$, those who would exist if assigned treatment but would not exist if assigned control; $Y_{ij}(1) \in \mathbb{R}$ and $Y_{ij}(0) = *$;[6]

[6]For $S_{ij}(D_j) = 0$, we define the outcome as "$*$," following the notation used in the

- $?N = \{i : S_{ij}(1) = 1 \text{ or } 0 \text{ and } S_{ij}(0) = 0\}$, those who may or may not exist depending on subsequent situations (mediators) when assigned treatment (they would not exist if assigned control); $Y_{ij}(1) \in \mathbb{R}$ or $Y_{ij}(1) = *$ and $Y_{ij}(0) = *$;
- $NE = \{i : S_{ij}(1) = 0 \text{ and } S_{ij}(0) = 1\}$, those who would exist if assigned control but would not exist if assigned treatment; $Y_{ij}(1) = *$ and $Y_{ij}(0) \in \mathbb{R}$;
- $N? = \{i : S_{ij}(1) = 0 \text{ and } S_{ij}(0) = 1 \text{ or } 0\}$, those who may or may not exist depending on subsequent situations (mediators) when assigned control (they would not exist if assigned treatment); $Y_{ij}(1) = *$ and $Y_{ij}(0) \in R$ or $Y_{ij}(0) = *$;
- $?? = \{i : S_{ij}(1) = 1 \text{ or } 0 \text{ and } S_{ij}(0) = 1 \text{ or } 0\}$, those who may or may not exist depending on subsequent situations (mediators) under both treatment arms; $Y_{ij}(1) \in \mathbb{R}$ or $Y_{ij}(1) = *$ and $Y_{ij}(0) \in \mathbb{R}$ or $Y_{ij}(0) = *$.

Because principal strata are not affected by treatment assignment (although defined by a post-treatment variable), it is possible to identify the causal effects within each stratum. However, because causal effects are defined as comparisons of potential outcomes on a common set of units (e.g., Rubin (1974, 2005)), the individual-level causal effect is well defined on $\mathbb{R}$ only for the "EE" group.

In reality, we cannot directly observe the principal strata for the individuals because we cannot observe both $S_{ij}(1)$ and $S_{ij}(0)$ at the same time. We can only observe the following two groups based on the observed treatment assignment and the observed existence indicator ($OBS(D_j, S_{ij}^{obs})$):

- $OBS(1,1) = \{i : D_j = 1, S_{ij}^{obs} = 1\}$, those who existed in the treatment group;

truncation-by-death literature (e.g., Zhang and Rubin (2003)).

- $OBS(0,1) = \{i : D_j = 0, S_{ij}^{obs} = 1\}$, those who existed in the control group.

Each individual is observed to fall into one of the two groups but also belongs to an unobserved principal stratum. Their relationship is summarized in Table 1.

Table 1: Observed Data Pattern and Unobserved Principal Strata

$OBS(D_j, S_{ij}^{obs})$	D_j	S_{ij}^{obs}	Y_{ij}^{obs}	Unobserved Principal Strata
$OBS(1,1)$	1	1	$\in \mathbb{R}$	$EE, EN, ?N, ??$
$OBS(0,1)$	0	1	$\in \mathbb{R}$	$EE, NE, N?, ??$

Table 1 reveals that $OBS(1,1)$ and $OBS(0,1)$ consist of a mixture of the EE, EN, $?N$, ?? groups and the EE, NE, $N?$, ?? groups, respectively; these two groups involve different combinations of principal strata, suggesting that a comparison of the two outcomes is not an "apples-to-apples" comparison. Therefore, the average observed difference, $E[Y_{ij}^{obs}|S_{ij}^{obs} = 1, D_j = 1] - E[Y_{ij}^{obs}|S_{ij}^{obs} = 1, D_j = 0]$, is not the average causal effects, $E[Y_{ij}(1)] - E[Y_{ij}(0)]$.

To compare the outcomes for a common set of groups, which is a causal inference, one can assume that each individual would always exist regardless of his or her treatment assignment: $S_{ij}(D_j) = 1$ for all D_j. This existence assumption reduces the six principal strata only to EE and thus allows us to identify the causal effects for the EE group, where both $Y_{ij}(1)$ and $Y_{ij}(0)$ are well defined in $\mathbb{R}$. This assumption would also require that all potential individuals always exist. Thus, in this case, the population is identical to the superpopulation and its distribution is not affected by the treatment assignment (or mediators).

D. An Alternative Approach

The assumptions discussed above to justify the causal inference for historical

persistence with microdata cannot be empirically examined (even partially) regarding their validity. To avoid imposing such untenable assumptions, as a simple alternative approach, it might be better to make the causal inference with clusters or groups (i.e., regions), rather than individual units. This is because, in many cases, the requirements of the Rubin causal model are met: (1) The clusters or groups stably exist throughout a specified time frame and (2) the action of treatments and the measurement of outcomes take place on a common unit.[7] Although we are required to change the causal question of interest to that at the cluster or group level, we can identify the causal effects through a more transparent analysis, as follows.

Data Structure. Figure 1B depicts the data structure. The individual-level data in Figure 1A are now aggregated at the cluster (regional) level. We use C_j, where $j = 1, \ldots, G$, to denote each cluster, the size of which is denoted by N_j, where $\sum_{j=1}^{G} N_j = N$. We assume that regional-level pre-treatment variables, Z_j^{pre}, and post-treatment variables, Z_j^{post}, are available. Individual characteristics, X_{ij}, and outcome, Y_{ij}, are observed at the individual level. For simplicity, we assume that all individual units (who actually existed) within all clusters are sampled.

Essential Elements. The three essential elements of the Rubin causal model are as follows: (1) The population of units is G clusters, $U = \{C_1, \ldots, C_G\}$, (2) the set of treatments is $D_j = \{1, 0\}$, where $D_j = 1$ if cluster j protected property rights during the colonial era and $D_j = 0$ if cluster j did not, and (3) the response variable, Y_j, is regional income, which is defined using the individual-level outcome, Y_{ij}. We can also define various response variables using other post-treatment variables in Z_j^{post} and X_{ij}.

[7] In some cases, the formation of geographic units (e.g., state formation) may be affected by historical events (e.g., Alesina and Spolaore (2003)).

Assumptions. The three key assumptions (SUTVA, unconfoundedness, and overlap) are assumed to hold. The plausibility of the unconfoundedness and overlap assumptions is assumed to be improved by innovative quasi-experimental designs, as generally done in existing studies (see Nunn (2009, 2014)). Formally, the unconfoundedness assumption is described as

$$\big(Y_j(1), Y_j(0)\big) \perp D_j | Z_j^{pre}.$$

The overlap assumption is given as

$$0 < Pr(D_j = 1 | Z_j^{pre}) < 1.$$

Identification. Under the three assumptions, the conditional average treatment effect, $E[Y_j(1) - Y_j(0)|Z_j^{pre}]$, can be identified as follows:

$$\begin{aligned} E[Y_j(1) - Y_j(0)|Z_j^{pre}] &= E[Y_j(1)|Z_j^{pre}] - E[Y_j(0)|Z_j^{pre}] \\ &= E[Y_j(1)|Z_j^{pre}, D_j = 1] - E[Y_j(0)|Z_j^{pre}, D_j = 0] \\ &= E[Y_j^{obs}|Z_j^{pre}, D_j = 1] - E[Y_j^{obs}|Z_j^{pre}, D_j = 0]. \end{aligned}$$

Estimation. For simplicity, we suppose that the treatment effect is constant and the outcome is linear in D_j and Z_j^{pre}, where Z_j^{pre} is a K-dimensional column vector. Provided $G > K + 2$, we estimate the following regression equation to identify the average treatment effect:

$$\overline{Y_j} = \alpha + \rho D_j + Z_j^{pre'} \gamma + \epsilon_j, \quad \text{where } \overline{Y_j} = \frac{\sum_{i=1}^{N_j} Y_{ij}}{N_j}. \tag{2}$$

Here we consider the group average, $\overline{Y_j}$, as the dependent variable. The unconfoundedness assumption implies that the exogenous assumption, $\epsilon_j \perp D_j | Z_j^{pre}$, holds and the clusters are assumed to be independent of each other. Under the assumptions, the unweighted between-groups estimator consistently estimates the average treatment effect.

Here let us touch on the difference between this approach and that proposed in Donald and Lang (2007). Both approaches estimate between-groups estimators despite individual-level data being available, specifically, when an outcome variable varies among individual units and the variable of interest varies only at the cluster level. The two approaches have different motivations. Our motivation is simple: The population of our interest is clusters or groups, not individual units. In contrast, their motivation is to obtain a valid inference in the context of cluster sampling with a small number of clusters, as is typically the case for difference-in-differences estimation. Their population of interest is still individual units, not clusters or groups. They are motivated by the cluster-robust inference being valid when the number of clusters is large (Hansen (2007)) but not when it is small. To solve the inference problem, they propose estimating the between-groups estimator (see Donald and Lang (2007), Wooldridge (2010, Chapter 20) for details).

III. An Empirical Example

This section provides an empirical example based on the alternative approach as well as the standard one using the influential Dell (2010) paper, which follows the potential outcomes framework for studying historical persistence at the micro level. Dell examines the long-run impacts of the *mita*, an extensive forced mining labor system the Spanish government instituted in Peru and Bolivia between 1573 and 1812, on contemporary individual outcomes. Focusing on a sharp change in the *mita* boundary, she uses a regression discontinuity (RD) approach to examine the historical persistence.[8] Although she also examines the underlying mechanisms, we focus on estimation of the causal effects

[8]See, e.g., Imbens and Lemieux (2008), Lee and Lemieux (2010) for a description of regression discontinuity designs using the potential outcomes framework.

of the *mita* on current living standards (equivalent household consumption in 2001 and the prevalence of stunting among children aged 6-9 in 2005).

Estimation. We additionally estimate the following regression equation:

$$\begin{aligned} \overline{c_{db}} = \alpha + \gamma mita_d &+ X_d'\beta + f(geographic\ location_d) + \phi_b + \epsilon_{db}, \\ \text{where } \overline{c_{db}} &= \frac{\sum_{i=1}^{N_d} c_{idb}}{N_d}. \end{aligned} \tag{3}$$

In this estimation, unlike Dell's (2010) approach, we use clusters (districts) as the causal units. The population of our interest is districts, not individual units: Our interest is in the causal effects of the *mita* on current living standards for districts, not individuals. The clusters are assumed to meet the requirements of the Rubin causal model.

N_d denotes the number of individual units in district d. $\overline{c_{db}}$ is the mean outcome for district d along segment b of the *mita* boundary. $mita_d$ is an indicator variable equal to 1 if district d contributed to the *mita* and 0 otherwise. X_d is a vector of covariates that includes elevation and slope for district d. $f(geographic\ location_d)$ is the RD polynomial, which controls for smooth functions of geographic location. ϕ_b is a set of boundary segment fixed effects. The descriptive statistics are presented in Appendix Table A1; see Dell (2010) for detailed data information.

Assumptions. The first key identifying assumption in the RD approach is that all relevant factors are continuous at the *mita* boundary. Based on Dell's careful work in checking the validity of the assumption (see Dell (2010, Section 3.2)), we assume that the smoothness assumption holds.[9] The second key identifying assumption is that the functional form of the regression model is correct. Because she considers various functional forms regarding the RD

[9]Although Dell uses individuals as causal units, she considers the validity of the smoothness assumption for district-level pre-treatment variables.

polynomial, we simply follow her three specifications: (1) cubic polynomial in latitude and longitude, (2) cubic polynomial in distance to Potosí (km), and (3) cubic polynomial in distance to the *mita* boundary (km).

An additional assumption to validate the RD design is no selective sorting around the *mita* boundary. Since we use geographic units (clusters) as causal units, it is plausible to assume that the no-manipulation assumption holds. If we consider this assumption from the individual-unit point of view, the assumption is relevant to the individual units having been directly exposed to the treatments, not the contemporary individual units. For that reason, Dell (2010), using contemporary individual units as causal units, provides an unusual discussion regarding the validity of the assumption,[10] which is not generally found in the standard regression discontinuity design literature (e.g., Imbens and Lemieux (2008), Lee and Lemieux (2010)).

Results. Table 2 reports the estimated impacts of the *mita* on equivalent household consumption (panel A) and prevalence of stunting in children aged 6-9 (panel B). We report the estimates based on individual-level data in columns 1-3 and district-level data in columns 4-6. Panels A/B-1, -2, and -3 report the estimates based on the three different specifications mentioned above. In the former estimation, based on our discussions above, to justify the causal inference, we simply assume that the population is identical to the superpopulation: Potential individuals always existed.

[10] "(A)n additional assumption often employed in RD is no selective sorting across the treatment threshold. This would be violated if a direct *mita* effect provoked substantial outmigration of relatively productive individuals, leading to a larger indirect effect. Because this assumption may not be fully reasonable, I do not emphasize it. Rather, I explore the possibility of migration as an interesting channel of persistence, to the extent that the data permit." (Dell (2010, p. 1876)). This implies that the assumption of intact clusters regarding SUTVA might be violated, although it seems that migration was low.

Table 2: Impacts of the *Mita* on Living Standards

	A. Log Equivalent Household Consumption (2001)					
Units:	Households			Districts		
Sample within:	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.
	(1)	(2)	(3)	(4)	(5)	(6)
	A-1. Cubic Polynomial in Latitude and Longitude					
Mita	-0.282	-0.217	-0.335	-0.166	-0.115	-0.192
	(0.201)	(0.210)	(0.220)	(0.196)	(0.217)	(0.236)
R-squared	0.059	0.059	0.068	0.391	0.370	0.413
	A-2. Cubic Polynomial in Distance to Potosí					
Mita	-0.337***	-0.308***	-0.330***	-0.339***	-0.300***	-0.318***
	(0.088)	(0.102)	(0.098)	(0.092)	(0.102)	(0.103)
R-squared	0.046	0.035	0.045	0.276	0.214	0.283
	A-3. Cubic Polynomial in Distance to *Mita* Boundary					
Mita	-0.278***	-0.232**	-0.225**	-0.295***	-0.230**	-0.223**
	(0.079)	(0.090)	(0.093)	(0.089)	(0.098)	(0.102)
R-squared	0.044	0.041	0.038	0.277	0.249	0.194
Clusters	71	60	52	71	60	52
Observations	1,478	1,161	1,013	71	60	52
	B. Children Aged 6-9 Having Stunted Growth (2005)					
Units:	Children			Districts		
	B-1. Cubic Polynomial in Latitude and Longitude					
Mita	0.070	0.084*	0.087*	-0.012	-0.008	-0.021
	(0.043)	(0.046)	(0.048)	(0.025)	(0.027)	(0.029)
R-squared	0.051	0.020	0.017	0.388	0.298	0.211
	B-2. Cubic Polynomial in Distance to Potosí					
Mita	0.080***	0.078***	0.078***	0.046***	0.031*	0.025
	(0.021)	(0.022)	(0.024)	(0.016)	(0.016)	(0.018)
R-squared	0.049	0.017	0.013	0.330	0.261	0.156
	B-3. Cubic Polynomial in Distance to *Mita* Boundary					
Mita	0.073***	0.061***	0.064***	0.047***	0.025*	0.021
	(0.023)	(0.022)	(0.023)	(0.015)	(0.015)	(0.018)
R-squared	0.040	0.015	0.013	0.293	0.236	0.147
Clusters	289	239	185	289	239	185
Observations	158,848	115,761	100,446	289	239	185

Notes: The table reports ordinary least squares (OLS) estimates where the unit of observation is the household (child) in columns 1-3 and the district in columns 4-6. Robust standard errors, adjusted for clustering by district, are reported in parentheses in columns 1-3 and robust standard errors are reported in parentheses in columns 4-6. The dependent variable in panel A is log equivalent household consumption in columns 1-3 and the district mean of log equivalent household consumption in columns 4-6. The dependent variable in panel B is an indicator variable equal to 1 if the child has stunted growth and 0 otherwise in columns 1-3 and the district mean of children aged 6-9 having stunted growth in columns 4-6. *Mita* is an indicator variable equal to 1 if the (household's/child's) district contributed to the *mita* and 0 otherwise. Panel A/B-1 includes a cubic polynomial in the latitude and longitude of the observation's district capital. Panel A/B-2 includes a cubic polynomial in Euclidean distance (km) from the observation's district capital to Potosí. Panel A/B-3 includes a cubic polynomial in Euclidean distance (km) to the nearest *mita* boundary. All regressions include controls for elevation, slope, and boundary segment fixed effects. The sample in columns 1 and 4 includes observations whose district capitals are located within 100 km of the *mita*; this threshold is reduced to 75 km in columns 2 and 5 and 50 km in columns 3 and 6. *** = Significant at the 1% level. ** = Significant at the 5% level. * = Significant at the 10% level.

The main findings are summarized as follows. First, the estimates in columns 1-3 in panel A differ slightly from those in Dell (2010) (columns 1-3 of Table 2). This is because we avoid adjusting for demographic variables (the number of infants, children, and adults in the household), which are post-treatment variables. However, because these variables are little affected by the *mita* (not reported),[11] the two results are quite similar in regard to the magnitude of the impacts and the level of statistical significance. We also perform the same exercises for the district-level regressions and find similar results,[12] implying that the adverse impacts of the *mita* are not driven by its potential effects on the post-treatment variables.

Second, the estimated impacts of the *mita* on the equivalent household consumption are similar in columns 1-3 and 4-6 (panel A). However, the estimated impacts of the *mita* on the prevalence of stunting in children differ between columns 1-3 and 4-6 (panel B). Unlike Dell's results, we find positive impacts for the specification of a cubic polynomial in latitude and longitude; these impacts are, however, not statistically significant. In addition, although negative impacts are found for the specifications of a cubic polynomial in distance to Potosí and the *mita* boundary, the estimated impacts become smaller when the sample is limited to that closer to the *mita* boundary. Also, the significant impacts vanish when the sample is limited to that within 50 km of the *mita* boundary. These results imply that more conservative discussions regarding the *mita* impacts might be needed because the untenable assumptions imposed in the analyses might be violated.

Although the population of interest as well as the units of analysis differ

[11] Significant impacts are found only for the number of children in the specification of a cubic polynomial in latitude and longitude. The results are available from the author upon request.

[12] The results are available from the author upon request.

between the estimations based on individual-level data and aggregated data, we touch on the reasons why the estimated impacts of the *mita* are similar in panel A, but different in panel B. First, we note that the two estimators are identical when all clusters have the same number of observations (see, e.g., Donald and Lang (2007) and Wooldridge (2010, Chapter 20) for related discussions). Given this, the following two facts mainly cause the different results: (1) The number of individual units is relatively similar in panel A, but quite different in panel B and (2) the covariate distribution for the specification of a cubic polynomial in latitude and longitude is sensitive to the difference in group size, while that for the specifications of a cubic polynomial in distance to Potosí and the *mita* boundary is not (see Appendix Table A1).

IV. Concluding Remarks

In this paper, we have raised issues inherent in the causal inference for historical persistence with microdata following the potential outcomes framework. When microdata are available, it is tempting to directly use such microdata for the analysis to utilize the information most effectively. However, in this distinct context, the choice of individual units as causal units generally introduces bias because their existence is potentially affected by the treatments. Also, covariate selection guided by relevant economic theories and empirical findings often contains post-treatment variables, which may introduce another potential bias. Using an empirical example, we have discussed a simple alternative approach to avoid such problems that makes the causal inference with clusters or groups, not individual units. The approach is coherent with the Rubin causal model.

The discussion presented here is relevant not only to the causal inference for historical persistence with microdata, but also to the long-run impacts

of treatments if the existence of the causal units of interest is potentially affected by treatments more generally (e.g., across generations). We believe our discussion can help in designing/analyzing future relevant observational and experimental studies/data and lead to more transparent research. We will also development a general framework to address such causal inference at the micro level in future work.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 2001, *91* (5), 1369–1401.

_ , _ , and _ , "Institutions as a Fundamental Cause of Long-Run Growth," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A, New York: North Holland, 2005, chapter 6, pp. 385–472.

Acharya, Avidit, Matthew Blackwell, and Maya Sen, "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects," *American Political Science Review*, 2016, *110* (3), 512–529.

Alesina, Alberto and Enrico Spolaore, *The Size of Nations*, Cambridge: The MIT Press, 2003.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press, 2009.

Barnow, Burt S., Glend G. Cain, and Arthur S. Goldberger, "Issues in the Analysis of Selectivity Bias," in Ernst W. Stromsdorfer and George

Farkas, eds., *Evaluation Studies Review Annual*, Vol. 5, California: Sage, 1980, pp. 43–59.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, New York: Cambridge University Press, 2005.

Dell, Melissa, "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 2010, *78* (6), 1863–1903.

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

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit," in T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics*, Vol. 4, Amsterdam and New York: North Holland, 2008, chapter 61, pp. 3895–3962.

Elwert, Felix and Christopher Winship, "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable," *Annual Review of Sociology*, 2014, *40*, 31–53.

Fisher, Ronald A., *The Design of Experiments*, 1st ed., London: Oliver and Boyd, 1935.

Frangakis, Constantine E. and Donald B. Rubin, "Principal Stratification in Causal Inference," *Biometrics*, 2002, *58* (1), 21–29.

Hansen, Christian B., "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T is Large," *Journal of Econometrics*, 2007, *141* (2), 597–620.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

_ **, Robert J. LaLonde, and Jerey A. Smith**, "The Economics and Econometrics of Active Labor Markets Programs," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Elsevier, 1999, chapter 31, pp. 1865–2097.

Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986, *81* (396), 945–960.

_ **and Donald B. Rubin**, "Causal Inference in Retrospective Studies," *Evaluation Review*, 1988, *12* (3), 203–231.

Hong, Guanglei and Stephen W Raudenbush, "Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data," *Journal of the American Statistical Association*, 2006, *101* (475), 901–910.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, New York: Cambridge University Press, 2015.

_ **and Jeffery M. Wooldridge**, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86.

_ **and Thomas Lemieux**, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

LaLonde, Robert J., "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 1986, *76* (4), 604–620.

Lechner, Michael, "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption," in Michael Lechner and Friedhelm Pfeiffer, eds., *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica, 2001, pp. 43–58.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Lietrature*, 2010, *48* (2), 281–355.

Manski, Charles F., Gary D. Sandefur, Sara McLanahan, and Daniel Powers, "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," *Journal of the American Statistical Association*, 1992, *87* (417), 25–37.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres, "How Conditioning Post-Treatment Variables Can Ruin Your Experiment and What To Do About It," 2017. unpublished manuscript.

Moulton, Brent R., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 1986, *32* (3), 385–397.

Neyman, Jerzy, "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.," Translated in *Statistical Science*, 1923, *5* (4), 465–480, 1990.

Nunn, Nathan, "The Importance of History for Economic Development," *Annual Review of Economics*, 2009, *1* (1), 65–92.

_ , "Historical Development," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 2A, New York: North Holland, 2014, chapter 7, pp. 347–402.

_ **and Leonard Wantchekon**, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 2011, *101* (7), 3221–3252.

Rosenbaum, Paul R., "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society. Series A (General)*, 1984, *147* (5), 656–666.

_ **and Donald B. Rubin**, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

Rubin, Donald B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 1974, *66* (5), 688–701.

_ , "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 1977, *2* (1), 1–26.

_ , "Discussion of "Randomization Analysis of Experimental Data in the Fisher Randomization Test" by Basu," *Journal of the American Statistical Association*, 1980, *75* (371), 591–593.

_ , "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers," *Journal of the American Statistical Association*, 1986, *81* (396), 961–962.

_ , "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, 2005, *100* (469), 322–331.

Wooldridge, Jeffrey M., "Violating Ignorability of Treatment by Controlling for Too Many Factors," *Econometric Theory*, 2005, *21* (5), 1026–1028.

_ , *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge: MIT Press, 2010.

Zhang, Junni L. and Donald B. Rubin, "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death"," *Journal of Educational and Behavioral Statistics*, 2003, *28* (4), 353–368.

For Online Publication

Table A1: Descriptive Statistics

	A. Log Equivalent Household Consumption (2001)					
Units:	Households			Districts		
Sample Within:	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Number of households				20.817	19.350	19.481
				(12.312)	(9.251)	(9.373)
Log equivalent	5.877	5.799	5.848	5.839	5.805	5.848
household consumption	(1.010)	(0.915)	(0.855)	(0.401)	(0.362)	(0.358)
Mita	0.752	0.716	0.674	0.718	0.700	0.654
	(0.432)	(0.451)	(0.469)	(0.453)	(0.462)	(0.480)
Elevation	3.841	3.824	3.827	3.792	3.786	3.794
	(0.378)	(0.389)	(0.383)	(0.408)	(0.400)	(0.391)
Slope	7.130	8.319	8.548	7.784	8.615	8.742
	(4.124)	(3.699)	(3.649)	(4.106)	(3.771)	(3.798)
Longitude	-0.335	0.046	0.106	-0.105	0.123	0.132
	(1.203)	(0.921)	(0.777)	(1.110)	(0.885)	(0.767)
Latitude	-0.054	-0.340	-0.412	-0.202	-0.393	-0.447
	(0.820)	(0.638)	(0.578)	(0.765)	(0.621)	(0.586)
Longitude^2	1.559	0.849	0.614	1.226	0.785	0.594
	(1.742)	(0.900)	(0.491)	(1.473)	(0.822)	(0.510)
Latitude^2	0.675	0.522	0.503	0.618	0.534	0.537
	(0.533)	(0.368)	(0.335)	(0.488)	(0.387)	(0.363)
Longitude*Latitude	-0.617	-0.252	-0.113	-0.435	-0.242	-0.133
	(1.071)	(0.689)	(0.527)	(0.883)	(0.652)	(0.535)
Longitude^3	-1.956	-0.142	0.188	-1.008	0.046	0.221
	(4.497)	(2.041)	(0.836)	(3.724)	(1.794)	(0.839)
Latitude^3	0.163	-0.200	-0.284	-0.015	-0.242	-0.309
	(1.004)	(0.584)	(0.479)	(0.893)	(0.600)	(0.535)
Longitude^2*Latitude	0.900	-0.012	-0.188	0.385	-0.096	-0.195
	(2.407)	(1.147)	(0.519)	(1.874)	(1.030)	(0.540)
Longitude*Latitude^2	-0.451	0.014	0.114	-0.156	0.063	0.122
	(1.384)	(0.703)	(0.416)	(1.062)	(0.659)	(0.441)
Distance to Potosí	8.964	9.484	9.587	9.262	9.586	9.632
	(1.450)	(1.036)	(0.814)	(1.300)	(0.983)	(0.815)
Distance to Potosí^2	82.453	91.017	92.570	87.450	92.836	93.430
	(24.911)	(18.779)	(15.357)	(22.829)	(18.004)	(15.372)
Distance to Potosí^3	775.093	882.584	899.934	839.133	907.258	912.200
	(328.449)	(259.657)	(218.859)	(307.124)	(251.264)	(219.335)
Distance to *mita* bound.	0.406	0.281	0.233	0.380	0.290	0.243
	(0.286)	(0.174)	(0.126)	(0.263)	(0.170)	(0.127)
Distance to *mita* bound.^{2}	0.247	0.109	0.070	0.213	0.113	0.075
	(0.288)	(0.122)	(0.065)	(0.262)	(0.118)	(0.066)
Distance to *mita* bound.^{3}	0.181	0.051	0.024	0.147	0.052	0.026
	(0.264)	(0.079)	(0.031)	(0.239)	(0.075)	(0.031)
Bound. segm. dummy I	0.086	0.078	0.083	0.099	0.083	0.077
	(0.280)	(0.269)	(0.276)	(0.300)	(0.279)	(0.269)
Bound. segm. dummy II	0.289	0.138	0.100	0.197	0.100	0.077
	(0.453)	(0.345)	(0.300)	(0.401)	(0.303)	(0.269)
Bound. segm. dummy III	0.384	0.477	0.484	0.451	0.517	0.519
	(0.487)	(0.500)	(0.500)	(0.501)	(0.504)	(0.505)
Observations	1,478	1,161	1,013	71	60	52

Continue

Table A1: Descriptive Statistics

	B. Children Aged 6-9 Having Stunted Growth (2005)					
Units:	Children Aged 6-9			Districts		
Sample Within:	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.	< 100 km of Bound.	< 75 km of Bound.	< 50 km of Bound.
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Number of children				549.647	484.356	542.951
				(1365.381)	(678.517)	(736.596)
Children having	0.346	0.391	0.403	0.380	0.392	0.412
stunted growth	(0.476)	(0.488)	(0.491)	(0.125)	(0.120)	(0.116)
Mita	0.780	0.707	0.684	0.702	0.678	0.665
	(0.414)	(0.455)	(0.465)	(0.458)	(0.468)	(0.473)
Elevation	3.911	3.908	3.896	3.864	3.899	3.908
	(0.388)	(0.427)	(0.411)	(0.482)	(0.475)	(0.447)
Slope	6.414	7.724	7.890	8.021	8.210	8.245
	(3.917)	(3.489)	(3.478)	(3.800)	(3.585)	(3.591)
Longitude	-0.547	-0.149	-0.080	0.014	0.009	-0.003
	(1.225)	(0.903)	(0.798)	(1.077)	(0.921)	(0.808)
Latitude	0.017	-0.312	-0.418	0.029	-0.037	-0.154
	(0.822)	(0.636)	(0.552)	(0.763)	(0.744)	(0.688)
Longitude^2	1.800	0.837	0.643	1.157	0.845	0.650
	(1.846)	(0.880)	(0.600)	(1.266)	(0.902)	(0.639)
Latitude^2	0.676	0.501	0.479	0.580	0.553	0.494
	(0.584)	(0.441)	(0.393)	(0.538)	(0.499)	(0.405)
Longitude*Latitude	-0.694	-0.187	-0.089	-0.285	-0.168	-0.018
	(1.159)	(0.662)	(0.549)	(0.822)	(0.732)	(0.604)
Longitude^3	-2.451	-0.280	0.007	-0.372	-0.151	-0.033
	(4.842)	(1.898)	(1.097)	(3.160)	(1.993)	(1.206)
Latitude^3	0.230	-0.187	-0.301	0.213	0.117	-0.063
	(1.050)	(0.677)	(0.527)	(0.901)	(0.831)	(0.648)
Longitude^2*Latitude	1.186	-0.013	-0.172	0.310	0.175	0.059
	(2.606)	(1.047)	(0.636)	(1.529)	(1.158)	(0.734)
Longitude*Latitude^2	-0.657	-0.037	0.039	-0.136	-0.080	0.016
	(1.491)	(0.698)	(0.499)	(0.926)	(0.806)	(0.569)
Distance to Potosí	8.735	9.301	9.434	9.204	9.247	9.320
	(1.486)	(1.022)	(0.858)	(1.201)	(1.035)	(0.847)
Distance to Potosí^2	78.513	87.559	89.738	86.156	86.566	87.580
	(25.386)	(18.606)	(16.060)	(21.373)	(18.555)	(15.410)
Distance to Potosí^3	723.445	833.387	860.361	818.516	819.519	829.211
	(333.028)	(257.969)	(227.659)	(290.890)	(253.693)	(213.211)
Distance to *mita* bound.	0.447	0.292	0.244	0.417	0.326	0.243
	(0.299)	(0.183)	(0.141)	(0.273)	(0.202)	(0.144)
Distance to *mita* bound.^{2}	0.290	0.119	0.079	0.248	0.147	0.079
	(0.302)	(0.127)	(0.076)	(0.263)	(0.146)	(0.074)
Distance to *mita* bound.^{3}	0.218	0.057	0.030	0.173	0.076	0.029
	(0.277)	(0.083)	(0.037)	(0.238)	(0.099)	(0.035)
Bound. segm. dummy I	0.097	0.109	0.109	0.225	0.230	0.249
	(0.296)	(0.311)	(0.311)	(0.418)	(0.422)	(0.433)
Bound. segm. dummy II	0.342	0.156	0.093	0.225	0.192	0.119
	(0.475)	(0.362)	(0.290)	(0.418)	(0.395)	(0.325)
Bound. segm. dummy III	0.278	0.348	0.369	0.329	0.310	0.297
	(0.448)	(0.476)	(0.483)	(0.471)	(0.463)	(0.458)
Observations	158,848	115,761	100,446	289	239	185

Notes: Panels A and B present the means and standard deviations for the variables used in the regressions with equivalent household consumption and children aged 6-9 having stunted growth, the latter of which are reported in parentheses. The unit of observation is the household (child) in columns 1-3 and the district in columns 4-6. The sample in columns 1 and 4 includes observations from those whose district capitals are located within 100 km of the *mita*; this threshold is reduced to 75 km in columns 2 and 5 and 50 km in columns 3 and 6.