Università Politecnica delle Marche

DIPARTIMENTO DI ECONOMIA

Endogeneity and sample selection in a model for remittances

Giulia Bettin, Riccardo Lucchetti, Alberto Zazzaro

QUADERNI DI RICERCA n. 361

June 2011

Abstract

We estimate a remittance model in which we address endogeneity and reverse causality relationships between remittances, pre-transfer income and consumption. In order to take into account the fact that a large share of individuals do not remit, instrumental variable variants of the double-hurdle and Heckit selection models are proposed and estimated by Limited Information Maximum Likelihood (LIML). Our results for a sample of recent immigrants to Australia show that endogeneity is substantial and that estimates obtained by the methods previously employed in the literature may be very misleading if given a behavioural interpretation.

JEL Class.: F24, F22, C21, C24
Keywords: Selection models, migration, remittances

Indirizzo: Dipartimento di Economia - Università Politecnica delle Marche - P.le Martelli 8, 60121 Ancona – Italy.
g.bettin@univpm.it -- r.lucchetti@univpm.it -- a.zazzaro@univpm.it

Endogeneity and sample selection in a model for remittances*

Giulia Bettin, Riccardo Lucchetti, Alberto Zazzaro

1 Introduction

The history of industrialization and economic development intertwines inextricably with the history of migration and remittance flows. In some emigration countries, the industrial take-off was financed, directly and indirectly, by the remittances of their emigrants (see for example Esteves and Khoudur-Castéras (2009)). On the other hand, the immigrants' labor supply, consumption, investments and the stimulus they produced to national savings and foreign capital were primary growth engines in immigration countries (Hatton and Williamson, 1998; Solimano, 2003).

Nowadays, the situation is little different from the past. Remittance flows represent a major source of income and foreign exchange revenue for many developing countries. According to the Migration and Remittances Factbook 2011 published by the World Bank, remittance flows to developing countries increased from 81.3 billion US dollars in 2000 to 226.7 billion in 2006 and 307.1 billion in 2009, representing 2 percent of developing countries GDP, 85 percent of foreign direct investments (towards developing countries) and almost 255 percent of official development aid. At the same time, for immigration countries remittances can be a sizeable outflow of capital. For example, remittance outflows from Australia rose from 1,053 million US dollars in 2000 to 3,000 million in 2009 (equivalent to 0.3 percent of GDP)[1].

* *We wish to thank Tommaso Frattini, Olga Nottmeyer, Oded Stark and the participants at the conferences "Third TOM Conference" (Hamburg), "Poverty Traps" (Naples), "DIW End of Year Summit" (Berlin), "Migration: A World in Motion" (Maastricht) and at seminars at CSEF (University of Naples Federico II), University of Bologna and University of Ferrara for comments and suggestions.*

Giulia Bettin gratefully acknowledges financial support from the "Transnationality of Migrants" (TOM) Marie Curie Research Training Network.

We also thank the Australian Social Science Data Archive (ASSDA) for making the LSIA dataset available. All errors are ours.

[1]Similarly, among the top immigration countries in 2009, remittance outflows were 1.5

In this perspective, a proper understanding of the individual motivations and determinants of remitting behavior is a key step in analyzing the dynamics of remittance flows at the aggregate level and in designing policies to attract the savings of migrants. In their path-breaking study on motivations to remit, Lucas and Stark (1985) suggest a taxonomy distinguishing three main drives of remittances: "pure altruism", when migrants derive utility from the utility of family and friends at home, "pure self-interest", when migrants are moved by the desire to acquire material and immaterial (reputation, prestige) assets at home; "tempered altruism or enlightened self-interest", when remittances are the result of contractual arrangements between migrants and parents left at home enforced by a mix of altruistic and self-interested forces[2].

Discriminating among remittance motivations empirically has proved to be a very challenging task. A major reason is that datasets are scarce and often largely incomplete, devoid of a time profile and collected either at the sending or receiving end, hence lacking important information to match migrants' and recipients' households[3].

However, identification and data problems aside, existing literature has typically treated the statistical procedure for modelling the remittance behaviour of migrants as neutral to its theoretical analysis. As a consequence, the possibility of endogeneity of the major determinants of remittances has often been overlooked. Moreover, the existence of non-remitting migrants has usually been treated as a merely statistical problem, to be dealt with by using standard endogenous selection models such as Heckman's, rather than a possible outcome (a corner solution) of the choice process by the individual. We argue that these aspects must be taken into account and a truly behavioural analysis of the determinants of remittances can only be attempted in the context of a statistical model in which these features are made fully endogenous.

Recently, a few studies have addressed the issues of reverse causation and omitted variables with regard to the wealth and income of relatives back home (Osili, 2007; Yang and Choi, 2007) and the migrants' intention

percent of GDP in the Russian Federation, 0.48 percent in Germany, 0.34 percent in the United States, and 0.17 percent in the United Kingdom (Ratha, Mohapatra, and Silwal, 2011).

[2]Rapoport and Docquier (2006), Hagen-Zanker and Siegel (2007), Carling (2008) and Stark (2009) provide exhaustive and updated reviews of modern microeconomic theoretical and empirical literature on remittances.

[3]An exception is represented by the paper by Osili (2007), where migrants are considered together with their respective origin-families. Such complete information, on the other hand, comes together with a very limited number of observations (61 pairs).

to return to the home country (Dustmann and Mestres, 2010). Surprisingly enough, however, no previous studies have been concerned with the endogeneity of immigrants' income[4] and consumption/saving behavior[5]. Yet it is highly conceivable that people who wish to remit a larger sum of money are willing to increase the number of hours worked per week. Moreover, immigrants' income may be, at least in part, boosted by an unobserved investment in their education by the family of origin, which may in turn cause a (gratitude or money) debt for immigrants that remittances repay. Similar concerns hold for consumption: immigrants could choose to reduce consumption in the host country in order to increase transfers to the home country or could prefer to invest their savings in earning assets rather than in buying property or other durables. Were immigrants' income and consumption truly endogenous, non-IV estimates of the effect of income and other explanatory variables on remittances would merely amount to a conditional mean, with no possible behavioral interpretation, especially in terms of altruism versus selfishness[6].

Apart from the difficulty of finding suitable instruments, addressing the issue of the endogeneity of income and consumption is made more challenging by the fact that many migrants decide not to remit money at all. Typically, in the remittance literature the problem of non-remitters has been dealt with by using either Tobit (with censoring at zero) or Heckit models (Funkhouser, 1995; Brown, 1997; Aggarwal and Horowitz, 2002; Dustmann and Mestres, 2010). With these models, however, the researcher implicitly introduces extreme assumptions about the motives why migrants are non-remitters, which are unduly restrictive in the context of remittances, especially when transfer costs are non-negligible (Freund and Spatafora, 2008). The Tobit model assumes that the entire population of migrants is formed by potential remitters and that non-remitters are such because of their low propensity and finan-

[4]A notable exception is Hoddinott (1994). However, in that paper the problem is dealt with simply by showing that for the studied data set the Hausman exogeneity test fails to reject the null hypothesis that migrants' earnings are exogenous.

[5]A number of papers have analyzed remittance and saving decisions of migrants jointly: Merkle and Zimmermann (1992), Amuedo-Dorantes and Pozo (2006), Sinning (2011), Dustmann and Mestres (2010). However, these papers proceed by estimating separate reduced-form models for the different types of transfers and savings, among which the income of migrants can be allocated.

[6]For example, the neglect of a proper treatment of endogeneity might explain why estimates of income elasticity on microeconomic data are surprisingly smaller than one (Menjivar, Da Vanzo, Greenwell, and Valdez, 1998; Sinning, 2011; Yang, 2009), which would suggest that remittances are a basic good, while empirical models which proxy immigrants' income by per capita GDP or industrial hourly wage rate in the destination country provide elasticities greater than one (Lianos, 1997).

cial resources. In addition, if the censoring point is fixed at zero, transfer costs are implicitly assumed to be zero or so low as to not discourage any migrants from remitting money home. On the contrary, the Heckit model distinguishes between potential remitters and non-remitters in the migrant population. The former are assumed to always send some money home and therefore zero remittances should arise out of the unwillingness by a share of migrants to remit, regardless of their income and the level of transfer costs.

Quite obviously, however, both possibilities should be considered in the empirical analysis: remittances might be zero either because the migrant's earnings are not sufficient to afford to send money home or because the migrant, even if financially able to remit, chooses not to. In this view, we follow the double-hurdle approach, introduced by Cragg (1971), which, by featuring both censoring and selection mechanisms, allows for both cases. In particular, in order to set remittances in the broader context of work and consumption decisions, we develop a LIML estimator for the double-hurdle and Heckit models with endogenous regressors and analyze the remittance strategy of a sample of recent immigrants to Australia from 125 different countries.

By way of preview, our results indicate that migrants' income and consumption cannot be considered exogenous to migrants' decisions on remittances. Once endogeneity is taken into account, the elasticity of remittances to pre-transfer income is positive and statistically not different from one while individual consumption is negatively correlated to remittances. Moreover, we find some evidence of the fact that the appropriate selection mechanism is of the double-hurdle type, rather than a simple Heckman-style selection model.

The rest of the paper is organized as follows. In Section 2 we explain our empirical strategy, focussing on the implications of different censoring mechanisms for modelling remittances in the presence of transfer costs in Section 2.1, and introducing the double-hurdle and the Heckit maximum likelihood estimator with instrumental variables in Section 2.2. In Section 3, we provide a detailed description of our dataset, the variables and the models we estimate. Our results are presented in Section 4, while Section 5 concludes.

2 Econometric methods

The task of building an empirical model for migrants' remittances is a complex one. On the one hand, one has to take into account the fact that the decision of whether to remit money at all may be partly separated from the decision on the amount of the remittance and the latter decision may be influenced by the presence of transfer costs. On the other hand, the desired amount of remittances is likely to influence other choices on labour supply and consumption.

2.1 The selection mechanism

Suppose that each migrant i decides on remittances by solving the following standard maximisation program:

$$\max U_i\left(C_i, R_i\right) \tag{1}$$
$$\text{s.t.} \quad Y_i = C_i + R_i + \tau \cdot \mathbf{I}(R_i > 0)$$

where U_i is a (possibly individual-specific) continuous and differentiable utility function, C_i, R_i and Y_i indicate, respectively, consumption, remittances and income of migrant i, τ the fixed cost of transferring remittances and $\mathbf{I}(\cdot)$ the indicator function.

Assume that the marginal utility of consumption $U_i^C \equiv \frac{\partial U_i}{\partial C_i}$ is strictly positive for any i, while $U_i^R \equiv \frac{\partial U_i}{\partial R_i} \geq 0$, where the non-negativity inequality indicates that some individuals may gain no utility from remitting. In addition, if $\tau > 0$, by continuity we have that $\lim_{R_i \to 0} U\left(C_i, R_i\right) < U\left(C_i, 0\right)$: that is, when sending money home is costly, there is a minimum amount of remittances $\underline{R} > 0$ under which the additional utility that the migrant derives from remitting is lower than the utility-sacrifice of transfer costs. Therefore, the optimal amount of remittances may be zero for two reasons: (i) given the utility weight attached to remittances, the migrant's income is so low that the value of R solving the maximisation program (1) is $R_i^* < \underline{R}$; (ii) the migrant does not attach any utility to remittances ($U_i^R = 0$).

Obviously, the choice of the statistical tool for modelling remittance decisions depends on the assumptions made on the migrants' behaviour and on the characteristics of the sample of migrants under scrutiny. If perfect homogeneity across individuals as to the value of U_i^R is assumed, then the two choices on whether to remit at all and how much to remit are both governed by one and the same mechanism; as a consequence, zero-remittances can only be caused by a budget constraint (case (i)), and the appropriate model for remittances is the Tobit model with censoring point at $\underline{R}$.

On the contrary, allowing for the possibility that $U_i^R = 0$ for some individuals implies that two different mechanisms govern the decisions on whether to be a remitter and on the amount of money to send home. If we further assume that migrants who gain utility from remittances always remit, then zero-remittances all belong to case (ii) and the appropriate statistical approach would be the Heckman selection model.

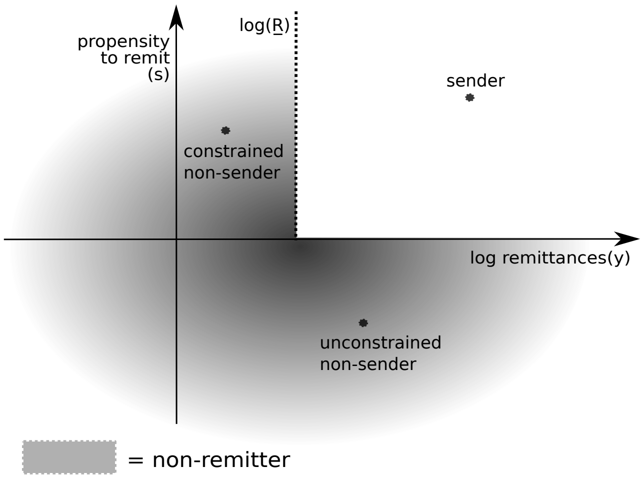

Figure 1: The selection mechanism

However, if cases (i) and (ii) are both considered possible, we would have, at least in principle, three categories of households: those who are not interested in sending money; those who wish to make remittances, and do; and finally, households who would like to remit some money, but are prevented from doing so by a budget constraint. Figure 1 illustrates the three categories graphically.

Assuming that the remittance variable is modelled in natural logarithms[7], a statistical procedure which takes into account the potential for "corner solutions" is called for and the empirical model could be written as

$$y_i^* = \ln R_i^* = \beta' X_i + \varepsilon_i \tag{2}$$

$$s_i^* = \alpha' Z_i + u_i \tag{3}$$

$$y_i = \begin{cases} \ln R_i^* & \text{if } R_i^* > \underline{R} \text{ and } s_i^* > 0 \\ NA & \text{otherwise} \end{cases} \tag{4}$$

where s_i^* is the unobservable propensity to remit. The appropriate statistical model for this situation is the double-hurdle model.

The double-hurdle model is a commonly employed technique for estimating models with double censoring. In the literature, the earliest reference is Cragg (1971), in which the two disturbances ε_i and u_i were assumed to be

[7]This is, in most cases, a very natural choice, since remittances typically exhibit a very skewed distribution, which makes the assumption of conditional normality difficult to justify. Moreover, the interpretation of coefficients as elasticities is usually more natural.

independent Gaussian variates (of course, $V(u_i) = 1$ is assumed for identification). Subsequent literature has extended the base model in a number of ways, the most notable of which is doubtlessly the so-called "dependent" model, in which the zero correlation constraint between u_i and ε_i is removed and the two disturbance terms are assumed distributed as

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N \left[0, \begin{pmatrix} \sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & 1 \end{pmatrix} \right] \tag{5}$$

This model has been used in countless applications, such as labour market studies (the classic reference here is Blundell, Ham, and Meghir (1987), but recent examples are Cardoso, Fontainha, and Monfardini (2010) or Zaiceva and Zimmermann (2007)) or, most notably, demand analysis for certain types of goods such as tobacco or alcohol, from Jones (1989) onwards. Estimation is typically carried out by maximum likelihood[8].

The use of a double-hurdle model in the empirical modelling of transfer decisions, although uncommon, is not new: a double-hurdle model is used in a similar context by Cox, Eser, and Jimenez (1998), via a somewhat *ad-hoc* two-step method instead of maximum likelihood. A recent paper by Sinning (2011) also uses a double-hurdle model, but only in its restricted independent version. Neither article tackles the problem of endogeneity of the explanatory variables.

The limiting case when $\underline{R} = 0$ is interesting to analyse: here, budget-constrained individuals are only those for whom $R_i^* < 0$ (that is, the utility function has no maximum for $R > 0$) and it is possible, at least in principle, to observe any amount of remittance, however small. Of course, this scenario is purely hypothetical, but serves the purpose of modelling a situation in which transfer costs are so small that they may be considered irrelevant in practice.

In this case, the dependent variable in equation (2) has no lower bound and the selection model is a Heckman-style one, in which no potential senders are prevented from sending by transfer costs. Such a selection mechanism has been used in the context of remittance modelling in, eg, Hoddinott (1994) and Aggarwal and Horowitz (2002). In this case, the selection mechanism should be modelled as

$$y_i = \begin{cases} \ln R_i^* & \text{if } s_i^* > 0 \\ NA & \text{otherwise} \end{cases} \tag{6}$$

The two selection mechanisms (6) and (4) may produce significantly different results or not. In the former case, we would take it as an indication

[8]A common reference is Jones (1992), who popularised a reasonably simple technique for maximising the log-likelihood numerically.

that transfer costs affect a significant share of potential remitters; in the latter, we would conclude that the evidence for a corner solution in the remitter's allocation problem is not visible in our dataset. In Section 4 we will employ both selection mechanisms and compare the results[9].

In the dataset at our disposal there is no indication on the transfer costs involved and $\underline{R}$ is unobservable. In fact, the transfer costs would be very difficult to quantify exactly, especially because $\underline{R}$ is probably individual-specific: for a start, some intermediaries take a percentage, while some others do not. Moreover, transfer costs are likely to vary by destination country. Finally, it is impossible to take non-monetary costs into account. However, while there is no reason to assume that $\underline{R}$ should be equal to τ, it sounds reasonable that they should be of the same order of magnitude. For these reasons, we estimate several models in which $\underline{R}$ was set *a priori* to different values within reasonable bounds, with a view to establishing the robustness of our results.

2.2 Selection models with endogenous regressors

In the context of remittance modelling, the above model must be modified to handle endogenous regressors in equation (2). The general problem of estimating simultaneous-equation systems with censoring has been comprehensively analyzed in Blundell and Smith (1994). However, our problem is considerably simpler: as an explicit reduced form exists[10], estimation can be carried out via a control function approach, which lends itself to a LIML-like strategy quite naturally.

Assume you have

$$\left(\begin{array}{c} s_i^* \\ Y_i \end{array} \right) \Bigg| Z_i \sim N \left[\left(\begin{array}{c} \alpha' Z_i \\ \Pi Z_i \end{array} \right), \left(\begin{array}{cc} 1 & \lambda' \\ \lambda & \Sigma \end{array} \right) \right]$$

where s_i^* is the latent propensity to remit, Y_i is the vector of endogenous regressors (in our context, pre-transfer immigrants' income and consumption) and $Z_i' = (X_{1i}', X_{2i}')$ is the vector of exogenous variables: X_{1i} are the exogenous variables appearing as regressors in the remittance equation; X_{2i}' are the instruments.

[9]Given the structure of the immigrants' sample we study, we do not report Tobit estimates: 70 percent of immigrants included in our sample do not remit (see Section 3.2) and we can confidently exclude that this is only due to constraints in the immigrants' family budget. Anyway, we tested this conjecture by running a Vuong test comparing double-hurdle and Tobit models (see Section 2.3) which clearly favours the former type of censoring (results are available upon request).

[10]See Blundell and Smith (1994), footnote 1.

The reduced form for (s_i^*, Y_i') is

$$\begin{aligned} s_i^* &= \alpha' Z_i + u_i & (7) \\ Y_i &= \Pi Z_i + \eta_i = \Pi_1 X_{1i} + \Pi_2 X_{2i} + \eta_i & (8) \end{aligned}$$

It is worth noting that we are not estimating a structural form for the selection equation (7), but rather its unrestricted reduced form. In our opinion, the theoretical arguments behind the over-identifying restrictions we should use for IV estimation would be much less convincing if applied to the selection mechanism: equation (7) models the *ex ante* psychological propensity of an individual to send money abroad and imposing over-identifying restrictions derived from economic theory would be rather adventurous.

As u_i has unit variance and $u_i|\eta_i \sim N(\lambda'\Sigma^{-1}\eta_i, \omega^2)$, with $\omega^2 \equiv 1-\lambda'\Sigma^{-1}\lambda$, we can write

$$s_i^* = \alpha' Z_i + \lambda'\Sigma^{-1}\eta_i + w_i$$

where $w_i \equiv u_i - E(u_i|\eta_i)$ and $V(w_i) = \omega^2$.

We then have

$$y_i^* = \gamma' Y_i + \beta' X_{1i} + \varepsilon_i \tag{9}$$

which is our structural relation. Note that ε_i may be correlated to η_i (causing endogeneity) and/or to u_i. Assume now that the correlation between ε_i and η_i can be modelled as

$$\varepsilon_i = \theta'\eta_i + v_i \tag{10}$$

which, for example, would be the case under joint normality. This way, v_i is linearly independent of η_i, although it may be correlated with u_i. As a consequence, we have

$$y_i^* = \gamma' Y_i + \beta' X_{1i} + \theta'\eta_i + v_i.$$

With the joint normality assumption, one may write

$$y_i^*|(Z_i, \eta_i) \sim N(\gamma' Y_i + \beta' X_{1i} + \theta'\eta_i, \sigma^2)$$

Conditionally on η_i (that is, treating Y_i as given), the censoring mechanism works exactly as in an ordinary selection model (see Subsection 2.1): define a binary variable d_i as

$$d_i = \begin{cases} \mathbf{I}\left[s_i^* > 0\right] & \text{for the Heckit model} \\ \mathbf{I}\left[(y_i^* > y_{\min}) \wedge (s_i^* > 0)\right] & \text{for the double} - \text{hurdle model} \end{cases}$$

where $\mathbf{I}(A) = 1$ if A is true and 0 if A is false. The observed amount of (log) remittances is $y_i = d_i y_i^*$, which reads as: an individual will send a positive amount overseas only if he/she intends to do so in the first place ($s_i^* > 0$) and, given the level of transfer costs, has enough money to make it worthwhile ($y_i^* > y_{\min}$). This variable may be rewritten as

$$d_i = \mathbf{I}\left[(w_i > -(\alpha' Z_i + \lambda' \Sigma^{-1} \eta_i)\right] \times \\ \times \mathbf{I}\left[(v_i > -(\gamma' Y_i + \beta' X_{1i} + \theta' \eta_i + y_{\min})\right]$$

Hence,

$$P_i^r = P(d_i = 1) = \\ = \Phi_2 \left(\frac{\alpha' Z_i + \lambda' \Sigma^{-1} \eta_i}{\omega}, \frac{\gamma' Y_i + \beta' X_{1i} + \theta' \eta_i + y_{\min}}{\sigma}, \rho \right)$$

where $\Phi_2(\cdot)$ is the cumulative distribution function of the double normal, $\sigma^2 = V(v_i)$ and ρ is the correlation between w_i and v_i.

From a practical viewpoint, if $y_{\min}$ is known, it is convenient to redefine the dependent variable for the double hurdle model as $y_i = d_i(y_i^* - y_{\min})$ and use 0 as the boundary for the corner solution. Of course, the estimated coefficient for the constant in equation (2) must be adjusted accordingly, but apart from that no other modification to an ordinary double-hurdle model is needed. This is what we do in Section 4.

Hence, the log-likelihood (conditional on η_i) for individual i can be written as

$$\ell_i^c = (1 - d_i) \ln(1 - P_i^r) + \\ + d_i \ln \left[P(s_i^* > 0 | v_i) \times \frac{1}{\sigma} \varphi \left(\frac{\gamma' Y_i + \beta' X_{1i} + \theta' \eta_i}{\sigma} \right) \right]$$

The full log-likelihood can be recovered by adding to ℓ_i^c the marginal log-likelihood for η_i, which is

$$\ell_i^m = \text{const} - 1/2 \left[\ln |\Sigma| + (Y_i - \Pi Z_i)' \Sigma^{-1} (Y_i - \Pi Z_i) \right].$$

2.3 Hypothesis testing

Once estimation is carried out, it becomes possible to test for several hypotheses: the hypothesis of exogeneity of Y_i, which is particularly interesting for the interpretation of the results, can be easily checked via a Wald test on θ, leading to an equivalent of the Hausman test in an ordinary IV linear model, and poses no particular problems.

Another test of interest is the one for the over-identification restrictions implied by equation (9). This has no obvious economic interpretation, but is nevertheless important to judge the appropriateness of the choice of instruments. To see how the test is carried out, substitute (8) and (10) into (9) to get the restricted reduced form

$$y_i^* = (\gamma'\Pi_1 + \beta)X_{1i} + \gamma'\Pi_2 X_{2i} + (\gamma + \theta)'\eta_i + v_i = \mu_1' X_{1i} + \mu_2' X_{2i} + \mu_3'\eta_i + v_i;$$

once the structural parameters are estimated by LIML, a test for the over-identifying restriction may be performed by computing the score matrix of the unrestricted model for the corresponding values of μ_1, μ_2 and μ_3 and perform a score (conditional moment) test via an OPG artificial regression (see Davidson and MacKinnon (1984)).

Testing for the nature of the selection mechanism is a more complex matter: a comparison between the two sample selection models we are considering here would shed light on the nature of censoring. If a Heckit-type censoring occurred, then each individual who chooses to remit would generate a non-zero figure for the actual remittances sent; put differently, zero remittances would indicate unambiguously that the individual has no intention to remit money home whatever his/her income, and no individuals are constrained. By contrast, if the censoring mechanism is double-hurdle, zero remittances could result from potential, but income-constrained remitters.

Since the Heckit-type censoring arises as a limiting case of the general censoring scheme when $y_{\min} \to -\infty$, this does not lead to a restriction testable via an ordinary likelihood-based test procedure. These difficulties can be circumvented by using Vuong's test (see Vuong (1989)): this test is used for comparing non-nested models in terms of the difference in their respective Kullback-Leibler distance from the (unknown) "true" model. Define the log-likelihood-ratio for observation i as

$$LR_i = \ell_i^A - \ell_i^B$$

where ℓ_i^A (ℓ_i^B) is the i-th contribution to log-likelihood for model A (B). Under the null hypothesis that the two models offer an equivalent representation of the data, the statistic

$$VT = \frac{1}{\sqrt{n}\sqrt{V(LR_i)}} \sum_{i=1}^{n} LR_i$$

is asymptotically distributed as a standard normal random variable. Large positive (negative) values are taken as evidence in favor of model A (B).

Finally, the possibility of weak instruments should also be considered. The theory of IV regression with weak instruments has attracted considerable attention in the past decade and a sizeable body of literature has developed, from Staiger and Stock (1997) onwards (see for example Andrews and Stock (2005)). However, most contributions focus on the linear case and there appears to be no equivalent of statistics like the Cragg-Donald statistic[11] for non-linear models such as ours. As a consequence, we will report a few statistics which we consider informative, despite the fact that they cannot be considered test statistics in a technical sense:

1. F-tests for the instruments in the reduced forms of the endogenous variables, computed via OLS;

2. for each model, a Wald test for the hypothesis $\Pi_2 = 0$ in eq. (8).

3 Data and variables

3.1 The Longitudinal Survey of Immigrants to Australia

The dataset we use is the Longitudinal Survey of Immigrants to Australia (LSIA), a longitudinal study of recently arrived visaed immigrants undertaken by the Commonwealth Department of Immigration and Multicultural and Indigenous Affairs.

We consider the first cohort of the LSIA (LSIA1), that was selected from visaed immigrants aged 15 years and over, who arrived in Australia between September 1993 and August 1995[12]. The questionnaires cover various topics: the migrant's family in Australia, the immigration process, the initial settlement, financial assets and transfers (remittances), working status, income, consumption expenditure, education and English knowledge, health, citizenship and return visits to the former country. All this information gives an incomparable socio-economic picture of immigrants, that is essential to understand their remittance behaviour.

Individuals were interviewed three times between six months and three years from their arrival in Australia. In the first two waves remittances are designed as a discrete ordered variable, while in the third wave they are continuous. This prevented us from exploiting the panel dimension of the dataset since the remittance variable is incomparable across the different waves; as a consequence, we only employ data from the third wave.

[11] See Cragg and Donald (1993).

[12] The sampling unit is the Primary Applicant (PA), the person upon whom the approval to immigrate was based. The population for the survey consisted of about 75,000 PAs and was stratified by the major visa groups and by individual countries of birth.

The initial sample included 5,192 individuals, but due to sample attrition immigrants interviewed in the third wave fell to 3,752 (2,160 men and 1,592 women). In principle, such a great deal of attrition in a panel dataset on immigrants might be endogenous due to return strategies closely linked to income levels and remittance behaviour. However, since we consider recently arrived immigrants who have spent less than 3 years in Australia, it is reasonable to think that the time interval is still too short to put return strategies into effect. Therefore sample attrition is assumed as exogenous in this context.

The interviewees came from 125 countries. The most represented region is Asia, followed by Europe, Africa and the Middle East, which taken together represent the origin for almost 90 percent of immigrants in the sample. Table 3 shows that five out of the ten main countries of origin are located in South-East Asia, suggesting that geographic distance plays an important role in migration choices. However, the largest number of immigrants come from the United Kingdom (8.40 percent of the sample), which suggests that cultural affinities, common language and past colonial relationships also affect the locational choice of migrants.

3.2 Some descriptive statistics on remittances

The number of immigrants who stated they had remitted money is 1,154 (31 percent). Out of these, 10 observations may well be regarded as misreporting cases, since individuals declared an amount of remittances lower than 1 AUS $, and were therefore excluded. In the remaining sample (1,144 immigrants), the minimum amount reported is around 25 AUS $. This is consistent with the existence of non-negligible transaction costs in making remittances[13]: if, given the immigrant's disposable income, the optimal amount of remittances is lower than a certain threshold, the utility-sacrifice of transfer costs exceeds the utility that immigrants derive from sending money home and therefore we should not observe any remittance. The average amount of money remitted since the previous interview is around 1,250 AUS $, although the median is

[13] It is interesting to note that the fees immigrants in Australia pay to send money overseas are in a narrow range: for example, the Commonwealth Bank applies a 22 AUS $ fee for each transfer (`http://www.commbank.com.au/help/faq/netbank/netbank-charges.aspx`), Westpac applies a 20 AUS $ fee for Internet transfers, while telegraphic transfers from a branch cost between 30 AUS $ (`http://www.westpac.com.au/personal-banking/services/overseas-services/sending-receiving-money/`). With ANZ the fees vary between 24 and 32 AUS $ according to the different transfer methods (`http://www.anz.com/personal/travel-foreign-exchange/international-payments/international-money-transfers/`). Western Union fees vary according to the country of destination and the amount of money transferred.

about 500 AUS $.

By analysing the number of remitters by country of birth, several interesting facts emerge: the Iraqis are the most likely to send money back home (62.5 percent of Iraqi immigrants are remitters) followed by the Afghans (58.7 percent) and the Filipinos (58 percent); the share of remitters is much lower for immigrants from high-income countries (as an example, the figure is 10.7 percent for the USA and 14.3 percent for the UK). Conversely, the average amount remitted by Japanese immigrants (11,100 AUS $) is the largest in the sample, and is also much larger than the amount that immigrants coming from similar countries (in terms of per capita GDP) like the UK, Germany or Italy send back home. By contrast, among poor countries Cambodians' remittances (1,061 AUS $) are lower than the average. Such figures clearly suggest that immigrants' remittance behaviour may be consistent with the exchange motive (Cox, Eser, and Jimenez, 1998): immigrants from richer (poorer) countries are less (more) likely to send remittances, but if they do, they send larger (smaller) amounts (see Table 3 and Figure 2).

3.3 The empirical model

3.3.1 Remittance equation variables

The main remittance equation in the model we estimate[14] is:

$$y_i = \ln R_i = \alpha + \beta_y Y_i + \beta_c C_i + \delta' X_{1i} + \gamma' Z_{j(i)} + \varepsilon_i \tag{11}$$

where R_i is the amount of money sent home every year by the immigrant household[15], Y_i is the yearly pre-transfer income of the immigrants' household in Australia and C_i the yearly household's non-durable consumption expenditure (all variables are in logarithm). As immigrants' income and consumption are recorded by intervals in the survey, we take the mid-points of the intervals[16]. Evidently, the mid-point imputation introduces an unavoidable measurement-error bias. However, with regard to consumption the problem is mitigated by the fact that it is calculated as the sum of various non-durable consumption items singled out in the questionnaire: food,

[14] In order to control for outliers, we computed the Mahalanobis distance from the centroid on the following variables: income, consumption, gender, age, origin country per capita GDP, time since arrival and household members. We then considered outliers (and excluded from the subsequent analysis) the top 0.5 percent, which amounted to dropping 10 observations.

[15] In the questionnaire, immigrants were originally asked about the amount of money sent back home from the previous interview. However, since data concerning income and consumption are reported on a yearly basis, we also transformed remittances accordingly.

[16] For the top class, we double the lower bound.

transport, clothes, health, expenditures for the children and for the house (gas, electricity etc.)[17]. Hence, the distribution of the resulting aggregate variable is rather granular.

X_1 is a vector of exogenous immigrant characteristics that theory and previous empirical literature indicate as possible explanatory variables of remittance behaviour: (i) the immigrant gender (*MALE*); (ii) the age of the immigrant and its square (*AGE*, *AGE2*); (iii) the time (in years) elapsing from arrival in Australia (*TIME*)[18]; (iv) a citizenship indicator that takes the value of 1 if the immigrant has obtained, applied to obtain or has the intention to apply for Australian citizenship (*CITIZENSHIP*); (v) the formal qualification of the immigrant (*EDUCATION*) proxied by four dummies corresponding to PhD/MA degree, BA degree or diploma, 10/12 years of schooling, and 9 or less years of schooling (the base category is the immigrant with a PhD or an MA degree); (vi) a dummy for the presence of close relatives in the country of origin[19] (*RELATIVES*).

The set of explanatory variables $Z_{j(i)}$ includes variables relative to the home country j of individual i, which aim to capture parental family characteristics for which the LSIA survey does not keep records. In the baseline specification, we consider the log of the mean per capita GDP over the period 1992-2000 (*GDP_PC*) as a proxy for the economic conditions of relatives at home[20]. Furthermore, the log of the distance between Australia and the country of origin (*DISTANCE*) is considered because of its influence on the relations with the home country. Being far away raises the cost of visiting home and also reduces the frequency of contacts at a distance, due to different time zones, thus weakening the strength of altruistic feelings. At the same time both costs associated to migrating to the host country and to transferring money back to the origin family increase with the distance from the home country.

Ideally, one may want to control for unobserved characteristics of the country of origin by using a complete set of country dummies. In our case, however, this is not possible for two reasons. The most important reason is that, after adjusting for missing values, we have 102 different origin coun-

[17]Therefore, the "consumption" variable does not exhaust all consumption possibilities and does not include any durable item.

[18]The relationship between time since immigration and the remittance behaviour might be non-monotonic, calling for a quadratic specification (Brown, 1997). Since our sample includes only recently arrived immigrants that have all spent less than three years in Australia we prefer not to include $TIME^2$ in our basic model. However, the augmented specification has been tested and both variables proved insignificant.

[19]It refers to partner/spouse, children, parents and siblings.

[20]GDP data are from the World Development Indicators database.

tries, but most of these countries are represented by a very small number of households, often with no remitting households at all. Clearly, the inclusion of dummy variables for such countries would cause unsurmountable identification problems and would make little sense anyway. Secondly, even if we had a large enough sample size for each country, as illustrated above we already have a few country-specific variables in our baseline specification and including a complete set of country dummies would obviously force us to drop those variables because of collinearity problems.

Hence, we decided to include a limited set of country-specific dummies for those countries from which we had at least 15 households and 5 remitting households[21]. In practice, this is equivalent to grouping all the remaining countries into a residual "other countries"category. This choice led us to include 30 country dummies; it is worth noting that since these 30 countries are the origin country for 1,635 households out of 2,170, our limited set of country dummies actually covers 75.3 percent of our sample.

3.3.2 Instrumental variables

The set of instruments for pre-transfer immigrants' income Y_i and consumption C_i includes X_{1i}, $Z_{j(i)}$ and X_{2i}. The first two subsets contain the regressors illustrated above for the remittance equation. X_{2i}, instead, is a vector of variables that ideally should exert a strong influence on income and consumption and be independent of the amount of money remitted and more in general of any "remittance-oriented" migration strategy[22].

Using the survey data, we build a set of seven instruments. First, we include two dummy variables which are related to events affecting the immigrants' conditions after their arrival in Australia. The former, *MOVE*, takes value 1 if the intended State of residence upon the immigrant's arrival in Australia is different from the State of residence at the time of the interview. The second dummy variable, *ILLNESS*, refers to the immigrant's health status and takes value 1 if an immigrant who was healthy at the time of immigration reports to have recently been affected by long-term health problems which restrict him/her in physical activities or work[23].

[21]Other thresholds were tried, with no appreciable differences.

[22]As explained in Section 2.2, we estimate the selection equation in its reduced form. This means that we assume the variables in X_{2i} to be uncorrelated with the error term of the remittance equation (8), but they may well exert a direct impact on the decision on whether to remit.

[23]The long-term diseases listed in the questionnaire include arthritis, hearing problems, deafness, blindness, nerves or stress problems, heart disorder, loss of limb or any other part of the body, diabetes, asthma and any permanent loss of memory or loss of mental ability.

Since, by construction, the decision to move to a different location and recent health problems were not part of the information set of immigrants up at the the time of their arrival in Australia, they can doubtlessly be considered exogenous with respect to the motivations behind migration and remittance strategies. In addition, there is little doubt that *MOVE* and *ILLNESS* influence the immigrants' earning capacity and propensity to consume, even if the direction in which they exert such influence is in principle ambiguous. For example, when the decision to move the place of residence is driven by the job market conditions, by the possibility of getting better jobs in the new place or by the fact that the area where immigrants settled at first does not offer adequate employment opportunities, *MOVE* should affect income positively. However, if there are other unforeseen life events that drive the change of residence, it is reasonable to expect that it may hurt the immigrant's earning capacity at least in the short run. Similarly, *MOVE* may be associated with lower or higher consumption according to the precautionary attitude of the immigrant and the level of the unexpected expenses due to the change of residence. With regard to *ILLNESS*, the expected impact on earned income is clearly negative. However, poor health conditions should increase expenses on some consumption items linked to the treatment of the disease and reduce expenses on other items linked to the disabilities created by the disease, with ambiguous total effects on the budget share allocated to consumption.

A third instrument included in X_{2i} refers to the knowledge of English (*ENG*) and takes the value of 1 if the first spoken language of the immigrant is English. This variable identifies those immigrants whose mother tongue is English and can be safely considered exogenous with respect to the decisions concerning the amount of money to remit[24]. We expect that a high level of English proficiency should increase the chances to get more remunerative jobs, facilitate the consumption of local items and, through such positive effects on income and consumption, affect remittances.

The last four instruments refer to the composition of the immigrant household in Australia: a dummy for the presence of children in the immigrant household (*CHILD*), a dummy for the presence of the partner in the immigrant household (*SPOUSE*), the number of members in the immigrant household (*FAMILY_N*) and its square ($FAMILY\_N^2$). We expect the composition of the family in Australia to contribute to determining earned income and consumption patterns, and via such variables, remittances. In partic-

[24] A variable on the overall level of English proficiency was also available, but we decided not to use it as an instrument because it is conceivable that individuals who choose to migrate in order to remit may improve their knowledge of the language by taking courses prior to the move.

ular, we expect income to increase with the number of household members and with the presence of a partner and to decrease in the presence of children. Consumption, of course, should increase with the number of household members. The square term is included to account for any non-linearity in the relationship between income, consumption and household size which is expected to enter both income and consumption with a negative sign.

Admittedly, some of the latter instruments might be suspected of not being fully exogenous with respect to remittances. For example, the choice of bringing the whole family (partner and children) along could be linked to the plan to settle in Australia on a long-term basis and this could have a noticeable direct impact on the remittance strategy. However, since we can rely on the dummies *MOVE*, *ILLNESS* and *ENG* that satisfy the requirements to be valid and relevant instruments, we prefer to keep a sufficiently large degree of over-identification and rely on a test of over-identifying restrictions to assess the coherence of the whole set of instruments.

4 Results

4.1 The selection mechanism

in Tables 1 and 2 we report the results from our preferred specification for the remittance model with the two different selection mechanisms.

As illustrated in Section 2.2, the double-hurdle model accommodates both the case of immigrants who do not remit because they are unwilling to send money back home and the case of immigrants who are financially constrained, while the Heckit model considers all non-remitters as people who choose not to remit whatever their income. In particular, when using a Heckman-style selection mechanism, we model a situation in which any amount of remittance, however small, might be observed ($\underline{R} = 0$) and no immigrants would be prevented from remitting by the presence of transfer costs and budget constraints. In a double-hurdle setting, instead, there is a minimum amount of remittances $\underline{R} > 0$ below which the costs migrants need to cover to send money back home are not offset by the additional utility they derive from the transfer.

In the baseline specification $\underline{R}$ is set equal to 20 AUS \$, which we consider a plausible value given that the minimum transfer in our sample is around 25 AUS \$ and transaction fees applied by the major banks in Australia are never lower than 20 AUS \$[25]. To test the robustness of our results, however,

[25]See footnote 13.

Table 1: IV dependent double-hurdle model

Main equation					
	coeff.	std.err.	z-stat	p-value	non-IV estimates
const	14.407	4.246	3.393	0.001 ***	-0.578 ***
MALE	0.230	0.111	2.079	0.038 **	0.195 *
AGE	0.367	0.458	0.802	0.423	-0.410
AGE2	-0.058	0.058	-0.998	0.318	0.044
TIME	0.222	0.411	0.540	0.589	0.036
CITIZENSHIP	-0.439	0.318	-1.379	0.168	-0.384
RELATIVES	0.526	0.291	1.806	0.071 *	0.336
EDUCATION_2	-0.319	0.150	-2.131	0.033 **	-0.236 *
EDUCATION_3	-0.530	0.197	-2.687	0.007 ***	-0.468 ***
EDUCATION_4	-0.753	0.231	-3.261	0.001 ***	-0.578 **
GDP_PC	0.203	0.122	1.666	0.096 *	0.264 **
DISTANCE	-0.443	0.351	-1.262	0.207	-0.089
INCOME	1.057	0.371	2.850	0.004 ***	0.171 *
CONSUMPTION	-2.151	0.618	-3.480	0.001 ***	0.439 ***

Selection equation					
	coeff.	std.err.	z-stat	p-value	non-IV estimates
const	1.561	1.301	1.200	0.230	1.609
MALE	0.187	0.067	2.802	0.005 ***	0.188 ***
AGE	0.352	0.239	1.471	0.141	0.410 *
AGE2	-0.067	0.029	-2.292	0.022 **	-0.073 ***
TIME	0.094	0.277	0.339	0.735	0.069
CITIZENSHIP	0.245	0.137	1.787	0.074 *	0.247 *
RELATIVES	0.322	0.161	2.007	0.045 **	0.326 **
EDUCATION_2	-0.090	0.085	-1.061	0.289	-0.083
EDUCATION_3	0.003	0.109	0.026	0.979	0.008
EDUCATION_4	0.214	0.144	1.492	0.136	0.223
GDP_PC	-0.281	0.057	-4.977	0.000 ***	-0.275 ***
DISTANCE	-0.441	0.241	-1.833	0.067 *	-0.466 *
ENG	-0.176	0.101	-1.750	0.080 *	-0.175 *
MOVE	-0.018	0.077	-0.232	0.817	0.031
FAMILY_N	-0.003	0.104	-0.027	0.979	-0.031
FAMILY_N^2	-0.007	0.012	-0.545	0.586	-0.005
CHILD	-0.168	0.095	-1.770	0.077 *	-0.167 *
SPOUSE	0.258	0.093	2.773	0.006 ***	0.217 **
ILLNESS	-0.134	0.151	-0.886	0.376	-0.150
σ	1.106	0.034	32.920	0.000 ***	1.179 ***
ρ	0.094	0.192	0.489	0.625	-0.340

Note: QMLE standard errors (see White (1982)). Country-specific fixed effects included. The log-likelihood is equal to -4214.66 in the IV estimation and to -2123.17 in the simple double-hurdle model. The total number of cases is 2170 with 1518 censored observations. The χ^2_2 Wald test statistic for exogeneity for income and consumption is 22.807 (p-value: 1.12e-5). The χ^2_5 overidentifying restriction test statistic is 9.859 (p-value: 0.079). First-stage F-tests: 59.841 (income), 71.391 (consumption). Wald test for $\Pi_2 = 0$: 682.272 (p-value: 1.56E-136).

Table 2: IV Heckit model

Main equation

	coeff.	std.err.	z-stat	p-value	non-IV estimates
const	14.283	3.772	3.786	0.000 ***	-0.112
MALE	0.225	0.130	1.734	0.083 *	0.196 *
AGE	0.360	0.447	0.806	0.421	-0.370
AGE2	-0.056	0.058	-0.967	0.334	0.038
TIME	0.221	0.493	0.448	0.654	0.045
CITIZENSHIP	-0.426	0.286	-1.490	0.136	-0.362
RELATIVES	0.505	0.328	1.542	0.123	0.340
EDUCATION_2	-0.313	0.154	-2.027	0.043 **	-0.235 *
EDUCATION_3	-0.521	0.205	-2.547	0.011 **	-0.462 ***
EDUCATION_4	-0.732	0.260	-2.818	0.005 ***	-0.553 ***
GDP_PC	0.194	0.146	1.332	0.183	0.240 *
DISTANCE	-0.436	0.443	-0.985	0.325	-0.107
INCOME	1.027	0.374	2.745	0.006 ***	0.154
CONSUMPTION	-2.100	0.599	-3.509	0.001 ***	0.412 **

Selection equation

	coeff.	std.err.	z-stat	p-value	non-IV estimates
const	1.549	1.396	1.110	0.267	1.584
MALE	0.189	0.069	2.726	0.006 ***	0.191 ***
AGE	0.352	0.249	1.415	0.157	0.395
AGE2	-0.067	0.029	-2.282	0.023 **	-0.071 **
TIME	0.093	0.298	0.311	0.756	0.071
CITIZENSHIP	0.244	0.170	1.438	0.150	0.246 *
RELATIVES	0.330	0.170	1.938	0.053 *	0.330 **
EDUCATION_2	-0.092	0.091	-1.020	0.308	-0.087
EDUCATION_3	-0.001	0.123	-0.011	0.991	0.002
EDUCATION_4	0.205	0.152	1.346	0.178	0.211
GDP_PC	-0.280	0.063	-4.470	0.000 ***	-0.273 ***
DISTANCE	-0.443	0.283	-1.563	0.118	-0.465 **
ENG	-0.177	0.106	-1.664	0.096 *	-0.175 *
MOVE	-0.019	0.078	-0.247	0.805	0.024
FAMILY_N	-0.001	0.111	-0.010	0.992	-0.024
FAMILY_N^2	-0.007	0.013	-0.510	0.610	-0.005
CHILD	-0.171	0.098	-1.743	0.081 *	-0.168 *
SPOUSE	0.256	0.099	2.573	0.010 **	0.228 *
ILLNESS	-0.136	0.158	-0.856	0.392	-0.149
σ	1.091	0.053	20.710	0.000 ***	1.141 ***
ρ	0.108	0.494	0.219	0.827	-0.273

Note: QMLE standard errors (see White (1982)). Country-specific fixed effects included. The log-likelihood is equal to -4217.56 in the IV estimation and to -2162.162 in the simple Heckit model. The total number of cases is 2170 with 1518 censored observations. The χ^2_2 Wald test statistic for exogeneity for income and consumption is 22.060 (p-value: 1.62e-5). The χ^2_5 overidentifying restriction test statistic is 8.719 (p-value: 0.121). First-stage F-tests: 59.841 (income), 71.391 (consumption). Wald test for $\Pi_2 = 0$: 719.824 (p-value: 1.51E-144).

Table 7 shows alternative specifications for the double-hurdle model where $\underline{R}$ takes different values between 10 and 30 AUS \$.

The nature of the selection mechanism is tested by using Vuong's test, whose statistic definitely favors the double-hurdle versus the Heckit model regardless of the value of $\underline{R}$[26]. Hence, from a statistical point of view there is an indication that a corner solution exists and transaction costs seem to play a crucial role in the remittance strategy of a significant share of immigrants.

However, the two rival models provide very similar descriptions as to what the determinants of the selection mechanism are: the estimated coefficients are extremely close in magnitude, with only slight differences in the z-statistics, which are generally larger for the double-hurdle. Hence, whether or not transaction costs are considered does not affect the determinants of the *a priori* probability of a household being a remitter. Consequently, our discussion of the estimation results focusses on the double-hurdle model but is also valid for the Heckit model.

4.2 Income and consumption

First-stage regressions for income and consumption are reported in the Appendix in Tables 5 and 6. On the whole, instrumental variables have a very significant impact on income and consumption and estimated coefficients have the expected signs[27]. In addition, the Wald test strongly rejects the hypothesis of exogeneity for income and consumption, thus suggesting the importance of accounting for reverse causality and simultaneity between such variables and remittances, while the over-identification test supports the validity of the chosen instruments. F-tests and the Wald test finally allow us to discard the issue of weak instruments (with the caveats we mentioned in Section 2.3).

Moving on to the main equation of the remittance model, in the last column of tables 1 and 2, we report results from the non-IV estimates. The comparison with coefficients from IV estimates provides clear confirmation of the need to address the endogeneity of immigrants' income and consumption. First, when *INCOME* and *CONSUMPTION* are not instrumented, the two coefficients never appear to be simultaneously significant. Moreover, the income elasticity of remittances is suspiciously much lower than 1[28].

[26]In the baseline specification, Vuong's test statistic is equal to 3.88, with a p-value of 0.99.

[27]Only *MOVE* and *ILLNESS* are not significant in the consumption equation; this may be because of heterogeneous effects across individuals by these variables on consumption.

[28]Interestingly, a very similar result was found by Sinning (2011). Using a non-IV double-hurdle model he reported that the elasticity of remittances to income ranges from

On the other hand, the elasticity to consumption, which is significant at the 5 percent level in the non-IV Heckit model, displays a rather puzzling positive sign, the natural prediction being that, given income, remittances diminish as the level of consumption expenditure of the immigrants' household increases. However, this result is rather easy to interpret if we view this coefficient not as a behavioral parameter, but rather as a mere statistical predictor (which is exactly what a non-IV estimate yields): consumption may be a better predictor of remittances than income simply because it is closer to the household's perceived "permanent income".

When controlling for endogeneity, the amount of money transferred to the family of origin depends positively on migrants' pre-transfer income and negatively on consumption expenditure. Both coefficients are statistically significant at the 1 percent confidence level but the coefficient for consumption is almost twice as large (in modulus) than that for income. In particular, the coefficient on income, measuring the elasticity of remittances to the immigrant's pre-transfer income, is never significantly different from 1. It is worth noting that pure altruistic models of remittance behavior suggest that remittances are a superior good, while in a pure exchange motivation scenario remittances can be a normal good; consequently, our findings are at odds with predictions of pure altruistic models of remittances, while they are reasonably consistent with exchange-driven models where optimal remittances are broadly proportional to income (see Appendix A).

By contrast, the elasticity of remittances to consumption is approximately equal to -2. In addition, a joint test for $\beta_y = -\beta_c$ rejects the null with a p-value of 0.003 for the Heckit model and a p-value of 0.006 for the double-hurdle model.

4.3 Other determinants of remittances

The effects of immigrants' gender and age on remittances are broadly in line with the majority of previous studies. Male immigrants tend to transfer back home significantly greater amounts of money than females and are also more likely to remit[29], while the age of immigrants is not significantly correlated with the intensity of transfers to the family of origin (Funkhouser, 1995; Osili, 2007; Dustmann and Mestres, 2010). However, older immigrants are more

0.09 to 0.53.

[29]The greater propensity to remit of males is consistent with the findings of Funkhouser (1995), Aggarwal and Horowitz (2002), Amuedo-Dorantes and Pozo (2006) and Dustmann and Mestres (2010). By contrast, Lucas and Stark (1985), Osaki (2003) and Naufal (2008) found that it is females who remit more.

likely to remit, albeit at a decreasing pace[30].

Time elapsing from the arrival of immigrants to Australia has no significant effect on remittance decisions (Brown, 1997; Amuedo-Dorantes and Pozo, 2006; Dustmann and Mestres, 2010) consistently with the altruistic motivation to remit. However, it should be noted that our sample includes only recently arrived immigrants.

The citizenship status of immigrants can be relevant to explaining remittance decisions, even if its effects are ambiguous. For example, to the extent that application for citizenship indicates the willingness of immigrants to reside permanently, or for a long time, in the host country we should expect a negative relationship with transfers to the home country. However, if citizenship status comes with more stable and protected occupations in the host country and with access to wider forms of social protection, we might observe that 'citizen' immigrants are more inclined to remit. Our findings show that immigrants who did apply for Australian citizenship are more likely to remit but, on average, they seem to remit smaller sums (even if the coefficient is not significant). This suggests that by being well integrated in the host society, immigrants who have obtained Australian citizenship can afford to send money back home with higher frequency than non-citizen immigrants[31].

As for schooling, we find that the probability of immigrants remitting money is not significantly correlated with their level of educational attainment. However, in the main equation the coefficients on schooling dummies are significantly negative and increasing in modulus, suggesting that, income being equal, the more educated the immigrants, the greater is the amount of money they transfer home. The positive correlation between education and remittances has already been documented in the literature by, for example, Lucas and Stark (1985), Hoddinott (1994), Funkhouser (1995) and Bollard, McKenzie, Morten, and Rapoport (2011), and has been taken as an important piece of evidence in favor of the exchange (repayment) motivations to remit.

Unsurprisingly, as any remittance theory predicts, we find that the presence of close relatives in the country of origin affects both the decisions on

[30]A joint test of significance of AGE and $AGE2$ coefficients shows that the relation with the probability to remit is non-monotonic. An inverted U relationship between immigrants' age and remittances is also documented in Hoddinott (1994) and Clark and Drinkwater (2007).

[31]In a similar vein, the effects of legal status of the immigrant on remittances has been analyzed with results that are not consistent across the literature. For example, Konica and Filer (2009) report that legal Albanian emigrants remit more than their illegal counterparts; by contrast, Amuedo-Dorantes and Pozo (2006) and Markova and Reilly (2007) found an opposite correlation for the case, respectively, of documented Mexican and Bulgarian emigrants.

whether to remit and the amount actually remitted in a positive way[32].

As stated before, unfortunately the LSIA survey lacks information on the immigrants' family of origin. In order to circumvent this deficiency to some degree, we exploit the cross-sectional nature of the dataset by introducing country-level variables such as the per capita GDP in the country of origin and the distance from Australia.

Regression results show that GDP exerts significant effects only in the selection equation: the income of immigrants being equal, the probability of origin-households receiving a transfer is higher if they live in a poor country. To the extent that per-capita GDP in the home country captures the living condition of the family of origin, our evidence does not show any prevalence of the altruistic motivation to remit since the relative coefficient is not significant in the main equation. However, the fact that country-specific variables do not exert significant effects may well be a side effect of the inclusion of very significant country dummies[33], due to obvious collinearity problems. As a matter of fact, after removing country dummies GDP exerts significant effects both in the main equation and the selection equation, but with opposite signs[34]. This result would be consistent with predictions of the exchange bargaining-type model by Cox, Eser, and Jimenez (1998). They show, in contrast to the pure altruism hypothesis, that for exchange motivated immigrants an increase in the income of recipients may raise the amount transferred, as the bargaining power of the latter is higher, but decreases the likelihood of remittances, as the benefits of participating in the exchange with migrants are lower for recipients.

The distance from the home country and the cost of contacts with the origin family could have inconsistent effects on remittances. For example, altruistic senders might be discouraged from transferring money back home if their sentiment decreases with the distance from the family. Similarly, selfish senders might be discouraged from remitting far away, if we assume that the enforcement of exchanges with relatives at home can become more difficult with distance. However, by saving on visit expenses, immigrants, whether altruistic or selfish, can afford to send more money back home. Further, if remittances repay migration costs, transfers to the home country

[32] Again in Funkhouser (1995) family relationships are shown to be relevant both to the decision whether to remit and to the amount of money remitted. See also Clark and Drinkwater (2007).

[33] For the double-hurdle model, the LR statistic for the null hypothesis of joint non-significance of the country dummies is 281.45, which has 120 degrees of freedom and leads to rejecting the null at any significance level.

[34] In the remittance equation, the GDP coefficient is equal to 0.162 (p-value: 0.058) while in the selection equation the coefficient is -0.449 (p-value: 1.18e-7).

should increase with geographic distance. Our results show that distance plays a significant role only in the selection process. The farther the country of origin from Australia, the lower the propensity to remit[35].

The interpretation of the coefficients for the selection equation needs a preliminary *caveat*: since we are estimating the selection equation in reduced form, our endogenous variables (income and consumption) are not explicitly included among the determinants of the decision to remit: this is not to say that they are irrelevant. Their effect, however, is not separately discernible from that of the exogenous variables, instruments included.

For example, the positive sign of the variable *SPOUSE* is probably due to the super-imposition of two effects[36]: trivially, the presence of a working spouse leads to a higher income and to a higher probability to remit. However, a separate effect may come from the fact that when both wife and husband migrate, relatives of the enlarged family in the country of origin are left devoid of the help of both the adult sons, which increases their financial needs.

The variable *CHILD* is negatively associated with the likelihood of immigrants sending money back home: it has already been suggested elsewhere (Clark and Drinkwater, 2007; Bollard, McKenzie, and Morten, 2011; Dustmann and Mestres, 2010) that the presence the presence of children may reduce the probability of immigrants remitting. Finally, immigrants whose best spoken language is English are less likely to transfer money back home than others. To the extent that *ENG* captures the strength of social and cultural links to the country of origin, people who still do not consider English as their first language, after three years spent in Australia, probably view their homeland as the main center of their interests and are therefore more likely to remit.

5 Conclusions

In this paper, we provided empirical evidence about remittance strategies of a panel of immigrants from 125 different countries who recently settled in Australia. Our contribution adds to both the literature on remittances and applied econometrics. First, by means of a simultaneous-equation model, we

[35]When removing country dummies, a negative effect also emerges in the main equation. A similar result is found in Lueth and Ruiz-Arranz (2008), where the aggregate amount of remittances between pairs of countries is explained by means of a gravity model, showing that altruism is less of a factor than commonly considered.

[36]This is in contrast with all the previous literature (see Hagen-Zanker and Siegel (2007)).

set remittances in the broader context of households' work and consumption decisions, hence addressing the problem of endogeneity of both pre-transfer income and consumption, that, to the best of our knowledge, has never been dealt with previously in the literature. Second, we address the censored nature of the remittance variable by estimating IV Heckit and double-hurdle models via a LIML strategy. The importance of employing the former technique rather than the latter is intrinsically related to the hypothesis made about the presence (and the size) of transfer costs that may seriously affect remittance behavior.

Estimation results reveal that a proper treatment of endogeneity is very important: our estimates show that the elasticity of remittances to pre-transfer income is positive (as shown by previous studies) but is considerably larger than previously thought, while elasticity to consumption is negative.

This evidence seems to go in favor of exchange motives governing remittances, even if altruistic motivation cannot be excluded. First, the hypothesis of unit elasticity of remittances to pre-transfer income (consistent with the selfish motive) cannot be rejected. Second, the amount of money sent to the country of origin increases with immigrants' education but the likelihood of remittances does not, which is consistent with the loan repayment hypothesis. Third, the home country per capita GDP, that we read as a proxy for the pre-transfer recipients' income, has a negative impact on the propensity to remit but not on the amount transferred (in fact, there is some weak evidence of a positive effect, if anything).

As far as the selection mechanism is concerned, our tests favor the double-hurdle censoring mechanism over a Heckman-style one: it follows that transfer costs are not irrelevant from a statistical viewpoint. Nevertheless, the estimated coefficients of the IV Heckit and double-hurdle models are very similar; in our sample, therefore, the main determinants of the *a priori* probability of a household being a remitted are the same whether or not transfer costs are taken into account.

References

AGGARWAL, R., AND A. W. HOROWITZ (2002): "Are International Remittances Altruism or Insurance? Evidence from Guyana Using Multiple-Migrant Households," *World Development*, 30(11), 2033–2044.

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press.

AMUEDO-DORANTES, C., AND S. POZO (2006): "Remittances as insurance: evidence from Mexican immigrants," *Journal of Population Economics*, 19(2), 227–254.

ANDREWS, D. W., AND J. H. STOCK (2005): "Inference with Weak Instruments," NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.

BLUNDELL, R., J. HAM, AND C. MEGHIR (1987): "Unemployment and Female Labour Supply," *Economic Journal*, 97(388a), 44–64.

BLUNDELL, R., AND R. J. SMITH (1994): "Coherency and estimation in simultaneous models with censored or qualitative dependent variables," *Journal of Econometrics*, 64(1–2), 355–373.

BOLLARD, A., D. MCKENZIE, AND M. MORTEN (2011): "The Remitting Patterns of African Migrants in the OECD," *Journal of African Economies*, 20(3), 605–634.

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.

BROWN, R. P. (1997): "Estimating Remittance Functions for Pacific Island Migrants," *World Development*, 25(4), 613–626.

CARDOSO, A., E. FONTAINHA, AND C. MONFARDINI (2010): "Children's and parents' time use: empirical evidence on investment in human capital in France, Germany and Italy," *Review of Economics of the Household*, 8, 479–504.

CARLING, J. (2008): "The Determinants of Migrant Remittances," *Oxford Review of Economic Policy*, 24(3), 581–598.

Clark, K., and S. Drinkwater (2007): "An Investigation of Household Remittance Behaviour; Evidence from the United Kingdom," *The Manchester School*, 75(6), 717–741.

Cottrell, A., and R. J. Lucchetti (2009): *Gretl User's Manual*, available at `http://gretl.ecn.wfu.edu/pub/gretl/manual/en/gretl-guide-a4.pdf`.

Cox, D., Z. Eser, and E. Jimenez (1998): "Motives for private transfers over the life cycle: An analytical framework and evidence for Peru," *Journal of Development Economics*, 55(1), 57–80.

Cragg, J., and S. Donald (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

Cragg, J. G. (1971): "Some statistical models for limited dependent variables with application to the demand for durable goods," *Econometrica*, 39, 829–844.

Davidson, R., and J. G. MacKinnon (1984): "Model specification tests based on artificial linear regressions," *International Economic Review*, 25, 485–502.

Dustmann, C., and J. Mestres (2010): "Remittances and temporary migration," *Journal of Development Economics*, 92(1), 62–70.

Esteves, R., and D. Khoudur-Castéras (2009): "Remittances, Capital Flows adn Financial Development during the Mass Migration Period, 1870-1913," Working Paper 12, CEPII.

Freund, C., and N. Spatafora (2008): "Remittances, transaction costs, and informality," *Journal of Development Economics*, 86(2), 356 – 366.

Funkhouser, E. (1995): "Remittances from International Migration: a Comparison of El Salvador and Nicaragua," *The Review of Economics and Statistics*, 77(1), 137–146.

Hagen-Zanker, J., and M. Siegel (2007): "The Determinants of Remittances: A Review of the Literature," Working Paper 07/03, Maastricht Graduate School of Governance, Maastricht University.

Hatton, T. J., and J. G. Williamson (1998): *The Age of Mass Migration. Causes and Economic Impact*. Oxford University Press, Oxford.

Hoddinott, J. (1994): “A Model of Migration and Remittances Applied to Western Kenya,” *Oxford Economic Papers*, 46, 459–476.

Jones, A. M. (1989): “A Double-Hurdle Model of Cigarette Consumption,” *Journal of Applied Econometrics*, 4(1), 23–39.

——— (1992): “A Note on Computation of the Double-Hurdle Model with Dependence with an Application to Tobacco Expenditure,” *Bulletin of Economic Research*, 44(1), 67–74.

Konica, N., and R. K. Filer (2009): “Albanian Emigration: Causes and Consequences,” *South-Eastern Europe Journal of Economics*, 7(1), 75–98.

Lianos, T. P. (1997): “Factors Determining Migrant Remittances: The Case of Greece,” *International Migration Review*, 31(1), 72–87.

Lucas, R. E., and O. Stark (1985): “Motivations to remit: evidence from Botswana,” *Journal of Political Economy*, 93(5), 901–918.

Lueth, E., and M. Ruiz-Arranz (2008): “Determinants of Bilateral Remittance Flows,” *The B.E. Journal of Macroeconomics*, 8(1).

Markova, E., and B. Reilly (2007): “Bulgarian migrant remittances and legal status: some micro-level evidence from Madrid,” *South-Eastern Europe Journal of Economics*, 5(1), 55–69.

Menjivar, C., J. Da Vanzo, L. Greenwell, and R. B. Valdez (1998): “Remittance behaviour among Salvadoran and Filippino immigrants in Los Angeles,” *International Migration Review*, 32(1), 97–126.

Merkle, L., and K. F. Zimmermann (1992): “Savings, Remittances and Return Migration,” *Economics Letters*, 38(1), 77–81.

Naufal, G. S. (2008): “Why Remit? The Case of Nicaragua,” IZA Discussion Paper Series 3276, Institute for the Study of Labor.

Osaki, K. (2003): “Migrant Remittances in Thailand: Economic Necessity or Social Norm?,” *Journal of Population Research*, 20(2), 203–222.

Osili, U. O. (2007): “Remittances and Savings from International Migration: Theory and Evidence using a Matched Sample,” *Journal of Development Economics*, 83(2), 446–465.

RAPOPORT, H., AND F. DOCQUIER (2006): "The Economics of Migrants' Remittances," in *Handbook on the Economics of Giving, Altruism and Reciprocity*, ed. by S. Kolm, and J. Mercier Ythier, vol. 2, pp. 1135–1198. Elsevier.

RATHA, D., S. MOHAPATRA, AND A. SILWAL (2011): *Migration and Remittances Factbook 2011*. World Bank.

SINNING, M. (2011): "Determinants of Savings and Remittances: Empirical Evidence from Immigrants to Germany," *Review of Economics of the Household*, forthcoming.

SOLIMANO, A. (2003): "Development Cycles, Political Regimes and International Migration: Argentina in the 20th century," Macroeconomia del Desarrollo 22, CEPAL, Naciones Unidas.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STARK, O. (2009): "Reasons for Remitting," Discussion Papers 52800, University of Bonn, Centre for Development Research (ZEF).

VUONG, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57(2), 307–33.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.

YANG, D. (2009): "International Migration, Remittances and Investment: Evidence From Philippine Migrants' Exchange Rate Shocks," *Economic Journal*, 118(4), 591–630.

YANG, D., AND H. CHOI (2007): "Are Remittances Insurance? Evidence from Rainfall Shocks in the Philippines," *The World Bank Economic Review*, 21(2), 219–248.

ZAICEVA, A., AND K. F. ZIMMERMANN (2007): "Children, Kitchen, Church: Does Ethnicity Matter?," CEPR Discussion Papers 6491, CEPR, London.

A Two simple models of pure altruistic and selfish remittances

In this appendix we present two very simple models for immigrants' pure-unilateral-altruistic and pure-exchange remittances and savings decisions in which for simplicity we exclude transfer costs. Consider an immigrant who lives for two periods. During the first period the immigrant is assumed to earn an income Y_m that he/she can allocate over consumption, savings in the host country and remittances to the origin family. The immigrant maximises a time-separable and log-linear utility function defined over first- and second-period consumption and, alternatively, origin-family utility in the first period (altruistic transfers) or services acquirable from remittances back home (selfish transfers).

A.1 Pure altruism

In this case, the maximisation problem is:

$$\max_{C_1,C_2,R} \quad U_m = \ln(C_1) + \rho\ln(C_2) + \gamma\ln(C_p)$$

s.t.

$$\begin{aligned} C_p &= Y_p + R \\ Y_m &= C_1 + \delta C_2 + R \\ R &\geq 0 \end{aligned}$$

where C and Y stand for consumption and income, m and p for immigrant and parental family, R is the amount of money remitted, ρ and δ are the individual and market inter-temporal discount factors, and γ is the relative degree of altruism. From the first order conditions, it is easy to verify that:

$$R = \max\left\{0, \frac{\gamma}{1+\gamma+\rho}Y_m - \frac{1+\rho}{1+\gamma+\rho}Y_p\right\}$$

and the elasticity of R with respect to Y_m

$$\epsilon = \frac{\gamma Y_m}{\gamma Y_m - (1+\rho)\,Y_p} > 1$$

A.2 Pure selfishness

$$\max_{C_1,C_2,R} \quad U_m = \ln(C_1) + \rho\ln(C_2) + \beta\ln(S)$$

s.t.

$$\begin{aligned} S &= \alpha R \\ Y_m &= C_1 + \delta C_2 + R \\ R &\geq 0 \end{aligned}$$

where S is the amount of services in the home country acquired by remittances, β is its relative utility and α is the relative price. From the first order conditions, it is easy to verify that:

$$R = \frac{\beta}{1 + \beta + \rho} Y_m$$

and the elasticity of R with respect to Y_m is equal to 1.

B Numerical issues in the computation of maximum likelihood estimates

Since the first-order conditions for a maximum cannot be solved analytically, a numerical maximization procedure is needed[37]. Although our estimation technique is a fairly straightforward application of numerical maximum likelihood, in some cases standard numerical procedures may not yield optimal

[37] We used the BFGS implementation provided by gretl; see Cottrell and Lucchetti (2009). In order to ensure that parameter σ remains positive during the numerical search, the log-likelihood is reparametrized in terms of $\ln \sigma$. For similar reasons, the unconstrained parameters on which the marginal log-likelihood function ℓ_i^m is based are not the elements of Σ itself, but rather those of the Cholesky factorization of Σ^{-1}. In practice, ℓ_i^m, the second component of the log-likelihood, is computed as

$$\ell_i^m = \text{const} + \ln |K| - \frac{\xi_i' \xi_i}{2}$$

where K is a lower-triangular matrix such that $KK' = \Sigma^{-1}$ and $\xi_i = K'(R_i - \Pi' Z_i)$. This has two advantages: not only is a matrix inversion avoided, but the determinant of K (which is by construction $|\Sigma|^{-1/2}$) is trivial to compute since K is triangular, via

$$-0.5 \ln |\Sigma| = \sum_{i=1}^{m} \ln K_{ii}.$$

Finally, the correlation coefficient between u_i and v_i was reparametrized via the hyperbolic tangent transformation as

$$a = 0.5 \ln \left(\frac{1 + \rho}{1 - \rho} \right).$$

results, since the log-likelihood function may have multiple maxima. Consistency of maximum likelihood estimators is known to stem from the fact that the expected value of the log-likelihood has a unique maximum at θ_0 and uniform convergence of the observed log-likelihood to its expectation[38]. However, the observed log-likelihood may well have multiple maxima in finite samples. In these cases, it is reasonable to take the global maximum as the ML estimator. Numerical methods, however, do not guarantee that the algorithm stops at the global maximum, since they may get stuck in a local maximum.

With our dataset, we found that in several instances this was indeed the case. For some specifications, there were two maxima, corresponding to two different values of the correlation coefficient ρ.[39] In order to circumvent this problem, we used the following computational strategy: given a value of ρ, carry out the maximum likelihood estimation of the remaining parameters, thus obtaining a restricted estimate $\hat{\psi}(\rho)$ (where ψ is a vector gathering all the other parameters). This procedure was repeated over a grid of values for ρ from -0.9 to 0.9 with increments of 0.1; the value of $\hat{\psi}(\rho)$ yielding the maximum likelihood was then used as the starting point for the maximization of the unrestricted log-likelihood. We found this procedure to be mildly time-consuming, but very effective.

[38] A classic exposition of the argument is found in Amemiya (1985).

[39] This appears to be a little-known feature of the dependent double-hurdle model. To our knowledge, no systematic investigation has been carried out on this matter.

C Auxiliary tables and figures

Table 3: Main countries of origin of immigrants in LSIA 1

Country of Origin	Number of households	Percentage of remitters	Avg. remittance (AUS $)
United Kingdom	315	14.3	3,650
China	148	37.8	3,882
India	135	37.0	3,583
Vietnam	134	47.8	1,320
Philippines	131	58.0	2,205
Former Yugoslavia	114	48.2	2,243
Hong Kong, China	109	22.0	6,267
South Africa	106	14.2	5,690
Sri Lanka	96	50.0	1,464
Iraq	88	62.5	3,296
Malaysia	82	35.4	4,278
Indonesia	79	30.4	1,473
Fiji	77	35.1	2,033
Germany	70	10.0	2,129
Myanmar	69	40.6	1,788
Ukraine	68	27.9	1,403
Lebanon	67	23.9	1,425
Korea, Rep.	64	10.9	1,379
Afghanistan	63	58.7	2,389
Cambodia	62	37.1	1,061
Japan	62	8.1	11,100
Russian Federation	62	32.3	2,935
Poland	61	19.7	1,900
Iran, Islamic Rep.	60	16.7	3,580
Italy	60	6.7	1,925
Bosnia and Herzegovina	56	53.6	1,605
Egypt, Arab Rep.	56	23.2	2,908
Turkey	56	37.5	1,390
United States	56	10.7	2,150
Romania	55	41.8	1,761

Note: the average amount of remittances is computed on remitters only.

Figure 2: Share of remitters and average log remittances by log GDP of country of origin

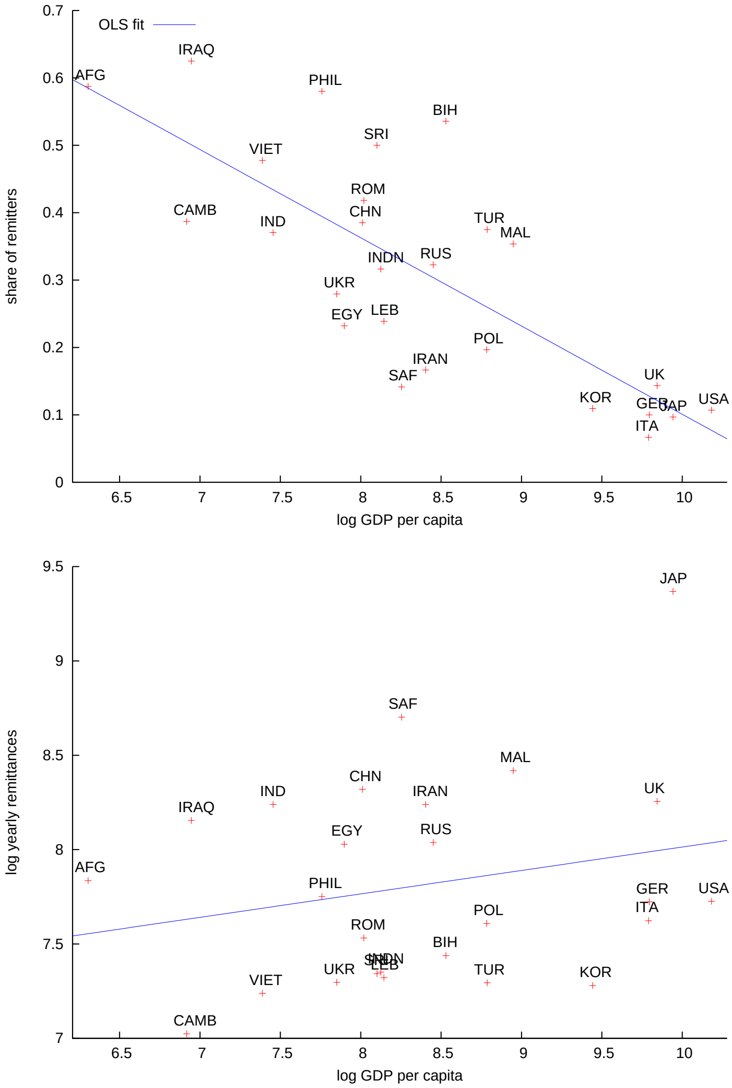

Table 4: Descriptive statistics

Variable	Mean	Minimum	Maximum	Std. Dev.
REMITTANCES	1.919	0	9.953	2.974
INCOME	10.505	8.232	11.177	0.689
CONSUMPTION	9.907	7.122	10.955	0.395
MALE	0.416	0	1	0.493
AGE	3.793	1.800	8.700	1.158
TIME	3.444	3.039	4.868	0.127
CITIZENSHIP	0.927	0	1	0.260
RELATIVES	0.946	0	1	0.227
EDUCATION_2	0.505	0	1	0.500
EDUCATION_3	0.221	0	1	0.415
EDUCATION_4	0.114	0	1	0.318
GDP_PC	8.814	6.208	10.594	0.986
DISTANCE	9.269	7.825	9.777	0.390
MOVE	0.475	0	1	0.499
ILLNESS	0.055	0	1	0.228
ENG	0.187	0	1	0.390
CHILD	0.543	0	1	0.498
SPOUSE	0.750	0	1	0.433
FAMILY_N	3.408	1	16	1.564
FAMILY_N^2	14.062	1	256	14.065

Table 5: First-stage estimates for Heckit model

First stage for log income				
	coeff.	std.err.	z-stat	p-value
const	8.177	0.501	16.330	0.000 ***
MALE	0.028	0.027	1.016	0.309
AGE	0.137	0.063	2.167	0.030 **
AGE2	-0.029	0.007	-4.344	0.000 ***
TIME	0.162	0.116	1.399	0.162
CITIZENSHIP	-0.086	0.057	-1.510	0.131
RELATIVES	0.011	0.060	0.183	0.855
EDUCATION_2	-0.214	0.038	-5.688	0.000 ***
EDUCATION_3	-0.322	0.046	-6.943	0.000 ***
EDUCATION_4	-0.243	0.059	-4.100	0.000 ***
GDP_PC	0.074	0.023	3.218	0.001 ***
DISTANCE	0.143	0.092	1.558	0.119
ENG	0.261	0.046	5.655	0.000 ***
MOVE	0.111	0.030	3.713	0.000 ***
FAMILY_N	0.418	0.042	9.992	0.000 ***
FAMILY_N$\hat{2}$	-0.032	0.005	-6.794	0.000 ***
CHILD	-0.459	0.041	-11.090	0.000 ***
SPOUSE	0.201	0.038	5.287	0.000 ***
ILLNESS	-0.154	0.051	-3.038	0.002 ***
First stage for log consumption				
	coeff.	std.err.	z-stat	p-value
const	8.301	0.279	29.740	0.000 ***
MALE	0.007	0.015	0.433	0.665
AGE	0.115	0.040	2.859	0.004 ***
AGE2	-0.018	0.004	-3.944	0.000 ***
TIME	0.165	0.064	2.583	0.010 ***
CITIZENSHIP	-0.076	0.028	-2.757	0.006 ***
RELATIVES	0.002	0.034	0.054	0.957
EDUCATION_2	-0.083	0.020	-4.213	0.000 ***
EDUCATION_3	-0.136	0.026	-5.155	0.000 ***
EDUCATION_4	-0.165	0.033	-4.962	0.000 ***
GDP_PC	0.047	0.013	3.689	0.000 ***
DISTANCE	0.020	0.047	0.431	0.667
ENG	0.127	0.022	5.773	0.000 ***
MOVE	0.014	0.015	0.930	0.352
FAMILY_N	0.197	0.020	9.604	0.000 ***
FAMILY_N^2	-0.015	0.002	-6.359	0.000 ***
CHILD	-0.068	0.021	-3.295	0.001 ***
SPOUSE	0.170	0.019	8.803	0.000 ***
ILLNESS	-0.029	0.029	-0.996	0.319

Note: QMLE standard errors (see White (1982)). Country-specific fixed effects included. First-stage F-tests: 59.841 (income), 71.391 (consumption). Wald test for $\Pi_2 = 0$: 719.824 (p-value: 1.51E-144).

Table 6: First-stage estimates for double-hurdle model

First stage for log income					
	coeff.	std.err.	z-stat	p-value	
const	8.176	0.533	15.350	0.000	***
MALE	0.028	0.025	1.105	0.269	
AGE	0.137	0.094	1.450	0.147	
AGE2	-0.029	0.011	-2.583	0.010	***
TIME	0.162	0.108	1.505	0.132	
CITIZENSHIP	-0.086	0.048	-1.811	0.070	*
RELATIVES	0.011	0.056	0.194	0.846	
EDUCATION_2	-0.214	0.031	-6.972	0.000	***
EDUCATION_3	-0.322	0.044	-7.403	0.000	***
EDUCATION_4	-0.243	0.056	-4.363	0.000	***
GDP_PC	0.074	0.023	3.265	0.001	***
DISTANCE	0.143	0.113	1.271	0.204	
ENG	0.261	0.035	7.550	0.000	***
MOVE	0.111	0.026	4.221	0.000	***
FAMILY_N	0.418	0.044	9.595	0.000	***
FAMILY_N$\hat{2}$	-0.032	0.005	-6.048	0.000	***
CHILD	-0.459	0.035	-13.180	0.000	***
SPOUSE	0.201	0.035	5.761	0.000	***
ILLNESS	-0.155	0.060	-2.593	0.010	***
First stage for log consumption					
	coeff.	std.err.	z-stat	p-value	
const	8.301	0.298	27.890	0.000	***
MALE	0.007	0.014	0.459	0.646	
AGE	0.115	0.046	2.486	0.013	**
AGE2	-0.018	0.005	-3.358	0.001	***
TIME	0.165	0.062	2.659	0.008	***
CITIZENSHIP	-0.076	0.031	-2.504	0.012	**
RELATIVES	0.002	0.032	0.057	0.954	
EDUCATION_2	-0.083	0.019	-4.423	0.000	***
EDUCATION_3	-0.136	0.024	-5.653	0.000	***
EDUCATION_4	-0.165	0.032	-5.159	0.000	***
GDP_PC	0.047	0.013	3.754	0.000	***
DISTANCE	0.020	0.061	0.337	0.736	
ENG	0.127	0.021	5.998	0.000	***
MOVE	0.014	0.015	0.948	0.343	
FAMILY_N	0.197	0.025	7.746	0.000	***
FAMILY_N^2	-0.015	0.003	-4.774	0.000	***
CHILD	-0.068	0.021	-3.198	0.001	***
SPOUSE	0.170	0.021	8.302	0.000	***
ILLNESS	-0.029	0.031	-0.953	0.341	

Note: QMLE standard errors (see White (1982)). Country-specific fixed effects included. First-stage F-tests: 59.841 (income), 71.391 (consumption). Wald test for $\Pi_2 = 0$: 682.272 (p-value: 1.56E-136).

Table 7: Double-hurdle model with alternative choices of $\underline{R}$

	$\underline{R}$=10	$\underline{R}$=15	$\underline{R}$=25	$\underline{R}$=30
		Main equation		
const	14.311 ***	14.352 ***	14.475 ***	14.551 ***
MALE	0.226 **	0.228 **	0.234 **	0.238 **
AGE	0.361	0.364	0.372	0.378
AGE2	-0.056	-0.057	-0.059	-0.060
TIME	0.221	0.221	0.223	0.225
CITIZENSHIP	-0.429	-0.433	-0.447	-0.458
RELATIVES	0.510 *	0.517 *	0.538 *	0.551 *
EDUCATION_2	-0.314 **	-0.316 **	-0.323 **	-0.328 **
EDUCATION_3	-0.523 ***	-0.526 ***	-0.535 ***	-0.541 ***
EDUCATION_4	-0.737 ***	-0.744 ***	-0.764 ***	-0.777 ***
GDP_PC	0.196	0.199 *	0.208 *	0.214 *
DISTANCE	-0.437	-0.440	-0.448	-0.455
INCOME	1.033 ***	1.042 ***	1.075 ***	1.098 ***
CONSUMPTION	-2.110 ***	-2.127 ***	-2.181 ***	-2.218 ***
		Selection equation		
const	1.550	1.554	1.572	1.589
MALE	0.188 ***	0.188 ***	0.185 ***	0.183 ***
AGE	0.352	0.352	0.351	0.351
AGE2	-0.067 **	-0.067 **	-0.067 **	-0.066 **
TIME	0.093	0.093	0.094	0.095
CITIZENSHIP	0.244 *	0.244 *	0.246 *	0.248 *
RELATIVES	0.328 **	0.326 **	0.317 **	0.311 *
EDUCATION_2	-0.092	-0.091	-0.088	-0.085
EDUCATION_3	-0.001	0.001	0.006	0.010
EDUCATION_4	0.206	0.209	0.221	0.230
GDP_PC	-0.280 ***	-0.280 ***	-0.283 ***	-0.285 ***
DISTANCE	-0.442 *	-0.442 *	-0.440 *	-0.438 *
ENG	-0.177 *	-0.177 *	-0.176 *	-0.176 *
MOVE	-0.019	-0.018	-0.017	-0.017
FAMILY_N	-0.001	-0.002	-0.004	-0.005
FAMILY_N^2	-0.007	-0.007	-0.006	-0.006
CHILD	-0.171 *	-0.170 *	-0.165 *	-0.161 *
SPOUSE	0.256 ***	0.257 ***	0.260 ***	0.262 ***
ILLNESS	-0.135	-0.135	-0.133	-0.131
σ	1.094 ***	1.099 ***	1.115 ***	1.125 ***
ρ	0.104	0.100	0.087	0.080
Endog test	0.00	0.00	0.00	0.00
Overid test	0.12	0.11	0.04	0.01
Vuong's test statistic	2.856	3.421	4.276	4.621
Log-likelihood	-4217.12	-4216.19	-4212.49	-4209.68
Total cases	2170	2170	2170	2170
Uncensored	652	652	652	652

Note: QMLE standard errors (see White (1982)). Country-specific fixed effects included.