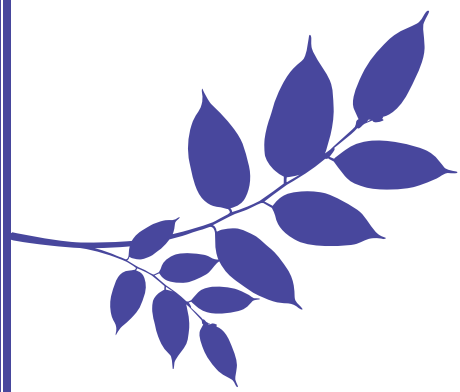
Some Remarks on the Causal Inference for Historical Persistence

Katsuo Kogure



Japan-ASEAN Transdisciplinary Studies Working Paper Series No.3
January 2019

Some Remarks on the Causal Inference for Historical Persistence *

Katsuo Kogure†

January 20, 2019

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 empirical issues inherent in the causal inference following the potential outcomes framework. Using an empirical example, we also illustrate a simple alternative approach to avoid these issues that is coherent with the potential outcomes framework.

JEL Codes: B52, N01, C18

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 Grants-in-Aid for Scientific Research No.15K17044, Japan Society for the Promotion of Science.

†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.

1. Introduction

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 empirical issues inherent in the causal inference following the Rubin causal model, a framework for causal inference based on potential outcomes (Rubin (1974, 1977), Imbens and Rubin (2015)). The discussion presented 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 adopt the potential outcomes framework because it enables us to raise most effectively the issues inherent in the causal inference for historical persistence with microdata and many modern empirical works that examine causal questions adopt this framework (see, e.g., Angrist and Pischke (2009), Duflo et al. (2008), Imbens and Wooldridge (2009), Imbens and Rubin (2015)) (we do not adopt the econometric approach to causal modelling (see, e.g., Heckman and Vytlacil (2007), Heckman (2008)), because historical impacts often go beyond the scope of relevant microeconomic theories and the existing studies do not follow it (see Nunn (2009, 2014))). Using an empirical example, we also illustrate a simple alternative approach with aggregated data to avoid certain issues that is coherent with the Rubin causal model.

The paper proceeds as follows. For expository purposes, Section 2 provides a brief description of the Rubin causal model, specifying the major premises. Section 3 considers the causal inference for historical persistence with contemporary micro-

data and discusses fundamental problems inherent in the causal inference. We also discuss a simple alternative approach to address these problems. Section 4 provides an empirical example based on the approach and Section 5 concludes.

2. 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)),[1]

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

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,

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

[1]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)).

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)).

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:

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

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 (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 non-compliance and "truncation-by-death" (e.g., Frangakis and Rubin (2002), Zhang and Rubin (2003))). Holland (1986) discusses the 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 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).

3. Causal Inference for Historical Persistence

Following the potential outcomes framework, this section considers the causal inference for historical persistence with contemporary microdata.

3.1 The Standard Approach

Let us first look at the standard approach used in existing studies. These studies consider historical events such as colonial institutions and Africa's slave trade (e.g., Dell (2010), Nunn and Wantchekon (2011)). 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 (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 following a standard regression framework (many existing studies follow it (see Nunn (2009, 2014))). For convenience, we assume no missing values on the income and other relevant variables.

To examine the causal relationship, 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.

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.

3.2 Remarks on the Standard Approach

The standard approach, however, may face 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 empirical 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 including the characteristics of ancestors and previous regional characteristics (i.e., intermediate variables (mediators) prior to the emergence of the individual units), most of which are unobserved to researchers). In other words, the distribution of the population (and samples drawn from the population) can be affected by the treatments (and intermediate variables). Thus, with an ill-defined population, it is not feasible to apply the potential outcomes framework to those individual units and 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 findings in related literature (e.g., Cameron and Trivedi (2005)). However, such covariate selection may not always apply in this context, 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, the validity of SUTVA may also 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 treat-

ment may be violated. In addition, potential outcomes of an individual might be affected by others with different versions of treatment through social interactions. The assumption of no interference among individual units may also be violated.

Inference. Fourth, with an *ill-defined* population, statistical inference for parameters of the population may not be valid, because it is premised on a *well-defined* population; the use of cluster-robust standard errors for inference may be irrelevant.

3.3 The Identification Problems

We now more formally consider the identification problems following the potential outcomes framework. Here we 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 problems have long been recognized and discussed in the existing literature (e.g., Rosenbaum (1984), Wooldridge (2005), Lechner (2008), Angrist and Pischke (2009), Elwert and Winship (2014), Huber (2015), Acharya et al. (2016), Montgomery et al. (2018)) but the former have 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). Importantly, its distribution is not affected by the treatments (and intermediate variables). Thus, it is feasible to apply the potential outcomes framework to those potential individuals.

An ideal experiment (hypothetical one) would be that all potential individuals in the 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. Let $Y_{ij}(D_j)$ denote the potential income of potential individual i in region j given D_j. The pa-

rameter of our interest is the average treatment effect (ATE) for the superpopulation,

$$\tau_{sp} = E[Y_{ij}(1) - Y_{ij}(0)] = E[Y_{ij}(1)] - E[Y_{ij}(0)],$$

(we use the subscript "sp" ("fs" (given below)) for the superpopulation (finite sample)). However, what we can actually observe (in the real world) 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. For the sake of brevity, we assume that the existence of each potential individual depends only on his or her treatment assignment, although it potentially depends on both his or her treatment assignment and intermediate variables. SUTVA is assumed to hold. $S_{ij}(1) = 1$ ($S_{ij}(0) = 1$) 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,

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

We again 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 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 depends on his or her treatment assignment and we *cannot* always define $Y_{ij}(1)$ and $Y_{ij}(0)$ for all potential individuals.

To see more details about this, we follow the idea of the principal stratification approach, a statistical framework for making causal inferences with intermediate variables, developed to address problems regarding noncompliance, censoring-by-death, and surrogate outcomes (Frangakis and Rubin (2002)). The approach first partitions individual units into *latent* classes or subgroups ("principal strata") defined by the joint potential values of an intermediate variable (e.g., a censoring indicator variable) under both the observed and the counterfactual treatment conditions and then makes causal inferences within the principal strata. The key idea is that because principal strata are not affected by treatment assignment, it is possible to make a causal inference within each stratum; the stratum variable can be treated as a pre-treatment covariate. This appraoch generalizes the "local average treatment effect" (LATE) framework (Imbens and Angrist (1994), Angrist et al. (1996)); a causal effect within a principal stratum can be interpreted as a LATE.

In our case, we can classify the individuals who actually existed ($S_{ij}^{obs} = 1$) into the following three latent subgroups 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 $Y_{ij}(0) = *$ (for $S_{ij}(D_j) = 0$, we define the outcome as "$*$" ("missing" or "censored") following the truncation-by-death literature (e.g., Zhang and Rubin (2003)));
- $NE = \{i : S_{ij}(1) = 0 \text{ and } S_{ij}(0) = 1\}$, those who would exist if assigned control but would not exist if assigned treatment; $Y_{ij}(1) = *$ and $Y_{ij}(0) \in \mathbb{R}$.

As discussed above, since principal strata are not affected by the treatment assignment (although defined by a post-treatment variable), it is possible to make a causal inference 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;
- $OBS(0,1) = \{i : D_j = 0, S_{ij}^{obs} = 1\}$, 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 reveals that $OBS(1,1)$ and $OBS(0,1)$ consist of a mixture of the EE, EN groups and the EE, NE 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, but an "apples-to-oranges" 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 three 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. In this case, the population is identical to the superpopulation and its distribution is not affected by the treatment assignment: The population is well defined for contemporary individual units.

3.4 An Alternative Approach

The assumptions discussed above to justify the causal inference cannot be empirically examined regarding their validity. To avoid imposing such untenable assumptions, this subsection considers a simple alternative approach that makes the causal inference with clusters or groups (i.e., regions), rather than individual units. 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.[2] 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: The unit of analysis is cluster, not individual. We use C_j, where $j = 1, \ldots, G$, to denote each cluster, the size

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

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. While individual characteristics, X_{ij}, and outcome, Y_{ij}, are still observed at the individual level, the aggregated data are used for analysis. 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}.

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 (it is noted that the number of individual units at each cluster (regional) level is potentially affected by the treatments).

Here we 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; unlike our case, the population is assumed to be well defined for individual units. 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).

4. 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 explicitly 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 (see, e.g., Imbens and Lemieux (2008), Lee and Lemieux (2010) for a description of regression discontinuity designs using the potential outcomes framework). Although she also examines the underlying mechanisms, we focus on estimation of the causal effects 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:

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

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 (although Dell uses individuals as causal units, she considers the validity of the smoothness assumption for district-level pre-treatment variables). 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 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,[3] which is not generally found in the standard regression discontinuity design literature

[3] "(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 out-migration 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 SUTVA (discussed above) might be violated, although it seems that migration was low.

(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.

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* (columns 1-3 of Appendix Table A2),[4] 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 (columns 4-6 of Appendix Table A2), 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

[4]Significant impacts are found only for the number of children in the specification of a cubic polynomial in latitude and longitude (panel B).

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 careful consideration 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 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).

5. 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 illustrated 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 develop a general framework to address such causal inference at the micro level in our future work.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1A, Chapter 6, pp. 385–472. New York: North Holland.

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review 110*(3), 512–529.

Alesina, A. and E. Spolaore (2003). *The Size of Nations*. Cambridge: The MIT Press.

Angrist, J. D., G. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*, 444–472.

Angrist, J. D. and J. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Barnow, B. S., G. G. Cain, and A. S. Goldberger (1980). Issues in the analysis of selectivity bias. In E. W. Stromsdorfer and G. Farkas (Eds.), *Evaluation Studies Review Annual*, Volume 5, pp. 43–59. California: Sage.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Dell, M. (2010). The persistent effects of peru's mining *mita*. *Econometrica 78*(6), 1863–1903.

Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics 89*(2), 221–233.

Duflo, E., R. Glennerster, and M. Kremer (2008). Using randomization in development economics research: A toolkit. In T. P. Schultz and J. Strauss (Eds.), *Handbook of Development Economics*, Volume 4, Chapter 61, pp. 3895–3962. Amsterdam and New York: North Holland.

Elwert, F. and C. Winship (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology 40*, 31–53.

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics 58*(1), 21–29.

Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics 141*(2), 597–620.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. (2008). Econometric causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 70, pp. 4780–4874. Amsterdam and Oxford: Elsevier.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Holland, P. W. and D. B. Rubin (1988). Causal inference in retrospective studies. *Evaluation Review 12*(3), 203–231.

Huber, M. (2015). Causal pitfalls in the decomposition of wage gaps. *Journal of Business & Economic Statistics 33*(2), 179–191.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

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

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*(4), 604–620.

Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M. Lechner and F. Pfeiffer (Eds.), *Econometric Evaluation of Labour Market Policies*, pp. 43–58. Heidelberg: Physica.

Lechner, M. (2008). A note on endogenous control variables in causal studies. *Statistics & Probability Letters 78*(2), 190–195.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Lietrature 48*(2), 281–355.

Manski, C. F., G. D. Sandefur, S. McLanahan, and D. Powers (1992). Alternative estimates of the effect of family structure during adolescence on high school graduation. *Journal of the American Statistical Association 87*(417), 25–37.

Montgomery, J. M., B. Nyhan, and M. Torres (2018). How conditioning on post-treatment variables can ruin your experiment and what to do about it. *American Journal of Political Science 62*(3), 760–775.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*(3), 385–397.

Nunn, N. (2009). The importance of history for economic development. *Annual Review of Economics 1*(1), 65–92.

Nunn, N. (2014). Historical development. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 2A, Chapter 7, pp. 347–402. New York: North Holland.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in africa. *American Economic Review 101*(7), 3221–3252.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General) 147*(5), 656–666.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics 2*(1), 1–26.

Rubin, D. B. (1980). Discussion of "randomization analysis of experimental data in the fisher randomization test" by basu. *Journal of the American Statistical Association 75*(371), 591–593.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association 81*(396), 961–962.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association 100*(469), 322–331.

Wooldridge, J. M. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory 21*(5), 1026–1028.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge: MIT Press.

Zhang, J. L. and D. B. Rubin (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics 28*(4), 353–368.

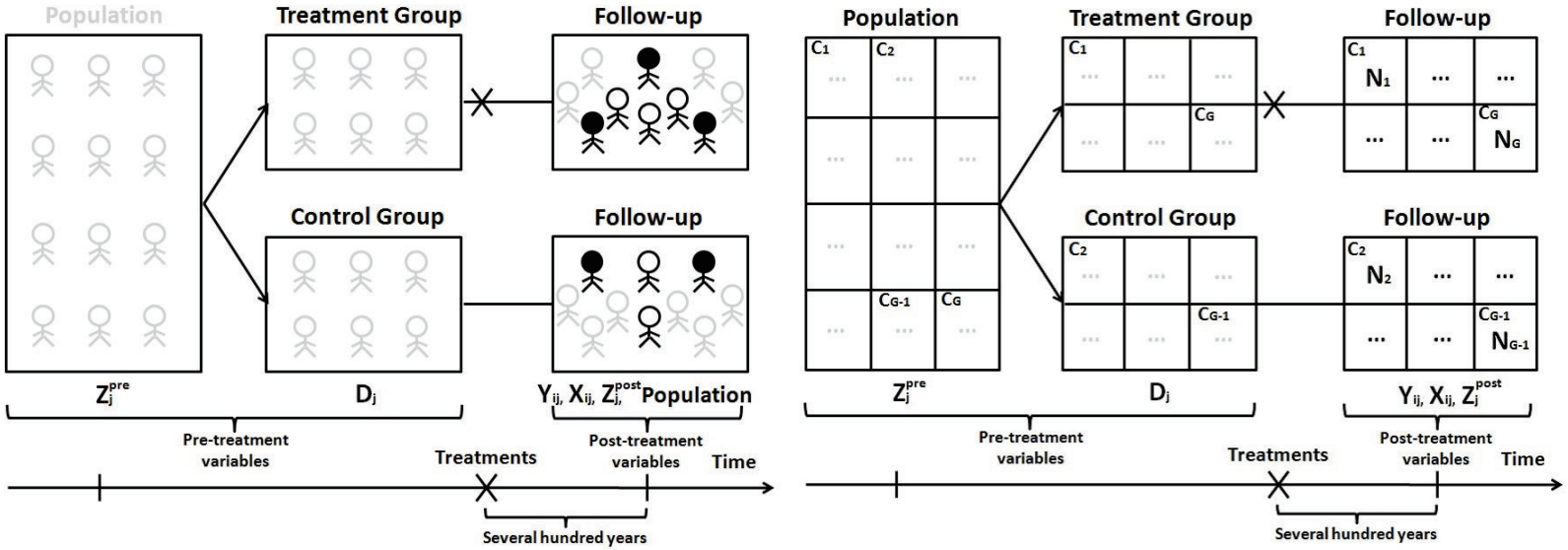
Table 1: Observed Data Pattern and Unobserved Principal Strata

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

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. *** $p < 0.01$; ** $p < 0.05$; and * $p < 0.1$.



(a) The Standard Approach

(b) An Alternative Approach

Figure 1: Two Approaches to Causal Inference for Historical Persistence

Appendix

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.
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.
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.

Table A2: Impacts of the *Mita* on Demographic Characteristics

Units:	A. Number of Infants (2001)					
	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	0.033	0.105	0.096	0.075	0.147	0.149
	(0.087)	(0.104)	(0.118)	(0.096)	(0.113)	(0.127)
R-squared	0.028	0.026	0.031	0.430	0.416	0.483
	A-2. Cubic Polynomial in Distance to Potosí					
Mita	0.020	0.014	0.015	0.041	0.044	0.030
	(0.050)	(0.059)	(0.057)	(0.056)	(0.062)	(0.063)
R-squared	0.023	0.018	0.024	0.343	0.303	0.376
	A-3. Cubic Polynomial in Distance to *Mita* Boundary					
Mita	0.018	0.013	0.020	0.034	0.033	0.041
	(0.050)	(0.054)	(0.054)	(0.054)	(0.060)	(0.062)
R-squared	0.021	0.019	0.023	0.341	0.308	0.305
Clusters	71	60	52	71	60	52
Observations	1,478	1,161	1,013	71	60	52
Units:	B. Number of Children (2001)					
	Households			Districts		
	B-1. Cubic Polynomial in Latitude and Longitude					
Mita	0.264*	0.326**	0.340*	0.290*	0.377*	0.435**
	(0.138)	(0.157)	(0.175)	(0.169)	(0.196)	(0.209)
R-squared	0.038	0.028	0.032	0.537	0.501	0.527
	B-2. Cubic Polynomial in Distance to Potosí					
Mita	0.055	0.074	0.072	0.014	0.046	0.062
	(0.084)	(0.104)	(0.110)	(0.095)	(0.116)	(0.130)
R-squared	0.026	0.017	0.021	0.338	0.300	0.339
	B-3. Cubic Polynomial in Distance to *Mita* Boundary					
Mita	0.029	0.075	0.076	-0.002	0.049	0.053
	(0.077)	(0.089)	(0.090)	(0.089)	(0.102)	(0.104)
R-squared	0.027	0.019	0.019	0.362	0.314	0.298
Clusters	71	60	52	71	60	52
Observations	1,478	1,161	1,013	71	60	52

Continue

Table A2: Impacts of the *Mita* on Demographic Characteristics

	C. Number of Adults (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)
	C-1. Cubic Polynomial in Latitude and Longitude					
Mita	-0.100	0.086	0.164	0.105	0.299	0.343
	(0.188)	(0.203)	(0.217)	(0.233)	(0.229)	(0.233)
R-squared	0.026	0.026	0.045	0.214	0.279	0.452
	C-2. Cubic Polynomial in Distance to Potosí					
Mita	-0.039	-0.015	-0.030	-0.011	0.017	-0.009
	(0.089)	(0.098)	(0.093)	(0.134)	(0.148)	(0.138)
R-squared	0.026	0.020	0.026	0.175	0.122	0.219
	C-3. Cubic Polynomial in Distance to *Mita* Boundary					
Mita	-0.054	-0.080	-0.070	-0.055	-0.110	-0.098
	(0.082)	(0.090)	(0.091)	(0.138)	(0.153)	(0.150)
R-squared	0.024	0.019	0.024	0.147	0.075	0.150
Clusters	71	60	52	71	60	52
Observations	1,478	1,161	1,013	71	60	52

Notes: The table reports OLS estimates where the unit of observation is the household 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/B/C is the number of infants/children/adults in the household in columns 1-3 and the district mean of the number of infants/children/adults in the households in the district in columns 4-6. *Mita* is an indicator variable equal to 1 if the (household's) district contributed to the *mita* and 0 otherwise. Panel A/B/C-1 includes a cubic polynomial in the latitude and longitude of the observation's district capital. Panel A/B/C-2 includes a cubic polynomial in Euclidean distance (km) from the observation's district capital to Potosí. Panel A/B/C-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. *** $p < 0.01$; ** $p < 0.05$; and * $p < 0.1$.