

Piracha, Matloob; Tani, Massimiliano; Tchuente, Guy

Working Paper

Immigration Policy and Remittance Behaviour

GLO Discussion Paper, No. 94 [rev.]

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Piracha, Matloob; Tani, Massimiliano; Tchuente, Guy (2021) : Immigration Policy and Remittance Behaviour, GLO Discussion Paper, No. 94 [rev.], Global Labor Organization (GLO), Essen

This Version is available at:
<http://hdl.handle.net/10419/232515>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Immigration Policy and Remittance Behaviour^{*}

Matloob Piracha[†] Massimiliano Tani[‡] Guy Tchuente[§]

March 2021

Abstract

This paper analyses the remittance behaviour of two cohorts of migrants who entered Australia before and after a policy change implemented in the 1990s, which tightened the entry requirements for a subgroup of applicants. We use a mix of a conditional difference-in-differences and OLS estimator accounting for the presence of interactive fixed-effects to address the challenge of evaluating the impact of policy change using data drawn from two distinct migrant samples, deriving the conditions to obtain a consistent estimator. We show two results: one due to policy change and the other due to change in the composition of migrants. The two results capture different aspects of remittance behaviour.

Keywords: Immigration, average treatment effect on the treated, difference-in-differences.

JEL Classification: C13, F22, F24, J61.

^{*}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, it is significantly different from the other one as our main contribution in this paper is the identification methodology needed for the kind of data available to us. The authors thank participant to the SCSE 2018, CEA 2018 and, CESG 2018 for their useful comments.

[†]University of Kent. Corresponding author: M.E.Piracha@kent.ac.uk

[‡]University of New South Wales and IZA

[§]University of Kent and MaGHIC

1 Introduction

The main objective of this paper is to contribute to the literature on the remittance behaviour of migrants. Using longitudinal data from Australia, we analyse the effect of a change in immigration policy on the probability (extensive margin) and amount (intensive margin) of remittances. However, because the data were collected on migrant cohorts settling just before and after the year in which the policy change took place, evaluating the impact of policy on remittance behaviour with the standard estimation procedure (difference-in-differences) incurs the risk of omitting variables that are relevant only to one cohort. If this occurs, the resulting estimates are likely to be biased. We address this challenge by extending the conditional difference-in-differences methodology, and apply the derived unbiased estimator to measure the effect of the policy change.

Migrants' remittance flows to their families and friends in the sending countries reflect several influences.¹ 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 migration turns out to be a failure - akin to taking insurance against bad economic outcome. Although migrants' remittances have received widespread attention because of their importance for the economies of developing nations, their patterns have been analysed with respect to migrants' legal status, labour market conditions, and/or characteristics of the family left behind (see Rapoport and Docquier (2006); Piracha and Zhu (2012) and; Batista and Umblijs (2016)). In contrast, the influence of the host country's immigration policy on remittance pattern has been under-researched. Yet, this change in institutional settings can have widespread implications for the underlying factors determining migrants' remitting choices, as it sets the main characteristics of immigrants who can enter the country. For instance, most of the migrant-receiving countries have instituted stringent immigration policies to attract relatively higher skilled migrants, which has raised concerns among many that the resultant brain drain will not be properly compensated for,

¹For a detailed survey of remittances and motivations to remit, see Rapoport and Docquier (2006); Yang (2011)

as the more educated tend to remit less.²

To the best of our knowledge, the only paper in the existing literature that connects immigration policy to remittance behaviour is Docquier, Rapoport, and Salomone (2012).³ They study the effect of 'restrictive' and 'selective' immigration policies in the destination countries using a bilateral remittances database. They find that, for a given country pair, relatively more skilled migrants remit less if the destination country uses a selective immigration policy. Our focus is different from theirs since we are interested in exploring how a change in immigration policy impacts migrant's remittance behaviour for a country that already admits economic migrants using a selection mechanism. More precisely, we analyse the migration policy change implemented in the mid-1990s in Australia. In 1996 Australia tightened economic migration policy with reference to skills (education and work experience) and knowledge of English for two applicant groups: those on 'independent' visa ('Skilled Independent') and the family-sponsored ('Concessional' family). The Skilled Independent are admitted via a point system favouring higher education, work experience, and younger demographics while Concessional family are those who have insufficient skills to be admitted as skilled independent but have family relations living in Australia providing a guarantee of financial support for their relative for a certain period of time. The remaining visa streams, which cover close family members ("Preferential" family), employer-sponsored workers ("Business") and Humanitarian migrants were unchanged. Besides raising the skill level of prospective

²There are a number of theoretical arguments for this (see Bollard, McKenzie, Morten, and Rapoport (2011)): that more educated are generally from wealthy families and therefore don't need to send remittances; that they tend to bring their families with them; and that they are less likely to return. However, there is a counter theoretical argument: that higher skilled tend to earn more and therefore remit a higher amount than the low skilled. Empirical evidence on the relationship between education and remittances is not clear. At a micro level, using household survey data for 11 destination countries, Bollard, McKenzie, Morten, and Rapoport (2011) show a positive relationship between education and amount remitted while they find mixed results for the likelihood of remittances. Using GSOEP data, Dustmann and Mestres (2010) find a negative effect of education on remittances after controlling for intentions to return and household composition at destination. At a macro level, Faini (2007) and Niimi, Ozden, and Schiff (2010) have found a negative relationship between remittance flows and education.

³For restrictive policies they use three proxies, namely bilateral guest worker program, proportion of refugees among migrants and proportion of females among migrants; and for selectivity they use the point system.

applicants, the policy change also removed income support for the first two years after migration as well as subsidies to attend English classes after settlement. As a result, new migrants affected by the policy change tended to be more educated and possessed better English language skills than previous cohorts (see Cobb-Clark (2003)).

The available longitudinal data (Longitudinal Survey of Immigrants in Australia – LSIA), however, are not suitable for applying a difference-in-differences approach. The cohorts entering Australia before and after the policy change have different observable, and likely unobservable, characteristics. Hence, even if in each cohort we have migrant subgroups that are affected (Skilled Independent and Concessional family) and not affected by the change (Preferential family, Business and Humanitarian), the use of a simple difference-in-differences imposes assumptions that are unlikely to be satisfied.⁴ Since the policy change was partly instituted to select more educated migrants with better host country language skills, homogeneity of the population before and after the policy is violated. One way to deal with the problem is to use conditional difference-in-differences⁵, which is the appropriate approach when a suitable control group cannot be identified directly from the observed variables. However, the conditional difference-in-differences approach works under the assumption of similar unobserved macro shocks for the control and the treatment group (or absence of selection on unobservables). This may not be the case in the Australian context as labour market conditions could be different for the treated and the control groups of different cohorts. A propensity score matching is likely to be problematic as well since matching methods do not account for variation in time trends between the control and the treatment groups.

To account for the presence of difference in macroeconomic trend in the treated and control groups as well as for the change in groups’ composition induced by the policy, we propose a methodology that bridges two distinct econometric approaches – conditional difference-in-differences and the evaluation of treatment effects with panel data in the presence of differential macroeconomic shocks – by allowing for an interactive fixed-effects. More precisely, we propose a correlated random slope model with interactive

⁴Cohort 1 arrived in Australia between September 1993 and August 1995 and contains three waves, with interviews conducted at 5, 17 and 41 months after arrival; Cohort 2 arrived between September 1999 and August 2000 and consists of two waves with interviews conducted at 5 and 17 months after arrival.

⁵Based on the propensity score as suggested by Heckman, Ichimura, and Todd (1997).

fixed-effects.⁶ This model encapsulates better the policy evaluation problem faced by an empirical researcher in the context where the change in policy induced a change in the treatment groups (captured by the random slope), and the groups may face heterogeneous unobserved macroeconomic shocks (captured by the interactive fixed-effects).

The correlated random slope model with interactive fixed-effects is used to discuss the identification condition of the average treatment effect on the treated (ATT). The practical question faced by an empirical researcher is whether to use only conditional difference-in-differences⁷, and/or only a method that accounts for interactive fixed-effects⁸ or a combination of both. We answer this question by analysing the generic bias of the difference-in-differences estimator when the true data generating process has a correlated (with treatment status) random slope and interactive fixed-effects. Our derivation shows that the bias of the classic difference-in-differences estimator is a non-separable combination of the biases coming from the correlated random slope and the interactive fixed-effects. This result implies that using an estimation method that removes only the interactive fixed-effect component may not always eliminate the bias of the difference-in-differences estimator. Moreover, using methods that only account for the composition effect (for example conditional difference-in-differences) may also fail to always eliminate the bias of the difference-in-differences. However, combining them will eliminate the bias. We evaluate the relevance of the use of a mix of conditional difference-in-differences with methods that could eliminate interactive fixed-effect component via a Monte Carlo experiment.

We provide estimates for the policy as well as the cohort effects. The effect attributable to the policy change, i.e., when we account for time-varying heterogeneous trends and use matching individuals on their propensity score, shows that the policy change influences positively and statistically significantly the probability to remit but not the amount of remitted. The cohort effect, which captures the changes in the com-

⁶The use of a correlated random slope relates our work to Wooldridge (2005) with the difference that the random slope relationship with the treatment variable is not restricted and we are in the presence of interactive fixed-effects. The interactive fixed-effects is added in the same spirit as in Gobillon and Magnac (2016).

⁷Following Heckman, Ichimura, and Todd (1997) or Hong (2013).

⁸For example estimating the effect using an OLS following Bai (2009).

position of migrants that are not attributable to the policy change shows that migrants in the second cohort are less likely to remit but those who do remit send a higher amount. This result reconciles some of the contradictory results found in the existing literature. We provide some intuition in Section 5.

The rest of the paper is organized as follows. A brief description of the 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 is explained in Section 4 while Section 5 presents the data and empirical results. Concluding remarks appear in the last section.

2 Background of Australian immigration policy

In 1996 the Australian government introduced a number of significant changes to the migration policy, affecting the skilled independent and concessional family reunification visa but not the humanitarian, employer-sponsored (business) and close family reunification (preferential family) 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 skilled 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 had the power to disqualify applicants from eligibility for skilled migration.

3 Identification of the Impact of the Mid-1990 Australian Migration Policy Changes.

To analyse the effect of the policy change we adopt a theoretical set-up that can be used to evaluate the effects of the change in policy outlined in Section 2. The model proposed corresponds to a panel data structure with correlated random slope and interactive fixed-effects. The model can be used to address two potential sources of violation of the parallel trend assumption that invalidates a simple application of the diff-in-diffs approach.

3.1 Institutional Framework Implication

The first challenge that we address is to restore the validity of the crucial parallel trend assumption, which underpins the diff-in-diffs approach, and which is likely violated in our data for two main reasons: (1) differential macroeconomic shocks faced by treated and control groups and/or (2) asymmetric changes in the composition of the treatment and control group as a result of the policy implementation.

Possible differential macroeconomics shocks can be accounted for in the classical diff-in-diffs model by adding interactive fixed-effects (see Gobillon and Magnac (2016)). A panel data model with interactive fixed-effects is the one where the unobserved heterogeneity is modeled as a product of an individual-specific factor loading and a set of unobserved factors. It aims to account for time-varying individual specific effects that could come from the individual reaction to macroeconomic shocks. The effect of time-varying exogenous variables is estimated in models with interactive fixed-effects by using estimators that account for presence of unobserved factors such as OLS procedure proposed by Bai(2009).

The unlikely stability of the treated and control groups is normally addressed in cross-section data with the use of propensity score matching (PSM), as it helps to construct appropriate treatment and control groups. In the context of diff-in-diffs, Heckman, Ichimura, and Todd (1997) showed that the conditional diff-in-diffs can achieve consistent estimation of the treatment effects of interest, but their results are obtained without interactive fixed-effects. We develop an estimation strategy for the proposed

correlated random slope model with interactive fixed-effects. Our estimation approach is a mix of estimation techniques accounting for the presence of unobserved factors and those accounting for the presence of composition-effects. Our identification strategy is complementary to the work on identification in cases of deviation from the classical diff-in-diffs assumption (for instance, Abadie, Diamond, and Hainmueller (2012), Gobillon and Magnac (2016), Ouyang and Peng (2015), Heckman, Ichimura, and Todd (1997), Abadie (2005), Hong (2013), and Sant’Anna and Zhao (2020) and Abadie, Diamond, and Hainmueller (2015)).⁹

3.2 Model set-up and Identification

We outline a formal presentation of the model using a potential outcome framework. Specifically, we consider a sample composed of N individuals observed at dates $t = 1, \dots, T$. Some of the individuals, $i = 1, \dots, N_1$, are observed only for $t = 1, \dots, T_D$ while others, $i = N_1 + 1, \dots, N$, are observed only for $t = T_D + 1, \dots, T$. A treatment, $D_i \in \{0, 1\}$, is implemented at date $t > T_D$. After the treatment, some units $i = N_1 + 1, \dots, N_2$ 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 potential outcome of the individual i at time t if his treatment status is d .¹⁰ The effect of the

⁹Our work on identification also relies on the literature on 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 individuals. 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 account for all sources of heterogeneity. See also Heckman, Ichimura, and Todd (1997), Abadie (2005), Hong (2013), and Sant’Anna and Zhao (2020) for work on the extension of the seminal difference-in-difference methodology.

¹⁰ $d = 1$ 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.

policy on the individual i at time t is:

$$\alpha_{it} = Y_{it}(1) - Y_{it}(0) \quad (1)$$

At each time period, an interesting parameter is the average treatment effect on treated; for $t > T_D$, it is given by:

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

A natural estimator of $E(Y_{it}(1)|D_i = 1)$ is its empirical counterpart. However, we do not observe $E(Y_{it}(0)|D_i = 1)$, which is the counterfactual expected outcome 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 or parallel trends assumption, for $t \geq T_D$:

$$E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0) = E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1). \quad (2)$$

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. In other words, individuals in the treated group are similar enough to individuals in the control group. The equal trends assumption could also be obtained conditional on exogenous characteristics ($X_i = (X_{i1}, \dots, X_{iT})'$).

$$E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0, X_i) = E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, X_i) \quad (3)$$

almost surely.

Note that the support of X_i may be different for treated and control groups. If the two supports do not overlap, the conditional parallel trend cannot be written. We assume that there is a correlation between D_i and X_i and allow the support to overlap. To account for the possibility that the treatment changes the X_i in a non-symmetric way, for both groups, we model the effect of X_{it} on the outcomes as being heterogenous. We assume that the difference in the group is captured by the difference in the distribution of the heterogenous effects in the treatment and the control groups.

We put structure in our context by assuming that the outcome in the absence of treatment is presented as:

$$Y_{it}(0) = X_{it}\beta_i + \delta_i\gamma_i + U_{it} \quad (4)$$

where β_i and γ_i are real random variables representing, respectively, the heterogeneity of the effects of exogenous characteristic and macroeconomic shocks.

The model presented in Equation (4) has two differences with the classical diff-in-diffs set-up. First, it has heterogeneity in the effect of the observed characteristics i.e., β_i (random slope). Second, there is an unobserved time-varying heterogeneity γ_i .¹¹ The introduction of $\delta_t\gamma_i$ enables us to control for selection on the unobservables while controlling for heterogeneity in macroeconomic shocks over time. The use of β_i allows us to account for the instability of the treatment group resulting from the treatment. The policy intervention generates an imbalance in the time-varying exogenous characteristic. However, we assume that the variables have the same support.¹² Thus, the inclusion of random slopes (β_i) enables us to account for the impact of the imbalance on our causal estimation.

The data generating process in Equation (4) is obtained under the following assumptions.

Assumption 1: $E(U_{it}|\beta_i, \gamma_i, X_i) = 0$ almost surely.

Assumption 2: $E(U_{it}|\beta_i, \gamma_i, X_i) = E(U_{it}|D_i, \beta_i, \gamma_i, X_i)$ almost surely.

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

The first assumption represents exogeneity of observed and unobserved characteristics with respect to the error term. Assumption 2 implies that the treatment status is conditionally 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 changes in our main results.

¹¹The set-up considered in this paper extends the set-up in Moon and Weidner (2018), Gobillon and Magnac (2016), Li (2018) and Bai (2009).

¹²It is important to note that with the common support assumption on the exogenous variables X , if we assume constant slopes, the imbalance may not be a problem as long as the conditional parallel trend assumption holds. Using a random slope thus allows us to capture the composition effect even after controlling for exogenous variables.

3.3 Biases of the classic Diff-in-Diffs, conditional Diff-in-Diffs and Diff-in-Diffs with interactive fixed-effects

This subsection shows that in the presence of time-varying group specific heterogeneity and random slope, the classical diff-in-diffs estimator is biased. Moreover, we show that a conditional diff-in-diffs or a diff-in-diffs accounting for individual specific heterogeneous temporal shocks will not solve the bias problem. In both cases, we derive the form of the bias.

The parameter of interest is the ATT

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

It is identified under the equal trends and exogeneity assumptions and can be estimated using a classical diff-in-diffs strategy. When parallel trends assumption does not hold, the *ATT* is not identified. Our aim is to investigate the effect of the introduction of a new heterogeneity on the identification of the *ATT*. We, therefore, need to characterize these heterogeneities.

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:

- $dG_0(\gamma_i, \beta_i | D_i = 1)$ and $dG_0(\gamma_i, \beta_i | D_i = 0)$
- $dG_1(\beta_i | D_i = 1)$ 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, $dG_k(\cdot | D_i = 1)$ is absolutely continuous with respect to $dG_k(\cdot | D_i = 0)$, $k = 0, 1, 2$. This corresponds to a common support assumption for the random parameters.

The Radon-Nikodym derivatives are given as follows:

- $dG_0(\gamma_i, \beta_i | D_i = 1) = r_0(\gamma_i, \beta_i) dG_0(\gamma_i, \beta_i | D_i = 0)$
- $dG_1(\beta_i | D_i = 1) = r_1(\beta_i) dG_1(\beta_i | D_i = 0)$
- $dG_2(\gamma_i | D_i = 1) = r_2(\gamma_i) dG_2(\gamma_i | D_i = 0)$

Proposition 1 *Consider the potential outcomes model presented in Section 3.2. Under Assumptions 1 and 2, the parallel trend assumption does not always hold.*

1. The bias in the parallel trend is

$$\begin{aligned} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1] &- E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0] \\ &= Cov[Y_{it}(0) - Y_{iT_D-1}(0), r_0(\gamma_i, \beta_i)|D_i = 0](5) \end{aligned}$$

2. If in addition to Assumptions 1 and 2 we also have Assumption 3, then;

$$\begin{aligned} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1] &- E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0] \quad (6) \\ &= Cov[Y_{it}(0) - Y_{iT_D-1}(0), r_1(\beta_i)r_2(\gamma_i)|D_i = 0] \end{aligned}$$

Proof of Proposition 1 in Appendix A.

Proposition 1 shows that there is systematic deviation from the parallel trends equation. This implies that the *ATT* is not identified under Assumptions 1 to 3 as the parallel trend assumption is not always true. Moreover, the deviation seems to depend on both sources of heterogeneity through the way in which the distribution of the heterogenous parameters differ in the treatment and control groups. In a situation where both types of heterogeneity are independent, the expression of the bias in the parallel trend is as follows,

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

or

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

Note that, in both cases, the bias is an aggregation of the bias coming from the time-varying macroeconomic shock and the part coming from heterogeneity of the effects from observed characteristics. The non-separable nature of the bias suggests that we need to use an estimation strategy that accounts for both sources of bias.

It is worth noting that all equations in Proposition 1 can be obtained conditional on X_i . Note that under assumptions 1 to 3, D_i is allowed to be correlated with β_i , γ_i ,

and X_i . This means that the conditional distributions of β_i , γ_i and X_i knowing $D_i = 1$ can be different from the conditional distributions of β_i , γ_i and X_i knowing $D_i = 0$.

In the special case when all individuals in the treated or the control group share the same exogenous characteristic value, the correlation between D_i and X_i is very close to one or is equal to one. Therefore, the common support assumption of the conditional distributions of X_i knowing D_i is not verified in this case. Thus, if the X_i affects the level of outcome Y_{it} , the parallel trend assumption in Equation (6) cannot be written conditional on X_i , meaning that X_i will not enter a diff-in-diffs regression.

In many empirical situations, whilst X_i and D_i could be correlated, the support of X_i is assumed to be the same in the control and the treatment groups. Therefore, the parallel trend assumption in Equation (6) can be written conditional on X_i . This paper considers empirical applications where X_i , β_i and γ_i are assumed to be correlated with D_i . The following sub-section gives additional conditions on X_i , β_i and γ_i under which the ATT is identified.

3.4 Identification and Estimation of the ATT

As shown in Proposition 1, using a simple diff-in-diffs with a models that account for the presence of factor or solely for the composition effects may lead to a biased estimator of the ATT. This section proposes the conditions under which the ATT is identified. It also shows that an estimation method that mixed both approaches can deliver a consistent estimator of the ATT.

In general, 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

$$Y_{it} = \alpha_{it}D_iI_t + \beta_iX_{it} + \gamma_i\delta_t + U_{it} \tag{7}$$

where D_i is the treatment group indicator and $I_t = \mathbf{1}\{t \geq T_D\}$ is the treatment period indicator. The following assumption is needed for the discussion on the identification of the *ATT*.

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 $P(X) = P(D_i = 1|X)$.

Assumption 4 (i) 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.

Proposition 2 *Let $\tilde{Y}_{it}(0) = X_{it}\beta_i + U_{it}$ be the potential outcomes from the model presented in Section 3.2 after correction of the unobserved time-varying heterogenous effects; under Assumptions 1 to 4, the parallel trend holds conditional on the propensity i.e.*

$$E[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0)|D_i = 1, P(X)] = E[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0)|D_i = 0, P(X)].$$

Moreover, an estimation strategy mixing a least squared in presence of interactive fixed-effects as in Bai (2009) and conditional diff-in-diffs provides a consistent estimator on the ATT.

Proof of Proposition 2 in Appendix A.

Proposition 2 shows that if β_i and γ_i are random and correlated with D_i , the *ATT* is identified under Assumptions 1, 2 and 4. The identification strategy is constructive and suggests the following procedure for consistent estimation. In a first stage, we partial-out the interactive fixed-effect using Bai (2009) and obtain \tilde{Y}_{it} as a result. In the second stage, we use a propensity score to match treated and control groups. After the matching procedure, we estimate the *ATT* using a diff-in-diffs style regression on the matched sample.

$$\tilde{Y}_{it} = \alpha D_i I_t + \beta D_i + \delta I_t + u_{it} \tag{8}$$

Equivalently, we can apply an OLS as in Bai (2009) on an equation similar to Equation (8) with the appropriately matched sample. In other words, we can run an estimation a la Bai (2009) on the following model $Y_{it} = \alpha D_i I_t + \beta D_i + \delta_t \gamma_i + v_{it}$, with a sample of appropriately matched individuals.

The following section discusses the implementation of this procedure to the specific case of the evaluation of the change in Australian migration policy.

4 Application to the Effect of Immigration Policy

Since the data consists of immigrants who entered before and after the policy change, 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 showing if an individual is in the control or the treated group. The time periods correspond to the waves i.e. $t = 1, \dots, 5$ and $T_D = 4$. The remittances are hence represented by:

$$Y_{it} = \alpha D_i I_t + \beta_i X_{it} + \gamma_i \delta_t + V_{it}. \quad (9)$$

To evaluate ATT captured by α , we implement a three steps algorithm. The first step allows us to remove the time-varying macroeconomic shocks ($\gamma_i \delta_t$). The second step is the stage of the construction of the appropriate control group, for which we use a propensity score matching method. The last step is the estimation stage in which we use the outcomes of the first step and estimate α using the appropriate control and treatment groups.

ATT Estimation Algorithm:

Step: 1 Estimate the effect of macroeconomics shocks.

For each cohort and each treatment group run the panel data model estimation with time (waves) dummies and individual fixed effects. Then compute the adjusted for time effect \tilde{Y}_{it} , which are free of heterogeneous time-varying group effects. This procedure eliminates $\gamma_i \delta_t$ under the assumption that all individuals in a group have the same time-varying fixed-effects.¹³

Step: 2 Create appropriate treatment and control groups.

Compute the propensity score using exogenous characteristics as covariates.¹⁴

¹³We are in an empirical context with a small number of period $T = 5$, therefore, we can't use a factor structure for $\gamma_i \delta_t$. Allowing for a factor structure will mean that both the factor and the loading factor should be estimated. Li (2018) proposes a procedure to estimate the number of factors, the factor loading as well as the factor when we are interested in the evaluation of the treatment effect. After following this procedure adjusted values for Y_{it} are obtained.

¹⁴The parameters of the propensity score function are obtained using exogenous characteristics of the pre-

Then match the individuals in the treated group with those in the control groups.
The immigrant who does not get a match is left out of the sample.

Step: 3 Estimate α on the appropriate sample.

We run an OLS estimation

$$\tilde{Y}_{it} = \alpha D_i I_t + \beta_0 D_i + \delta I_t + \beta X_{it} + u_{it} \quad (10)$$

where t corresponds to the wave-cohort ($t = 1, \dots, 5$) number with the appropriately matched sample of individuals, kept in Step 2 and $\hat{\alpha}$ is an estimator of the *ATT*.

The small sample behaviour of this estimation procedure is studied using a Monte Carlo experiment and the results are presented in Appendix B. The procedure we are proposing performs well when the control and treatment groups are of similar size; this is a feature of our empirical application.

5 Data and Estimation Results

The Longitudinal Survey of Immigrants to Australia (LSIA) is a comprehensive survey focusing on the migration experience and early settlement process of new migrants. It is sourced from a representative sample of new intakes and it provides unique information on migrants' characteristics, including their labour market outcomes, expectations, and motivations before and after migration, as well as valuable insights on the experience of migrating. An additional unique feature of the LSIA is the timing of its implementation, as it surveyed cohorts entering Australia just prior and subsequent to a substantive change in immigration policy. The new policy raised the criteria for admission for two categories of applicants: the skilled independent and concessional family reunification visa. However, those who entered on humanitarian, employer-sponsored (business) visa and close family reunification (preferential family) streams were not affected by the policy change. The first cohort entered Australia between 1993 and 1995, just before the policy change, while the second one entered in 1999-2000, after the policy

treatment period. For each individual, the propensity score is predicted using his observed characteristics. This explains the importance of conditioning on the whole data set.

change.¹⁵ Table 1 provides a breakdown of the composition of the two cohorts, by type of visa. This enables us to disentangle the visa categories affected by the policy change (skilled independent and concessional family) from those that were not affected (family preferential, business, humanitarian), which we refer to as ‘affected’ and ‘not affected’, respectively. Cohort 2 includes fewer primary migrants in working age than cohort 1, reflecting both worse macroeconomics conditions and changes in migration policy, whose effect we aim at disentangling empirically.

Table 2 reports the unconditional mean and standard deviation of affected and not affected migrants in the working sample, before and after the policy change, across a variety of demographic, regions of origin and other characteristics. These two groups are different with respect to several demographic indicators (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. The affected group is admitted through the point system (Tani (2014)), 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, the not affected comprise a more heterogeneous group of settlers with a high incidence of family reunification. The difference in the means of ‘before’ and ‘after’ columns is separately reported for both affected and not-affected groups. In general, those admitted after the stricter immigration policy is introduced are less likely to remit (by about 5%) but remit larger amount. The gender composition and education level of primary applicants in the affected visa categories in cohort 2 is also considerably different than the corresponding earlier group (more female and more university educated migrants), as well as their probability of employment (higher), as noted by (Cobb-Clark (2003), Chiswick and Miller (2006), Thapa and Gorgens (2006) and, Mahuteau and Junankar (2008)). Much more subdued is the corresponding difference between ‘before’ and ‘after’ subgroups among the non-affected categories.

Tables 3 and 4 present the estimation results of the cohort and policy effects on the

¹⁵The LSIA consists of two distinct longitudinal datasets, three waves covering cohort 1 (LSIA1) and two waves for cohort 2 (LSIA2). Immigrants in cohort 1 were interviewed at 5 months, 17 months and 41 months after arrival (between March 1994–December 1995) while those in cohort 2 were interviewed at 5 months and 17 months after arrival (June 1999, June 2000).

extensive and intensive margins of remittances, respectively. The effect is estimated using diff-in-diffs (OLS and RE with Mundlak), diff-in-diffs with heterogeneous time varying trends by groups (OLS C.) and conditional diff-in-diffs with heterogeneous time-varying trend by groups (OLS C. and Match). The first two columns (diff-in-diffs) report the estimates obtained under the (implicit) assumption of parallel trends, while the remaining two columns report the results obtained when macroeconomic trends in the treated and control groups are allowed to differ. As shown in Table 3, there is a significant difference in the point estimates about the probability to remit, revealing the importance to account for different macroeconomic conditions.

The effect attributable to the policy change, measured by the interaction of cohort and policy variables, is reported in the first row of Tables 3 and 4. The results show that the policy change influences positively and statistically significantly the probability to remit (4.3%), but not the amount of remittances sent (which is negative but insignificant). Because of the stringent policy in terms admission criteria, the affected migrants had to achieve those skills before entering Australia and therefore might have borrowed money from extended family to acquire those skills, which is probably why they are more likely to remit (see Ilahi and Jafarey (1999)). However, since we control for income, education and a wide range of individual characteristics (see full results in Appendix C), the results on the intensive margin support the hypothesis that the policy change captures an increase in the perceived level of riskiness of settlement among immigrants in the second cohort. Since the policy change included the delay of financial assistance for up to two years after arrival, there might be a tendency for the migrants in the affected group to keep their savings in Australia, for precautionary reasons.

Another aspect of the policy change is the cohort effect. Even though the policy didn't affect all visa streams, it's still possible that migrants in the second cohort were on average better skilled than those in the first cohort (perhaps with the exception of those on humanitarian visa). Therefore the cohort effects are due to changes in the composition of migrants. Our results on cohort effect, which are reported in the second row of Tables 3 and 4, show that migrants in the second cohort are less likely to remit but those who do remit send a higher amount (76.7%). This result reconciles the conflicting estimates found in the literature on brain drain and remittances. It shows that the policy change that brought better educated to the country had a dual

effect: they were less likely to remit which meant the brain drain compensation is not fulfilled, but since those who remit sent a higher amount, the brain drain effect may not be too detrimental. The higher intensive margin result is in essence the same as in Bollard, McKenzie, Morten, and Rapoport (2011), since those who entered in the second cohort are relatively better skilled than those who entered in 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 or ease their upskilling via further education in the country of origin. 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. However, since they come with their families and with the intention to move permanently, they have a lower tendency to remit, which in essence is what Ratha and Mohapatra (2011) find at the micro level and what Faini (2007) found at the macro level.

6 Conclusions

Using the Longitudinal Survey of Immigrants to Australia, we studied the impact of a change in migration policy on migrant's remittance behaviour. We built on current Diff-Diff methods to address time-varying individual heterogeneity and hence control for the change in migration conditions between the two LSIA cohorts. The new policy, which was announced in 1996, dramatically changed the conditions of entry making them more stringent both in terms of initial requirements (relatively higher skills and English language proficiency) and in terms of curbing the level of support offered upon entry for two visa categories: concessional family and skilled independent. Our two main results are the change in remittance behaviour of the affected due to change immigration policy and the change in remittance behaviour due to change in the composition of migrants, which we call the cohort effect. The policy and cohort effect show opposite results. In the former case the extensive margin is positive and significant with no effect on the intensive margin whereas the latter shows that those who remit send a significantly higher amount compared to the first cohort. We therefore conclude that the policy change that brought better educated to the country had a dual effect: they were less likely to remit which meant the brain drain compensation is not fulfilled, but since those

who remit sent a higher amount, the brain drain effect may not be too detrimental.

We also advance the methodology appropriate to deal with data sampled from distinct cohorts – in our case settling in Australia before and after new migration conditions were put in place. In particular, we extended the literature on conditional difference-in-differences and evaluation of treatment effect with panel data in the presence of interactive fixed effects. We allow for heterogeneity in how observable characteristics affect the outcome and for the presence of interactive fixed effects (or time-varying individual effects). We show that the classical difference-in-difference is biased and the presence of both sources of heterogeneity need to be taken into account via a mix of conditional difference-in-differences and methods accounting for time-varying individual effects.

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.
- BAI, J. (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.
- DOCQUIER, F., H. RAPOPORT, AND S. SALOMONE (2012): “Remittances, migrants’ education and immigration policy: Theory and evidence from bilateral data,” *Regional Science and Urban Economics*, 42(5), 817–828.
- DUSTMANN, C., AND J. MESTRES (2010): “Remittances and temporary migration,” *Journal of Development Economics*, 92(1), 62–70.

- FAINI, R. (2007): “Remittances and the Brain Drain: Do more skilled migrants remit more?,” *The World Bank Economic Review*, 21(2), 177–191.
- GOBILLON, L., AND T. MAGNAC (2016): “Regional policy evaluation: Interactive fixed effects and synthetic controls,” *Review of Economics and Statistics*, 98(3), 535–551.
- 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.
- HONG, S.-H. (2013): “Measuring the effect of napster on recorded music sales: difference-in-differences estimates under compositional changes,” *Journal of Applied Econometrics*, 28(2), 297–324.
- 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.
- ILAHY, N., AND S. JAFAREY (1999): “Guestworker migration, remittances and the extended family: evidence from Pakistan,” *Journal of development Economics*, 58(2), 485–512.
- LI, K. (2018): “Inference for factor model based average treatment effects,” *Available at SSRN 3112775*.
- 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).
- MOON, H. R., AND M. WEIDNER (2018): “Nuclear Norm Regularized Estimation of Panel Regression Models,” *arXiv preprint arXiv:1810.10987*.
- NIIMI, Y., C. OZDEN, AND M. SCHIFF (2010): “Remittances and the brain drain: skilled migrants do remit less,” *Annals of Economics and Statistics/Annales de Economie 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.
- RAPOPORT, H., AND F. DOCQUIER (2006): “The economics of migrants’ remittances,” *Handbook of the economics of giving, altruism and reciprocity*, 2, 1135–1198.
- RATHA, D., AND S. MOHAPATRA (2011): “Preliminary estimates of diaspora savings,” .
- SANT’ANNA, P. H., AND J. ZHAO (2020): “Doubly robust difference-in-differences estimators,” *Journal of Econometrics*, 219(1), 101–122.
- TANI, M. (2014): “Using a point system for selecting immigrants,” *IZA World of Labor*.
- 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.
- WOOLDRIDGE, J. M. (2005): “Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models,” *Review of Economics and Statistics*, 87(2), 385–390.
- YANG, D. (2011): “Migrant remittances,” *Journal of Economic perspectives*, 25(3), 129–52.

Table 1: Composition of cohorts in wave 1, by type of visa

| | Cohort 1 | Cohort 2 |
|---------------------|-----------------|-----------------|
| Affected | | |
| Skilled Independent | 17.3 | 13.6 |
| Concessional Family | 15.8 | 10.3 |
| Not Affected | | |
| Preferential Family | 40.1 | 45 |
| Business Skills | 10.9 | 12.3 |
| Humanitarian | 15.9 | 18.8 |
| Total | 100 | 100 |
| N | 4,922 | 2,808 |

Source: LSIA's Primary applicants only. The age group is truncated only working age migrants (20-65 years) included.

Table 2: Means and differences: Balance tests

| Variables | Affected ^a | | | Not Affected ^b | | |
|-------------------|-----------------------|--------------|------------|---------------------------|--------------|------------|
| | 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 3: Probability of remittances¹⁶

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

t statistics are in parentheses.

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

Table 4: Value of remittances¹⁷

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

t statistics are in parentheses.

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

A Proof of Propositions

Equalities and propositions presented in these proofs are true almost surely in the corresponding probability space.

Proof of Proposition 1

Under Assumptions 1 and 2, for $t \geq T_D$ we can show that

$$E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0, \beta_i, \gamma_i] = E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \beta_i, \gamma_i] \quad (11)$$

Note that the assumption allow for arbitrary correlation between β_i, γ_i, D_i and X_i .

$$\begin{aligned} E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1) &= E[E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \beta_i, \gamma_i)] \quad (12) \\ &= \int_{\gamma, \beta} E(Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \beta_i, \gamma_i) dG_0(\gamma_i, \beta_i|D_i = 1) \end{aligned}$$

From Equation (11) and by noting that $\int_{\gamma, \beta} r_0(\gamma_i, \beta_i) dG_0(\gamma_i, \beta_i|D_i = 0) = 1$; we show that

$$\begin{aligned} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1] &= \int_{\gamma, \beta} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1, \beta_i, \gamma_i] dG_0(\gamma_i, \beta_i|D_i = 1) \\ &= \int_{\gamma, \beta} E[(Y_{it}(0) - Y_{iT_D-1}(0))r_0(\gamma_i, \beta_i)|D_i = 0, \beta_i, \gamma_i] dG_0(\gamma_i, \beta_i|D_i = 0) \\ &= E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0] \\ &+ \int_{\gamma, \beta} E[(Y_{it}(0) - Y_{iT_D-1}(0))r_0(\gamma_i, \beta_i)|D_i = 0, \beta_i, \gamma_i] dG_0(\gamma_i, \beta_i|D_i = 0) \\ &- A \times \int_{\gamma, \beta} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 0, \gamma_i, \beta_i] dG_0(\gamma_i, \beta_i|D_i = 0) \end{aligned}$$

with $A = \int_{\gamma, \beta} r_0(\gamma_i, \beta_i) dG_0(\gamma_i, \beta_i|D_i = 0) = 1$.

The above equality is equivalent to saying that

$$\begin{aligned} E[Y_{it}(0) - Y_{iT_D-1}(0)|D_i = 1] &= E[(Y_{it}(0) - Y_{iT_D-1}(0))|D_i = 0] \\ &+ Cov[(Y_{it}(0) - Y_{iT_D-1}(0)), r_0(\gamma_i, \beta_i)|D_i = 0] \end{aligned}$$

Under Assumption 3, the density of the join distribution can be written as the product of marginal distribution. We have $r_0(\gamma_i, \beta_i) = r_1(\beta_i)r_2(\gamma_i)$ and the second result follows by replacing r_0 in Equation (5) of the main paper. This ends the proof

of Proposition 1. \square

Proof of Proposition 2 If we have $\tilde{Y}_{it}(0) = X_{it}\beta_i + U_{it}$; under assumptions 1 to 3 we can show that for $t > T_D$

$$\begin{aligned} E[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0)|D_i = 1, P(X)] &- E[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0)|D_i = 0, P(X)] \\ &= Cov[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0), r_2(\beta_i)|D_i = 0, P(X)] \end{aligned}$$

Using assumption 4 (ii), we have $Cov[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0), r_2(\beta_i)|D_i = 0, P(X)] = Cov[\tilde{Y}_{it}(0) - \tilde{Y}_{iT_D-1}(0), 1|D_i = 0, P(X)] = 0$. The conditional parallel trend follows.

Now we discuss estimation of the ATT using a mixed strategy. We assume that $\beta = E(\beta_i)$, equation (7) then becomes

$$Y_{it} = \alpha D_i I_t + X_{it}\beta + \delta_t \gamma_i + U_{it} + (\alpha_{it} - \alpha) D_i I_t + X_{it}(\beta_i - \beta) \quad (13)$$

We assume a fixed number of time periods. And that, the time effects δ_t can be estimated as in Bai (2009). For each individual the observed outcome in vector notation is

$$Y_i = \alpha D_i I_{[1:T]} + X_i \beta + \Delta \gamma_i + U_i + \Omega_i D_i I_{[1:T]} + X_i(\beta_i - \beta) \quad (14)$$

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

Let $M_\Delta = I - \Delta'(\Delta\Delta')\Delta$ and multiplying equation (14) by M_Δ on both sides, we get

$$M_\Delta Y_i = \alpha D_i M_\Delta I_{[1:T]} + M_\Delta X_i \beta + M_\Delta U_i + M_\Delta \Omega_i D_i I_{[1:T]} + M_\Delta X_i(\beta_i - \beta) \quad (15)$$

$M_\Delta Y_i$ is a quantity where the time-varying effects are partial-out.

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

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

When we substitute D_i by its value as a function of X_i , Equation (15) then becomes

$$M_\Delta Y_i = \alpha D_{iX} M_\Delta I_{[1:T]} + M_\Delta \tilde{U}_i + M_\Delta \Omega_i D_i I_{[1:T]} + M_\Delta X_i(\beta_i - \beta) \quad (16)$$

where $\tilde{U}_i = U_i + X_i \beta + \alpha \text{vec}(X_i)' \rho I_{[1:T]}$. 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.

Now we discuss the assumptions under which Condition 2 holds.

$$Cov(\varepsilon_i, D_{iX}) = Cov(\tilde{U}_i, D_{iX}) + Cov(\Omega_i D_i I_{[1:T]}, D_{iX}) + Cov(X_i(\beta_i - \beta), D_{iX}) \quad (17)$$

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 last term in the correlation 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 helps us to recover the equality, because under it,

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

The last term on the right hand side of Equation (17) 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 the first and second terms. If we do not apply matching, the term $Cov(X_i(\beta_i - \beta), D_{iX})$ will not vanish and the estimator on the *ATT* is not identify.

In conclusion, we have shown that Conditions 1 and 2 are verified. Thus the *ATT* is identified under Assumptions 1, 2 and 4. This ends the proof of Proposition 2.

B 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:

$$Y_{it} = \alpha D_i I_t + \beta_i X_{it} + \gamma_i \delta_t + U_{it} \quad (18)$$

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. $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \dots)$ and $\delta_t = (1; \delta_{t1}; \delta_{t2} \dots)'$. 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 $\alpha = 0.3$.

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 and no change in composition ($c = 0$, no heterogeneity in β), column 1 results show that the estimated treatment effects 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 are larger. However,

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 |

with a treatment group larger than the control group the bias of methods using matching becomes smaller than the other estimators.

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 (change group composition) with overlap support ($c = 0.5$), which is case we are interested in, 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 ($N_1 = 120$), Matching and Bai and Matching methods have the smallest biases.

As the difference between the treated and control group increases ($c = 1$), the biases of diff-in-diffs and Bai (2009) methods also increase. However, Matching, Synthetic Control and Bai and Matching all have good bias properties for relatively large treated group ($N_1 = 80$).

In all the cases of interest ($c = 0.5$ or 1), the Bai and Matching estimator has the smallest or the second smallest bias. Thus, if we have an empirical application that has a control and the treatment groups are not similar enough and if the treated population is large relative to the total population, Bai and Matching should be used to estimate the effect of the treatment. We are , therefore, going to present results similar to using Bai and Matching combine with diff-in-diffs in our empirical application.

C Full tables of results

Table B1: Value of remittances

| | OLS | Mundlak | OLS C. | OLS Match. | OLS C. and Match. |
|-------------------------------|----------------------|----------------------|---------------------|---------------------|---------------------|
| Affected x post-reform cohort | -0.151 (-1.50) | -0.151 (-1.52) | 0.0277 (0.28) | -0.190 (-1.11) | -0.0595 (-0.35) |
| Post-reform cohort | 0.443*** (7.09) | 0.452*** (7.47) | 0.768*** (12.32) | 0.487*** (4.04) | 0.767*** (6.14) |
| Affected group | -0.0130 (-0.21) | -0.00636 (-0.10) | -0.111 (-1.79) | 0.104 (0.84) | -0.0229 (-0.19) |
| income per week <\$155 | 0.125* (1.98) | 0.115 (1.84) | 0.169** (2.65) | 0.559** (2.97) | 0.601*** (3.39) |
| income per week [\$385-\$675] | 0.187** (2.62) | 0.170* (2.47) | 0.167* (2.34) | 0.349 (1.76) | 0.359 (1.76) |
| income per week > \$675 | 0.592*** (7.25) | 0.580*** (7.32) | 0.546*** (6.70) | 0.563** (2.71) | 0.544** (2.61) |
| cmals1==2 | -0.270*** (-3.62) | -0.292*** (-4.09) | -0.238** (-3.18) | -0.227 (-1.19) | -0.169 (-0.90) |
| cmals1==6 | -0.265** (-2.91) | -0.277** (-3.12) | -0.288** (-3.16) | -0.279 (-1.17) | -0.478* (-1.96) |
| Education: BA | 0.0544 (0.65) | 0.0404 (0.48) | 0.0486 (0.57) | 0.0822 (0.42) | 0.0966 (0.48) |
| Postgraduate | 0.163** (2.67) | 0.147* (2.45) | 0.131* (2.14) | 0.0402 (0.32) | -0.00881 (-0.07) |
| Higher education | 0.101 (1.68) | 0.0987 (1.66) | 0.0754 (1.26) | -0.118 (-0.85) | -0.173 (-1.23) |
| language interview is English | -0.0757 (-1.75) | -0.0812 (-1.90) | -0.0740 (-1.72) | -0.290* (-2.40) | -0.280* (-2.26) |
| SE Asia | -0.0597 (-0.57) | -0.0526 (-0.52) | -0.0773 (-0.72) | 0.165 (0.69) | 0.153 (0.64) |
| E Asia | 0.112 (0.97) | 0.127 (1.15) | 0.0775 (0.66) | 0.425 (1.65) | 0.398 (1.57) |
| S Asia | 0.0846 (0.82) | 0.0992 (1.00) | 0.0566 (0.54) | 0.265 (1.17) | 0.264 (1.17) |
| N America | 0.538*** (4.04) | 0.574*** (4.47) | 0.515*** (3.86) | 1.050*** (3.98) | 1.046*** (3.93) |
| Latin America | 0.148 (1.33) | 0.163 (1.51) | 0.127 (1.12) | 0.550* (2.24) | 0.523* (2.15) |
| Africa | 0.00441 (0.04) | 0.0140 (0.12) | 0.0258 (0.20) | 0.0605 (0.26) | 0.123 (0.53) |
| Oceania | -0.0180 (-0.14) | 0.00198 (0.02) | -0.0387 (-0.30) | 0.0480 (0.21) | 0.0465 (0.20) |
| age at migration | 0.00387 (1.30) | 0.00367 (1.25) | 0.00360 (1.19) | 0.0150 (1.94) | 0.0118 (1.50) |
| Constant | 6.562*** (47.65) | 6.570*** (48.68) | 6.335*** (45.30) | 5.987*** (15.75) | 5.918*** (15.37) |
| Observations | 2335 | 2335 | 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

| | OLS | Mundlak | OLS C. | OLS C. and Match. |
|-------------------------------|------------------------|------------------------|------------------------|-----------------------|
| Affected x post-reform cohort | 0.00379 (0.26) | 0.00246 (0.17) | 0.0563*** (4.34) | 0.0426*** (4.98) |
| Post-reform cohort | -0.112*** (-15.83) | -0.111*** (-15.85) | -0.0810*** (-9.16) | -0.0654*** (-9.30) |
| Affected group | -0.0136 (-1.38) | -0.0148 (-1.52) | 0.106*** (6.99) | -0.00133 (-0.04) |
| income per week <\$155 | -0.0189* (-2.39) | -0.0173* (-2.27) | -0.0148* (-1.99) | 0.0140** (3.02) |
| income per week [\$385-\$675] | 0.0587*** (5.74) | 0.0611*** (6.17) | 0.0559*** (5.59) | 0.0768*** (11.07) |
| income per week > \$675 | 0.0275* (2.53) | 0.0341** (3.25) | 0.0159 (1.57) | 0.0515*** (8.58) |
| emalFs1==2 | -0.0715*** (-7.35) | -0.0768*** (-8.21) | -0.0613*** (-6.74) | -0.0212*** (-3.87) |
| emalFs1==6 | -0.0594*** (-5.65) | -0.0607*** (-5.97) | -0.0579*** (-5.88) | -0.0341*** (-5.21) |
| Education: BA | 0.00348 (0.25) | 0.00494 (0.37) | 0.00690 (0.57) | -0.0399*** (-4.66) |
| Postgraduate | -0.0110 (-1.25) | -0.00986 (-1.15) | -0.00778 (-1.02) | -0.0320*** (-4.53) |
| Higher education | -0.00278 (-0.29) | -0.00134 (-0.14) | -0.00196 (-0.23) | -0.0411*** (-5.22) |
| language interview is English | -0.00837 (-1.34) | -0.00900 (-1.49) | -0.00689 (-1.14) | -0.0241*** (-5.43) |
| SE Asia | -0.130*** (-5.05) | -0.128*** (-5.11) | -0.128*** (-6.14) | -0.0280* (-2.36) |
| E Asia | -0.0702** (-2.58) | -0.0654* (-2.47) | -0.0730** (-3.28) | -0.0227 (-1.74) |
| S Asia | -0.0144 (-0.53) | -0.0137 (-0.52) | -0.0148 (-0.68) | 0.0296* (2.29) |
| N America | -0.126*** (-4.76) | -0.123*** (-4.77) | -0.125*** (-5.82) | -0.0450*** (-3.61) |
| Latin America | 0.0181 (0.64) | 0.0185 (0.67) | 0.0166 (0.71) | 0.00541 (0.42) |
| Africa | -0.102*** (-3.66) | -0.100*** (-3.71) | -0.0998*** (-4.42) | -0.0247 (-1.78) |
| Oceania | -0.0743** (-2.59) | -0.0730** (-2.61) | -0.0728** (-3.12) | -0.0396** (-3.03) |
| age at migration | -0.00198*** (-5.70) | -0.00194*** (-5.72) | -0.00202*** (-6.83) | 0.0000529 (0.22) |
| Group specific trends | No | No | Yes | Yes |
| Constant | 0.371*** (12.59) | 0.364*** (12.70) | 0.367*** (14.86) | 0.141*** (9.71) |
| Observations | 15436 | 15436 | 15436 | 15436 |
| \bar{R}^2 | 0.066 | | 0.083 | 0.048 |

t statistics in parentheses

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