ESTIMATING INTERACTION EFFECTS WITH PANEL DATA

CHRIS MURIS AND KONSTANTIN M. WACKER

Abstract. A common task in empirical economics is to estimate interaction effects that measure how the effect of one variable $X$ on another variable $Y$ depends on a third variable $H$. This paper considers the estimation of interaction effects in linear panel models with a fixed number of time periods. There are at least two ways to estimate interaction effects in this setting, both common in applied work. Our theoretical results show that these two approaches are distinct, and only coincide under strong conditions on unobserved effect heterogeneity. Our empirical results show that the difference between the two approaches is large, leading to conflicting conclusions about the sign of the interaction effect. Taken together, our findings may guide the choice between the two approaches in empirical work.

1. Introduction

In empirical work in economics, we often want to estimate how the effect of one variable $X$ on another variable $Y$ depends on a third variable $H$. The standard approach to answering such a question with cross-sectional data is to estimate interaction effects based on a linear model $Y = \alpha + X\beta + XH\kappa + U$.\(^1\) In this case, $\kappa$ is called the interaction term coefficient, and it measures how the effect of $X$ on $Y$ varies with $H$.

With panel data, there are at least two ways to estimate interaction effects in linear models. The most common approach, which we will call the interaction term estimator (ITE), is based on a regression of $Y_{it}$ on $X_{it}$ and $X_{it}H_i$ (and additive fixed effects).\(^2\) The OLS estimate of the coefficient on $X_{it}H_i$ is the ITE for $\kappa$. A second approach, which we will call the correlated interaction term

\(^1\)Alternative approaches for modeling effect heterogeneity include those based on random coefficients (cf. Lewbel and Pendakur, 2017), and those based on tools from machine learning (see Athey and Imbens, 2015; Bordt, Farbmacher, and Kogel, 2020; Chernozhukov, Hansen, Liao, and Zhu, 2011; Epifani and Gancia, 2011; Herrera, Ordoñez, and Trebesch, 2020; List and Sturm, 2006; Manacorda and Tesei, 2020; Shambaugh, 2004; Spilimbergo, 2009; Storeygard, 2016).

\(^2\)See, for example, Alsan and Goldin (2019); Amiti and Konings (2007); Berman, Martin, and Mayer (2012); Bloom, Draca, and Van Reenen (2016); Bloom, Sadun, and Van Reenen (2012); Burnside and Dollar (2000); Duflo, Dupas, and Kremer (2011); Epifani and Gancia (2009); Herrera, Ordoñez, and Trebesch (2020); List and Sturm (2006); Manacorda and Tesei (2020); Shambaugh (2004); Spilimbergo (2009); Storeygard (2016).
estimator (CITE), consists of two steps. First, regress $Y_{it}$ on $X_{it}$ separately for each panel unit $i$, obtaining $b_i$. In a second step, project $b_i$ onto $H_i$ in a cross-sectional regression. The result of the second step is the CITE for $\kappa$. CITE has been used in applied work but appears to be less popular than ITE.\(^3\)

One goal of this paper is to show that these two approaches – ITE and CITE – are distinct, and will only recover the same object under strong conditions on unobserved effect heterogeneity. We derive conditions under which CITE is consistent for the interaction effect. These conditions are not sufficient to guarantee consistency of ITE. In two empirical applications, we show that ITE and CITE lead to conflicting conclusions about the sign of interaction effects. The difference in interaction effect estimates can be large. Based on our results, since CITE requires weaker assumptions on unobserved effect heterogeneity for consistency, we recommend it for typical applications in economics where there may be substantial, correlated unobserved effect heterogeneity.\(^4\) In most existing work, the choice between ITE and CITE is typically not made explicit, and is done without motivation. We believe that we are the first to provide a rigorous analysis of these two estimators.

Fundamental to our results is that, in applications where interaction effects matter, there is likely additional unobserved effect heterogeneity in the effect of $X$ on $Y$. To fix ideas, consider the following outcome equation, which is a special case of our general framework:\(^5\)

$$Y_{it} = \alpha_i + X_{it}\beta_i + \kappa X_{it}H_i + \epsilon_{it}. \quad (1)$$

In (1), $\beta_i$ captures unobserved effect heterogeneity in addition to the observed effect heterogeneity due to the interaction term $\kappa H_i$.

That the consistency of ITE relies on restrictive exogeneity conditions is evident from equation (1): the error term in the linear model underlying a regression of $Y$ on $X$ and $XH$ will have a composite error term $\epsilon_{it} + (\beta_i - \beta)X_{it}$. This error term is correlated with $X_{it}$ and $H_i$, unless the unobserved effect heterogeneity is unrelated to $X_{it}$ and $H_i$.

CITE, on the other hand, allows the unobserved effect heterogeneity $\beta_i$ to be arbitrarily correlated with $X_{it}$, by treating the $\beta_i$ as parameters. We show that it is consistent for the interaction effect under conditions that do not guarantee the consistency of ITE. Importantly, this approach does not require a large number of time periods. Our consistency result is obtained under fixed-$T$.

If $H_i$ is endogenous, neither ITE nor CITE recover a causal effect of $H_i$ on the effect of $X$ on $Y$. The best one can hope for is to recover the correlation between $\beta_i$ and $H_i$. CITE recovers this correlation, but ITE does not.

To demonstrate the consistency of CITE, we build on existing work from the literature on correlated random coefficient models for linear panel models (see Arellano and Bonhomme, 2012; Chamberlain, 1992; Graham and Powell, 2012; Laage, 2020; Sasaki and Ura, 2021). The object of

\(^3\)Existing papers that use this approach include Couttenier, Petrencu, Rohner, and Thoenig (2019) and MaCurdy (1981).

\(^4\)In a sense that will become clear later, this preference for CITE over ITE corresponds to the preference for fixed effects over (correlated) random effects approaches for linear panel models with additive unobserved heterogeneity.

\(^5\)Our general framework in Section 2 additionally has vector-valued $X_{it}$, additional controls, time-varying interaction variables, and additional sources of unobserved effect heterogeneity.
interest in this literature is (some feature of) the distribution of \( \beta_i \) in \( Y_{it} = X_{it} \beta_i + U_{it} \). Techniques from this literature can also be used in the context of difference-in-difference estimation (cf. de Chaisemartin and D’Haultfoeuille, 2022; Verdier, 2020). Our work differs from the papers in this literature because we are interested in the estimation of interaction effects, while the distribution of the random coefficients is a nuisance parameter.

Previous work on interaction effects has made mention of using CITE. However, we are not aware of existing work that demonstrates that ITE and CITE are distinct, and that analyzes their asymptotic properties.

Organization. Section 2 introduces our framework for the estimation of interaction effects with panel data. Section 3 formally defines the two estimators. Sections 4 and 5 present our theoretical results. Section 6 contains the results for two empirical applications. Appendix A contains proofs of the main results. Appendix B has additional data and details for the empirical applications.

2. Setup

We are interested in the estimation of interaction effects in a static linear panel model. Typically, the outcome equation for this purpose is specified as

\[
Y_{it} = X_{it} H_i \kappa + X_{it} G_{it} \phi + Z_{it} \gamma + U_{it},
\]

which allows the effect of the explanatory variable \( X_{it} \) on the dependent variable \( Y_{it} \) to depend on observable, time-invariant interaction variables \( H_i \) and observable, time-varying interaction variables \( G_{it} \). The parameters of interest are \( \kappa \), the interaction term coefficient (ITC) on \( H_i \), and \( \phi \), the ITC on \( G_{it} \).

Once we admit that the effect of \( X \) on \( Y \) may depend on the observable \( G \) and \( H \), we may wonder whether there are additional, unobserved, sources of effect heterogeneity. Our framework explicitly introduces unobserved heterogeneity in the effect of \( X \) on \( Y \) via the following three equations for cross-section unit \( i = 1, \ldots, n \) at time \( t = 1, \ldots, T \):

\[
Y_{it} = X_{it} \beta_{it} + Z_{it} \gamma + U_{it},
\]

\[
\beta_{itk} = \delta_{ik} + G_{it} \phi_k + V_{itk}, \quad k = 1, \ldots, K_x,
\]

\[
\delta_{i1} = H_i \kappa + \epsilon_i.
\]

The outcome equation (3) describes how \( Y_{it} \in \mathbb{R} \) responds to a change in \( X_{it} \in \mathbb{R}^{K_x} \), allowing for control variables \( Z_{it} \in \mathbb{R}^{K_z} \). The effect of \( X_{itk} \) on \( Y_{it} \), denoted \( \beta_{itk} \), can vary across \( i \) and \( t \).

The coefficient equation (4) describes the heterogeneity in the effect of \( X_{itk} \) on \( Y_{it} \). It decomposes the heterogeneous effect \( \beta_{itk} \) in:

(i) a part that depends on \( G_{it} \in \mathbb{R}^{K_g} \);

(ii) a time-invariant, unit-specific \( \delta_{ik} \in \mathbb{R} \); and

\footnote{Balli and Sørensen (2013) observe that unobserved effect heterogeneity can lead to inconsistency in the ITE, and suggest that “if the time-series dimension of the data is large, one may directly allow for country-varying slopes”. We show that such slopes should generally be preferred regardless of the time-series dimension. Giesselmann and Schmidt-Catran (2020) discuss CITE, but prefer their double-demeaned estimator, citing computational concerns, a concern that interaction effects involving time-invariant variables cannot be estimated, and observing that the minimum number of time periods is higher for CITE than for their ITE.}
(iii) an idiosyncratic effect heterogeneity $V_{itk}$.

The heterogeneity equation (5) relates the unit-specific effect heterogeneity in the effect of the first regressor, $\delta_{i1}$, to $H_i \epsilon_i \in \mathbb{R}^{K_i}$, so that $\epsilon_i$ captures the unobserved part of $\delta_{i1}$. We do not model the time-invariant coefficients on the other regressors, see Remark 1.

In our framework, the effect of $X$ on $Y$ may vary even when holding constant $G$ and $H$, due to the presence of unobserved effect heterogeneity $\epsilon_i$ and $V_{itk}$. If we shut down unobserved effect heterogeneity, i.e.

$$V_{itk} = \epsilon_i = 0 \text{ for all } i, t, k,$$

and if all observables are scalar, then we obtain the standard outcome equation (2). Thus, our framework is a natural generalization of the standard way of thinking about interaction effects in linear models that takes seriously the role of unobserved effect heterogeneity.

In a model with scalar $X, G, H, Z$, and with unobserved effect heterogeneity, the reduced form of (3)-(5) is

$$Y_{it} = X_{it} H_i \kappa + X_{it} G_{it} \phi + Z_{it} + (U_{it} + X_{it} V_{it} + X_{it} \epsilon_i).$$

It is clear from (6) that, under our setup, the error term $\hat{U}_{it}$ in (2) contains the unobserved effect heterogeneity terms. If those terms are correlated with the observables, a regression of $Y$ on $XG$ and $XH$ is inconsistent due to endogeneity.

**Remark 1.** Our specification allows for additive fixed effects in the outcome equation by setting $X_{it2} = 1$. Then $\delta_{i2}$ is the conventional additive fixed effect.

**Remark 2.** The index $t$ need not refer to time, but can refer to students $t$ for a given classroom $i$, counties $t$ within a given state $i$, employees $t$ within a given firm $i$, etc.

### 3. Two estimators

There are at least two estimators for the ITCs $\phi$ and $\kappa$ introduced in equations (3)-(5). We call them the interaction term estimator (ITE) and the correlated interaction term estimator (CITE). The ITE, defined formally in Section 3.2, is a regression of $Y$ on $(X, Z)$ augmented with interactions of $X$ with $G$ and $H$. The CITE, defined formally in Section 3.1, is a two-step estimator. The first step is a regression of $Y$ on $(X, Z)$ augmented with interactions of $X$ with $G$ and dummy variables for each $i$. The second step is a regression of the estimated coefficients on the dummy variable interaction terms on $H$.

Both approaches can accommodate additive fixed effects in the outcome equation by setting $X_{it2}$ equal to a constant.

As a particular example of our framework and the two estimators, consider the special case with scalar $X$ and $H$, no $Z$ and $G$, and $V_{it1} = 0$. Then our setup simplifies to

$$Y_{it} = X_{it} \delta_{i1} + U_{it},$$

$$\delta_{i1} = H_i \kappa + \epsilon_i,$$
with reduced form

\[ Y_{it} = (X_{it} H_i) \kappa + (U_{it} + X_{it} \epsilon_{it}) . \]  

Then ITE is the regression of \( Y \) on \( X \times H \) suggested by (9). In contrast, CITE is based on (7) and (8): first, regress \( Y \) on \( X \) for each \( i \) to obtain \( \hat{\delta}_{i1} \). Then regress \( \hat{\delta}_{i1} \) on \( H_i \) to obtain \( \kappa \).

In the remainder of this section, we formally define the two estimators. From now on, we will assume access to a random sample consistent with the framework in Section 2.

Assumption 1 (Random sampling). For each \( i = 1, \cdots, n \), the observed data is \( W_i = (Y_{i1}, \cdots, Y_{iT}, X_{i1}, \cdots, X_{iT}, G_{i1}, \cdots, G_{iT}, Z_{i1}, \cdots, Z_{iT}, H_i) \), generated by equations (3)–(5). \( \{W_i\}_{i=1}^n \) is an i.i.d. random sequence.

3.1. Correlated interaction term estimator (CITE). Substituting (4) into (3) obtains the reduced form

\[
Y_{it} = \sum_k X_{itk} \beta_{itk} + Z_{it\gamma} + U_{it} \\
= \sum_k X_{itk} (\delta_{ik} + G_{it\phi_k} + V_{itk}) + Z_{it\gamma} + U_{it} \\
= X_{it} \delta_i + \Psi_{it} \theta + \zeta_{it},
\]

where we have introduced the following objects (dimensions in square brackets):

\[
\delta_i = (\delta_{i1}, \cdots, \delta_{iK}) \quad [K_x \times 1], \\
\Psi_{it} = (X_{it} \otimes G_{it}, Z_{it}) \quad [1 \times (K_x K_g + K_z)], \\
\theta = (\phi_1, \cdots, \phi_K, \gamma) \quad [(K_x K_g + K_z) \times 1], \\
\zeta_{it} = U_{it} + \sum_k V_{itk} X_{itk} \quad [1 \times 1].
\]

For a given \( i \), collecting this relationship across \( t = 1, \cdots, T \) obtains

\[
Y_i = X_i \delta_i + \Psi_i \theta + \zeta_i,
\]

where \( Y_i, X_i, \Psi_i, \) and \( \zeta_i \) are the counterparts of the objects in (10), with \( T \) rows.

For the CITE to be well-defined, we assume sufficient variation in \( X_i \).

Assumption 2. There exists a \( h > 0 \) such that

\[
\inf_i \det \left( X'_i X_i \right) \geq h.
\]

Assumption 2 avoids the identification issues in Graham and Powell (2012) and Arellano and Bonhomme (2012). If it does not hold for all \( i \), our analysis should be thought of as conditioning on a sub-population for which it does hold, cf. Arellano and Bonhomme (2012). Assumption 2 allows us to define the residual maker matrix

\[
M_i = I - X_i \left( X'_i X_i \right)^{-1} X'_i.
\]
and write

\[ M_i Y_i = M_i X_i \delta_i + M_i \Psi_i \theta + M_i \zeta_i \]

\[ = M_i \Psi_i \theta + M_i \zeta_i, \]

using that \( M_i X_i = 0 \). We also require sufficient variation in \( \Psi_i \) after projecting out \( X_i \).

**Assumption 3.** The matrix \( E \left( \Psi_i' M_i \Psi_i \right) \) is invertible.

Assumptions 1, 2 and 3 guarantee that CITE for \( \theta \) is well-defined for large enough \( n \).

**Definition 1.** CITE for \( \theta \) is given by

\[ \hat{\theta}_n = \left( \sum_{i=1}^{n} \Psi_i' M_i \Psi_i \right)^{-1} \sum_{i=1}^{n} \Psi_i' M_i Y_i. \]

CITE for \((\phi, \gamma)\) can be extracted from \( \hat{\theta}_n \). For \( \kappa \), we need variation in \( H_i \).

**Definition 2.** Then the CITE for \( \kappa \) is

\[ \hat{\kappa}_n = \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \delta_{i,1}. \]

**Remark 3.** In the special case that \( K_x = 1 \) and \( X_{i,1} = 1 \), CITE is the linear fixed effects estimator based on

\[ Y_{it} = \delta_{i,1} + Z_{it} \gamma + U_{it}. \]

In contrast, ITE is a correlated random effects estimator, see Remark 4.

### 3.2 Interaction term estimator (ITE)

Define \( X_{i,-1} \) to be \( X_i \) with the first column removed, and define \( \delta_{i,-1} \) analogously. Then rewrite (11) as

\[ Y_i = X_i \delta_i + \Psi_i \theta + \zeta_i \]

\[ = X_{i,-1} \delta_{i,-1} + X_{i,1} \delta_{i,1} + \Psi_i \theta + \zeta_i \]

\[ = X_{i,-1} \delta_{i,-1} + X_{i,1} (H_i \kappa + \epsilon_i) + \Psi_i \theta + \zeta_i \]

\[ = X_{i,-1} \delta_{i,-1} + \underbrace{X_{i,1} H_i \kappa + \Psi_i \theta}_{\Psi_i \hat{\theta}} + \underbrace{\zeta_i + X_{i,1} \epsilon_i}_{\hat{\zeta}}. \]

The ITE requires sufficient variation in \( X_{i,-1} \).

**Assumption 5.** There exists a \( h > 0 \) such that

\[ \inf_i \det \left( X_{i,-1}' X_{i,-1} \right) \geq h. \]
This assumption is similar to assumption 5, but is slightly weaker because of excluding the first column of \( X_i \). Under assumption 5, we can define
\[
M_{i,-1} = I - X_{i,-1} (X'_{i,-1} X_{i,-1})^{-1} X'_{i,-1} \]
and premultiply (13) by \( M_{i,-1} \) to obtain
\[
M_{i,-1} Y_i = M_{i,-1} X_{i,-1} \delta_{i,-1} + M_{i,-1} \tilde{\Psi}_i \tilde{\theta} + M_{i,-1} \tilde{\zeta}_i \]
(14)

**Assumption 6.** The matrix \( E(\tilde{\Psi}_i M_{i,-1} \tilde{\Psi}_i) \) is invertible.

Assumption 6 guarantees sufficient variation in \( \tilde{\Psi}_i \) after projecting out \( X_{i,-1} \). Given Assumptions 5 and 6, the ITE is well-defined.

**Definition 3.** The ITE for \( \tilde{\theta} = (\kappa, \phi, \gamma) = (\kappa, \theta) \) is
\[
\tilde{\theta}_n \equiv \left( \sum_{i=1}^n \tilde{\Psi}_i M_{i,-1} \tilde{\Psi}_i \right)^{-1} \sum_{i=1}^n \tilde{\Psi}_i M_{i,-1} Y_i. 
\]

ITE estimates for \( \kappa, \phi \) and \( \gamma \) can be extracted from \( \tilde{\theta}_n \).

**Remark 4.** In the special case that \( K_x = 1 \) and \( X_{i1} = 1 \), the ITE is a correlated random effects estimator (cf. Mundlak (1978) and Wooldridge (2010), 10.7.3). It approximates the unobserved heterogeneity by \( \delta_{i1} = H_i \kappa + \epsilon_i \), so that
\[
Y_{it} = \delta_{i1} + Z_{it} \gamma + U_{it} = H_i \kappa + Z_{it} \gamma + (U_{it} + \epsilon_i). 
\]
Consistency of the ITE would require assumptions about orthogonality of \( \epsilon \) and \( (H, Z) \).

**Remark 5.** In the special case that \( K_x = 2 \) and \( X_{i2} = 1 \), we have a linear fixed effects model with interaction terms. With a scalar \( H \) and no \( G \),
\[
Y_{it} = \delta_{i2} + X_{it} H_i \kappa + (U_{it} + X_{it} V_{it1} + X_{it} \epsilon_i). 
\]
In this case, the transformation \( M_{i,-1} \) is the within transformation that produces
\[
\tilde{Y}_{it} = Y_{it} - \frac{1}{T} \sum_{s=1}^T Y_{is}, \quad \tilde{X}_{it} = X_{it} - \frac{1}{T} \sum_{s=1}^T X_{is} 
\]
and the ITE is obtained from linear regression of \( \tilde{Y}_{it} \) on \( H_i \times \tilde{X}_{it} \).

### 4. Main Results

We now show that a set of exogeneity conditions that is sufficient for the consistency of CITE (Section 4.2) is not sufficient for the consistency of ITE (Section 4.3).

These results are derived without exogeneity restrictions involving \( \epsilon_i \). This allows the heterogeneity equation to be misspecified. We discuss this modeling choice and its consequences in Section 4.1. In Section 5, we provide results under correct specification.
4.1. **Misspecification.** Throughout this section, we will not make exogeneity assumptions that involve $\epsilon_i$.\(^{7}\) This leaves room for *misspecification of the heterogeneity equation*, in the sense that we may have

\[
E[\epsilon_i | H_i] \neq 0.
\]

Such misspecification arises if relevant variables are omitted from the heterogeneity equation. To see this, let there be a vector $H_i^*$ such that $\delta_{i1} = H_i^* \kappa + \epsilon_i$ and $E[\epsilon_i^* | H_i^*] = 0$. If the researcher only uses (or has access to) a partial list $H_i \subsetneq H_i^*$ and uses the alternative specification

\[
\delta_{i1} = H_i \kappa + \epsilon_i,
\]

then generally $E[\epsilon_i | H_i] \neq 0$. Misspecification also arises if $H_i$ is measured with error, if there is functional form misspecification, etc.

Our results below show that CITE is robust against such misspecification, but that ITE is not.

Under misspecification of the heterogeneity equation, $\kappa$ is too ambitious of a target even if $\delta_{i1}$ were known. Consider instead the infeasible regression of $\delta_{i1}$ on $H_i$, which tends to the *population projection coefficient* of $\delta_{i1}$ on $H_i$,

\[
\tilde{\kappa} = \left( E\left[H_i' H_i\right]\right)^{-1} E\left[H_i' \delta_{i1}\right]
\]

Then generally $E[\epsilon_i | H_i] \neq 0$. Misspecification also arises if $H_i$ is measured with error, if there is functional form misspecification, etc.

4.2. **Consistency of CITE.** Our analysis proceeds under the following exogeneity assumption, where $V_i$ collects all the coefficient equation error terms $V_{itk}$.

**Assumption 7** (Strict exogeneity). *The error terms in equations (3) and (4) satisfy*

\[
0 = E( U_{it} | X_{it}, Z_{it}, G_{it}, H_i) \\
E(V_{it} | X_{it}, Z_{it}, G_{it}, H_i).
\]

This assumption is similar to the strict exogeneity assumption that is standard in literature on correlated random coefficient panel models.\(^{8}\) It is restrictive, but not much stronger than what is necessary for accommodating additive fixed effects in a linear model with fixed $T$. Our version of strict exogeneity requires $H$ — along with $(X, Z, G)$ — to be orthogonal to the error terms in the outcome and coefficient equations. This is not necessary for estimation of $\phi$ and $\gamma$, which would only require orthogonality with respect to $(X, Z, G)$, not $H$.

We first state the result for $\theta = (\phi, \gamma)$.

\(^{7}\)See Section 5 for an analysis under correct specification.

\(^{8}\)See Arellano and Bonhomme (2012); Chamberlain (1992); Graham and Powell (2012). A notable exception is Laage (2020).
Theorem 1 (Consistency of CITE for $\theta$). If Assumptions 1, 2, 3, and 7 hold, and if 

$$E \|M_i Y_i\|^2 < \infty, \ E \|M_i \Psi_i\|^2 < \infty$$

then as $n \to \infty$, $\hat{\theta}_n \xrightarrow{P} \theta$.

Proof. See Appendix A. \qed

The second result is for $\kappa$. In what follows, $e'_1$ is the column vector of length $K_h$, with the first element equal to 1, and all other elements equal to zero.

Theorem 2 (Consistency for $\kappa$). Under the conditions of Theorem 1, and if Assumption 4 holds, and if (i) $E \|H_i\|^2 < \infty$, (ii) $E \|\delta_{1i}\|^2 < \infty$, (iii) $E \|H_i e'_1 (X'_i X_i)^{-1} X'_i \Psi_i\| < \infty$, and (iv) $E \|H_i e'_1 (X'_i X_i)^{-1} X'_i \kappa\| < \infty$, then, as $n \to \infty$, $\hat{\kappa} \xrightarrow{P} \kappa$.

Proof. See Appendix A. \qed

Remark 6. Theorem 2 adapts (Arellano and Bonhomme, 2012, Corollary 1) to the present context. Note that our Assumption 2 implies that we can drop their conditioning on $S$. Then, set their $F_i = H_i$, their $y_i = Y_i$, their $\gamma_i = \delta_{1i}$, their $Z_i = \Psi_i$, their $\delta = \theta$.

4.3. Inconsistency of the ITE. Without assumptions on $\epsilon_i$, the ITE is inconsistent. To show this, we consider the special case with $X \in \mathbb{R}$, $H \in \mathbb{R}^2$, and without $G, Z$. We consider a setting with omitted variables: the ITE in this section uses the first element of $H$ but not the second. Other forms of endogeneity lead to similar conclusions. For example, letting $H_{i2} = H^2_{i1}$ in what follows is functional form misspecification. The case of measurement error in $H_{i1}$ is similar.

Partition $H_i = (H_{i1}, H_{i2})$ and rewrite the model equations for the special case under consideration:

$$Y_{it} = X_{it1} \beta_{it1} + U_{it}$$
$$\beta_{it1} = \delta_{1i} + V_{it1}$$
$$\delta_{1i} = H_{i1} \kappa + \epsilon_i,$$
$$= H_{i1} \kappa_1 + H_{i2} \kappa_2 + \epsilon_i.$$

Assume strict exogeneity for the full set of covariates $H_i$

$$0 = E (U_i | X_i, Z_i, G_i, H_{i1}, H_{i2}) ,$$
$$0 = E (V_i | X_i, Z_i, G_i, H_{i1}, H_{i2})$$

and additionally assume that the heterogeneity equation is correctly specified if both $H_{i1}$ and $H_{i2}$ are included, i.e.

$$0 = E (\epsilon_i | X_i, Z_i, G_i, H_{i1}, H_{i2}).$$

These assumptions are stronger than necessary for consistency of CITE (using $H_{i1}$ only) for $\kappa_1$,

the projection coefficient of $\delta_{1i}$ on $H_{i1}$, which for this special case equals:

$$\kappa_1 = \kappa_1 + \left( E [H_{i1}^2] \right)^{-1} E [H_{i1} H_{i2}] \kappa_2.$$

9This follows from Theorem 2.
However, they are not sufficient for consistency for \( \tilde{\kappa}_1 \) of the ITE that uses only \( H_{i1} \),

\[
\tilde{\kappa}_1 = \left( \sum_i \sum_t X_{it1}^2 H_{i1}^2 \right)^{-1} \left( \sum_i \sum_t X_{it1} H_{i1} Y_{it} \right)
\]

\[
= \kappa_1 + \left( \sum_i \sum_t X_{it1}^2 H_{i1}^2 \right)^{-1} \left( \sum_i \sum_t X_{it1} H_{i1} (X_{it1} (H_{i2} \kappa_2 + \epsilon_i + V_{it1}) + U_{it}) \right)
\]

To see that it is not consistent under the maintained assumptions, assume that the relevant laws of large numbers apply, so that

\[
\tilde{\kappa}_1 \xrightarrow{p} \kappa_1 + \left( \sum_i \sum_t E [X_{it1}^2 H_{i1}^2] \right)^{-1} \left( \sum_i \sum_t E [X_{it1} H_{i1} (X_{it1} (H_{i2} \kappa_2 + \epsilon_i + V_{it1}) + U_{it})] \right)
\]

\[
= \kappa_1 + \left( \sum_i \sum_t E [X_{it1}^2 H_{i1}^2] \right)^{-1} \left( \sum_i \sum_t E [X_{it1}^2 H_{i1} H_{i2} \kappa_2] \right)
\]

\[
\neq \kappa_1 + (E [H_{i1}^2])^{-1} E[H_{i1} H_{i2} \kappa_2].
\]

\[
= \tilde{\kappa}_1.
\]

This shows that CITE with \( H_{i1} \) converges to \( \tilde{\kappa}_1 \), whereas ITE does not. From the expression of the probability limit of \( \tilde{\kappa}_1 \), it is clear that the inconsistency can be made arbitrarily large.

5. Results under correct specification

The results in the previous section were derived under misspecification. We now add the assumption of correct specification to the model.

**Assumption 8.** The heterogeneity equation is correctly specified,

\[
(18) \quad E (\epsilon_i | H_i) = 0.
\]

This assumption is likely too strong for most empirical applications in economics, see Section 4.1. Together, Assumptions 7 and 8 require that all variables are exogeneous in the outcome and coefficient equation, and that \( H_i \) is exogenous in the heterogeneity equation. Assumption 8 does not require \( \epsilon_i \) to be orthogonal to the other regressors \((X, G, Z)\).

If the heterogeneity equation is correctly specified, then CITE is consistent for the causal parameter \( \kappa \) (Theorem 3) but ITE is not (Section 5.2). We also discuss stronger conditions that restore consistency of ITE (Section 5.3).

5.1. Consistency of CITE. Theorem 1 applies without modification: CITE is consistent for \((\phi, \gamma)\). Under correct specification, we additionally have that CITE is consistent for \( \kappa \), instead of for the projection coefficient \( \tilde{\kappa} \).

**Theorem 3** (Consistency of DVAITE for \( \kappa \)). If the conditions for Theorem 2 hold, and if assumption 8 holds, then, as \( n \to \infty \), \( \tilde{\kappa} \xrightarrow{p} \kappa \).

**Proof.** Under assumption 8, \( \tilde{\kappa} = \kappa \). Other than that, the proof is identical to that of Theorem 2. \( \square \)
5.2. Inconsistency of ITE. Assumption 8, in conjunction with Assumption 7, is likely too strong for most applications in economics. But it is not sufficient for consistency of ITE for \( \kappa \). Consider the special case that \( G, X, H \) are scalar, and that there are no \( Z \). Then the reduced form simplifies to:

\[
Y_{it} = \Psi_{it} \tilde{\theta} + (U_{it} + X_{it}\epsilon_i + X_{it}V_{it}),
\]

where

\[
\tilde{\Psi}_{it} = (X_{it}G_{it}, X_{it}H_{it}),
\]

and \( \tilde{\theta} = (\phi, \kappa) \). The ITE simplifies to

\[
\hat{\theta} = \left( \sum_i \sum_t \tilde{\Psi}_{it}' \tilde{\Psi}_{it} \right)^{-1} \sum_i \sum_t \tilde{\Psi}_{it}' Y_{it},
\]

so that, if an appropriate law of large numbers holds,

\[
\hat{\theta} - \tilde{\theta} \overset{p}{\rightarrow} \left( \sum_t E \left[ \tilde{\Psi}_{it}' \tilde{\Psi}_{it} \right] \right)^{-1} \sum_t E \left[ \tilde{\Psi}_{it}' (U_{it} + X_{it}V_{it} + X_{it}\epsilon_i) \right]
= \left( \sum_t E \left[ \tilde{\Psi}_{it}' \tilde{\Psi}_{it} \right] \right)^{-1} \sum_t E \left[ \tilde{\Psi}_{it}' X_{it}\epsilon_i \right].
\]

The equality follows from Assumptions 7 and 8, which eliminate the first two components of the composite error terms. However, the remaining component, \( X_{it}\epsilon_i \), is not orthogonal to \( \Psi_{it} = (X_{it}G_{it}, X_{it}H_{it}) \).

5.3. A sufficient condition for consistency of ITE. The strongest exogeneity assumptions we have imposed so far is

\[
0 = E(U_i | X_i, Z_i, G_i, H_i),
0 = E(V_i | X_i, Z_i, G_i, H_i),
0 = E(\epsilon_i | H_i).
\]

It does not restrict the distribution of \( \epsilon_i \) conditional on the time-varying observables. It is easy to show that the ITE is consistent if we impose the following stronger correlated random effects assumption:

\[
0 = E(U_i | X_i, Z_i, G_i, H_i)
= E(V_i | X_i, Z_i, G_i, H_i)
= E(\epsilon_i | X_i, Z_i, G_i, H_i).
\]

This requires \( \epsilon_i \) to be orthogonal to \( (X, G, Z) \) in addition to \( H \).

6. Empirical applications

We use two empirical applications to show that ITE and CITE can lead to meaningfully different conclusions about interaction effects.
Table 1. Point estimates and standard errors (in parentheses) for interaction effects in Stock and Watson’s “beer tax”. First column has results for ITE, second column has results for CITE. Time-varying interaction variables are the unemployment rate and the existence of minimum punishment. Time-invariant interaction variables are the proportion of a state’s population that is mormon/southern baptist. *** p < 0.01, ** p < 0.05, * p < 0.1.

| Interaction Variable               | ITE     | CITE     |
|-----------------------------------|---------|----------|
| unemployment rate                 | 0.003   | -0.045***|
|                                   | (0.015) | (0.018)  |
| minimum punishment                | 0.260***| 0.139    |
|                                   | (0.119) | (0.125)  |
| proportion mormon                 | 0.001   | 0.111    |
|                                   | (0.008) | (0.100)  |
| proportion southern baptist       | -0.041***| 0.065    |
|                                   | (0.019) | (0.101)  |

6.1. **Stock and Watson, 2015, Chapter 10.** In their textbook example, Stock and Watson study the relationship between a U.S. state’s traffic fatality rate ($Y_{it}$) and an alcohol tax ($X_{it}$) using data from 48 U.S. states over a period of 7 years. We extend this example to explore whether this relationship depends on time-varying interaction variables $G$ (state’s unemployment rate; state’s minimum punishment for drunk driving). We also consider time-invariant interaction variables $H$, namely the period-1 values of the proportion of a state’s population that is mormon or southern baptist.

Table 1 reports the ITCs estimated with ITE (column 1) and CITE (column 2). CITE suggests that the effect of alcohol taxes on traffic fatalities depends negatively on the unemployment rate. In contrast, the ITE does not find evidence for the presence of an interaction effect. For the presence of minimum punishment, the ITE estimates a positive interaction effect. CITE also estimates a positive effect, but it is not statistically significant at conventional levels.

The estimates for time-invariant interaction variables are also not in agreement. Using the ITE, we find a statistically significant relationship between the alcohol tax effect and the proportion of southern baptist. When repeating the analysis with CITE, we do not find evidence for such a relationship.

In conclusion, the two estimators for interaction effects yield contrasting conclusions for three of the four interaction variables. For two variables, the point estimate changes sign.

6.2. **Couttenier, Petrencu, Rohner, and Thoenig (2019).** Empirical research increasingly uses higher-dimensional panel data and interactions among those dimensions (e.g., Berman, Martin, and Mayer, 2012; Bloom, Draca, and Van Reenen, 2016; Manacorda and Tesei, 2020). To show the applicability of CITE in such a setting and contrast it with the performance of ITE, we revisit the results in Couttenier, Petrencu, Rohner, and Thoenig (2019), CPRT hereafter, who use micro-level panel data from Switzerland that are aggregated to a time ($t$), age cohort ($a$), migrant nationality ($n$), and Swiss regional ($i$) dimension (see Appendix B for details).
CPRT are interested in the relationship

\[ CP_{n,a,t,i} = \beta_i \text{kid}_{n,a,t} + Z_{n,a,t} \gamma_a + FE_{i,t} + FE_{n,t} + FE_a + \varepsilon_{n,a,t,i}, \]

to estimate the effect \( \beta_i \) of migrants’ past exposure to conflict (such as civil war) as a kid in their origin country on current crime propensity \( CP \) in Switzerland. They refer to \( \beta_i \) as the crime premium. Interaction effects in this setting can therefore be interpreted as the impact of interaction variables on that crime premium.

CPRT are interested in whether policies in host regions are related to the crime premium. In particular, they are interested in the role of the time-invariant interaction variable \( openjobacc_i \), which equals 1 for regions where asylum seekers can start working in all sectors of activity three months after arrival. To that end, they use CITE, i.e. they project their region-specific coefficient estimates \( \hat{\beta}_i \) from (21) onto \( openjobacc_i \):

\[ \hat{\beta}_i = a + b \text{openjobacc}_i + \varepsilon_i. \]

We reexamine their CITE results by including additional, time-varying interaction variables and compare them to the ITE results. The additional interaction variables that we include are two demographic variables: a region’s share of middle-aged persons \( \text{pop}_{middle} \) and a region’s share of urban population \( \text{urbanpop} \). Appendix B has more details on the data and specifications.

Table 6.2 reports our interaction effect estimates. First, we show in Appendix B that the mean effect of conflict exposure on crime propensity according to CITE is 0.46. This is very close to CPRT’s results, but not identical to the inclusion of additional time-varying interaction variables in our specification.

Second, CITE reveals that the crime premium is lower if urbanization increases (column 1, \( \text{urbanpop} \)), possibly reflecting better psycho-social support and integration opportunities. Note that the negative interaction coefficient \( -0.027 \) is not only statistically significant but also economically large: it suggests that at an urbanization rate of 92% (the 75th percentile), the crime premium is essentially 0 (equal to 0.023), while it is estimated to be 0.79 at an urbanization rate of 64% (the 25th percentile) – a magnitude that is close to the higher-end baseline result in CPRT.

Third, CITE suggests that states with \( openjobacc_i = 1 \), i.e. states with a generous labour market policy for immigrants, have much lower crime premiums.

In contrast, ITE does not find any statistically significant interaction terms. The point estimates for \( openjobacc \) and \( \text{urbanpop} \) are about half those of CITE.\(^{10}\)

In conclusion, CITE can be used in complex panel data settings that are increasingly common in the applied literature. Furthermore, CITE reveals two interesting interaction effects related to the crime premium that ITE does not find.

References

Alsan, M., and C. Goldin (2019): “Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880–1920,” Journal of Political Economy, 127(2), 586–638.

\(^{10}\)If the ITE for urban population of -0.011 is taken at face value, it implies a crime premium of 0.62 and 0.31 at the 75th and 25th percentile of \( \text{urbanpop} \), respectively, and hence a much lower implied difference for this interaction than the reported CITE results.
|     | (1)     | (2)     |
|-----|---------|---------|
| CITE | 0.805   | -0.171  |
| ITE  |         |         |
| pop\_middle | (0.758) | (0.216) |
| urbanpop  | -0.0274*** | -0.0112 |
|         | (0.00919) | (0.0110) |
| openjobacc | -1.146* | -0.490  |
|         | (0.622) | (0.447) |
| nT     | 48272   | 48272   |

Table 2. Point estimates and standard errors (in parentheses) for CPRT. First column has results for CITE, second column has results for ITE. Time-varying interaction variables are the share of a region’s population that is middle-aged, pop\_middle, and the share of a region’s population that is urban urbanpop. The time-invariant interaction variable openjobacc is a dummy variable indicating whether asylum seekers can start working in all sectors of activity three months after arrival. *** p<0.01, ** p<0.05, * p<0.1.

Amiti, M., and J. Konings (2007): “Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia,” *American Economic Review*, 97(5), 1611–1638.

Arellano, M., and S. Bonhomme (2012): “Identifying Distributional Characteristics in Random Coefficients Panel Data Models,” *The Review of Economic Studies*, 79(3), 987–1020.

Athey, S., and G. W. Imbens (2015): “Machine Learning Methods for Estimating Heterogeneous Causal Effects,” *Stat*.

Balli, H. O., and B. E. Sørensen (2013): “Interaction Effects in Econometrics,” *Empirical Economics*, 45(1), 583–603.

Berman, N., P. Martin, and T. Mayer (2012): “How Do Different Exporters React to Exchange Rate Changes?,” *Quarterly Journal of Economics*, 127(1), 437–492.

Bloom, N., M. Draca, and J. Van Reenen (2016): “Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity,” *Review of Economic Studies*, 83(1), 87–117.

Bloom, N., R. Sadun, and J. Van Reenen (2012): “Americans Do IT Better: US Multinationals and the Productivity Miracle,” *American Economic Review*, 102(1), 167–201.

Bordt, S., H. Farbmacher, and H. Kogel (2020): “Estimating Grouped Patterns of Heterogeneity in Repeated Public Goods Experiments,”

Burnside, C., and D. Dollar (2000): “Aid, Policies, and Growth,” *American Economic Review*, 90(4), 23.

Chamberlain, G. (1992): “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60(3), 567–596.

Chernozhukov, V., C. Hansen, Y. Liao, and Y. Zhu (2019): “Inference for Heterogeneous Effects Using Low-Rank Estimation of Factor Slopes,”

Couttenier, M., V. Petrencu, D. Rohner, and M. Thoenig (2019): “The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland,” *American Economic Review*, 109(12), 4378–4425.
de Chaisemartin, C., and X. D’Haultfoeuille (2022): “Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey,” Discussion paper.
Duflo, E., P. Dupas, and M. Kremer (2011): “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101(5), 1739–1774.
Epifani, P., and G. Gancia (2009): “Openness, Government Size and the Terms of Trade,” *The Review of Economic Studies*, 76(2), 629–668.
Gesselmann, M., and A. W. Schmidt-Catran (2020): “Interactions in Fixed Effects Regression Models,” *Sociological Methods & Research*, p. 004912412091493.
Graham, B. S., and J. L. Powell (2012): “Identification and Estimation of Average Partial Effects in ”Irregular” Correlated Random Coefficient Panel Data Models,” *Econometrica*, 80(5), 2105–2152.
Herrera, H., G. Ordoñez, and C. Trebesch (2020): “Political Booms, Financial Crises,” *Journal of Political Economy*, 128(2), 507–543.
Laage, L. (2020): “A Correlated Random Coefficient Panel Model with Time-Varying Endogeneity,” .
Lewbel, A., and K. Pendakur (2017): “Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients,” *Journal of Political Economy*, 125(4), 1100–1148.
List, J. A., and D. M. Sturm (2006): “How Elections Matter: Theory and Evidence from Environmental Policy*,” *Quarterly Journal of Economics*, 121(4), 1249–1281.
MacCurdy, T. E. (1981): “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *Journal of Political Economy*, 89(6), 1059–1085.
Manacorda, M., and A. Tesei (2020): “Liberation Technology: Mobile Phones and Political Mobilization in Africa,” *Econometrica*, 88(2), 533–567.
Mundlak, Y. (1978): “On the Pooling of Time Series and Cross Section Data,” *Econometrica*, 46(1), 69–85.
Sasaki, Y., and T. Ura (2021): “Slow Movers in Panel Data,” .
Shambaugh, J. C. (2004): “The Effect of Fixed Exchange Rates on Monetary Policy,” *Quarterly Journal of Economics*, 119(1), 301–352.
Spilimbergo, A. (2009): “ Democracy and Foreign Education,” *American Economic Review*, 99(1), 528–543.
Stock, J. H., and M. W. Watson (2015): *Introduction to Econometrics*, The Pearson Series in Economics. Pearson, Boston Columbus Indianapolis New York San Francisco Hoboken Amsterdam Cape Town Dubai London, updated third edition, global edition edn.
Storeygard, A. (2016): “Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa,” *Review of Economic Studies*, 83(3), 1263–1295.
Verdier, V. (2020): “Average Treatment Effects for Stayers with Correlated Random Coefficient Models of Panel Data,” *Journal of Applied Econometrics*, 35(7), 917–939.
APPENDIX A. PROOFS

Proof of Theorem 1. Recall from Definition 1 that \( \text{CITE} \) is the coefficient on the explanatory variables \( M_i \Psi_i \) in a linear regression with dependent variable \( M_i Y_i \), i.e.

\[
\hat{\theta}_n = \left( \sum_{i=1}^{n} \Psi_i'M_i \Psi_i \right)^{-1} \sum_{i=1}^{n} \Psi_i'M_i Y_i.
\]

The reduced form for \( M_i Y_i \), see (12), is

\[
M_i Y_i = M_i \Psi_i \theta + M_i \zeta_i.
\]

Therefore,

\[
\hat{\theta}_n - \theta = \left( \sum_{i=1}^{n} \Psi_i'M_i \Psi_i \right)^{-1} \sum_{i=1}^{n} \Psi_i'M_i Y_i - \theta
= \left( \frac{\sum_{i=1}^{n} \Psi_i'M_i \Psi_i}{n} \right)^{-1} \frac{\sum_{i=1}^{n} \Psi_i'M_i \zeta_i}{n}.
\]

By the boundedness assumptions in the statement of the theorem, the weak law of large numbers (WLLN) for random vectors ensures that both terms converge to their expectations.

We will now show that \( E(\Psi_i'M_i \zeta_i) = 0 \), which completes the proof. Recall that the elements of \( \zeta_i \) are

\[
\zeta_{it} = U_{it} + \sum_k V_{itk} X_{itk}
\]

and that, because of Assumption 7, \( E(\zeta_{it} | X_i, G_i, Z_i) = 0 \) so that, by the law of iterated expectations (LIE), \( E(\Psi_i'M_i \zeta_i) = 0 \), since \( M_i \) and \( \Psi_i \) are transformations of \( (X_i, G_i, Z_i) \) only. \( \square \)

Proof of Theorem 2. Recall that \( \hat{\delta}_{11} \) be the first element of

\[
\hat{\delta}_i = \left( X_i'X_i \right)^{-1} X_i' \left( Y_i - \Psi_i \hat{\theta}_n \right).
\]

and that

\[
Y_i - \Psi_i \hat{\theta}_n = \Psi_i (\theta - \hat{\theta}_n) + X_i \delta_i + \zeta_i
\]
so that the CITE for $\kappa$ in definition 2 is

$$
\hat{\kappa}_n = \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \hat{\delta}_{i1} \\
= \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \left( Y_i - \Psi_i \hat{\theta}_n \right) \\
= \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \Psi_i (\theta - \hat{\theta}_n) \\
+ \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \delta_i \\
+ \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \zeta_i \\
= A_{1n} + A_{2n} + A_{3n}
$$

For the second term,

$$
A_{2n} = \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \hat{\delta}_{i1} \\
\xrightarrow{p} (E(H_i' H_i))^{-1} E(H_i' \hat{\delta}_{i1}) \\
= \hat{\kappa}
$$

where the first line simplifies the expression for $A_{2n}$ from the previous display; the convergence on the second line follows from the WLLN, which applies because of Assumption 1 and conditions (i) and (ii) in the statement of the result; and the final equality is (17).

We will now show that $A_{1n}$ and $A_{3n}$ converge in probability to zero, which completes the proof. First,

$$
A_{1n} = \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \left( \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \Psi_i \right) (\theta - \hat{\theta}_n) \\
= o_p(1),
$$

because a WLLN applies to the first and second term because of conditions (i) and (iii) in the statement of the result, so that they are $O_p(1)$ (note that the inverse exists for $E(H_i' H_i)$ because of Assumption 4), and the final term is $o_p(1)$ from Theorem 1. Second,

$$
A_{3n} = \left( \sum_{i=1}^{n} H_i' H_i \right)^{-1} \sum_{i=1}^{n} H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \zeta_i \\
\xrightarrow{p} (E(H_i' H_i))^{-1} E \left( H_i' \epsilon_i' \left( X_i' X_i \right)^{-1} X_i' \zeta_i \right)
$$

because WLLNs apply in light of conditions (i) and (iv) in the statement of the result. That the second expectation is zero follows from the last step of the proof of Theorem 1.

\[\square\]
Appendix B. Data and additional details for section 6

To investigate whether past exposure to conflict in origin countries (such as civil conflict or mass killing) makes migrants more violence-prone in their host country, Couttenier, Petrecu, Rohner, and Thoenig (2019, CPRT hereafter) use data aggregated to the age cohort \((a)\) and immigrant nationality \((n)\) level and observed for each of the years \((t)\) between 2009 and 2016.\(^{11}\) In their baseline result, they find that cohorts exposed to civil conflict or mass killing in their origin country during childhood are 35 percent more prone to commit a violent crime in Switzerland than the average cohort. They then move on to explore heterogeneity in public policies across 26 Swiss regions, so-called ‘cantons’ \((c)\), to investigate how host country institutions modulate the impact of past exposure to conflict on current crime propensity \((CP)\).\(^{12}\) Therefore, in a first step, they run the regression:

\[
CP_{n,a,t,c} = \beta_c \text{kid}_{n,a,t} + Z_{n,a,t} \gamma_a + FE_{c,t} + FE_{n,t} + FE_a + \varepsilon_{n,a,t,c}
\]

for their sub-sample of male asylum seekers, where \(\text{kid}_{n,a,t}\) is a binary measure of early-age exposure to violence, \(Z\) contains binary control variables, and the \(FE\)s are fixed effects for canton\(\times\)years, nationality\(\times\)years, and age. Note that \(\beta_c\) is a canton-specific parameter capturing the violent ‘crime premium’ of conflict exposure over crime propensity of the average cohort and that its identification relies on within-canton, within-nationality, between-cohort variation.

In a second step, this canton-specific coefficient \(\hat{\beta}_c\) is projected on a set of canton-specific policy and control variables, which resembles the heterogeneity equation in the context of our CITE framework:

\[
\hat{\beta}_c = \alpha \times Policy_c + X'_c \beta + \varepsilon_c.
\]

To take into account the precision with which the cantonal-specific effects \(\hat{\beta}_c\) are estimated in the first step, CPRT estimate this second equation by GLS using the inverse of the standard errors estimated in the first stage as weights.\(^{13}\) Concerning those policy variables, CPRT mostly focus on ‘openjobacc’, which is a binary variable equal to 1 for cantons where asylum seekers can start working in all sectors of activity three months after arrival.

B.1. Replication of CPRT’s second stage results. We replicate those results in the first column of table 3, which is identical to column 1 of table 9 in CPRT. To facilitate comparison with our CITE framework, column 2 of table 3 reports the result based on OLS estimation, which leads to nearly identical results (although with somewhat higher standard errors).\(^{14}\)

\(^{11}\)We follow the notation of CPRT in this Appendix to facilitate comparison with their paper. The notation in section 6 is slightly different to facilitate consistency with the setup in our paper.

\(^{12}\)CP is measured as the share (scaled in percentage points) of individuals in the cohort who perpetrate at least one violent crime in a given year. The main analysis we rely on limits the sample to the years 2011-2016 and to 25 cantons (by dropping Appenzell-Innerrhoden).

\(^{13}\)CPRT refer to Bertrand and Schoar (2003 QJE) and Bandiera, Prat, and Valletti (2009 AER), who have previously taken this approach.

\(^{14}\)Note that the coefficient \(\times \text{mean}('\text{openjobacc}') + \text{constant}, -0.62 \times 0.64 + 0.91 = 0.506\), provides an estimate of the mean of the canton-specific parameters \(\hat{\beta}_c\). Those 0.506 are identical to regressing the \(\hat{\beta}_c\) on a constant (up to the third decimal) and almost identical to the estimate we obtain for a homogeneous relationship in equation (21): 0.498 (SE: 0.470).
Table 3. Second stage results in CPRT (2019)

| VARIABLES       | (1)     | (2)    |
|-----------------|---------|--------|
| openjobacc      | -0.640* | -0.624 |
|                 | (0.366) | (0.439) |
| constant        | 0.796** | 0.905**|
|                 | (0.303) | (0.352) |
| Observations    | 25      | 25     |
| R-squared       | 0.246   | 0.253  |
| Estimation      | GLS     | OLS    |
| Standard errors | in parentheses |          |
| *** p<0.01, ** p<0.05, * p<0.1 | |

Table 4. Descriptive statistics

| Variable          | Obs   | Mean  | SD   | Min  | Max  |
|-------------------|-------|-------|------|------|------|
| CP                | 48,272| 3.92  | 17.28| 0.00 | 100.00|
| x1_kid012         | 48,272| 0.62  | 0.49 | 0.00 | 1.00  |
| h1_openjobacc     | 48,272| 0.62  | 0.48 | 0.00 | 1.00  |
| g4_pop_middle     | 48,272| 61.87 | 1.20 | 58.36| 63.90 |
| g6_urbanpop       | 48,272| 75.77 | 19.64| 0.00 | 100.00|

B.2. Addition of time-varying G variables. To explore the behavior of CITE vs. ITE in the framework developed in our paper, we add time-varying canton-specific data on the share of middle-aged (20-64 years) and urban population. Those data come from annual ‘Kantonsporträts’ provided by the Swiss ‘Federal Statistical Office’ and capture the G variables in the framework of our paper. Table 4 provides descriptive statistics of the sample data used in our application.

B.3. Calculation of mean effect of conflict exposure. The mean ‘crime premium’ of conflict exposure can be calculated from table 6.2 by multiplying all interaction term coefficients with the sample means of the respective interaction variables and adding those terms up. For example, for the CITE estimate, this gives $-1.146 \times 0.64 - 46.5 \text{ (from column 2)} + 0.805 \times 61.867 - 0.0274 \times 75.774 \text{ (from column 1)} = 0.462$. This is very similar to the ‘crime premium’ result of 0.498 one obtains from CPRT when not including any interaction terms (see footnote 14).

---

15We initially included the share of young (0-19 years) and old (65 years and above) population but since both of them led to nearly identical parameter estimates (around -1 for CITE and around 0.25 for ITE), we decided to opt for a simpler model with the share of middle-aged population (which is the residual of young and old).

16Note that for CITE, the constant of the second-step regression needs to be taken into account, while for ITE the coefficient for x1_kid needs to be taken into account.