I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10927

Immigration Policy and Remittance Behaviour

Matloob Piracha
Massimiliano Tani
Guy Tchuente

JULY 2017

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10927

Immigration Policy and Remittance Behaviour

Matloob Piracha
University of Kent and IZA

Massimiliano Tani
University of New South Wales and IZA

Guy Tchuente
University of Kent and MaGHiC

JULY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Immigration Policy and Remittance Behaviour*

This paper analyses the impact of a change in Australia's immigration policy, introduced in the mid-1990s, on migrants' remittance behaviour. More precisely, we compare the remittance behaviour of two cohorts who entered Australia before and after the policy change, which consists of stricter entry requirements. Our empirical strategy uses conditional difference-in-differences in the presence of interactive fixed-effects. We first show that Bai's (2009) least squares estimator and conditional difference-in-differences are biased if used on their own. We then derive conditions that are required to obtain a consistent estimator using a combination of conditional difference-in-differences and Bai's (2009) least squares estimator. The results indicate that those who entered under more stringent conditions – the second cohort – have a higher probability to remit than those in the first cohort, though the policy change has no discernible effect on the level of remittances.

JEL Classification: C13, F22, F24, J61
Keywords: immigration, treatment effect, difference-in-differences

Corresponding author:
Matloob Piracha
School of Economics
Keynes College
University of Kent
Canterbury, Kent CT2 7NP
United Kingdom
E-mail: M.E.Piracha@kent.ac.uk

* This paper uses the same idea and motivation as one of the earlier papers by the first two authors, from which some of the material is borrowed (Mahuteau, S; Piracha, M and Tani, M, (2010). Selection policy and immigrants' remittance behaviour. IZA DP No. 4874). However, this paper is significantly different from the other one as our main contribution in this paper is the econometric methodology needed for the kind of data we have available.

1 Introduction

International migration from developing countries is often linked to expectations of mutual gains for the migrants and their countries of origin and destination. Migrants benefit if the net return to their skills is higher in the host country than in their home country. The immigration of workers allows receiving countries to fill domestic labour market shortages and provide host country employers the private benefit to meet productive capacity without large hikes in wages. This is particularly so when domestic labour shortages are skill-specific and when migration policy is tuned to select such skills to grant entry into the country.

Recent literature has highlighted the importance of remittances as ”compensation” for a potential brain drain from the sending countries (Bollard, McKenzie, Morten, and Rapoport (2011)). The aggregate value of remittances is larger than those of common commodities traded, and most sending countries appear to receive substantial inflows (e.g. Philippines, Mexico, the Pacific islands, North Africa). However, it has also been highlighted that more skilled migrants tend to remit less than those with lower levels of education, leading to the implication that the brain drain is not properly compensated for when more educated leave the country (Niimi, Ozden, and Schiff (2010)).

Although migrant remittances have received widespread attention because of their importance for the economies of developing nations, few studies have examined the conditions that affect migrants remittance pattern. These could be determined by the underlying motivations to remit. For instance, in the presence of uncertainty about their legal status and/or labour market conditions migrants could potentially remit more because of the insurance motive rather than, say, the altruistic motive (see Amuedo-Dorantes and Pozo (2006); Piracha and Zhu (2012); Batista and Umblijs (2016)). These motivations could indeed be affected by the change in conditions in the host country, particularly a change in the immigration policy as that has an implication for immigrants' legal status. In addition, immigration policy could affect the characteristics (e.g., skill level) of immigrants who enter the country and consequently have an effect on their remittance behaviour. The main argument here is that the net benefits of migration are subject to the policy shift in both sending and receiving countries.

While it is generally accepted that changing migration admission policy affects the skill composition of new immigrants as well as their economic assimilation process, it is unclear what is its effect on subsequent saving/remittance behaviour. Yet this information is critical to determine whether, prima facie, the compensation hypothesis following tighter migration criteria is empirically sustainable, especially in light of policy discussions about possible skill-biased restrictions to immigrants. We contribute to the literature by testing the hypothesis using the case of a migration policy change implemented in the mid-1990s in Australia. In 1996 Australia tightened non-family and non-humanitarian migration policy with reference to skills (education and work experience) and knowledge of English. The policy change also removed income support for the first two years after migration and subsidies to attend English classes. As a result, new independent migrants to Australia tended to be more educated and possessed better English language skills than previous cohorts Cobb-Clark (2003). We use Longitudinal Survey of Immigrants to Australia (LSIA) to estimate the effect of the policy change on the probability to remit (the extensive margin) and the level of remittances (the intensive margin) of two cohorts of migrants. The first cohort entered Australia between 1993 and 1995 (LSIA1), just before the policy change, while the second one entered in 1999-2000 (LSIA2), after the policy change.

The LSIA consists of two distinct longitudinal datasets, three waves covering Cohort 1 (LSIA1) and two for Cohort 2 (LSIA2). In each cohort we have two groups of migrants, the non-family and non-humanitarian group (treatment group) and the family and humanitarian migrants (control group). Given the structure of the data, we could use the difference-in-differences methodology. However, migrants in Cohort 1 are likely to have different observable and unobservable characteristics than those who entered the country after the policy change (Cohort 2). Indeed, the treatment and the control group could have different job market realities, and also the composition of the treatment group can be affected by the change in policy. These reasons also imply that the simple propensity score matching is likely to be problematic since matching methods do not account for variation in time trend between the control and the treatment group.

We tackle this problem by extending the literature on conditional difference-in-differences and evaluation of treatment effect with panel data in the presence of interactive fixed-effects. In their seminal paper Heckman, Ichimura, and Todd (1997) provide evidence suggesting that the conditional difference-in-differences is an effective method to evaluate the effects of a programme under the assumption of conditional parallel trend. We consider an extension of their framework by assuming that there is an interactive fixed-effect in the data generating process. We also extend Gobillon and Magnac (2016) by allowing difference in the support of observable characteristics or in how they affect the outcome in the treated and untreated sample.

We consider a theoretical model in which there is a random slope on observable characteristics that are assumed to be independent (or not) of the time-varying individual fixed-effect. We assume that our estimations are done on a sample with a small time dimension and a large number of treated and untreated individuals. Such a framework is common for micro panel data. We discuss the identification condition of the average treatment effect on the treated in a linear model with random slope and time-varying fixed-effects. We derive the generic bias of the difference-in-differences estimator when the true data generating process has a single random slope and time-varying fixed-effects. We show that using an estimation method that removes only the time-varying component will not eliminate the bias of the difference-in-differences estimator. Moreover, using only methods that account for the possibility of difference in the support of observable characteristics (for example conditional difference-in-differences) will also fail to eliminate the bias of the difference-in-differences. However, combining them eliminates the bias. We evaluate the relevance of the use of a mix of conditional difference-in-differences with methods that could eliminate time-varying component via a Monte Carlo experiment. We use various strategies for time-varying fixed-effect estimation.

Our work on identification is a contribution to the literature of estimation of treatment effect using panel data. Hsiao, Steve Ching, and Ki Wan (2012) propose to estimate the correlations between the treatment and control regions based on the pre-treatment data. Ouyang and Peng (2015) extend their work by allowing the conditional mean to have a semi-parametric form. However their approach focuses more on panel with large time dimension (T) and few treated individual. In the same context, Abadie and Gardeazabal (2003) have proposed the synthetic control method to estimate average treatment effects. However, the synthetic control method could fail to remove the two sources of violation of the parallel trend assumption.

Our results show that those who entered in the second cohort, regardless of the origin region or the type of job obtained, had a higher probability to remit compared to the first cohort, with no discernible impact on the level of remittances.

The rest of the paper is organized as follows. Discussion on the background of immigration policy in Australia is presented in Section 2. Section 3 presents the theoretical framework employed to deal with the type of data used in this paper. The implementation of our theoretical model in explained in Section 4 while Section 4 presents the data and empirical results. Concluding remarks appear in the last section.

2 Background of Australian immigration policy

To understand the development of policies leading to the changes in the mid-1990s, one needs to put some context to Australia's immigration policy. In 1972 Australia formally ended a migration policy based on ethnicity ('white Australia policy'), replacing it with a focus on economic conditions to a limited number of workers in occupations where there was unmet demand. Eliminating racial discrimination from immigration selection resulted in increasing numbers of applicants and refugees from non-European countries and consequently higher stocks of immigrants with non-English speaking background (NESB). As an example Australia has taken refugees from conflicts in Chile (1973), Northern Cyprus (1974), Lebanon (1976-1983), Vietnam (1976-1982), Thailand (1976), East Timor (1977), Sri Lanka and El Salvador (1983) and the former Yugoslavia (1994).

Until the policy change introduced in 1996 and analysed in this paper, two major trends had characterized immigration policy in Australia. First, the development of a systematic approach to immigration selection based on current labour market conditions. This took place during the late 1970s and throughout the 1980s with the introduction of the Numerical Multifactor Assessment System (NUMAS) during 1979-1982, which selected immigrants on the basis of family ties, occupational and language skills, and the points test (since 1988) which was used to set annually a minimum pass mark to be eligible for migration based on the skill level (qualifications and work experience), age and English language proficiency. Points could be gained if the applicant was qualified to work in one of the occupations listed in a Priority Occupation List, which summarises employers' views and recruitment difficulties. Pre-migration English-language testing for particular occupations (chosen annually) was also introduced in 1989, first in the medical and nursing professions and then extending to 114 professions between 1991 and 1993 (Hawthorne (2005)), with points provided for both speaking and writing language abilities. Since 1988 the migration program is divided into three streams - family reunification, skill and humanitarian - with the of skill stream contributing about one third of all migrants to Australia until 1995-96.

The second trend has been the development of policies to favour the settlement and participation of migrants, especially if from NESB, to Australia's economic activity. This was accompanied by the introduction of instruments, including ad hoc data collections, to analyse their economic outcomes. Even with higher skill levels than comparable natives or migrants with an English-speaking background (Watson (1996)), NESB immigrants were characterized by substantially lower economic outcomes. To overcome a linguistic disadvantage, Australia had put in place a publicly funded system to provide new adult NESB immigrants with free language courses (as well as locally-funded technical training). Immigrants were paid to attend these courses, which lasted between one and six months and led them to attain a level of language ability that was adequate for employment. By 1990, Australia's Adult Migrant English Program (AMEP) "was the largest government-funded English teaching program for migrants worldwide, catering to over 70,000 immigrants per year including large number of unemployed professionals" (Hawthorne (2005)). The government actively pursued the private sector to adopt Equal Opportunity principles towards NESB as well as facilitated to ease the admission of 'professionals' from NESB (managed by professional associations) with the funding of specialist labour market programs designed to prepare NESB professionals for mandatory entry exams in a range of professions like medicine, engineering and nursing.

In 1996 a newly elected government introduced a number of significant changes to the migration policy, affecting the skill and family reunification but not the humanitarian streams. This new policy: (1) Abolished the social security benefit to new immigrants in the first two years after their arrival, as well as access to the Adult Migrant English Program (whose costs were now to be met by the immigrant) and labour market programs (whose costs were to be repaid after securing work); (2) Allocated the highest points weighting to employability factors, namely occupational skills, education, age, and English language ability. Age-related points for applicants over the age of 45 were abolished while bonus points were awarded to those with relevant Australian or international professional work experience, a job offer, a spouse meeting the skill application criteria, an Australian sponsor who had to provide a guarantee, and carrying A$100,000 or more in capital. By 2001 most migrants to Australia were in the skill stream; (3) Introduced additional points for occupations in demand in addition to degree-level specific (as opposed to generic) qualifications, and bonus points for qualifications obtained recently in Australia; (4) Pre-migration qualification screening was effectively outsourced to professional bodies, which could now disqualify NESB applicants from eligibility for skill migration.

3 Theory

We consider in this section a theoretical set-up that can be used to evaluate the effect of the change in policy outlined in Section 2. The proposed theoretical model corresponds to a panel data structure with random slope and interactive fixed effects. We first show that diff-in-diffs is biased and we then derive conditions under which the conditional diff-in-diffs combined with Bai's (2009) least square method leads to a consistent estimator.

Consider a sample composed of N individuals observed at dates $t = 1, ..., T$. Some of the individuals, $i = 1, ..., N_1$, are observed only for $t = 1, ..T_D$ while others, $i = N_1 + 1, ..., N$, are observed only for $t = T_D + 1, ..., T$. A treatment, $D_i \in \{0, 1\}$, is implemented at date $t > T_D$. After the treatment, some units $i = N_1 + 1, ...N_2$ are treated ($D_i = 1$) while others are not. For each individual we observe the outcome, Y_{it}. The outcome depends on the treatment status and we are interested in the average treatment effect on treated (ATT).

We consider the Rubin's potential outcomes framework. $Y_{it}(d)$ is the individual i at time t if his treatment status is d.[1] The average treatment effect on treated for $t > T_D$ is given by:

$$ATT = E(Y_{it}(1) - Y_{it}(0)|D_i = 1) = E(Y_{it}(1)|D_i = 1) - E(Y_{it}(0)|D_i = 1)$$

A natural estimator of $E(Y_{it}(1)|D_i = 1)$ is its empirical counterpart. However,

[1] $d = 1$ in presence of treatment and $d = 0$ in the absence of treatment. D_i and d are different because D_i represents the actual treatment and d the hypothetical treatment.

we do not observe $E(Y_{it}(0)|D_i = 1)$ for treated individuals. The challenge for the econometrician is to construct a consistent empirical counterpart to $E(Y_{it}(0)|D_i = 1)$.

Under the equal trend assumption, for $t \geq T_D$:

$$E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0) = E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1).$$

The counterfactual can be written as:

$$E(Y_{it}(0)|D_i = 1) = E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0) + E(Y_{iT_D-1}(0)|D_i = 1). \quad (1)$$

The equal trend assumption implies that, in the absence of the treatment, the average outcomes for the treated and control groups would have followed parallel paths over time. This assumption also means perfect compliance. In other words, individuals in the treated group are similar enough to individuals in the control group. However, in practice, if pre-treatment characteristics that are thought to be associated with the dynamics of the outcome variable have heterogeneous effect in the treated and the control group, then the parallel trend assumption will be violated. For example, if the measurement instruments are different for treated and control groups or selection for treatment is influenced by past outcomes then our treatment and control groups will be different in observed characteristics.

We consider the case in which heterogeneity in the effect of pre-treatment observed characteristics can lead to non-parallel outcome dynamics between the treatment and the control groups. Our aim is to generalize the usual set-up, in which diff-in-diffs gives a consistent estimate of the ATT, by allowing for unobserved heterogeneity trend and heterogeneity in the effect of pre-treatment observed characteristics.

At the micro level, heterogeneity in the effect of pre-treatment observed characteristics is a problem. For instance, there is a possibility of self selection based on observed and unobserved characteristics. Also, most programs are designed for specific groups, which means non-participants are likely to have significant differences with participants in observed characteristics. In some applications, control group are in a different environment such that the treatment and control groups are then affected by different type of shocks over time. We represent these heterogeneity using two random coefficients, β_i and γ_i. The outcome in absence of the treatment could then be presented as:

$$Y_{it}(0) = X_{it}\beta_i + \delta_t\gamma_i + U_{it} \quad (2)$$

The model presented in equation (2) has two differences with the classic diff-in-diffs set-up. We have heterogeneity in the effect of observed characteristics i.e., β_i (random slope) and unobserved time varying heterogeneity γ_i. The outcome in case of treatment is given by

$$Y_{it}(1) = \alpha_{it} + Y_{it}(0) \tag{3}$$

The data generating process is obtained under the following assumptions.

Assumption 1: $E(U_{it}|\theta_i, X_i) = 0$ with $\theta_i = (\beta_i, \gamma_i)$

Assumption 2: $E(U_{it}|\theta_i, X_i) = E(U_{it}|D_i, \theta_i, X_i)$

Assumption 3: $\gamma_i \perp \beta_i$ for all i where $\perp$ means independent.

The first assumption represents exogeniety of observed and unobserved characteristics with respect to the error term. Assumption 2 implies that the treatment status is independent of the error term, though it allows for correlation between treatment status and other characteristics. We can, therefore, have selection into the program based on observables and unobservables. Finally, Assumption 3 is designed to account for situations in which the random slope and the time varying unobserved heterogeneity are independent. This assumption can be relaxed without loss of generality.

3.1 Bias of the Diff-in-Diffs

This subsection shows that in the presence of time-varying group specific heterogeneity and random slope, the classic diff-in-diffs methodology estimator is biased. We derive the form of the bias and discuss the conditions under which we can apply the classic diff-in-diffs.

The parameter of interest is the ATT

$$\alpha = E\left[\frac{1}{T - T_D + 1}\sum_{t=T_D}^{T} \alpha_{it}|D_i = 1\right]$$

We take the probability measures associated with γ_i and β_i to be dominated by the Lebesgue measure. Their treatment status conditional forms are defined as follows:

- $dG_1(\beta_i|D_i = 1)$ and $dG_1(\beta_i|D_i = 0)$
- $dG_2(\gamma_i|D_i = 1)$ and $dG_2(\gamma_i|D_i = 0)$

We furthermore assume that for all individuals, $dG_k(\beta_i|D_i = 1)$ is absolutely continuous with respect to $dG_k(\beta_i|D_i = 0)$,$k = 1, 2$. This corresponds to a common support assumption. The Radon-Nikodym derivatives are given as follows:

- $dG_1(\beta_i|D_i = 1) = r_1(\beta_i)dG_1(\beta_i|D_i = 0)$
- $dG_2(\gamma_i|D_i = 1) = r_2(\gamma_i)dG_2(\gamma_i|D_i = 0)$

Under Assumptions 1 and 2, for $t \geq T_D$:

$$E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0, \theta_i) = E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \theta_i) \tag{4}$$

We are going to show that equation (4) does not imply the classic equal trend assumption in general. Indeed,

$$\begin{aligned} E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1) &= E[E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \theta_i)] \qquad (5) \\ &= \int_\gamma \int_\beta E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \theta_i)dG_1(\beta_i|D_i = 1)dG_2(\gamma_i|D_i = 1) \end{aligned}$$

If we are in a situation of perfect compliance, i.e., treated group is very close to the control group such that $\beta_i = \beta_0$, then our set-up coincides with Gobillon and Magnac (2016). If, in addition $\delta_t = 1$, we are in the case of the classic diff-in-diffs. Another situation in which Assumptions 1 and 2 imply the parallel trend assumption is if we have a perfectly controlled experiment that enables us to have complete randomization of the treatment ie $dG_1(\beta_i|D_i = 1) = dG_1(\beta_i|D_i = 0)$ and $dG_2(\gamma_i|D_i = 1) = dG_2(\gamma_i|D_i = 0)$. Under the common support assumption for β_i and γ_i along with Assumption 3, we can show that equation (6) can be written as follow:

$$\begin{aligned} E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1) &= E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0) \\ &+ Cov(Y_{it}(0) - Y_{iT_D-1}(0), r_1(\beta_i)r_2(\gamma_i)|D_i = 0) \end{aligned}$$

The time varying unobserved effects and random slope bias are represented by the second term on the right hand side. If there is indeed a time-varying factor or random slope, the second term is not equal to zero except under special circumstances as seen above.

$$
\begin{aligned}
Cov(Y_{it}(0) - Y_{iT_D-1}(0), r_1(\beta_i)r_2(\gamma_i)|D_i = 0) &= \int_{\beta} Cov[Y_{it}(0) - Y_{iT_D-1}(0), r_2(\gamma_i)|D_i = 0, \beta_i] \\
&\times\ r_1(\beta_i)dG_1(\beta_i|D_i = 0) \\
&= \int_{\gamma} Cov[Y_{it}(0) - Y_{iT_D-1}(0), r_1(\beta_i)|D_i = 0, \gamma_i] \\
&\times\ r_2(\gamma_i)dG_2(\gamma_i|D_i = 0)
\end{aligned}
$$

It represents the bias of the diff-in-diffs estimates. Note that the bias can be viewed as an aggregation of the bias coming from the time varying component and the part coming from heterogeneity effect of observed characteristics. The form of the bias suggests that we need to use an estimation strategy that accounts for both sources of bias.

3.2 Estimation of the ATT

In this subsection we discuss conditions under which the ATT is identified and how it can be estimated. The observed outcome is given by

$$Y_{it} = Y_{it}(0)\mathbf{1}\{D_i = 0\} + Y_{it}(1)\mathbf{1}\{D_i = 1\}$$

Using the values of $Y_{it}(0)$ and $Y_{it}(1)$ we can rewrite the observed outcome as

$$Y_{it} = \alpha_{it} D_i I_t + \beta_i X_{it} + \gamma_i \delta_t + U_{it} \tag{6}$$

where D_i is the treatment indicator and $I_t = \mathbf{1}\{t \geq T_D\}$ is the treatment period indicator. Assumptions 1 to 3 are maintained. Note that D_i is allowed to be correlated with θ_i and X_i. A special case is when the correlation between D_i and X_i is very strong. This can occur in situations where the common support assumption for the conditional distributions of X_i knowing D_i is not verify. This paper considers empirical application where X_i, β_i and γ_i are correlated with D_i. The following assumption gives conditions on X_i, β_i under which the treatment effect is identified.

Assumption 4:

(i) $\eta < P(D_i = 1|X) < 1 - \eta$ for some $\eta > 0$. This is the overlap assumption.

(ii) $dG_1(\beta_i|D_i = 1, P(X)) = dG_1(\beta_i|D_i = 0, P(X))$ with $P(X) = P(D_i = 1|X)$.

This assumption provides characteristics of the data set and some parameters of the model for which the ATT is identified. Assumption 4 (i) is a classic assumption in propensity score matching literature. It says that the support of the propensity scores overlap conditional on a set of exogenous variables. The second part of Assumption 4 implies that when we have two individuals (one from the treated and one from the control group) with the same propensity score, then the distribution of the effects of the exogenous characteristics should be the same.

Now we discuss the identification and estimation of the ATT. We assume that $\beta = E(\beta_i)$, equation (6) then becomes

$$Y_{it} = \alpha D_i I_t + X_{it}\beta + \delta_t \gamma_i + U_{it} + (\alpha_{it} - \alpha) D_i I_t + X_{it}(\beta_i - \beta) \tag{7}$$

Because our objective is to investigate micro-level data, we are going to assume fixed number of periods. We assume the time effect δ_t is known. For each individual the observed outcome in vector notation is

$$Y_i = \alpha D_i I_{[1:T]} + X_i\beta + \Delta\gamma_i + U_i + \Omega_i D_i I_{[1:T]} + X_i(\beta_i - \beta) \tag{8}$$

where $Y_i = (Y_{i1}, ..., Y_{iT})'$, $X_i = (X_{i1}, ..., X_{iT})'$, $U_i = (U_{i1}, ..., U_{iT})'$, $I_{[1:T]} = (I_1, ..., I_T)'$, $\Delta = (\delta_1, ..., \delta_T)'$, $\Omega_i = diag(\alpha_{i1} - \alpha, ..., \alpha_{iT} - \alpha)$.

Let $M_\Delta = I - \Delta'(\Delta\Delta')\Delta$ and multiplying equation (8) by M_Δ on both sides, we get

$$M_\Delta Y_i = \alpha D_i M_\Delta I_{[1:T]} + M_\Delta X_i\beta + M_\Delta U_i + M_\Delta \Omega_i D_i I_{[1:T]} + M_\Delta X_i(\beta_i - \beta) \tag{9}$$

The prediction of D_i as a function of X_i can be given by:

$$D_i = vec(X_i)'\rho + D_{iX}.$$

Equation (9) then becomes

$$M_\Delta Y_i = \alpha D_{iX} M_\Delta I_{[1:T]} + M_\Delta \tilde{U}_i + M_\Delta \Omega_i D_i I_{[1:T]} + M_\Delta X_i(\beta_i - \beta) \tag{10}$$

where $\tilde{U}_i = U_i + X_i\beta + \alpha vec(X_i)'\rho I_{[1:T]}$. Denote the general error term by $\varepsilon_i = \tilde{U}_i + \Omega_i D_i I_{[1:T]} + X_i(\beta_i - \beta)$.

Following are the necessary conditions for identification of α:

1. $E(D_{iX}) > 0$ and $M_\Delta I_{[1:T]}$ has full rank column.
2. $Cov(\varepsilon_i, D_{iX}) = 0$ (exogeneity condition).

Condition 1 means that the probability of being treated is positive, which follows from Assumption 4. The second part of the first condition means that $I_{[1:T]}$ is not equal to a linear combination of time effect. To ensure that, time dummies are set to zero for both $t > T_D$ and $t < T_D$.

Now we discuss the assumptions under which Condition 2 holds.

$$Cov(\varepsilon_i, D_{iX}) = Cov(\tilde{U}_i, D_{iX}) + Cov(\Omega_i D_i I_{[1:T]}, D_{iX}) + Cov(X_i(\beta_i - \beta), D_{iX}) \quad (11)$$

There are three terms in this correlation that we analyse in turn. The first term is equal to zero by construction using Assumption 2 and the fact that X_i and $vec(X_i)$ are uncorrelated with D_{iX}. The second term of the correlation above is more interesting and can be written as:

$$\begin{aligned}
E(\Omega_i D_i I_{[1:T]} D_{iX}) &= E(E(\Omega_i I_{[1:T]} D_{iX}|D_i)D_i) \\
&= E(E(\Omega_i I_{[1:T]}|D_i)D_i) - E(E(\Omega_i I_{[1:T]} vec(X_i)'\rho|D_i)D_i) \\
&= 0 - E(E(\Omega_i I_{[1:T]}|D_i, X_i)vec(X_i)'\rho D_i) \\
&= 0
\end{aligned}$$

These results hold by construction of Ω_i and the definition of ATT under the sufficient condition given by

$$E(\alpha_{it}|D_i = 1, X_i) = E(\alpha_{it}|D_i = 1)$$

The last term in the correlation, under the assumption $E(\beta_i|X_i) = E(\beta_i) = \beta$, is given by:

$$\begin{aligned}
Cov(X_i(\beta_i - \beta), D_{iX}) &= E(X_i(\beta_i - \beta)D_{iX}) \\
&= E(X_i(\beta_i - \beta)D_i) - E(X_i(\beta_i - \beta)vec(X_i)'\rho) \\
&= E(X_i(\beta_i - \beta)|D_i = 1)
\end{aligned}$$

Under the assumption of heterogeneity in the effect of observed characteristic and correlation between β_i and D_i, $E(\beta_i|X_i, D_i = 1) \neq E(\beta_i|X_i, D_i = 0)$. However, Assumption 4 (ii) helps us to recover the equality. It implies that

$$E(\beta_i|P(X), D_i = 1) = E(\beta_i|P(X), D_i = 0).$$

Under Assumption 4 the last term on the right hand side of equation (11) is equal to zero, since

$$E(X_i(\beta_i - \beta)|D_i = 1, X_i) = E(X_i(\beta_i - \beta)|D_i = 1, P(X_i)).$$

The conditioning on the propensity score can be applied also in the proof of second term. The ATT is identified under Assumptions 1, 2 and 4. Its estimation follows by applying the OLS procedure to equation (9) on the appropriately matched sample. In other words we are going to do a two stage procedure. In the first stage, we can use estimate and remove the interactive fixed-effect. In the second stage, we use a propensity score matching procedure to estimate the ATT. The consistency properties of this method can be derived following Bai (2003), Bai (2009) and Abadie (2005).

4 Application to the effect of immigration policy

The aim of this paper is to analyse the effect of a more stringent migration policy on remittance behaviour. To achieve this aim, we apply a before-after approach. The application of a before-after methodology comes with the challenge of the choice of a control group or the construction of an artificial control group. In the case of changes in the immigration policy, the treatment group consists of those subject to a more restrictive migration policy (concessional family and skilled independent visa categories) while the control group consists of migrants not affected by the policy (humanitarian and preferential family visa categories).

Diff-in-diffs is one of the most commonly used before-after approach. However, one of the fundamental assumptions of diff-in-diffs (the parallel trend in the control and treatment groups) is likely not to hold for the type of data we have to evaluate the effect of policy change. Indeed, humanitarian and family visa seem to be different from concessional family and skilled independent visa category holders. This implies that the effect of exogenous characteristics on remittances is not the same for both groups. Moreover, the two groups of migrants will not necessarily work in the same sector of the economy and they face different aggregate time effects over time.

The use of propensity score matching (PSM) is an appealing, alternative, solution. The PSM does assume selection on observables, which means that the decision of the difference in people with concessional family and skilled independent and those with humanitarian and preferential family is based on observable criteria only. This assumption means that if we match people with the same propensity score, the difference in their remittances can be interpreted as the effect of the policy. This implies that we can find a counterfactual of the skill group in the humanitarian group, which is a strong assumption since the change in policy could have an effect on the composition of the two groups. Moreover, with this method, it is not possible to account for time varying heterogeneity.

We address the aforementioned limitations by using a methodology that combines diff-in-diffs with PSM (conditional diff-in-diffs) while controlling for time varying unobserved heterogeneity. The data consists of two cohorts of immigrants, those who entered before the policy change (Cohort 1) and those entered after the policy change (Cohort 2). In addition, for each cohort we can take advantage of the panel structure of our data to control for difference in trend by correcting our estimate for time and group specific fixed effects. These aspects allow us to construct an appropriate counterfactual while controlling for difference in time trend.

The PSM ensures that only those with very similar weights (propensity scores) in the treatment and control groups are compared. However, the change in immigration policy is not completely random as it can be linked to a specific economic situation (recession on aggregate output) or shortage in specific sectors leading to unobserved time fixed effect that could be correlated with both the treatment groups (visa categories) and the outcome variable (remittances). We use the panel structure of our data for both cohorts to control for group specific trend. This is similar to the use of interactive fixed effects and synthetic control group allowing the influence of an unobserved factor as in Abadie, Diamond, and Hainmueller (2012), Gobillon and Magnac (2016), Ouyang and Peng (2015) and Abadie, Diamond, and Hainmueller (2015). However, our approach differs from theirs for two reasons. First, we cannot use factor models given the short number of periods we have for each cohort. We are therefore restricted to the used of year fixed-effect to capture individual specific time-varying effects.[2] Second, our main

[2]See Ahn, Lee, and Schmidt (2013) Basic Assumptions for the conditions on the number of factor and the

focus is on how the difference in the control and treated groups (in terms of exogenous characteristics and their effect on outcome) influence the result. Our work can be viewed as a special case of fuzzy diff-in-diffs, as in de Chaisemartin and D'Haultfoeuille (2015), where the fuzziness comes from difference in the support of exogenous variable in the treatments groups. Given that our treated population is relatively large, a Monte Carlo simulation (presented in Appendix A) suggests that PSM or conditional diff-in-diffs with control for time varying fixed effects should work better than diff-in-diffs.

We consider Y_{it} to be the observed level of remittances of individual i in period t, X_{it} is the set of exogenous characteristics, D_i is an indicator to show if an individual is in the control or the treated group and d_{it} is an indicator of treatment.

$$Y_{it} = \alpha_{it} d_{it} + \beta_i X_{it} + \gamma_i \delta_t + U_{it}$$

ATT Estimation Algorithm:

1. For each cohort and each treatment group run Panel data model estimation with time dummies and individual fixed effects. Compute the adjusted for time effect $\tilde{Y}_{it}$

2. Compute the propensity score and match individuals in the treated group of cohort 1 and cohort 2 with those in the control group.

3. Evaluate

$$\widehat{ATT} = \hat{E}(\tilde{Y}_{it} - \tilde{Y}_{it_0} | P(X), D_i = 1, X) - \hat{E}(\tilde{Y}_{it} - \tilde{Y}_{it_0} | P(X), D_i = 0, X)$$

where $\hat{E}$ is the empirical counterpart of the expectation, t is a period after the treatment and t_0 is a period before the treatment.

5 Data and Estimation Results

The Longitudinal Survey of Immigrants to Australia (LSIA) was commissioned in the early 1990s to fulfil the need to have better information on the settlement of new migrants than those available through censuses. It is based on a representative sample of 5% of migrants/refugees and comprises three surveys over a period of almost 10

number of time period necessary for an efficient estimation of factors.

years. We use the responses of primary applicants from the LSIA1 and LSIA2 only, which collect information from migrants entering Australia just before and after a major immigration policy reform.

The comparison of migrants from the samples in cohort 1 and 2 reveals a slight increase in the business stream group, an increase in the humanitarian stream and substantial decreases in the Independent and Concessional Family streams (see Table 1). Migrants in the second cohort originate from a wider variety of countries relative to the first cohort, with increased arrivals from Oceania, New Zealand and the Pacific Islands, North East Asia (which includes China and Hong Kong), and Southern and Central Asia (see Table 2). The second cohort had in general better education levels than the first cohort, especially with higher university degrees at the time of arrival (Table 3), and higher proportions of migrants that are not in the labour force immediately after resettlement and migrants undertaking further education in Australia (Table 4).

Tables 5 and 6 provide summary statistics on migrants' remittance behaviour across cohorts. For each visa category the percentage of migrants remitting in the second cohort is less than that in the first cohort. Table 5 also shows that the average remittances are larger in cohort 1 upon arrival (except for those on concessional family visa) and this trend persists over time. Table 6 reports proportion remitting and the amount remitted according to the country of origin, highlighting substantial variations. The immigrants with the highest propensity to remit are from Oceania (20.3% in cohort 1 and 8.8% in cohort 2). They are also the immigrants remitting the highest average amount (A$343.0 in cohort 1 and A$241.7 in cohort 2). Those remitting least are from North East Asia, Europe and former Russia and the Middle East and North Africa (about 5%-7% in cohort 1 and 1%-2.5% in cohort 2). Cohort 1 migrants remit more on average upon arrival but cohort 2 migrants appear to catch up within 12 months.

Looking across waves, it appears that the percentage of migrants from Southern and Eastern Europe, South Asia and Middle East who remit increases faster in cohort 1 compared to those from other regions. In cohort 2, we observe the highest increases for individuals from the Middle East, South America or Africa, Southern Europe and South East Asia and the increase across waves is modest.

Table 7 reports the unconditional mean and standard deviation for the working sample, by visa type, before and after tighter immigration criteria were introduced.

By focusing momentarily on average values between affected and not-affected, the two groups are different with respect to several demographic characteristics (gender, marital status), education (the affected are on average better educated) and countries of origin (the affected come from a wider group of countries), highlighting the different motivations for migration.[3] Those on Independent and Concessional visa (the affected group) are overwhelmingly admitted through the point system, and are therefore economic migrants with high prospects of immediate employability but limited or no host country support from family, employers or local institutions. In contrast, those on Preferential and Humanitarian visas, as well as employer-sponsored migrants, comprise a more heterogeneous group of settlers with a high incidence of family reunification.

The effect of policy change is the difference in the means of before and after columns, and is separately reported for assessed and not-assessed. The tighter immigration policy appears to reduce the probability to remit (by about 5%) and substantially raise the amount remitted in both groups. As highlighted by previous literature (Cobb-Clark (2003), Chiswick and Miller (2006), Thapa and Gorgens (2006), Mahuteau and Junankar (2008)), it substantially affects the gender composition and education level of primary applicants in the affected visa categories (more female and more university-educated migrants), as well as their probability of employment (higher). Much more subdued is the corresponding policy effect among the non-affected categories.

Tables 8 and 9 present estimation results of the effect of the change in immigration policy on the extensive and intensive margins of remittances. The effect is estimated using diff-in-diffs (Model OLS and Model RE), diff-in-diffs with heterogeneous time-varying trends by groups (Model OLS C.) and conditional diff-in-diffs with heterogeneous time-varying trend by groups (Model OLS C. and Match). If we assume that all group specific trends are the same and the control and treatment are similar, then the

[3]The affected and not-affected groups are broadly similar with respect to the country of origin (about half are born in an English-speaking country and member state of the Commonwealth) and post-migration residential choices, with three quarters of migrants settling in or around Australia's two main cities, Sydney and Melbourne. They are also similar with respect to average age (early 30s) and English language skills (almost all respondents use English in the interview). Other demographic characteristics differ significantly between the two groups. The not-affected are characterised by a higher number of women and married individuals.

appropriate procedure is the Diff-in-Diffs. Model OLS and Model RE suggest that the change in immigration policy has no effect on the probability to remit (note that the interaction of affected and post-reform cohort is our main covariate as it captures the policy effect). However, when we account for time-varying heterogeneous trends and use matching individuals on their propensity score (which is our main model), the result shows that the policy change influences positively and statistically significantly the probability to remit (5.6%), but not the amount of remittances sent (which is negative but insignificant). As we control for income and education, the results on the effect of the policy change perhaps capture an increase in the perceived level of riskiness of settlement among immigrants in the second cohort. Given that we control for a wide range of individual characteristics, including education levels, labour force status, region of origin, and other covariates (see full results in Appendix B), this result partly captures elements related to individuals becoming more risk averse and saw remittances as a possible insurance policy.

Post-reform on its own - row 2 Tables 8 and 9 - captures the cohort effect. Therefore, ceteris paribus, the probability of second cohort to remit is negative, but those who do remit send higher amount (76.7%) than the first cohort. This might be related to repaying a higher amount of loan (due to higher education cost in the developing country) to the extended family. Alternatively, it could be that they are investing back home where the cost of setting up a business is higher in 2000s than it was in the early 1990s. The last row captures the effect on those who were affected by the policy change. This result is qualitatively the same as the interaction result in the first row. Looking at the full results in Tables B1 and B2 in Appendix B, it is clear that the lowest income group has the lowest probability to remit, but when they do remit they send higher level of remittances than their better paid counterparts. This strengthens our argument related to either them finding the host country more risky or perhaps they borrowed more to fulfil the stringent entry conditions and therefore have to pay back that loan. Education seems to have no impact on the amount remitted, though it is clear that the probability to remit is lowest for the highest educated.

6 Conclusions

Several aspects could influence the remittance flows from migrants to their families and friends in the sending countries. These range from repaying of loans to fund migration costs, altruism towards those who remain in the country of origin or indeed because of selfish reasons to curry favour with those remaining back home in case the migration experiment ends up in a failure (i.e., equivalent to taking insurance against bad economic outcome).

Our main contribution to the literature is the use of a novel econometric approach that tackles the challenges of measuring the effect of a policy change using longitudinal data from two cohorts of migrants who entered before and after new migration conditions were put in place. We extended the literature on conditional difference-in-differences and evaluation of treatment effect with panel data in the presence of interactive fixed effects by assuming that there is an interactive fixed effect in the data generating process. We also allowed for differences in the support of observable characteristics or in how they affect the outcome in the treated and untreated sample. Our results point towards the direction of an overall positive relationship between stringent entry policy and the incidence of remittance flows. Those who entered in the second cohort, regardless of the origin region or type of job obtained, had a higher probability to remit compared to the first cohort, with no discernible impact on the level of remittances.

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, 72(1), 1–19.

Abadie, A., A. Diamond, and J. Hainmueller (2012): "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association.*

——— (2015): "Comparative politics and the synthetic control method," *American Journal of Political Science*, 59(2), 495–510.

Abadie, A., and J. Gardeazabal (2003): "The economic costs of conflict: A case study of the Basque Country," *The American Economic Review*, 93(1), 113–132.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2013): "Panel data models with multiple time-varying individual effects," *Journal of Econometrics*, 174(1), 1–14.

Amuedo-Dorantes, C., and S. Pozo (2006): "Remittances as insurance: evidence from Mexican immigrants," *Journal of Population Economics*, 19(2), 227–254.

Bai, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71(1), 135–171.

——— (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77(4), 1229–1279.

Batista, C., and J. Umblijs (2016): "Do migrants send remittances as a way of self-insurance?," *Oxford Economic Papers*, 68(1), 108–130.

Bollard, A., D. McKenzie, M. Morten, and H. Rapoport (2011): "Remittances and the brain drain revisited: The microdata show that more educated migrants remit more," *The World Bank Economic Review*, 25(1), 132–156.

Chiswick, B. R., and P. W. Miller (2006): "Immigration to Australia during the 1990s: Institutional and labour market influences," in *Public Policy and Immigrant Settlement*, ed. by D. A. Cobb-Clark, and S. Khoo. Edward Elgar, Cheltenham, UK.

Cobb-Clark, D. A. (2003): "Public policy and the labor market adjustment of new immigrants to Australia," *Journal of Population Economics*, 16(4), 655–681.

de Chaisemartin, C., and X. D'Haultfoeuille (2015): "Fuzzy differences-in-differences," Discussion paper, cemmap working paper, Centre for Microdata Methods and Practice.

Gobillon, L., and T. Magnac (2016): "Regional policy evaluation: Interactive fixed effects and synthetic controls," *Review of Economics and Statistics*, 98(3), 535–551.

Hawthorne, L. (2005): "Picking Winners: The recent transformation of Australia's skilled migration policy," *International Migration Review*, 39(3), 663–696.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The Review of Economic Studies*, 64(4), 605–654.

Hsiao, C., H. Steve Ching, and S. Ki Wan (2012): "A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong kong with mainland China," *Journal of Applied Econometrics*, 27(5), 705–740.

Mahuteau, S., and P. R. Junankar (2008): "Do migrants get good jobs in Australia? The role of ethnic networks in job search," *Economic Record*, 84(s1).

Niimi, Y., C. Ozden, and M. Schiff (2010): "Remittances and the brain drain: skilled migrants do remit less," *Annals of Economics and Statistics/Annales dâĂŹÉconomie et de Statistique*, pp. 123–141.

Ouyang, M., and Y. Peng (2015): "The treatment-effect estimation: A case study of the 2008 economic stimulus package of China," *Journal of Econometrics*, 188(2), 545–557.

Piracha, M., and Y. Zhu (2012): "Precautionary savings by natives and immigrants in Germany," *Applied Economics*, 44(21), 2767–2776.

Thapa, P. J., and T. Gorgens (2006): "Finding Employment after Migration: How Long does it Take?," in *Public Policy and Immigrant Settlement*, ed. by D. A. Cobb-Clark, and S. Khoo. Edward Elgar, Cheltenham, UK.

Watson, I. (1996): *Opening the glass door: overseas-born managers in Australia.* Australian Government Publishing Service.

Table 1: Migrants by visa categories

Major Visa Category	**LSIA 1**	**LSIA2**
Business (%)	3.44	5.79
Family (%)	65.33	62.18
Humanitarian (%)	14.12	7.44
Independent (%)	17.12	24.59

Table 2: Composition of immigrant population for cohort 1 (1993-1995) cohort 2 (1998-2000); by region of birth

	Migrant population arrived with cohort 1		Migrant population arrived with cohort 1	
Region of Birth	Freq.	Percent	Freq.	Percent
Oceania	1,858	2.48	959	2.96
Great Britain	11,454	15.27	4,752	14.66
South Europe	6,53	8.71	2,293	7.07
Western Europe	2,918	3.89	1,187	3.66
Eastern Europe	3,505	4.67	993	3.06
Middle East	7,383	9.84	3,356	10.35
South East Asia	16,305	21.74	5,879	18.14
North East Asia	10,245	13.66	5,259	16.23
South Asia	7,208	9.61	3,437	10.6
North America	2,685	3.58	1,112	3.43
Central, South America	1,335	1.78	423	1.3
Africa	3,563	4.75	2,764	8.53
Total	**74,99**	**100**	**32,415**	**100**

Table 3: Highest level of education completed by arrival cohort (percent of the immigrant population)

Highest level of qualification completed	Cohort 1 Wave 1	Cohort 1 Wave2	Cohort 1 Wave 3	Cohort 2 Wave 1	Cohort2 Wave 2
Primary school	4.98	4.93	4.68	3.67	3.95
Secondary school	33.22	33.72	28.67	25.56	26.08
Trade	7.28	7.16	8.12	6.7	6.97
technical/professional	21.55	21.46	22.48	20.64	20.69
Undergraduate Degree	20.98	20.8	20.11	23.84	21.82
Post graduate degree	4.95	4.88	6.27	5.44	6.76
Higher degree	7.04	7.05	9.66	14.16	13.73
Total	**100**	**100**	**100**	**100**	**100**

Table 4: Labour Force Status in Australia by wave of interview and cohort (population)

	Cohort 1			Cohort 2	
	wave 1	**wave 2**	**wave 3**	**wave 1**	**wave 2**
Business Owner (employing others, self employed)	**2.97%**	**4.93%**	**6.40%**	**5.05%**	**8.42%**
Business owner, self employed	2.28%	3.54%	4.55%	3.69%	6.41%
Business owner employing others	0.68%	1.38%	1.86%	1.36%	2.02%
Wage earner	**31.83%**	**42.84%**	**48.44%**	**45.80%**	**53.55%**
Other employed	**0.48%**	**0.84%**	**0.13%**	**0.29%**	**0.26%**
Unemployed looking for full time or part time job	**22.63%**	**13.94%**	**10.33%**	**11.18%**	**7.20%**
Unemployed looking for full time job	20.43%	12.22%	8.39%	9.28%	5.53%
Unemployed looking for part time job	2.20%	1.72%	1.94%	1.89%	1.66%
Student	**16.21%**	**13.57%**	**6.74%**	**15.15%**	**8.09%**
Not in the labour force	**25.87%**	**23.88%**	**27.95%**	**22.53%**	**22.48%**

Table 5: Percent of migrant population remitting by interview wave (population)

	Percent remitting (population)	Average value of remittances in AUD (population)
cohort 1 wave 1	7.76%	93.11
cohort 1 wave 2	22.00%	348.28
cohort 1 wave 3	31.05%	770.71
cohort 2 wave 1	4.20%	76.08
cohort 2 wave 2	13.40%	355.26

Table 6: Amount of remittances sent abroad by time since arrival, AUD 2000 (population)

Remittances ($)	Observations	Mean	Std. Dev.
Cohort 1			
6 mths or less	4379	85.88	566.86
6 to 12 mths	794	146.89	755.74
12 to 18 mths	3491	350.12	1727.31
18 to 24 mths	980	337.45	1154.01
24 mths or more	3757	766.81	2583.35
Cohort 2			
6 mths or less	2329	69.62	472.09
6 to 12 mths	794	93.76	497.28
12 to 18 mths	1713	367.88	1612.56
18 to 24 mths	917	336.09	1145.83
24 mths or more	17	30.86	167.24

Table 7: Means and differences: Balance tests

	Affected[a]			*Not Affected*[b]		
Variables	**Before**	**After**	***Difference***	**Before**	**After**	***Difference***
Probability remit	.146 (.354)	.107 (.309)	*-.039****	.126 (.331)	.068 (.253)	*-.058****
Amount remitted	6.950 (.938)	7.590 (.735)	*.640****	6.764 (.857)	7.396 (.870)	*.632****
Age	33.2 (6.52)	33.0 (6.63)	*-.02*	33.8 (9.85)	35.5 (10.27)	*1.7****
Female	.286 (.452)	.347 (.476)	*.061****	.503 (.500)	.497 (.500)	*-.006*
Married	.594 (.491)	.611 (.488)	*.017*	.790 (.407)	.748 (.434)	*-.042****
N household	2.58 (.653)	2.55 (.647)	*-.03**	2.59 (.570)	2.60 (.561)	*.01*
N relatives HC	5.88 (2.84)	5.77 (2.87)	*-.11*	5.24 (2.89)	4.90 (2.91)	*-.34****
N relative AU	.776 (1.37)	.617 (1.16)	*-.159****	1.34 (2.11)	1.70 (2.24)	*.36****
Previous visits	.454 (.498)	.662 (.473)	*.208****	.490 (.500)	.504 (.500)	*.014*
Education HS-	.411 (.492)	.318 (.466)	*-.093****	.682 (.466)	.654 (.476)	*-.028***
BA	.339 (.473)	.337 (.473)	*-.002*	.181 (.385)	.191 (.392)	*.010*
Postgraduate	.099 (.299)	.111 (.314)	*.012*	.048 (.213)	.046 (.210)	*-.002*
Higher	.150 (.358)	.234 (.424)	*.084****	.090 (.286)	.109 (.312)	*.019****
Interview E	.697 (.460)	.679 (.467)	*-.018*	.684 (.465)	.652 (.476)	*-.032****
Participates	.809 (.393)	.882 (.323)	*.073****	.591 (.492)	.586 (.492)	*-.005*
Income: low	.125 (.331)	.143 (.350)	*.018*	.308 (.462)	.289 (.453)	*-.019**
Medium-L	.281 (.449)	.112 (.315)	*-.169****	.313 (.464)	.238 (.426)	*-.075****
Medium-H	.281 (.450)	.190 (.393)	*-.091****	.184 (.388)	.173 (.378)	*-.011*
High	.294 (.456)	.548 (.498)	*.254****	.178 (.383)	.280 (.449)	*.102****
COB: NW Europe	.215 (.411)	.219 (.413)	*.004*	.187 (.344)	.159 (.366)	*-.028****
SE Europe	.110 (.313)	.066 (.418)	*-.044****	.168 (.373)	.195 (.396)	*.027****
MENA	.062 (.242)	.016 (.127)	*-.046****	.132 (.338)	.086 (.281)	*-.046****
SE Asia	.140 (.347)	.156 (.363)	*.016*	.229 (.420)	.169 (.375)	*-.060****
E Asia	.186 (.389)	.179 (.384)	*-.007*	.120 (.325)	.151 (.358)	*.031****
S Asia	.148 (.355)	.154 (.362)	*.006*	.043 (.202)	.044 (.205)	*.001*
N America	.012 (.111)	.015 (.124)	*.003*	.038 (.191)	.059 (.235)	*.021****
Latin America	.060 (.238)	.023 (.150)	*-.037****	.061 (.239)	.053 (.225)	*-.008*
Africa	.049 (.216)	.094 (.292)	*.045****	.047 (.212)	.057 (.233)	*.010***
Oceania	.018 (.132)	.076 (.265)	*.58****	.025 (.158)	.025 (.156)	*.0*
Gini coefficient	.397 (.088)	.400 (.096)	*.003*	.386 (.082)	.378 (.086)	*-.008****
Network	.051 (.086)	.055 (.088)	*.004****	.029 (.059)	.033 (.066)	*.004****
GDP: low	.265 (.442)	.284 (.451)	*.019*	.241 (.428)	.273 (.445)	*.032****
Medium-L	.222 (.415)	.210 (.408)	*-.012*	.261 (.439)	.226 (.418)	*-.035****
Medium-H	.187 (.390)	.200 (.400)	*.013*	.240 (.427)	.222 (.415)	*-.018**
High	.326 (.469)	.305 (.460)	*-.021*	.258 (.438)	.279 (.449)	*.021***
N	2,491	1,093		4,347	3,023	

Notes: Only the first two waves of each cohort are used for before/after comparability.

Standard deviation in parentheses. a Includes (i) Family concessional and (ii) skilled independent visa categories.

b Includes: (i) Family preferential, (ii) employer nomination and (iii) humanitarian visa categories.

Table 8: Probability of remittances

	Model OLS	Model RE	Model OLS C.	Model OLS C. and Match
Affected x post-reform cohort	0.00379	0.00246	0.0563***	0.0426***
	(0.26)	(0.17)	(4.34)	(4.98)
Post-reform cohort	-0.112***	-0.111***	-0.0810***	-0.0654***
	(-15.83)	(-15.85)	(-9.16)	(-9.30)
Affected group	-0.0136	-0.0148	0.106***	-0.00133
	(-1.38)	(-1.52)	(6.99)	(-0.04)
Observations	15436	15436	15436	.
R^2	0.066		0.083	0.048

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Value of remittances

	Model OLS	Model RE	Model OLS C.	Model OLS C. and Match
Affected x post-reform cohort	-0.151	-0.151	0.0277	-0.0595
	(-1.50)	(-1.52)	(0.28)	(-0.35)
Post-reform cohort	0.443***	0.452***	0.768***	0.767***
	(7.09)	(7.47)	(12.32)	(6.14)
Affected group	-0.0130	-0.00636	-0.111	-0.0229
	(-0.21)	(-0.10)	(-1.79)	(-0.19)
Observations	2335	2335	2335	.
R^2	0.157		0.187	0.344

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A Monte Carlo experiments

This section proposes some Monte Carlo experiments to compare the use of classic diff-in-diffs, conditional diff-in-diffs and the mixture of conditional diff-in-diffs and factor estimation. Our simulation uses a data generating process that induce the breakup of the parallel trend assumption. The violation of the parallel trend assumption comes from two independent sources.

The data generating process is given by a linear model:

$$Y_{it} = \alpha D_i I_t + \beta_i X_{it} + \gamma_i \delta_t + U_{it} \tag{12}$$

where the time effect δ_t is assumed to be represented by a fixed number of factors L and the treatment effect, α, is homogeneous across individuals. We include additive individual and time effects, i.e. $\gamma_i = (\gamma_{i1}, \gamma_{i2}, ...)$ and $\delta_t = (1; \delta_{t1}; \delta_{t2}...)'$. Exogenous characteristics X_{it} are correlated with individual treatment group. The effect of these exogenous characteristics is also assumed to be heterogeneous across individuals. This representation enables us to account for the imperfect compliance between the treated and the control group. Residuals U_{it} are independently and identically distributed and each of them is drawn from a mean zero and variance 1 normal distribution.

The number of treated individuals is N_1 (respectively total, N) and the numbers of periods before treatment, T_D, (respectively total, T). In our baseline experiment, $(N_1; N) = (80; 200)$, $(T_D; T) = (4; 8)$. We also run experiments with $(N_1; N) = (10; 200)$ and $(N_1; N) = (120; 200)$.

Our main objective is to see how the difference between the control and the treatment groups, in terms of their exogenous characteristics and in terms of their effect on the outcome, affects the estimation of the causal effect. In our baseline experiment, we assume that the support of β_i and X_{it} are the same for the treated individuals as for the untreated individuals. The random variables are drawn respectively from a uniform distribution on $[0; 1]$ and from a normal distribution of mean 0 and variance 1. In an alternative experiment, we construct overlapping supports for treated and untreated individuals. A shift in the support of treated units by 0.5 helps to create the overlap. In another experiment, supports of treated and untreated individuals are almost disjoint by shifting the support of treated individuals by 1. Because the original support is $[0; 1]$, this means that the intersection of the supports of treated and non-treated individuals is now reduced to one point for β_i and to a low probability of common support for X_{it}.

Our experiments evaluate five procedures:

1. A classic diff-in-diffs: The estimator of the treatment effect is obtained by assuming parallel trend assumption.
2. An approach where we estimate parameter α using Bai's method on a linear model (Bai (2009)). Bai's method is used to estimate the time-varying trend and the treatment variable and the exogenous characteristics are used as regressors.
3. The synthetic control approach (Synthetic Control). The treatment effect is obtained by following the technique of synthetic controls proposed by Abadie and Gardeazabal (2003) and further explored by Gobillon and Magnac (2016).
4. A matching approach (Matching). We use individual's exogenous characteristics from which a propensity score, discriminating treated and untreated individuals, is computed. A probit specification for the score is used to construct the counterfactual outcome in the treated group in the absence of treatment at periods $t > T_D$ using the kernel method (see Gobillon and Magnac (2016) for details.)
5. An approach where the Bai's method is mixed with matching (Matching-Bai). We use the same matching method introduced in the matching approach, but now the outcome of interest is the residual of Y_{it} obtained by removing the time-varying effects estimated by Bai's method.

In our simulations, the number of iterations for the Monte Carlo is 1000 and $\alpha = 0.3$.

Table A1: Properties of Treatment effect estimators, $\alpha = 0.3$, replications 1000, $N_1 = 10$

	c=0		c=0.5		c=1	
	Mean	SD	Mean	SD	Mean	SD
Diff-in-Diffs	0.299	0.231	0.304	0.255	0.359	0.319
Bai 2009	0.299	0.231	0.304	0.255	0.359	0.319
Synthetic Control	0.275	0.506	0.307	0.495	0.313	0.514
Matching	0.275	0.506	0.307	0.495	0.313	0.514
Bai and Matching	0.276	0.501	0.303	0.505	0.325	0.525

Table A2: Properties of Treatment effect estimators, $\alpha = 0.3$, replications 1000, $N_1 = 80$

	c=0		c=0.5		c=1	
	Mean	SD	Mean	SD	Mean	SD
Diff-in-Diffs	0.299	0.081	0.433	0.100	0.869	0.193
Bai 2009	0.299	0.082	0.433	0.101	0.870	0.195
Synthetic Control	0.303	0.231	0.291	0.215	0.299	0.236
Matching	0.303	0.231	0.291	0.215	0.299	0.236
Bai and Matching	0.306	0.228	0.291	0.209	0.294	0.226

Table A3: Properties of Treatment effect estimators, $\alpha = 0.3$, replications 1000, $N_1 = 120$

	c=0		c=0.5		c=1	
	Mean	SD	Mean	SD	Mean	SD
Diff-in-Diffs	0.302	0.067	0.376	0.078	0.661	0.163
Bai 2009	0.302	0.067	0.376	0.078	0.662	0.164
Synthetic Control	0.304	0.165	0.509	0.187	0.996	0.392
Matching	0.301	0.217	0.282	0.223	0.213	0.240
Bai and Matching	0.301	0.215	0.248	0.218	0.115	0.227

Simulation results are reported in Table A1, A2 and A3. We report the empirical mean and standard error of each estimator for each Monte-Carlo experiment. In all tables, in the case of perfect compliance ($c = 0$), column 1 results show that the estimated treatment parameter exhibits little bias for all methods controlling for difference in the control and treatment groups: Synthetic Control, Matching and Bai and Matching. On the other hand, diff-in-diffs and Bai 2009 are unbiased. Moreover, when the treatment group size is small, the bias of Synthetic Control, Matching and Bai and Matching is larger. However, with a treatment group larger than the control group the bias of methods using matching becomes smaller than the other.

The standard error of the estimator is larger when using Synthetic Control, Matching and Bai and Matching methods than when using the diff-in-diffs and Bai (2009) methods. The reason for this lies in the use of multiple estimation steps. Interestingly, the standard deviation of Bai and Matching is slightly smaller than that of Matching in all the cases.

In the case of imperfect compliance with overlap support($c = 0.5$), as expected, the diff-in-diffs and Bai (2009) become biased. The bias first increases with the number of treated individuals and slightly decreases when the treated population is larger than the untreated. Synthetic Control, Matching and Bai and Matching are unbiased for small ($N_1 = 10$) and relatively large treated group ($N_1 = 80$). But when the number of treated individuals is large ($N_1 = 120$), only Matching method has a small bias.

As the difference between the treated and control group increases (c=1), the biases of diff-in-diffs and Bai (2009) methods also increase. Matching, Synthetic Control and Matching-Bai all have good bias properties for relatively large treated group ($N_1 = 80$). In the extreme case of few treated individuals ($N_1 = 10$) and many treated individuals ($N_1 = 120$), all procedures are biased.

An interesting conclusion in this analysis, which could have empirical application, is that if the control and the treatment groups are not similar enough and if the treated population is large relative to the total population, Synthetic Control, Matching or Bai and Matching should be used to estimate the effect of the treatment. We are , therefore, going to present Bai and Matching and diff-in-diffs in our empirical application.

B Full tables of results

Table B1: Value of remittances

	Model OLS	Model RE	Model OLS C.	Model OLS Match	Model OLS C. and Match
Affected x post-reform cohort	-0.151	-0.151	0.0277	-0.190	-0.0595
	(-1.50)	(-1.52)	(0.28)	(-1.11)	(-0.35)
Post-reform cohort	0.443***	0.452***	0.768***	0.487***	0.767***
	(7.09)	(7.47)	(12.32)	(4.04)	(6.14)
Affected group	-0.0130	-0.00636	-0.111	0.104	-0.0229
	(-0.21)	(-0.10)	(-1.79)	(0.84)	(-0.19)
income per week <$155	0.125*	0.115	0.169**	0.559**	0.601***
	(1.98)	(1.84)	(2.65)	(2.97)	(3.39)
income per week [$385-$675]	0.187**	0.170*	0.167*	0.349	0.359
	(2.62)	(2.47)	(2.34)	(1.76)	(1.76)
income per week > $675	0.592***	0.580***	0.546***	0.563**	0.544**
	(7.25)	(7.32)	(6.70)	(2.71)	(2.61)
cmalfs1==2	-0.270***	-0.292***	-0.238**	-0.227	-0.169
	(-3.62)	(-4.09)	(-3.18)	(-1.19)	(-0.90)
cmalfs1==6	-0.265**	-0.277**	-0.288**	-0.279	-0.478*
	(-2.91)	(-3.12)	(-3.16)	(-1.17)	(-1.96)
Education: BA	0.0544	0.0404	0.0486	0.0822	0.0966
	(0.65)	(0.48)	(0.57)	(0.42)	(0.48)
Postgraduate	0.163**	0.147*	0.131*	0.0402	-0.00881
	(2.67)	(2.45)	(2.14)	(0.32)	(-0.07)
Higher education	0.101	0.0987	0.0754	-0.118	-0.173
	(1.68)	(1.66)	(1.26)	(-0.85)	(-1.23)
language interview is English	-0.0757	-0.0812	-0.0740	-0.290*	-0.280*
	(-1.75)	(-1.90)	(-1.72)	(-2.40)	(-2.26)
SE Asia	-0.0597	-0.0526	-0.0773	0.165	0.153
	(-0.57)	(-0.52)	(-0.72)	(0.69)	(0.64)
E Asia	0.112	0.127	0.0775	0.425	0.398
	(0.97)	(1.15)	(0.66)	(1.65)	(1.57)
S Asia	0.0846	0.0992	0.0566	0.265	0.264
	(0.82)	(1.00)	(0.54)	(1.17)	(1.17)
N America	0.538***	0.574***	0.515***	1.050***	1.046***
	(4.04)	(4.47)	(3.86)	(3.98)	(3.93)
Latin America	0.148	0.163	0.127	0.550*	0.523*
	(1.33)	(1.51)	(1.12)	(2.24)	(2.15)
Africa	0.00441	0.0140	0.0258	0.0605	0.123
	(0.04)	(0.12)	(0.20)	(0.26)	(0.53)
Oceania	-0.0180	0.00198	-0.0387	0.0480	0.0465
	(-0.14)	(0.02)	(-0.30)	(0.21)	(0.20)
age at migration	0.00387	0.00367	0.00360	0.0150	0.0118
	(1.30)	(1.25)	(1.19)	(1.94)	(1.50)
Constant	6.562***	6.570***	6.335***	5.987***	5.918***
	(47.65)	(48.68)	(45.30)	(15.75)	(15.37)
Observations	2335	2335	2335	.	.
R^2	0.157		0.187	0.282	0.344

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B2: Probability of remittances

	Model OLS	Model RE	Model OLS C.	Model OLS C. and Match
Affected x post-reform cohort	0.00379	0.00246	0.0563***	0.0426***
	(0.26)	(0.17)	(4.34)	(4.98)
Post-reform cohort	-0.112***	-0.111***	-0.0810***	-0.0654***
	(-15.83)	(-15.85)	(-9.16)	(-9.30)
Affected group	-0.0136	-0.0148	0.106***	-0.00133
	(-1.38)	(-1.52)	(6.99)	(-0.04)
income per week <$155	-0.0189*	-0.0173*	-0.0148*	0.0140**
	(-2.39)	(-2.27)	(-1.99)	(3.02)
income per week [$385-$675]	0.0587***	0.0611***	0.0559***	0.0768***
	(5.74)	(6.17)	(5.59)	(11.07)
income per week > $675	0.0275*	0.0341**	0.0159	0.0515***
	(2.53)	(3.25)	(1.57)	(8.58)
cmalfs1==2	-0.0715***	-0.0768***	-0.0613***	-0.0212***
	(-7.35)	(-8.21)	(-6.74)	(-3.87)
cmalfs1==6	-0.0594***	-0.0607***	-0.0579***	-0.0341***
	(-5.65)	(-5.97)	(-5.88)	(-5.21)
Education: BA	0.00348	0.00494	0.00690	-0.0399***
	(0.25)	(0.37)	(0.57)	(-4.66)
Postgraduate	-0.0110	-0.00986	-0.00778	-0.0320***
	(-1.25)	(-1.15)	(-1.02)	(-4.53)
Higher education	-0.00278	-0.00134	-0.00196	-0.0411***
	(-0.29)	(-0.14)	(-0.23)	(-5.22)
language interview is English	-0.00837	-0.00900	-0.00689	-0.0241***
	(-1.34)	(-1.49)	(-1.14)	(-5.43)
SE Asia	-0.130***	-0.128***	-0.128***	-0.0280*
	(-5.05)	(-5.11)	(-6.14)	(-2.36)
E Asia	-0.0702**	-0.0654*	-0.0730**	-0.0227
	(-2.58)	(-2.47)	(-3.28)	(-1.74)
S Asia	-0.0144	-0.0137	-0.0148	0.0296*
	(-0.53)	(-0.52)	(-0.68)	(2.29)
N America	-0.126***	-0.123***	-0.125***	-0.0450***
	(-4.76)	(-4.77)	(-5.82)	(-3.61)
Latin America	0.0181	0.0185	0.0166	0.00541
	(0.64)	(0.67)	(0.71)	(0.42)
Africa	-0.102***	-0.100***	-0.0998***	-0.0247
	(-3.66)	(-3.71)	(-4.42)	(-1.78)
Oceania	-0.0743**	-0.0730**	-0.0728**	-0.0396**
	(-2.59)	(-2.61)	(-3.12)	(-3.03)
age at migration	-0.00198***	-0.00194***	-0.00202***	0.0000529
	(-5.70)	(-5.72)	(-6.83)	(0.22)
Group specific trends	No	No	Yes	Yes
Constant	0.371***	0.364***	0.367***	0.141***
	(12.59)	(12.70)	(14.86)	(9.71)
Observations	15436	15436	15436	.
R^2	0.066		0.083	0.048

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$