How Important is Selection? Experimental Vs Non-experimental Measures of the Income Gains from Migration[1]

David McKenzie, *Development Research Group, World Bank*[*]
John Gibson, *University of Waikato*
Steven Stillman, *Motu Economic and Public Policy Research*

Abstract

Accurate measurement of the effect of migration on the income of potential migrants is a crucial factor in determining the impact that lowering barriers to migration would have on world income. However, measuring this effect is complicated by non-random selection of migrants from the general population, which makes it hard to obtain an appropriate comparison group of non-migrants. This paper uses a migrant lottery to experimentally estimate the income gains from migration, thus overcoming this problem. New Zealand allows a quota of Tongans to immigrate each year with a lottery used to choose amongst the excess number of applicants. A unique survey conducted by the authors in these two countries allows experimental estimates of the income gains from migration to be obtained by comparing the incomes of migrants to those who applied to migrate, but whose names were not drawn in the lottery, after allowing for the effect of non-compliance among some of those whose names were drawn. We also conducted a survey of individuals who did not apply for the lottery. Comparing this non-applicant group to the migrants enables assessment of the degree to which non-experimental methods can provide an unbiased estimate of the income gains from migration. We find evidence of migrants being positively selected in terms of both observed and unobserved skills. As a result, non-experimental methods other than instrumental variables are found to overstate the gains from migration by 20 to 82 percent, with difference-in-differences and bias-adjusted propensity-score matching performing best among the alternatives to instrumental variables.

Keywords: Migration, Selection, Natural Experiment
JEL codes: J61, F22, C21

---

[1] We thank the Government of the Kingdom of Tonga for permission to conduct the survey there, the New Zealand Department of Labour Immigration Services for providing the sampling frame, Halahingano Rohorua and her assistants for excellent work conducting the survey, and most especially the survey respondents. Mary Adams, Alan de Brauw, Deborah Cobb-Clark, Chirok Han, Manjula Luthria, Martin Ravallion, Ed Vytlacil and participants at seminars at BREAD, Columbia University, NEUDC, NZESG, DoL, the University of Canterbury, and the World Bank provided helpful comments. Financial support from the World Bank, Stanford University, the Waikato Management School and Marsden Fund grant UOC0504 is gratefully acknowledged. The views expressed here are those of the authors alone and do not necessarily reflect the opinions of the World Bank, the New Zealand Department of Labour, or the Government of Tonga.
[*] Corresponding author: E-mail: dmckenzie@worldbank.org. Address: MSN MC3-300, The World Bank. 1818 H Street N.W., Washington D.C. 20433, USA. Phone: (202) 458-9332, Fax (202) 522-3518.

1. Introduction.

Accurate measurement of the gain in income from migration is of fundamental importance for migration policy, since it is a major factor in determining the number of potential migrants from any easing of restrictions on movement and the welfare gains from such movement. For example, the large differences in wages and per capita income between the EU15 and Eastern Europe led to public concern about a flood of migrants once these countries joined the European Union, resulting in 12 of the 15 countries imposing transitional restrictions on immigration. However, such large income differentials also give rise to large global gains from more migration. In an influential study, Walmsley and Winters (2003) used wage differentials as a measure of the income gain from migration and estimated that a 3% increase in migration from developing countries would lead to a gain in world income greatly exceeding the gains to be had from removing all remaining barriers to goods trade.

Even estimates of the income gains from migration that go beyond simple cross-country comparisons of wage rates are likely to be misleading. Ideally, one must compare the earnings of the migrant to what they would have earned in their home country. The latter is unobserved, and is usually proxied by the earnings of stayers of a similar age and education to the migrant but if the two groups are really the same, they should have the same migratory behaviour (Lalonde and Topel, 1997). Simple comparisons of movers and stayers are therefore likely to be misleading, as income gains may just reflect unobserved differences in ability, skills, and motivation, rather than the act of moving itself. While statistical corrections for non-random selection are often used when modelling migration (Robinson and Tomes, 1982), there is some doubt about the assumptions behind these remedies for selectivity in non-experimental data (Deaton, 1997). These doubts persist because it is hard to know how well these remedies compare with the ideal of a randomized experiment.

The research reported here uses a unique random selection mechanism to overcome the interpretation difficulties posed by the non-random selection of migrants, and then compares experimental estimates of the gains from migration to results obtained using non-experimental estimation methods. The random selection mechanism we use is based on the Pacific Access Category (PAC) under New Zealand's immigration policy. The PAC allows an annual quota of Tongans to migrate to New Zealand. Many more applications are received than the quota allows, so a lottery is used by the New Zealand Department of Labour to randomly select from amongst the registrations. A survey administered by the authors was used to collect data on winners and losers in this lottery. Thus, we have a group of migrants and a comparison group who are similar to the migrants, but remain in Tonga only because they were not successful in the lottery.

By comparing the lottery winners and losers, we are able to obtain the only known experimental measure of the gain in income from migration. As not all individuals whose names were selected in the lottery had migrated by the time of our survey, this estimate accounts for non-compliance to the "treatment" of migration. We therefore consider both the intention-to-treat effect, which is the impact on expected income of having a winning ballot in the PAC lottery, and the average treatment effect on the treated, which is the average impact of migrating for individuals who migrate after winning the lottery. We estimate that there is an 84% increase in expected income from winning the lottery, and a 263% increase in income from migrating. This gain in income is only half of what a simple comparison of differences in per capita GDP would predict and only 43% of the difference in manufacturing wages between the two countries.

In addition to winners and losers in the PAC lottery, we also surveyed individuals who did not apply for the lottery. We use this sample of non-applicants along with the migrant sample to obtain non-experimental estimates of the income gains from migration. Five popular non-experimental methods for dealing with selectivity are considered. Instrumental variables using a good instrument (pre-migration distance to the New Zealand service office in Tonga) performs best, coming within 2% of the experimental estimate. Each of the other methods is found to overstate the gain in income from migration compared to the experimental estimate. A single-difference estimator overstates the gains by 25%, while difference-in-differences overstates the gains by 20%. Propensity-score matching overstates the gains by 19-33%, doing better when past income is included as a control and when the bias-adjusted methods of Abadie and Imbens (2005) are used. OLS overstates the gains by 31%, while a poor instrument (the size of the migrant network) overstates the gains by 82%.

The overstatement of the income gains from migration obtained from the non-experimental methods suggests that Tongan migrants are positively selected in terms of unobserved ability and skills. The Gini of weekly earnings from wage, salary and self-employment work in Tonga is 0.338, compared to a Gini of 0.374 in New Zealand,[2] so the Roy model used by Borjas (1987) would predict positive selection from Tonga. However, the existing empirical literature on migrant selectivity has focused exclusively on observable measures of skills, such as education (e.g. Chiquiar and Hanson, 2005).

[2] Tonga Gini calculated from our sample of workers in non-migrant households; Gini for New Zealand calculated from the 2002 New Zealand Income Survey.

We do indeed see positive selection of Tongan migrants in terms of observed skills, such as education. We then build on the existing literature by using pre-migration earnings to look at selection, finding that migrants are also positively selected in terms of unobserved components of labor earnings, after controlling for age, education and other observed characteristics of individuals.

The estimates we obtain of the income gains from migration and our finding of positive selection on unobservables apply to the specific case of 18 to 45 year olds migrating from Tonga to New Zealand. Nevertheless, Tongan migrants are not atypical of the average developing country migrants elsewhere in the world, suggesting that the results may apply more broadly. The average Tongan migrant in our sample has 11.7 years of education, compared to 11.0 years for the average 18-45 year old new arrival in the United States, and much less than the 15.1 years for the average 18-45 year old new arrival in highly skill-selective Canada.[3] Tongan migrants average 1.2 more years of schooling than non-migrants, a similar degree of positive selection on observables to the 0.8 years higher education of Mexican migrants moving to the U.S.

This paper also contributes to the literature started by the influential work of Lalonde (1986), which attempts to assess the ability of non-experimental estimators to obtain estimates similar to experimental results. To date, this literature has concentrated on a small number of labor market training programs.[4] After Lalonde's initial pessimistic assessment of non-experimental measures, there has been much recent debate as to the

[3] See Appendix 1 in the working paper (McKenzie, Gibson and Stillman 2006) for a comparison along other dimensions. The Tongan migrants we study are equally as likely to work as new migrants aged 18 to 45 in the United States and Canada, and lie somewhere between the U.S. and Canadian migrants in terms of average age, percent married, and percent female. This appendix also compares migrants to non-migrants, and Mexican new arrivals in the U.S. to non-migrants in Mexico.

[4] Glewwe, Kremer, Moulin and Zitzewitz (2004) is an exception, comparing regression and difference-in-difference estimates to the results of a randomized experiment on the effects of providing flip charts in schools in Kenya. They do not consider propensity-score matching or IV methods as alternatives.

ability of propensity-score matching methods to obtain better results (e.g. Heckman, Ichimura and Todd, 1997; Dehejia and Wahba 2002; Smith and Todd 2005; Dehejia 2005). The migration example we consider here offers many of the features identified by these studies as conducive to more accurate non-experimental estimation. Moreover, the size of the "treatment" considered here is large and strongly significant. This contrasts with the treatment effect in Lalonde's NSW male sample of only a 29% increase in earnings (with a t-statistic of only 1.82). Even with these favorable conditions, the non-experimental estimators still overstate the income gains. However, we find that the more recent refinements of propensity-score matching do enable more precision, and provide point estimates which are not statistically different from the experimental estimator.

The rest of this paper is structured as follows. Section 2 describes the immigration process used as the natural experiment and the sampling method and data from the Pacific Island-New Zealand Migration Study (PINZMS). Section 3 looks directly at selection into migration, Section 4 constructs the experimental estimates, Section 5 estimates five different types of non-experimental estimates, and Section 6 concludes.

2. The Pacific Access Category and PINZMS Data

The natural experiment we use is based on the Pacific Access Category (PAC) under New Zealand's immigration policy. The PAC was established in 2001 and allows an annual quota of 250 Tongans to migrate as permanent residents to New Zealand without going through the usual migration categories used for groups such as skilled migrants and business investors.[5] Specifically, any Tongan citizens aged between 18 and

[5] The Pacific Access Category also provides quotas for 75 citizens from Kiribati, 75 citizens from Tuvalu, and 250 citizens from Fiji to migrate to New Zealand.

45, who meet certain English, health and character requirements,[6] can register to migrate to New Zealand.[7] Many more applications are received than the quota allows, so a ballot is used by the New Zealand Department of Labour (DoL) to randomly select from amongst the registrations. The probability of success in the ballot is approximately 10%. Thus, we have a group of migrants and a comparison group who are similar to the migrants, but remain in Tonga only because they were not successful in the lottery. Once their ballot is selected in the lottery, applicants must provide a valid job offer in New Zealand within six months in order to have their application to migrate approved.

The other options available for Tongans to migrate are fairly limited, unless they have close family members abroad. Ninety-four percent of all Tongan migrants are located in New Zealand, the United States and Australia.[8] In the 2004/05 financial year New Zealand admitted 1482 Tongans, of which 58 entered through a business/skilled category, 549 through family sponsored categories and 749 through the Pacific Access Category.[9] Australia admitted 284 Tongans during the same financial year.[10] The United States admitted 324 Tongans in the 2004 calendar year, comprising only five under employment-based preferences and 290 under immediate relative or family-sponsored

[6] Data supplied by the DoL for residence decisions made between November 2002 and October 2004 reveals that out of 98 applications, only 1 was rejected for failure to meet the English requirement, and only 3 others were rejected for failing other requirements of the policy.

[7] The person who registers is a Principal Applicant. If they are successful, their immediate family (spouse and children under age 18) can also apply to migrate as Secondary Applicants. The quota of 250 applies to the total of Primary and Secondary Applicants, and corresponds to about 70 migrant households. During the period we study Tongans had to be in Tonga to make their residence application. The regulations have since changed so that Tongans lawfully in New Zealand (e.g. students) can also lodge applications for residence if successful in the PAC ballot.

[8] Source: GTAP database of Parsons et al. (2005).

[9] Source: Residence Decisions by Financial Year datasheet provided by New Zealand Department of Labour. Note that the high number of PAC approvals in the 2004/05 financial year reflects backlog from prior PAC ballots which were not approved until this time. Migrants under the family sponsored categories were mainly parents and spouses/domestic partners.

[10] Source: Settler Arrivals 2004-2005, Australian Government Department of Immigration and Multicultural Affairs.

categories.[11] Thus, the PAC accounted for 42% of all migration to these three countries, and over 90% of non-family category migration.

The Tongan component of the Pacific Island-New Zealand Migration Survey (PINZMS), is a comprehensive household survey designed to take advantage of the natural experiment provided by the PAC. The survey design and enumeration, which was overseen by the authors in the first half of 2005, covered random samples of four groups: (i) Tongan migrants to New Zealand, who were successful participants in the 2002/03 and 2003/04 PAC lotteries, (ii) successful participants from the same lotteries who were still in Tonga, either because their application for New Zealand residence was still being processed, or because it was not approved (typically because of lack of a suitable job offer) [12] (iii) unsuccessful participants from the same lotteries who were still in Tonga, and (iv) a group of non-applicants in Tonga.[13]

---

[11] Source: 2004 Yearbook of Immigration Statistics, U.S. Department of Homeland Security Office of Immigration Statistics.

[12] The initial sample frame for groups (i) and (ii) was a list of the names and addresses of the 278 (out of almost 3000 applicants) successful participants in the 2002/03 and 2003/04 migration lotteries, which was supplied under a contractual arrangement with the New Zealand Department of Labour, with strict procedures used to maintain the confidentiality of participants. Approximately 100 of these successful ballots had been approved for residence in New Zealand by the time of the survey, although some of those families had not yet moved to New Zealand. We managed to locate 65 of the families that had migrated, giving a sampling rate of over 70%. The data on the application forms is very limited, preventing a detailed comparison of the characteristics of individuals in our survey to those who we could not locate. However, we are able to check and confirm that the migrants who we did not locate do not differ significantly from those in our sample in terms of the proportion who are male, date at which the residence decision was approved, and last date of entry into New Zealand. It was easier to draw a random sample of 55 of the successful ballots that had not yet migrated, because the DoL records included postal and home addresses and telephone numbers in Tonga. This non-migrant group includes those whose applications were rejected and those whose applications were still being processed. We use the actual number of accepted and processed/rejected applications to weight our sample.

[13] The initial sample frame for the unsuccessful ballots in the 2002/03 and 2003/04 lotteries (group (iii)) was a list of names and addresses provided by the DoL. The details for this group were less informative than those for the successful ballots. Only a post office box address was supplied and there were no telephone numbers. Thus, it was not possible to determine whereabouts in Tonga those with unsuccessful ballots lived. We used two strategies to derive a sample of 78 unsuccessful ballots from this information: first, as part of our survey of the migrants in New Zealand we had obtained details about the location of remaining family (almost 60% of migrants still had family occupying their previous dwelling in Tonga). We used this information to draw a sample of unsuccessful ballots from the same villages (implicitly using the village of residence when the applicant entered the ballot as a stratifying variable). We also used the

Table 1 examines how random the sample we have is by comparing means of ex-ante characteristics for lottery winners and lottery losers among the principal applicants in our sample. The point estimates of the means are similar in magnitude for the two groups and we can not reject equality of means for any of the variables. This is as would be expected with the random selection of ballots among applicants in the Pacific Access Category.

The sample of non-applicants was obtained by selecting 60 households, with at least one member aged 18 to 45, in either the same villages that the migrants had been living in prior to migrating or in the same villages that unsuccessful ballots were found in. An initial screening question was used to check that no-one in the household had previously applied for the migration lottery. Data on employment, income, and demographics was collected on all members of these households. Additional questions on the reasons for not applying, the size of the family networks in New Zealand, and expectations, were asked of the oldest member aged 25-35 in the household, or of the oldest member aged 18-45 if no one was aged 25-35. We will refer to this group of individuals which received the extended questions as the group of pseudo-applicants.

Table 2 presents the proportion employed, mean hours worked, and mean work income among the different groups in our sample. The mean weekly income from work among migrants is NZ$425, compared to $81-104 for applicants for the Pacific Access Category (PAC) lottery who did not migrate, and $41 among all individuals aged 18 to 45

Tongan telephone directory to find contact details for people included in the list of names supplied by DoL. To overcome concerns that this would bias the sample to more accessible areas around the capital city of Nuku'alofa, who are more likely to have telephones, we deliberately included in the sample households from the Outer Islands of Vava'u and 'Eua.

in non-applicant households.[14] A t-test of equality of means strongly rejects the null hypothesis of equality of migrant income with any of the other groups. The point estimates suggest that migrants are more likely to be employed than non-migrants, and work slightly longer hours. However, these differences are not significant given our sample size.

3. Who applies for the PAC and who migrates under it? Direct evidence on selection

Most datasets on migrants lack information on earnings prior to migration, leading much of the literature to focus on comparing observable characteristics of migrants to those of non-migrants when examining selection into migration (e.g. Borjas 1987, Chiquiar and Hanson 2005). The average migrant in our sample has 11.7 years of education, compared to 10.5 years among the non-migrants, showing positive selection in terms of observable skill. However, the concern in using non-experimental estimators to measure the income gains from migration is that migrants also differ from non-migrants in terms of unobserved qualities. Using our data we can examine whether there is positive or negative selection on unobservables if, as in the existing literature, one were to only observe age, education, and other socioeconomic characteristics of migrants and non-migrants, but not pre-migration earnings.

We first examine the overall extent of selection by comparing the pre-migration income of migrants to that of observationally similar non-applicants via the following regression:

$$Income_{i,t-1} = \alpha + \beta * Migrant_{i,t} + \gamma' X_{i,t} + \varepsilon_{i,t-1} \quad (1)$$

[14] At the time of the survey, NZ$1=US$0.72.

where X consists of a set of controls, such as age, education, gender, marital status, height, and migrant network, and *Migrant* is a dummy variable taking the value one if person i applies for the lottery and migrates in the next period, and zero if they don't apply for the lottery.

We then consider selection into the lottery by using (1) to compare lottery applicants to non-applicants, replacing the *Migrant* dummy variable with a dummy variable for applying for the lottery. We compare the income for migrants in the 12 months prior to migration to the income of non-applicants in 2003, which corresponds to a similar reference period. The coefficient β then indicates whether migrants or applicants earned more or less prior to applying for the lottery than non-applicants, conditional on their observed characteristics. We carry out this analysis for the two groups of non-applicants: all individuals aged 18-45, and the set of pseudo-applicants.

The first two columns of Table 3 report the results of estimating equation (1), comparing migrants to all 18-45 year old non-applicants. The coefficient β is positive and highly significant. Migrants and non-applicants are seen to differ both in terms of observables and unobservables. Controlling for observables lowers the difference in lagged income from \$56 per week to \$29 per week. However, given that the average income of non-applicants in this group is \$33 per week, we see that migrants earned almost twice as much as observationally similar non-applicants prior to them migrating. Similar results are shown in columns (3) and (4), where we consider selection into the lottery and compare all principal applicants to non-applicants. We can not reject equality of the coefficient on migrating in column (2) with the coefficient on applying in column (4).

In Columns (5) and (6) we compare migrants to the pseudo-applicants. Despite the smaller sample, we still find a statistically significant positive coefficient on the migration dummy. The average income for the pseudo-applicants was $61 per week, so migrants are estimated to have earned over 35% more than observationally equivalent non-applicants in the pseudo-applicant group. Given this evidence of positive selection on unobservables, we therefore expect non-experimental estimators to overstate the income gains from migration: something that will be tested in Section 5.

In addition to selection into the PAC, we are also interested in whether there is selective compliance to the treatment we consider of winning the lottery. The last two columns of Table 3 examine this by modifying equation (1) to compare the pre-migration incomes of lottery winners who migrate to lottery winners who had not migrated at the time of the survey. The coefficient on migrating is found to be positive, but close to zero in magnitude ($13 per week without controls and $7 a week with controls), and insignificant, with t-statistics below 0.9 in absolute value. Migrants therefore do not appear to differ greatly from non-migrant lottery winners in terms of unobservable characteristics affecting pre-migration labor market earnings.

4. Experimental estimates of the income gain from migration

4.1. Estimating treatment effects using experimental data

To determine the income gains from migration, one must compare the earnings of the migrant to what they would have earned in their home country had they not migrated. Typically, it is not possible to readily identify this unobserved counterfactual outcome. However, the PAC lottery system, by randomly denying eager migrants the right to move to New Zealand, creates a control group of individuals that should have the same outcomes as what the migrants would have had if they had not moved. In our application, a comparison of mean income for lottery winners who migrate and lottery losers can be used to obtain an experimental measure of the gain in income from migration. This simple comparison of means at the bottom of Table 2 shows a $320 increase in weekly work income from migrating.

As discussed in Heckman et al. (2000), this simple experimental estimator of the treatment effect on the treated (SEE-TT) is biased if control group members substitute for the treatment with a similar program or if treatment group members drop out of the experiment. In our application, *substitution* bias will occur if PAC applicants who are not drawn in the lottery migrate to New Zealand through an alternative visa category such as the family or skills category or migrate to another country and *dropout* bias will occur if PAC applicants whose names are drawn in the lottery fail to migrate to New Zealand. We do not believe that substitution bias is of serious concern in our study, as individuals with the ability to migrate via other arrangements will likely have done so previously given the low odds of winning the PAC lottery.[15]

However, as shown in Table 2, dropout bias is a more relevant concern; only one-third of lottery winning principal applicants had migrated to New Zealand at the time of our survey. A number of the other individuals are in the process of moving, while others are unable to move due to the lack of a valid job offer in New Zealand.[16] The SEE-TT

[15] We did not come across any incidences where remaining family members told us that the unsuccessful applicant had migrated overseas during our fieldwork.

[16] Lottery winners have six months to lodge a formal residence application containing evidence of a job offer. It then typically takes three to nine months for applicants to receive a decision on their application, after which those who are approved have up to one year to move. Relatively few applications are rejected due to lack of a valid job offer, but lack of a job offer prevents many lottery winners from lodging residence applications.

estimate of a \$320 increase in weekly income from migrating will then only be a consistent estimate of the income gains from migration if there is no selection as to who migrates among those successful in the lottery. The previous section showed that those who migrated had slightly higher pre-migration income to lottery winners who didn't migrate, although this difference wasn't significant. However, selection may be stronger in terms of expected post-migration income, with those who expect higher income after migration being more likely to migrate. In such case, we would expect the SEE-TT to overstate the income gains from migration. We therefore turn our attention to measures of the effect of migration which are consistent even if there is selective migration among those with successful ballots.

4.2. Intention-to-treat effect

Experimental data, in the presence of substitution and dropout bias, can identify the mean impact of a program (eg. winning the lottery) on outcomes (eg. income for PAC applicants), also known as the intention-to-treat effect (ITT). This estimator can be computed by comparing the mean income for ballot winners to that for ballot losers. As shown at the bottom of Table 2, on average, winning the PAC lottery is estimated to increase weekly income by \$91.

While the results in Table 1 show that the lottery did indeed achieve reasonably comparable groups, the small size of our sample may have resulted in some differences between successful and unsuccessful ballots. To improve the efficiency of our ITT estimate, we re-estimate the ITT using the OLS regression model described in equation (2) to control for the observable pre-existing characteristics of the two groups:

$$\text{Income}_i = \alpha + \beta^{*}\text{BallotSuccess}_i + \delta' X_i + \omega_i \quad (2)$$

Column 1 of Table 4 first estimates this regression with no controls, repeating the estimate of $91 obtained as the difference in means. In Column 2 we add a set of controls for pre-existing characteristics of applicants. These include standard wage equation variables, such as age, sex, marital status, and years of education. In addition, we include height as a pre-existing measure of health, and whether or not the applicant was born on the main island of Tongatapu, as a measure of having more urban skills. The addition of these controls reduces the size of the estimated effect only slightly, to $90, which is not significantly different from that obtained without controls. Column 3 controls further for past income, which is expected to also capture the effect of a host of unobserved individual attributes that determine income. The addition of this term only marginally changes the estimated intent-to-treat effect, which is now estimated to be $87. The fact that the estimated program effect changes only slightly in magnitude as we add the controls is consistent with the result in Table 1, which showed that the lottery succeeded in randomizing these controls across successful and unsuccessful ballots.

4.3 Average treatment effects

The ITT gives the impact of receiving a successful ballot in the PAC lottery, rather than the impact of migration, which is the main object of interest. However, we can estimate the impact of migration by using the outcome of the PAC as an instrument for migration. This provides the local average treatment effect (IV-LATE), interpreted as the effect of treatment on individuals whose treatment status is changed by the instrument. In our application, this is the effect of migration on the income of individuals who migrate after winning the lottery. Angrist (2004) also demonstrates that in situations where no individuals who are assigned to the control group receive the treatment (eg. there is no substitution) then the IV-LATE is the same as the average treatment effect on the treated (IV-TT).

Having a successful ballot is of course strongly correlated with migration (the first stage F-statistic is 61.5). Validity of the exclusion restrictions then requires: (i) that success in the lottery is uncorrelated with individual attributes which might also affect income, which is provided by the randomization of the ballot draws; and (ii) that the lottery outcome does not directly affect incomes, conditional on migration status. One could conceive of stories such as that winning the lottery and not being able to migrate causes frustration and leads individuals to work less, or conversely, that winning the lottery acts as a spur to work harder in order to afford the costs of trying to find a job in New Zealand. No evidence of such stories was encountered in our field work, lending support to this identification assumption.[17]

Column 4 of Table 3 then reports the IV-TT estimator when no other controls are included in the regression model, and estimates a gain in weekly work income of almost \$274 from migrating. Column 5 then adds the same control variables used above when estimating the ITT; the estimate increases slightly to \$281. Column 6 adds past income as a further control, measured here as self-reported income from 2003. Past income is likely to capture a host of unobserved attributes of individuals which affect labor market

[17] As an additional check, we matched lottery losers to lottery winners who were still in Tonga using the same set of variables we include in the IV regression (except past income) and tested whether the difference in income before and after the ballot differed significantly between lottery winners in Tonga and lottery losers in Tonga. This difference-in-difference matching estimate finds lottery winners in Tonga to have slightly lower income growth than similar lottery losers, but the difference is not significant (p-value is 0.17). This therefore provides further support for our identification assumption.

performance and the likelihood of migrating conditional on winning the lottery, and is seen to be strongly significant. Each additional dollar of past income in 2003 is associated with 66 cents higher wage income today. Adding past income as a control results in an estimated income gain from migration of $274 per week. This is the same as was obtained in the model with no covariates, and confirms that randomization succeeded in making ballot success orthogonal to the other variables.

Therefore, after controlling for observable differences remaining after randomization, we estimate that a successful ballot increases expected income of PAC applicants by $87 per week, while migrating increases mean income by $274. Given that mean income of applicants with unsuccessful ballots is $104, this represents a 84% increase in expected income from winning the lottery, and a 263% increase in income from migrating.

5. Non-experimental estimators

The natural experiment provided by the use of a lottery to admit Pacific Islanders to New Zealand provides a unique opportunity to estimate the gain in income from migration. Other studies of migration are forced to use non-experimental methods to attempt to deal with the selectivity issues associated with migration, comparing the incomes of migrants to that of non-migrants of similar observable characteristics. In this section we explore how well such methods work in practice, comparing the results obtained from different non-experimental methods to the experimental results described above.

This approach for studying the validity of non-experimental methods has a long history in the labor program evaluation literature. For example, in perhaps the first attempt to do so, Lalonde (1986) compared experimental estimates from the National Supported Work (NSW) Demonstration to non-experimental results calculated using control groups created from household survey data. For this program and treatment, Lalonde found that non-experimental methods did a poor job of replicating the experimental results. Heckman, Ichimura and Todd (1997), Dehejia and Wahba (2002), and Smith and Todd (2005) each further exploit the data collected for the NSW to examine whether particular refinements to non-experimental methods can lead to a better replication of the experimental results.

In summary, these papers demonstrate that more accurate non-experimental estimates can be achieved if the treatment and non-experimental control groups are: i) compared over a common support (eg. the distribution of the likelihood of receiving the treatment is similar in both groups), ii) located in the same labour markets, and iii) administered the same questionnaire (eg. data is collected from both groups in an identical manner). A significant improvement can further be achieved if data is collected from both the pre- and post-treatment periods and a 'difference-in-differences' estimator is used to control for unobserved differences between the treatment and control groups by differencing out individual fixed effects which are correlated with both the outcome and the likelihood of being treated. Nonetheless, even with these refinements, Smith and Todd (2005, p.305) conclude, "Our analysis demonstrates that while propensity score matching is a potentially useful econometric tool, it does not represent a general solution to the evaluation problem."

Recall that PINZMS collects data for a sample of non-applicants to the lottery selected from either the same villages that the migrants had been living in prior to migrating or in the same villages that unsuccessful ballots were found in and administers them an identical questionnaire to the one given to other non-migrants in our sample (eg. the experimental control group). Thus, these individuals serve as an ideal non-experimental control group on which to test alternative methodologies for estimating the gains from migration. As discussed above, all individuals in our sample report their income from the previous year allowing us to also implement a 'difference-in-differences' estimator.

Before proceeding with microeconomic non-experimental estimators, it is worth comparing the experimental estimate of the income gains to the difference in per capita income and wages across countries, used as the basis for calculations of the sort undertaken by Walmsley and Winters (2003). In 2004, New Zealand's GDP per capita was NZ$30,469, while Tonga's was NZ$2,044.[18] This difference in GDP per capita therefore equates to NZ$546 per week, or twice as large as the actual gain experienced by the average migrant in our survey. The difference in manufacturing wages of NZ$635 is even larger, with the experimental income gain only 43% of this difference.[19]

[18] Source: World Bank GDF and WDI Central (August 2005 update) for population and GDP. The exchange rate of 1.372 pa'anga per NZ dollar prevailing at the time of our survey was used to convert Tongan GDP per capita to New Zealand dollars. We also collected prices in both countries and calculate the purchasing power parity rate to be extremely close to the actual exchange rate, at 1.368 pa'anga/NZ$ (see McKenzie, Gibson and Stillman (2006) for details)

[19] This calculation uses average manufacturing weekly income from the Tongan Manufacturing Census in 2002 (www.spc.int/prism/Country/To/stats/Economic/Production/Manufacturing/wages_salaries.htm) and the New Zealand Quarterly Employment Survey (averaged over 2002), converts Tongan pa'anga to New Zealand dollars at the 2002 average exchange rate, and then uses the New Zealand consumer price index to convert 2002 dollars to March 2005 dollars.

5.1. The Single Difference Estimator

We begin by examining whether a simple single difference estimate calculated using only information from the migrant group provides a good estimate of the income gains from migration. Several recent surveys of new immigrants (eg. the Longitudinal Immigrant Survey: New Zealand (LisNZ); and the New Immigrant Survey (NIS) in the U.S.) ask about income prior to migration. Thus, one approach to estimating the average income gain from migration is to calculate the mean difference between the migrant's pre-migration and post-migration incomes.

There are several possible sources of bias in such an estimate. The counterfactual one would ideally like is what a given individual would be earning in the current time period if he or she didn't migrate; this could be different from what they earned before migration due to macroeconomic factors such as aggregate growth or to changes in the income-earning potential of the individual over time, such as growing labor market experience. An additional potential form of bias when it comes to estimation is that recall of previous income may involve omissions or telescoping errors, leading to non-mean zero measurement error.

The first row of Table 5 provides the single-difference estimate, calculated as the difference between the current income of our migrant sample and what they reported earning prior to migration. Based on this method, we would estimate an income gain of $341. Comparing this to columns 4 and 6 of Table 4, we calculate that this method results in estimated income gains which are 25% higher than the experimental estimate. We can quantify one source of bias in this estimator by examining the increase in income that occurred for the unsuccessful ballots in Tonga. Mean income increased $28 per week for this group, which accounts for 42% of the difference in income gains estimated via this method compared to the experimental estimates.

5.2. OLS

A second non-experimental method commonly used to estimate the returns from migration is to assume that all differences between migrants and non-migrants which affect income are captured by the regressors in an OLS regression. One then estimates λ through the following regression:

$$Income_i = \kappa + \lambda^{*}Migrate_i + \pi' X_i + \upsilon_i \qquad (3)$$

We estimate equation (3) by combining the sample of migrants in New Zealand with the sample of non-applicants in Tonga. We do this for two samples in Tonga. One individual from each household of non-applicants was asked a longer set of questions, including information on their family networks in New Zealand, expectations about the future, and other broader issues. This was done for the group of pseudo-applicants, consisting of the oldest member aged 25 to 35 in the non-applicant household (or oldest member aged 18-45 if there was no 25 to 35 year-old). The first sample we use combines these individuals with the migrants. The second sample uses all individuals aged 18 to 45 in the non-applicant households.

We consider two sets of control variables when estimating equation (3). The basic specification includes the same controls as used for the experimental estimates, with the exception of past income (which we keep for the difference-in-differences estimator below). We also allow for a more flexible specification by interacting the male dummy variable with each of the other regressors in the base specification, and including fourth order polynomials in age and years of education, along with the interaction of age and years of education. An F-test of joint significance of these additional 12 regressors has a p-value of 0.415 for the restricted sample, and 0.056 for the sample using all individuals aged 18 to 45.

Table 5 shows that this results in an estimated income gain from migration of $384-391 using the restricted sample, and an income gain of $347-360 using the wider sample. Appendix 1 provides the full regression results for the base specifications. Comparing these with the experimental estimates, we see that the restricted sample overestimates the income gain by 40%. The full sample overestimates the income gain by 31% under the base specification, and by 27% under the specification with polynomials. The direction of this bias is consistent with the view that migrants have more drive or greater labor market ability than non-migrants.

Column 2 of Appendix 1 repeats this regression for the full-sample of 18 to 45 year olds without including any of the control variables in equation (3). The coefficient on migration is $386. Adding the observable characteristics as controls in column 3 reduces this to $360, showing positive selection on observables. However, the change in the migration coefficient from adding these controls is not significant, and their addition only reduces the overestimation of the income gains from 41% to 31%. It therefore seems that most of the OLS bias is due to selection on unobserved characteristics.

5.3. Difference-in-Differences

Using self-reported past income, we can also control for time invariant individual attributes which affect labor market income via difference-in-differences regression.

Since we do not have panel data on all the control variables, we estimate the following version of the difference-in-differences regression :

$$Income_i - PastIncome_i = \kappa + \lambda * Migrate_i + \pi' X_i + \upsilon_i \qquad (4)$$

Controlling for past income lowers the estimated income gain to $375 using the restricted sample and $328 using the wider sample. Columns 4 and 5 of Appendix 1 provide the full set of coefficients. These estimates are now respectively 37% and 20% higher than the experimental estimate, although given our sample sizes, we can only reject equality with the experimental estimate for the narrower sample. There are two main possible sources of remaining bias. The first is that unobserved characteristics like drive and ability may be rewarded differently in the New Zealand and Tongan labor markets, so that individual effects are time-varying. The second is that we are comparing migrants to not-very-similar non-migrants, and so the assumption of a common underlying trend in labor income is not tenable. The latter assumption is eased by using the wider sample, and can be relaxed further by ensuring that the migrants are compared to sufficiently similar non-migrants, which the next method attempts to do.

5.4. Propensity-Score Matching

Propensity-score matching is perhaps the non-experimental evaluation technique which has attracted most research interest in recent years, with proponents claiming that it can replicate experimental benchmarks when appropriately used (Dehejia and Wahba, 2002; Dehejia 2005). The standard approach first estimates a probit equation for the probability of migrating, and then matches each migrant to non-applicants with similar predicted probabilities of migration. This enables migrants to be compared to individuals who are similar in terms of observed characteristics. Once the matches are constructed, the gain in income is calculated as the mean income for migrants less the mean income for the matched sample. We use the nearest-neighbor matching, and following Abadie et al. (2001) match each migrant to the four nearest neighbors.

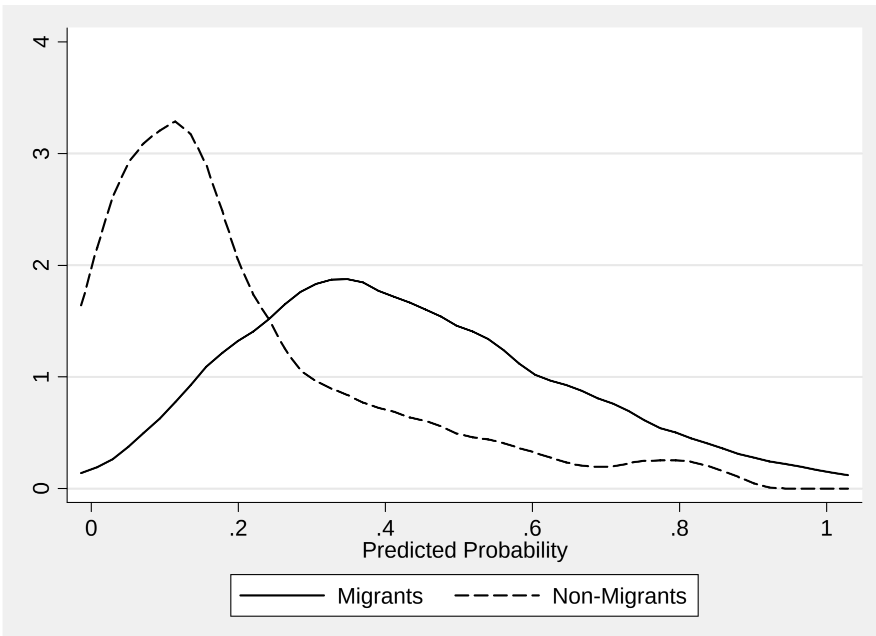
Our base variable specification uses the same set of control variables as used in the regression analysis to form the match. The existing literature (Heckman, Ichimura and Todd, 1997; Smith and Todd, 2005) have noted that difference-in-difference matching estimators can perform substantially better than cross-sectional matches[20]. While we do not have panel data on all matching variables, the inclusion of past income allows us to obtain estimates similar in spirit to difference-in-difference matching. Figure 1 then shows kernel densities of the propensity scores when past income is included alongside the other regression controls in forming the match. Note that there is considerable overlap in the distributions, with some migrants and some non-applicants in almost all the range. The propensity score for the migrant group ranges from 0.069 to 0.947, while that of the non-applicant comparison group ranges from 0.000 to 0.789. Estimation is restricted to the area of common support, where the two distributions overlap.

One potential criticism is that these base specifications are relatively parsimonious, using only 6 or 7 covariates to form the match. This is in large part due to the need to use retrospective questions and time invariant attributes to form the match,

[20] A concern in the evaluation of labor training programs is the possibility of a dip in earnings prior to participation in such programs (Ashenfelter, 1978), leading Dehejia (2005) to stress the use of two or more years of pre-treatment earnings when using matching to evaluate such programs. We only have income for one period prior to migration, but are able to check for a pre-migration-lottery dip by comparing labor histories for unsuccessful lottery applicants to those who do not apply for the lottery. In the working paper (McKenzie, Gibson and Stillman 2006) we show that individuals who applied for the lottery in early 2005 did not have statistically different income in 2004 than similar individuals, with similar incomes in 2002 and 2003, who did not apply for the lottery. We therefore rule out a pre-migration lottery earnings dip, and believe that using one rather than two year's pre-treatment earnings should not greatly affect the results.

since the survey was cross-sectional. To investigate the robustness of the matching results to a more flexible specification, we also estimated the propensity score allowing for interactions of sex with each of the other covariates, quartics in age and years of schooling, and an interaction between age and education, for a total of 19 covariates.

Figure 1: Propensity Scores for Migrants and Non-migrants



For each of these three specifications of variables used to form the match we calculate the sample average treatment effect (SATE) and sample average treatment effect for the treated (SATT) following Imbens (2004). Table 6 reports these estimates in rows A, B and C.[21] Once we control for past income, the SATE and SATT are very similar to one another. We focus on the SATT, since this is more directly comparable to the experimental treatment effect estimated using the migration lottery as an instrument for migration.

[21] Propensity-score matching was estimated in STATA using the routine described in Abadie et al. (2001).

Under the basic specification of variables to match on, the estimated income gain is \$364 per week, 33% higher and significantly different from the experimental estimate of \$274. Adding past income as a control lowers the bias to 28.5% and adding interactions reduces it to 27.4%. The t-statistic for testing equality of the treatment effect in model C with the experimental estimate is 1.61, close to the margin of being able to reject equality at the 10% level of significance.

Abadie and Imbens (2005) provide a bias-adjusted matching estimator which matches directly on the covariates rather than on the propensity-score, which has the advantage of not requiring an explicit choice of the propensity score functional form, such as the probit used above. They find that this bias-adjusted estimator performs well compared to the simple matching estimator and to regression in a simulation study. We carry out this bias-adjusted estimator to calculate the SATT for each of the specification sets of variables used above.

Table 6 shows that the bias-adjusted estimator brings the matching treatment effects closer to the experimental estimate, and we can no longer reject equality. In model C, with interactions, the bias is reduced from 27.4% to 19.9%. Dehejia (2005) notes that sensitivity of the matching estimator to small changes in the specification used is one diagnostic as to the quality of the comparison group. The bias-adjusted estimators are not that sensitive to the particular specification used for matching, ranging from \$329 to \$346 per week as the estimated income gain. Based on this, one would therefore be likely to conclude that the matching technique is working reasonably well in this context, even without reference to the experimental data.

Rows D and E of Table 5 conduct two other robustness tests suggested by the literature. The first is to not only estimate the matching estimator over the area of common support, but also to examine robustness to trimming observations in the support with very low or very high probabilities of being selected. Panel D trims propensity scores which are less than 0.01, 0.05, 0.10 and 0.15 or greater than 0.99, 0.95, 0.90 and 0.85 respectively. After the bias-adjustment, the estimated treatment effect is not very sensitive to such trimming, resulting in a bias of 18.9% to 20.1%. The second robustness test examines the sensitivity of the estimator to the number of neighbors used in forming the match. This trades bias for efficiency, which is seen in the smaller standard errors when more neighbors are used. Again the point estimates are very robust to this choice of specification, and result in a 20% higher income gain than is estimated by the experiment.

5.5. Instrumental Variables with a Non-experimental instrument

Like the regression approach, propensity score matching relies on selection on observables, so will overstate the income gains if migrants are more talented or have more drive than observationally similar non-migrants. An alternative approach to non-experimental estimation of the impact of migration explicitly recognizes that migrants are likely to be non-randomly selected, even conditioning on observables, and so attempts to find instruments for migration. An example is Munshi (2003), who uses rainfall in Mexican villages as an instrument for migration when looking at the effect of migration networks on job outcomes in the United States. Given the small size of Tonga, weather variation does not provide an instrument in our application. We instead consider two potential instruments, with varying likelihoods of the exclusion restriction being satisfied.

Several studies looking at the impact of migration on the sending country have employed historic migration networks (e.g. Woodruff and Zenteno 2006, McKenzie and Rapoport 2004). In our context it is likely that having a large network of relatives in New Zealand helps predict whether an individual migrates, and so we consider the effect of using the total number of types of relatives an individual has in New Zealand as an instrument for migration. This is strongly correlated with migration (first-stage F-statistic is 14.3). However, we would be highly concerned that the exclusion restriction is violated for this instrument, since many migrants in our survey said they found their first job in New Zealand through relatives.

We investigate this by estimating equation (3) using the migrant network as an instrument for migration, restricting analysis to the sub-sample for which we have information on their network. Table 5 and column 6 of Appendix 1 show that this results in an estimate of the income gain from migration of $499, which is 82% higher than the experimental estimate.[22] Using a poor instrument results in a bias almost as large as obtained when using the cross-country difference in incomes.

A valid instrument is a variable which predicts whether or not people apply to migrate, but doesn't otherwise affect their labor market outcomes if they move to New Zealand. Our survey asked eligible individuals who didn't apply for the Pacific Access Category why they didn't apply. The most important reason given for not applying was that they did not know the requirements, which 98% of non-applicants listed as a very

[22] One possibility is that the variation induced by network size is estimating a different local average treatment effect to the average treatment effect on the treated estimated by the experiment. However, when we estimate the experimental ATT effect separately for individuals with above and below the median network size, we obtain similar treatment effects for both groups: $281 for those below the median, $306 for those above the median (and we can't reject equality of the two). Hence it seems the overstatement of the gains using this instrument is a result of failure of the exclusion restriction rather than of comparison to a different treatment effect.

important reason for not applying. A distant second among the other reasons given was that they didn't think the chances of getting selected in the lottery were very high, which 12% listed as very important, and a further 60% gave as somewhat important.

This motivates our choice of an alternative potential instrument, which is based on how close the individual's house in Tonga is to the DoL service office in the capital city. Information about the requirements of the Pacific Access Category is obtained from this office, paperwork and help with the applications occurs there, and the applications have to be delivered there. GPS coordinates were taken of each of the households in our survey, and of the DoL office location, and based on these, the (log) of the distance between each household and the DoL office was calculated.[23]

Comparing migrants to non-applicants, we find log distance to be a very strong predictor of migration, with a first stage F-statistic of 21.9. Row 5 in Table 5 and column 7 of Appendix 1 show the resulting estimate of the income gains from migration using log distance as an instrument.[24] The estimated gain is $279, which is only 1.8% higher than the experimental estimate. Hence using a good instrument is able to give us estimates very close to those obtained from the experiment.

[23] A possible threat to the exclusion restriction needed for distance to be a valid instrument is that individuals who lived further away from the DoL office would have lived in more isolated, less urban areas. If this makes them less able to adapt to city life in New Zealand, then we would expect an upward bias in the IV estimator, since it would be in part capturing the returns to more urban experience. To investigate this possibility, for the group of migrants in New Zealand we regressed earnings on our set of controls, including past income, and log distance. The coefficient on log distance is positive (9.0) and insignificant (p-value of 0.76). Thus there is no strong effect of living in a location closer to the DoL office in Tonga on labor income, and if anything, migrants that lived further away earned slightly more in New Zealand. Based on this it is likely that one would conclude that this was a reasonable instrument, even without reference to the experimental comparison.

[24] Similar results were also obtained using a treatment effects model to take account of the fact that the treatment is binary (the estimated gain is $280 using the treatreg command in STATA).

6. Conclusions

The lottery used to select migrant applicants to New Zealand from the Pacific Islands provides a unique natural experiment which can be exploited to estimate the income gains from migration and to examine how successful are non-experimental methods in estimating these gains. We estimate that there is an 84% increase in expected income from winning the lottery, and a 263% increase in income from migrating. This increase is only half of that suggested by differences in per capita income between the two countries, and is also less than suggested by non-experimental comparisons of observationally similar migrants and non-migrants. Tongan migrants are found to be positively selected in terms of both observable characteristics and unobserved labor market attributes such as ability and drive.

If this positive selection on unobservables applies to other migrant groups, it would suggest that the income gains from migration can be considerably less than predicted, perhaps helping to explain the lower than anticipated movement of Eastern European workers to the EU15 following the lowering of migration barriers.[25] However, it also suggests that the global gain in income from an increase in migration may only be half as large as estimates based on wage differentials suggest.

The survey we use is unique in providing a natural experiment to estimate the gain from migration. In other circumstances economists will need to rely on non-experimental estimates of the income gain when making predictions, and so our comparison of experimental and non-experimental methods provides useful guidance as to the potential bias that can this can entail. Our results show that a good instrument (log distance to the office where ballots are deposited) works well, but also illustrate the perils

[25] See Traser (2005) for a preliminary assessment.

of using an instrument with good first-stage power but a priori questionable exclusivity (the migrant network here). If a good instrument is not available, difference-in-differences and propensity score matching with bias-adjustment work best, although both still overstate the income gains by 20%. However, both require collecting information on past income of migrants, which is possible in special surveys of migrants but not typically available when using receiving country census or labor force data. This highlights the importance of better data collection as a first step towards more accurate predictions of the impacts of policy change on migration.

References:

Abadie, A., D. Drukker, J.L. Herr and G.W. Imbens (2001) "Implementing Matching Estimators For Average Treatment Effects in Stata", *The Stata Journal* 1(1): 1-18.

Abadie, A. and G.W. Imbens (2005) "Large Sample Properties Of Matching Estimators For Average Treatment Effects", forthcoming, *Econometrica.*

Angrist, J.D. (2004) "Treatment Effect Heterogeneity In Theory and Practice", *Economic Journal* 502: C52-C83.

Ashenfelter, O. (1978) "Estimating the Effects of Training Programs On Earnings", *Review of Economics and Statistics* 60: 47-57.

Borjas, George J. (1987) "Self-selection and The Earnings Of Immigrants", *American Economic Review* 77(4): 531-53.

Chiquiar, D. and G. Hanson (2005) "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States", *Journal of Political Economy* 113(2): 239-81.

Clark, X., T.J. Hatton and J.G. Williamson (2002) "Where do U.S. Immigrants come from and why?", *NBER Working Paper 8998.*

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Johns Hopkins University Press, Washington DC.

Dehejia, R. (2005) "Practical Propensity Score Matching: A Reply To Smith and Todd", *Journal of Econometrics* 125(1-2): 355-64.

Dehejia, R. and Wahba, S. (2002) "Propensity Score Matching Methods for Non-Experimental Causal Studies", *Review of Economics and Statistics* 84(1): 151-161.

Glewwe, P., M. Kremer, S. Moulin and E. Zitzewitz (2004) "Retrospective vs. Prospective Analyses of School Inputs: The case of flip charts in Kenya", *Journal of Development Economics* 74(1): 251-268.

Heckman, J., Hohmann, N., Smith, J. and Khoo, M. (2000) "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment", *Quarterly Journal of Economics* 115(2): 651-694.

Heckman J., Ichimura H. and Todd P. (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economic Studies* 64 (4): 605-654.

Imbens, G. (2004) "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review", *Review of Economics and Statistics* 86(1): 4-29.

Lalonde R (1986). "Evaluating the Econometric Evaluations of Training Programs." *American Economic Review* 76: 604–620.

LaLonde, R. and Topel, R.. (1997) "Economic Impact Of International Migration And The Economic Performance Of Migrants", In M. Rosenzweig and O. Stark (ed.) *Handbook of Population and Family Economics*, Elsevier Science, pp. 799-850.

McKenzie, D., J. Gibson and S. Stillman (2006) "How Important is Selection? Experimental Vs Non-experimental Measures of the Income Gains from Migration", *World Bank Policy Research Working Paper No.*

McKenzie, D., and Rapoport, H. (2004) "Network Effects And The Dynamics Of Migration And Inequality: Theory And Evidence From Mexico", *Stanford University Center for International Development (SCID) Working Paper 191*.

Munshi, K. (2003) "Networks in the Modern Economy: Mexican Migrants in the United States Labor Market", *Quarterly Journal of Economics* 118(2): 549-597.

Parsons, Christopher, Ronald Skeldon, Terrie Walmsley and L. Alan Winters (2005) "Quantifying the International Bilateral Movements of Migrants", Mimeo. The World Bank

Robinson, C. and Tomes, N. (1982) "Self-selection and Inter-provincial Migration in Canada", *Canadian Journal of Economics* 15(3): 474-502.

Rosenbaum, P. and Rubin, D. (1983) "The Central Role of The Propensity Score In Observational Studies For Causal Effects", *Biometrika* 70: 41-55.

Smith J and Todd P (2005). "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators." *Journal of Econometrics* 125 (1-2): 305-353.

Traser, Julianna (2005) Report on the Free Movement of Workers in EU-25: Who's afraid of EU enlargement? Brussels: European Citizen Action Service.

Walmsley, T.L. and L.A. Winters (2003) "Relaxing the Restrictions on the Temporary Movements of Natural Persons: A Simulation Analysis", *CEPR Discussion Paper No. 3719.*

Woodruff, C. and Zenteno, R. (2006). "Remittances and Microenterprises in Mexico." Forthcoming, *Journal of Development Economics*.

TABLE 1: TEST FOR RANDOMIZATION

Comparison of Ex-ante characteristics of principal applicants in successful and unsuccessful ballots

	Sample Means APPLICANTS Successful Ballots	Sample Means APPLICANTS Unsuccessful Ballots	T-test of equality of means p-value
Age	33.6	33.7	0.91
Years of schooling	11.9	11.5	0.37
Proportion male	0.55	0.51	0.52
Proportion born on Tongatapu	0.75	0.79	0.54
Proportion who had been to NZ before 2000	0.39	0.35	0.63
Proportion who are married	0.60	0.62	0.77
Height	171.6	169.3	0.16
Proportion selling fish in 2003	0.03	0.06	0.40
Proportion selling crops in 2003	0.22	0.26	0.52
Income in 2003/before moving	103.7	88.0	0.32
Proportion with the following family members living in NZ at time of last application:			
Father/Father-in-law	0.38	0.44	0.45
Mother/Mother-in-law	0.40	0.35	0.46
Brother/Brother-in-law	0.72	0.71	0.78
Sister/Sister-in-law	0.64	0.60	0.63
Aunt or Uncle	0.65	0.55	0.17
Total Sample Size	**120**	**78**	

Average first quarter 2005 exchange rate of 1 Pa'anga = 0.729 NZ Dollars used for comparing mean incomes

TABLE 2: SAMPLE MEANS OF EMPLOYMENT, HOURS WORKED AND WAGES

	Observations	Proportion Employed	Mean hours worked per week	Mean weekly income from work (NZ Dollars)
APPLICANTS	198	0.723	27.3	108.9
Successful Ballots	120	0.662	28.4	194.7
Migrants	65	0.754	33.3	424.5
Non-migrants	55	0.618	26.0	81.1
Unsuccessful Ballots	78	0.731	27.1	104.1
NON-APPLICANTS	60	0.672	24.2	69.5
All Non-applicants 18-45	180	0.439	16.2	41.4
T-tests of equality of means				
Successful Ballots vs Unsuccessful Ballots		0.349	0.683	0.000
Migrants vs Non-migrant Successful Ballots		0.111	0.086	0.000
Migrants vs unsuccessful ballots		0.754	0.105	0.000
Pure Experimental Estimators of the Gain in Income from Migration				
Intention-to-treat effect			90.6	
SEE-TT			320.4	

Notes:
Average first quarter 2005 exchange rate of 1 Pa'anga = 0.729 NZ Dollars used for comparing mean incomes
SEE-TT is the simple experimental estimator of the effect of the treatment on the treated, and compares migrants to unsuccessful ballots.

Table 3: A Direct Look at Selection
Dependent Variable: Labour Income prior to applying/migrating

	Selection into the Lottery						Selection into Migration	
	All 18-45 year olds				Pseudo-applicants		Lottery Winners	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migration Dummy Variable	55.95	29.46			28.28	21.83	13.14	7.43
	(5.30)***	(3.13)***			(2.16)**	(1.97)*	(0.86)	(0.51)
Applicant Dummy Variable			47.12	23.06				
			(6.53)***	(3.06)***				
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.145	0.330	0.099	0.205	0.037	0.331	0.006	0.129
Number of Observations	234	221	366	350	120	117	119	117

Notes:
Absolute value of t-statistic in parentheses, using heteroskedasticity-consistent standard errors
*, **, and *** indicate significance at the 10%, 5% and 1% levels
Controls include gender, age, marital status, years of education, place of birth and height. The total number of relative types in New Zealand is also used as a control in the pseudo-applicant and lottery winner regressions.

TABLE 4: REGRESSION-BASED EXPERIMENTAL ESTIMATES

Dependent Variable: Weekly Income from Work in New Zealand Dollars

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV	(6) IV
Ballot Success Dummy	90.634	89.741	87.390			
	(3.68)**	(3.71)**	(3.89)**			
Male Dummy		-29.070	-23.855		-33.104	-27.772
		(1.19)	(1.08)		(1.43)	(1.33)
Married Dummy		-4.493	24.535		-10.695	18.376
		(0.16)	(1.05)		(0.40)	(0.82)
Age Dummy		0.558	-0.886		0.987	-0.462
		(0.34)	(0.71)		(0.64)	(0.41)
Years of Education		13.427	4.605		12.034	3.274
		(2.03)*	(1.18)		(1.99)*	(0.91)
Born on Tongatapu Dummy		29.167	27.600		29.594	28.005
		(1.55)	(1.87)		(1.64)	(2.04)*
Height		1.281	0.381		1.249	0.353
		(1.96)	(0.92)		(2.04)*	(0.93)
Past income			0.662			0.660
			(6.98)**			(7.31)**
Migration Dummy				273.996	281.050	273.736
				(4.46)**	(4.56)**	(4.99)**
Constant	104.051	-297.878	-60.422	104.051	-285.011	-48.595
	(8.85)**	(2.45)*	(0.74)	(8.90)**	(2.45)*	(0.66)
First stage F-statistic on instrument:				66.53	61.88	61.51
Observations	197	191	190	197	191	190
R-squared	0.04	0.14	0.27			

Robust t statistics in parentheses

* significant at 5%; ** significant at 1%

Average first quarter 2005 exchange rate of 1 Pa'anga = 0.729 NZ Dollars used for comparing mean incomes

TABLE 5: NON-EXPERIMENTAL ESTIMATES

Method:	Estimate	s.e.	Percent difference compared to experimental estimate 273.996	Testing equality with 273.996 T-stat
1) Using pre-migration income as the counterfactual	341.3	46.4	24.6	1.45
2) Selection on Observables: OLS regression				
oldest member aged 18-45: base specification	383.5	46.4	40.0	2.36
oldest member aged 18-45: polynomials	391.0	50.9	42.7	2.30
all members aged 18-45: base specification	360.0	41.2	31.4	2.09
all members aged 18-45: polynomials	346.9	41.8	26.6	1.75
3) Difference-in-Difference Regression				
oldest member aged 18-45	375.2	46.4	36.9	2.18
all members aged 18-45	328.5	42.8	19.9	1.27
4) Instrumental Variables using migrant network	498.8	209.6	82.0	1.07
5) Instrumental Variables using log distance to NZIS office	279.0	87.4	1.8	0.06

Notes: Experimental estimate is the IV estimate from column 6, Table 3.

TABLE 6: PROPENSITY-SCORE MATCHING ESTIMATES

	Estimate	s.e.	Percent difference compared to experimental estimate 273.996	Testing equality with 273.996 T-stat
A: Matching without using past income				
SATE	335.7	40.2	22.5	1.53
SATT	364.0	44.0	32.9	2.04
bias-adjusted SATT	346.3	45.4	26.4	1.59
B: Matching using past income				
SATE	355.8	43.6	29.9	1.88
SATT	352.2	45.4	28.5	1.72
bias-adjusted SATT	333.4	46.2	21.7	1.29
C: Matching using past income and interactions				
SATE	346.2	44.7	26.3	1.62
SATT	349.1	46.5	27.4	1.61
bias-adjusted SATT	328.6	47.3	19.9	1.16
D: Trimmed, bias-adjusted SATT using specification C				
Trimming 0.01 and 0.99	328.9	47.1	20.0	1.17
Trimming 0.05 and 0.95	329.1	47.4	20.1	1.16
Trimming 0.10 and 0.90	328.5	48.3	19.9	1.13
Trimming 0.15 and 0.85	325.7	49.9	18.9	1.04
E. bias-adjusted SATT in specification C with different numbers of matches				
Nearest neighbor	330.1	59.0	20.5	0.95
Nearest 2 neighbors	330.6	50.9	20.7	1.11
Nearest 5 neighbors	328.6	45.8	19.9	1.19
Nearest 10 neighbors	330.4	43.0	20.6	1.31

Notes: A matches on gender, age, marital status, years of education, place of birth and height. B also includes past income, while C adds interactions of sex with each covariate, quartics in age and years of schooling, and an interaction between age and years of education. Estimation uses the 4 nearest neighbours for matching each observation, except in E.

APPENDIX: NON-EXPERIMENTAL REGRESSIONS

Dependent Variable: Weekly work income for Columns 1-3 and 6-7.
Current Weekly work income - past weekly work income for columns 4 and 5.

	(1) OLS	(2) OLS	(3) OLS	(4) DD	(5) DD	(6) IV	(7) IV
Dummy for Migration	383.490	385.880	360.009	375.226	328.498	498.797	278.960
	(8.27)**	(8.41)**	(8.73)**	(8.09)**	(7.68)**	(2.38)*	(3.19)**
Male Dummy	176.214		75.998	162.739	80.517	200.683	97.966
	(2.62)*		(3.29)**	(2.29)*	(3.70)**	(2.36)*	(4.69)**
Married Dummy	-115.798		-31.548	-125.159	-45.883	-109.672	-40.083
	(1.87)		(1.15)	(1.94)	(1.77)	(1.84)	(1.59)
Age	4.706		2.797	2.314	0.083	4.807	0.692
	(1.38)		(1.57)	(0.67)	(0.04)	(1.44)	(0.38)
Years of Education	-2.056		-2.589	-17.712	-2.940	-10.519	3.226
	(0.18)		(1.03)	(1.44)	(1.30)	(0.60)	(1.05)
Born on Tongatapu	74.661		38.288	43.119	32.406	63.531	51.689
	(1.75)		(1.76)	(1.00)	(1.40)	(1.44)	(2.16)*
Height	7.094		3.589	7.404	5.964	6.397	6.040
	(2.38)*		(2.56)*	(2.42)*	(3.89)**	(2.07)*	(4.25)**
Past Income						0.031	0.351
						(0.10)	(1.22)
Constant	-1,393.759	41.906	-664.329	-1,211.553	-1,022.750	-1,253.125	-1,103.361
	(2.88)**	(9.94)**	(3.09)**	(2.36)*	(3.95)**	(2.47)*	(4.54)**
Sample	oldest 25-35	all 18-45	all 18-45	oldest 25-35	all 18-45	oldest 25-35	all 18-45
Instrument						network	log distance
First stage F-statistic						14.24	22.91
Number of Observations	118	230	230	116	226	116	226
R-squared	0.48	0.45	0.51	0.45	0.45		

Note: Oldest member 18-45 in the household is used when the household doesn't contain a 25-35 year old.
Robust t statistics in parentheses
Average first quarter 2005 exchange rate of 1 Pa'anga = 0.729 NZ Dollars used for comparing mean incomes
* and ** indicate significance at the 5% and 1% levels respectively.