# Selection bias, unobserved heterogeneity and state dependence in empirical remittance modelling: evidence from immigrants to Germany

Giulia Bettin
Università Politecnica delle Marche
g.bettin@univpm.it

Riccardo Lucchetti
Università Politecnica delle Marche
r.lucchetti@univpm.it

Claudia Pigini
Università Politecnica delle Marche
c.pigini@univpm.it

*Preliminary draft. Please do not quote without permission*

**Abstract**

The empirical modelling of remitting behaviour has been the object of a considerable amount of micro-level literature. Although there is a good deal of variety among the data employed for this purpose, some features are pervasive and general, such as the censored nature of international transfers. Moreover, the increasing availability of panel datasets makes it possible to explore the time dimension of remittance behaviour, but this poses the additional need to address permanent unobserved heterogeneity and state dependence.

However, these issues have not been jointly accounted for in the empirical migration literature so far. In this paper, we propose two dynamic, random-effects sample selection models: we combine the Maximum Likelihood estimators of the sample selection and double hurdle models (Heckman, 1974; Cragg, 1971) with the approach proposed by Heckman (1981) for dealing with state dependence and unobserved heterogeneity in a non-linear setting. Our objective is therefore two-fold: (i) we introduce feasible estimators for a potentially large class of empirical problems; (ii) we offer novel empirical evidence in the field of remittance modelling by applying our proposed models to the German SOEP dataset.

**JEL codes**: F22, F24, C23, C34, C35
**Keywords**: Migration, Remittances, Random effects, Sample selection, State dependence.

## 1 Introduction

International remittances have long been one of the most investigated issues in the migration research agenda. Transfers sent home by international migrants exceeded official development assistance and portfolio investment since the late 1990s and almost approached the magnitudes of FDI flows during the global financial crisis (Yang, 2011). At the aggregate level, the empirical literature has mainly focused on the development impact of remittances for developing countries, on the one hand, and on their cyclical properties and macroeconomic drivers on the other. At the microeconomic level, the analysis of remittances has been related to the decision making processes within transnational households by either investigating motives to remit or looking at the concrete use these resources were allocated to.

The aim of the present paper is to examine in detail the main methodological issues that arise in modelling individual remittance behaviour and discuss the different empirical strategies scholars may use to deal with them.

In particular, we will first consider a stylised fact encountered in most datasets: typically, a large share of migrants do not remit money to their country of origin and hence "zero" observations need to be handled appropriately. The selection problem has been long taken into account in the literature but our purpose is to shed further light on the appropriateness of different econometric models (Tobit, Heckman selection model, double hurdle model) in addressing it.

Second, issues arising with the use of longitudinal data will be discussed. Most of the empirical literature on remitting behaviour is based on cross-sectional information and the use of panel data is a relatively recent novelty in the field (Duval and Wolff, 2010, 2012; Dustmann and Mestres, 2010). Of course, dealing with unobserved heterogeneity becomes crucial in such context. In addition, the potentially intertemporal nature of remittance strategies introduces a second, but not less important, aspect that needs to be adequately modelled when using panel data. In fact, migrants may smooth remittances over time and adopt a forward-looking behaviour characterised by high levels of persistence (Bettin and Lucchetti, 2016).

Our paper therefore discusses how to address unobserved heterogeneity in remitting behaviour by static panel models, once selection mechanisms have been dealt with, and then shows possible extensions of these models to a dynamic setting where the intertemporal nature of remittance decisions arise at both the extensive (the choice to remit or not) and the intensive (the choice on the amount remitted) margin.

For this purpose, we develop two random-effects sample selection models that account for the presence of individual permanent unobserved heterogeneity and possible state dependence. First, we extend the Maximum Likelihood estimators of the sample selection (Heckman, 1974) and double hurdle (Cragg, 1971) models to accommodate the presence of individual unobserved effects relying on a two-step numerical integration procedure, in the spirit of Raymond et al. (2010). Then we propose a general dynamic version of these models that allows for the lags of both the dependent variables to be included in the main and selection equations. We follow Heckman (1981) and tackle the "initial conditions" problem by specifying additional equations that approximate the distribution of the initial realisations of the dependent variables conditional on the random effects.

The choice of a random-effects strategy is mainly driven by distinctive features of remittance data: determinant individual characteristics in modelling remittance behaviour, such as the migrant's family composition, typically exhibit little time variation and the decision to send remittances, the outcome of the selection equation, is highly persistent. With these data, estimation approaches based on differencing or conditioning on sufficient statistics for the individual intercepts may not allow for the identification of crucial determinants of the agent's behaviour and/or lead to a substantial information loss.

The general discussion is followed by an empirical analysis based on microlevel longitudinal data from the German Socio-Economic Panel (SOEP), which covers a large sample of immigrants from 1997 onwards and provides information on their characteristics, including remitting behaviour, both at the individual level and at the household level.

The paper is structured as follows: the main empirical issues in modelling remittance decisions and the way they have been addressed in the literature so far are discussed in depth in section 2. In section 3 we illustrate the static and dynamic random-effects sample selection models and survey the related econometric literature. Section 4 describes the GSOEP data and provides some descriptive evidence and the related empirical results are presented and discussed in section 5. Section 6 concludes.

## 2 Empirical issues in modelling remittance behaviour

Empirical literature investigating the drivers of individual remittance decisions by means of microlevel data has largely developed in the last decade (Rapoport and Docquier, 2006; Brown and Jimenez-Soto, 2015) Different motivations to remit might contribute to explain migrants' strategies, including altruistic feelings (Funkhouser, 1995; Aggarwal and Horowitz, 2002; Yang and Choi, 2007; Yang, 2008), inheritance motives (Hoddinott, 1994; de la Briere et al., 2002), insurance contracts (Lucas and Stark, 1985; Rosenzweig, 1988), exchange motives (Bernheim et al., 1985; Cox, 1987) and loan repayments (Cox et al., 1998; Poirine, 1997).

In general, empirical modelling of remittance behaviour poses a first main issue that needs to be addressed, that is the treatment of zeros. The share of remitting migrants is often not high in dedicated surveys[1] that have been employed in the literature to investigate remittance behaviour and might become even lower when using data from standard household surveys on either receiving or sending countries. Empirical analyses addressed the censoring of the dependent variable in different ways.

Banerjee (1984) pointed out that the appropriate regression model depends on the hypothesis made on the underlying decision scheme: the decision to remit can be explicitly modelled as a two-stage, sequential process (first to remit or not, then the amount remitted), or, as a one-stage simultaneous process. However, in neither case OLS is consistent.

In fact, remittances in most cases cannot be treated as a continuous variable but can be more accurately approximated by a mixture distribution, as they often present a non-negligible frequency mass on the value zero. Nevertheless, many studies have simply ignored the censoring problem, paying no special attention to the zero observations (Lucas and Stark, 1985).

The choice of the appropriate econometric model to deal with the large amount of zero–remittances depends on the interpretation of the individual behaviour.

Banerjee (1984) himself and Hoddinott (1992, 1994) were among the first to model the extensive (the choice of whether to remit or not) and the intensive margin (the decision on the amount remitted) separately and use the Heckman (1979) procedure to correct for the selection bias. The decision mechanism, in this case, envisages a sort of "reservation" amount, determined by transaction costs, below which the immigrant is unwilling to send remittances and the observed zeros, then, are a threshold for not sending.

Subsequent studies made large use of the same empirical methodology (e.g. Funkhouser, 1995; Cox et al., 1998; Aggarwal and Horowitz, 2002; Amuedo-Dorantes and Pozo, 2006; Bouyiour and Miftah, 2015) often relying also on the exclusion restrictions used in Hoddinott (1992) to correctly identify the two separate choices. However, it has been often recognised that results from such models are extremely sensitive to the choice of identification exclusions, whose suitability might be disputable due to the difficulties in conceiving factors that affect the decision to remit, but do not influence the amount remitted.

These problems might partly explain why the identification problem was circumvented in many cases by using a Tobit model (Tobin, 1958) that addresses the censored nature of the dependent variable in a single equation with a common set of regressors (Bouyiour and Miftah, 2015; Brown, 1997; de la Briere et al., 2002; Hoddinott, 1992). In this case, the observed zeros are the result of a cor-

[1] Amuedo-Dorantes and Pozo (2006) for example use the Encuesta sobre Migración en la Frontera Norte de México (EMIF) and show that approximately 53% of working immigrants in their sample does not remit.

ner solution of a constrained optimisation problem with left censoring at zero. Clearly, the disadvantage of the Tobit model is that the likelihood of remitting and the amount remitted are explained by one and the same mechanism and might not be separated from each other.

More recently, the double-hurdle model (Cragg, 1971) has been proposed in the empirical literature on remitting decisions as a further alternative to the Heckman (1979) selection model in order to take into account that both the above mechanism could be in place. Therefore, migrants who do not remit might not simply be individuals that are not interested in sending any money home (given transaction costs), but they might be willing to remit but prevented from doing so by a budget constraint. Sinning (2011) and Brown et al. (2014b) used a double-hurdle model in its restricted independent version, while Bettin et al. (2012) propose an instrumental variable extension of the dependent double-hurdle model, where the potential endogeneity of explanatory variables (migrants' income and consumption expenditure) is taken into account.

The censored nature of the remittance variable is an issue affecting all datasets, irrespective of their time dimension. The vast majority of migration and remittances surveys are cross–sectional surveys and empirical analyses based on them provide a snapshot of one point in time (Brown et al., 2014a).

Within the context of panel data, two additional issues arise: unobserved heterogeneity and, when accounting for the (possible) intertemporal nature of remittance choices, state dependence. If remittances were conceived as an alternative to consumption in the context of household's budget allocation, we might observe a smoothing process over time, according to the individual expectations on future income. This forward looking behaviour would imply a high level of persistence of remitting behaviour that directly depends on the stability of migrants' income over time, but also on (sudden) changes in other socioeconomic characteristics.

Evidence based on household panel surveys is still relatively scarce. Duval and Wolff (2010) adopted a static framework and estimated the probability to receive remittance for Albanian households using the Living Standard Measurement Study (LSMS) data for 2002-2004 and control for unobserved heterogeneity of recipient households via either a random-effects Probit model or a fixed-effects Logit model according to the different assumption on the correlation between covariates and individual effects.

A few other studies made use of the German Socio-Economic Panel (GSOEP) data which are available since 1984 and offer information on remittance behaviour of immigrant households living in Germany. Holst et al. (2011, 2012), for example, addressed both the censored nature of the amount remitted and unobserved heterogeneity at the individual level by means of a random-effects

Tobit model, thus assuming that the explanatory variables were uncorrelated with the unobserved individual effects.

Dustmann and Mestres (2010) use GSOEP data to investigate how return plans affect the decision on whether to remit and on the amount remitted, separately considered. Some dynamics was introduced in their model, but only by instrumenting the intention to return with past realisations of either the probability to remit or the size of the transfer.

The persistence in the decision to remit was instead the focus in Bettin and Lucchetti (2016) where different discrete choice dynamic models (random-effects Probit and fixed-effects Logit) were applied to GSOEP data and provided evidence in favour of an intertemporal strategy. True state dependence is highly significant, meaning that the propensity to remit at time $t$ depends on what the migrant actually did the year before, in $t-1$, even after controlling for persistence in observable and unobservable characteristics. The allocation of remittances seems to follow a multi-period scheme.

## 3 Econometric models

We discuss the specification and Maximum Likelihood estimation of the sample selection model proposed by Heckman (1974) and of the double hurdle model put forward by Cragg (1971). We consider a general formulation for the pooled models that can be easily extended to present our proposed random-effects static and dynamic models.

In order to pursue the censored nature of the data, let us consider the latent variables

$$y_{it}^{*} = \mu_{it}\left(\mathscr{F}_{it}, \alpha_i; \boldsymbol{\psi}\right) + \varepsilon_{it} \tag{1}$$

$$s_{it}^{*} = \nu_{it}\left(\mathscr{F}_{it}, \eta_i; \boldsymbol{\psi}\right) + u_{it}, \quad \text{for } i = 1, \ldots, n \quad t = 1, \ldots, T \tag{2}$$

where $y_{it}^{*}$ is the (latent) desired remitted amount and $s_{it}^{*}$ is the unobservable propensity to remit. Furthermore, $\mu(\cdot)$ and $\nu(\cdot)$ are index functions of the information set at time $t$ available to individual $i$, $\mathscr{F}_{it}$, of the individual time-invariant unobserved heterogeneity $\alpha_i$ and $\eta_i$, and of the model parameters $\boldsymbol{\psi}$. Finally, $\varepsilon_{it}$ and $u_{it}$ are iid error terms.

The decision to remit depends on the binary variable $s_{it} = \mathbb{I}\left(s_{it}^{*} > 0\right)$, where $\mathbb{I}\left(\cdot\right)$ is an indicator function, with $s_{it} = 1$ if the migrant sends remittances and zero otherwise. Let us define a binary variable $d_{it}$ indicating whether positive remitted amounts are observed

$$d_{it} = \begin{cases} \mathbb{I}\left(s_{it}^{*} > 0\right) & \text{sample selection} \\ \mathbb{I}\left(s_{it}^{*} > 0 \wedge y_{it}^{*} > 0\right) & \text{double hurdle} \end{cases}$$

so that $y_{it} = y^*_{it} d_{it}$. In the sample selection model, whether positive amounts are observed depends only on the decision to remit whereas, in the double hurdle model, the same observational rule is combined with the corner solution. The joint density of $(y_{it}, d_{it})$, for model (1)-(2) can therefore be written as

$$f\left(y_{it}, d_{it} | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}\right) = g\left(y_{it}, d_{it} = 1 | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}\right)^{d_{it}} \Pr\left(d_{it} = 0 | \mathscr{F}_{it}, \eta_i; \boldsymbol{\psi}\right)^{1-d_{it}}. \tag{3}$$

A Maximum Likelihood estimator of $\boldsymbol{\psi}$ can be obtained by specifying the density functions with suitable choices for $\mu(\cdot)$ and $\nu(\cdot)$ and distributional assumptions on $\alpha_i$, $\eta_i$, $\varepsilon_{it}$, and $u_{it}$. For $i = 1, \ldots, n$:

DD Conditional on $\mathscr{F}_{it}$, the terms $\varepsilon_{it}$ and $u_{it}$ are distributed as a bivariate normal with zero mean and variance-covariance matrix with elements $\mathrm{E}\left(\varepsilon_{it}\varepsilon_{is}\right) = \sigma^2_\varepsilon$, $\mathrm{E}\left(u_{it}u_{is}\right) = 1$, $\mathrm{E}\left(\varepsilon_{it}u_{is}\right) = \sigma_\varepsilon \rho$ if $t = s$, 0 otherwise, for $t, s = 2, \ldots, T$.

IED Conditional on $\mathscr{F}_{it}$, $\alpha_i$ and $\eta_i$ have degenerate distributions.

IS The information set $\mathscr{F}_{it}$ includes a set of individual covariates $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}]$ in (1), the same set of covariates plus suitable exclusion restrictions $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iT}]$ in (2).

Assumption (DD) is the standard distributional assumption for the sample selection and double hurdle models, Assumption (IED) leads to pooled models, and Assumption (IS) excludes lags of the dependent variables from the set of covariates. Following (IS), we further specify the usual linear index functions as

$$\mu_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}, \quad \nu_{it} = \mathbf{z}'_{it}\boldsymbol{\gamma}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are regression parameters, and $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{it}$ are vectors of explanatory variables, where $\boldsymbol{z}_{it}$ may contain additional exogenous variables with respect to $\boldsymbol{x}_{it}$.

Under Assumptions (DD)-(IS) and the linear index expressions, we can derive the joint density of $(y_{it}, d_{it} = 1)$. We follow Davidson and MacKinnon (2004) and write the joint density, omitting the relevant conditioning sets for brevity, as

$$g\left(y_{it}, d_{it} = 1\right) = \Pr\left(d_{it} = 1 | y_{it}\right) \times \frac{1}{\sigma_\varepsilon} \varphi\left(\frac{(y_{it} - \mu_{it})^2}{\sigma^2_\varepsilon}\right) \tag{4}$$

for $t = 1, \ldots, T$, where $\varphi(\cdot)$ is the standard normal density function. The probability of $d_{it}$ conditional on $y_{it}$ can easily be derived from the conditional distribution of $d^*_{it} | y^*_{it}$ under bivariate normality of $\varepsilon_{it}$ and $u_{it}$, that is

$$\Pr\left(d_{it} = 1 | y_{it}\right) = \Phi\left(c_\omega \nu_{it} + s_\omega (y_{it} - \mu_{it}) / \sigma_\varepsilon\right) \tag{5}$$

where $\Phi(\cdot)$ is the standard normal cdf, $c_\omega = \cosh(\omega)$, $s_\omega = \sinh(\omega)$, and $\omega = \text{atanh}(\rho)$. The probability $\Pr(d_{it} = 0)$ can be written as $1 - P_{it}$, where $P_{it}$ has a different specification according to whether a sample selection or a double hurdle model is estimated. In particular,

$$P_{it} = \begin{cases} \Phi(\nu_{it}) & \text{sample selection} \\ \Phi_2(-\mu_{it}/\sigma_\varepsilon, \nu_{it}, \rho) & \text{double hurdle} \end{cases} \tag{6}$$

where $\Phi_2(\cdot)$ is the bivariate standard normal distribution function.

Finally, we can specify the likelihood for individual $i$ as

$$\mathscr{L}_i(\boldsymbol{\psi}) = \prod_{t=1}^{T} \left[ \Phi\left(c_\omega \nu_{it} + s_\omega (y_{it} - \mu_{it})/\sigma_\varepsilon\right) \frac{1}{\sigma_\varepsilon} \phi\left(\frac{y_{it} - \mu_{it}}{\sigma_\varepsilon}\right) \right]^{d_{it}} \times (1 - P_{it})^{1-d_{it}}.$$

The empirical literature dealing with the estimation of the sample selection model has brought forward a great deal alternatives to Maximum Likelihood estimation under the assumption of bivariate normality of the error terms. In particular, the proposed approaches aim at either replacing the normality assumption, by specifying flexible bivariate distributions with copulae, or removing it, therefore relying on semi-parametric estimators. [2] Nevertheless, the fully parametric specification and the bivariate normality assumption allows for a general formulations that lends itself to a straightforward extension to include unobserved heterogeneity and state dependence.

Individual unobserved effects may be introduced by suitably modifying Assumption (IED). The joint density of $\boldsymbol{y}_i, \boldsymbol{d}_i$, where $\boldsymbol{y}_i = [y_{i1}, \ldots, y_{iT}]$ and $\boldsymbol{d}_i = [d_{i1}, \ldots, d_{iT}]$, for model (1)-(2) can be written as

$$f\left(\boldsymbol{y}_i, \boldsymbol{d}_i | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}\right) = \int_{\mathbb{R}} \int_{\mathbb{R}} \prod_{t=1}^{T} f\left(y_{it}, d_{it} | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}\right) h(\alpha_i, \eta_i) \mathrm{d}\alpha_i \mathrm{d}\eta_i$$

where $f\left(y_{it}, d_{it} | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}\right)$ is defined in (3).

The Maximum Likelihood estimator of $\boldsymbol{\psi}$ can be derived under additional distribution assumptions on $\alpha_i$ and $\eta_i$. For $i = 1, \ldots, n$:

IED′ Conditional on $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$, $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iT}]$, the terms $\alpha_i$ and $\eta_i$ are jointly distributed as a bivariate normal with zero mean and variance-covariance matrix $\Sigma$, where

$$\Sigma = \begin{bmatrix} \sigma_\alpha^2 & \\ \sigma_\alpha \sigma_\eta \kappa & \sigma_\eta^2 \end{bmatrix}$$

[2] See Pigini (2015) for a survey on alternative strategies for the estimation of the Heckman sample selection model, Escanciano et al. (2014) for a novel semi-parametric estimation approach to general double index models, and Schwiebert (2015) for the specification of double-hurdle models with bivariate copulae and flexible margins.

IED″ $(\alpha_i, \eta_i) \perp\!\!\!\perp (\varepsilon_{it}, u_{it})$ for all $i$ and $t$.

Assumption (IED′) is necessary to evaluate the double integral, by exploiting standard properties of the bivariate normal to derive the conditional distribution of $\eta_i$ on $\alpha_i$, that is $\eta_i|\alpha_i \sim N\left[\kappa\frac{\sigma_\eta}{\sigma_\alpha}\alpha_i \ ; \ \sigma^2_\eta(1-\kappa^2)\right]$. This means that the random effect of the selection equation can be written as $\eta_i = \kappa\frac{\sigma_\eta}{\sigma_\alpha}\alpha_i + \delta_i$ where $\delta_i \sim N\left[0 \ ; \sigma^2_\eta(1-\kappa^2)\right]$, and $\alpha_i \perp\!\!\!\perp \delta_i$, for $i = 1, \ldots, n$. Since the model has two random effects whose bivariate integral will have to be evaluated, specifying a bivariate normal distribution allows us to write the model as a function of two independent normally distributed random variables. Following Raymond et al. (2010), the marginalisation with respect to the random-effects can then easily be performed by two independent consecutive integrations. Furthermore, we re-specify the linear index functions to include the individual unobserved effects:

$$\begin{aligned} \mu_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i \\ \nu_{it} &= \mathbf{z}'_{it}\gamma + \kappa\frac{\sigma_\eta}{\sigma_\alpha}\alpha_i + \delta_i \end{aligned}$$

for $t = 1, \ldots, T$.

Under assumptions (DD), (IED′), (IED″) and (IS), and with the linear index expressions stated above, the joint density of $(y_{it}, d_{it} = 1)$ can be written using expressions (4)-(6); therefore, the likelihood function takes the form

$$\mathscr{L}_i(\boldsymbol{\psi}) = \int_{\mathbb{R}}\int_{\mathbb{R}} \prod_{t=1}^{T} \left[\Phi\left(c_\omega \nu_{it} + s_\omega(y_{it}-\mu_{it})/\sigma_\varepsilon\right)\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_{it}-\mu_{it}}{\sigma_\varepsilon}\right)\right]^{d_{it}} \times (1-P_{it})^{1-d_{it}} \\ \mathrm{d}\Phi\left(\frac{\alpha_i}{\sigma_\alpha}\right)\mathrm{d}\Phi\left(\frac{\delta_i}{\sigma_\eta\sqrt{1-\kappa^2}}\right) \tag{7}$$

The independence of $\alpha_i$ and $\delta_i$ makes it possible to evaluate the double integral sequentially, which in turn becomes a simple application of the Gauss-Hermite quadrature technique (Butler and Moffitt, 1982).

The availability of the time dimension also makes it possible to address the dynamic nature of the dependent variables so as to investigate the possibility that the migrant's remitting behaviour follows an intertemporal strategy, as argued in Bettin and Lucchetti (2016). In order to allow for state dependence in model (1)–(2), we modify Assumption (IS) to enlarge the information set of individual $i$ at time $t$ $\mathscr{F}_{it}$ to lags of the dependent variables, $\boldsymbol{y}_i^{t-1} = [y_{i1}, \ldots, y_{it-1}]$ and $\boldsymbol{d}_i^{t-1} = [d_{i1}, \ldots, d_{it-1}]$, together with the set of explanatory variables in $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$. In this case, the recursive nature of the joint density of $(y_{i1}, d_{i1})$ requires that the process is initialised, giving rise to the so-called "initial conditions" problem.

Therefore, accounting for a different conditioning set for the probability of the initial realisation $(y_{i1}, d_{i1})$, the joint density of $(\boldsymbol{y}_i, \boldsymbol{d}_i)$ is

$$f(\boldsymbol{y}_i, \boldsymbol{d}_i | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(y_{i1}, d_{i1} | \mathscr{F}_{i1} \alpha_i, \eta_i; \boldsymbol{\psi}) \times \prod_{t=2}^{T} f(y_{it}, d_{it} | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi}) \, h(\alpha_i, \eta_i) \mathrm{d}\alpha_i \mathrm{d}\eta_i$$

where $f(y_{it}, d_{it} | \mathscr{F}_{it}, \alpha_i, \eta_i; \boldsymbol{\psi})$ for $t = 1, \ldots, T$ is defined in (3). While the definition of $\mathscr{F}_{it}$ stated above is very general, we express the information set as to contain only the first lags of the dependent variables.

IS′ For $i = 1, \ldots, n$ and $t = 2, \ldots, T$, the information set is $\mathscr{F}_{it} = [y_{it-1}, d_{it-1}, \boldsymbol{X}_i]$ in (1), $\mathscr{F}_{it} = [y_{it-1}, d_{it-1}, \mathbf{Z}_i]$ in (2).

Under Assumption (IS′), we specify the linear index functions as

$$\mu_{it} = \phi_{11} y_{it-1} + \phi_{12} s_{it-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \alpha_i \tag{8}$$

$$\nu_{it} = \phi_{21} y_{it-1} + \phi_{22} s_{it-1} + \mathbf{z}_{it}' \gamma + \kappa \frac{\sigma_\eta}{\sigma_\alpha} \alpha_i + \delta_i \tag{9}$$

for $t = 2, \ldots, T$, where $\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$ are the state dependence parameters.

Finally, we deal with the conditional distribution of $(y_{i1}, d_{i1})$ following Heckman (1981), that is we specify two additional linearised reduced form equations with indices

$$\mu_{i1} = \mathbf{x}_{i1}' \boldsymbol{\pi} + \theta_1 \alpha_i + \theta_2 \delta_i \tag{10}$$

$$\nu_{i1} = \mathbf{z}_{i1}' \boldsymbol{\lambda} + \theta_3 \alpha_i + \theta_4 \delta_i \tag{11}$$

In the spirit of Heckman (1981), the linear index functions are merely an approximation of the conditional distribution of $(y_{i1}, d_{i1})$. Therefore, we allow for both the random effects to enter linearly each index multiplied by nuisance parameters. For the same reason, we leave the scale of $\varepsilon_{i1}$ unrestricted, so that $\mathrm{E}(\varepsilon_{i1}^2) = \theta_5$.

With Assumptions (DD), (IED′), (IED″), (IS) and expressions (8)-(11), the joint density of $(y_{it}, d_{it} = 1)$ can be written using expressions (4)-(6) and the likelihood function for individual $i$ can be written as in (7).

The sample selection and double hurdle models here proposed generalise the approach adopted by Raymond et al. (2010), by introducing a more general dynamic specification of the sample selection model, which allows lags of both dependent variables to appear either in the primary and the selection equations. In addition, Raymond et al. (2010) model initial conditions by parametrising the distribution of the random effects conditional on the initial realisations of the dependent variables as in Wooldridge (2005). However, Akay (2012) showed that, with short $T$, Heckman's estimator has superior finite sample properties. Our proposed models also represent a generalisation of other random-effects approaches so far brought forward by the econometric literature. Vella and Verbeek (1999) adopt a two-step estimation approach where they first derive estimates of the unobserved heterogeneity based on a random-effects estimation of the selection equation following Heckman (1981); this quantity is then used in the augmented primary equation to correct for the selection bias, estimated by OLS. However, they consider a model where the state dependence is included only in the selection equation. Recently, Semykina and Wooldridge (2013) proposed to perform the backward substitution for the lagged dependent variable in the main equation, so that the resulting equation of interest contains the lags of the explanatory variables and the initial realisation of the dependent variable.

Alternative estimation approaches to dynamic panel data sample selection models rely on differencing to remove the individual unobserved effects. Arellano et al. (1999) and Labeaga (1999) specified a sample selection model and double hurdle model, respectively, where the autoregressive specification is adopted only in the main equation. The two-step estimation strategy builds on Chamberlain (1984)'s specification of the conditional distribution of the unobserved effect for the selection equation. The estimation of the main equation parameters is carried out following Arellano and Bover (1995) and Bover and Arellano (1997). Similarly, Wooldridge (1995) developed a two-step fixed-effects estimator for testing and correcting for the presence of selection bias following the strategy of Chamberlain (1980). An extension was recently proposed by Semykina and Wooldridge (2010) to include endogenous explanatory variables along with a semi-parametric estimation strategy based on the two-step series estimator of Newey (2009). In the same line is the three-step semi-parametric series estimator of Gayle and Viauroux (2007) for the dynamic formulation of the sample selection models. Semi-parametric estimators of the static and dynamic sample selection model have also been developed by Kyriazidou (1997) and Kyriazidou (2001), where sample selectivity is eliminated by pair-wise comparison between similar observations, as in Powell (1987) and Ahn and Powell (1993): the parameters of the selection equation are estimated and then used to construct kernel weights to be used in the least squares/GMM estimation of the main equation's parameters.

## 4 Data

Our empirical analysis is based on data from the German Socio-Economic Panel (SOEP) for the period between 1996 and 2012[3]. SOEP is a representative longitudinal survey that includes yearly information on a large sample of households residing in Germany. Individual questionnaires are administered to each household member above 18 years together with a household-level one, which is usually answered by the head of the household. This allows for a perfect matching between information on demographic and socioeconomic individual characteristics and details on household composition and budget decisions for every person in the sample. Immigrant households were included in the sample from the first wave of the survey in 1984 but the nationality groups initially covered were only those with the longer tradition of immigration to Germany: Turks, Italians, Greeks, Spaniards and Yugoslavians[4]. The immigrant subsample was then significantly increased to include also other nationalities from 1995 onwards.

A detailed picture of the socio-economic conditions of relatives in the home country is missing in the SOEP dataset. Available information simply concerns the family structure, i.e. what relatives are still living abroad. This shortcoming might explain why, despite its longitudinal natural and the wide usage in the literature on migrants' assimilation and performance in the labour market, the SOEP has not been employed in many empirical contributions on remittance behaviour.[5] The sample used in the empirical analysis is restricted to the

[3]The data used in this paper was extracted using the Add-On package PanelWhiz for Stata. PanelWhiz was written by Dr. John P. Haisken-DeNew. See Haisken-DeNew and Hahn (2010) for details. The PanelWhiz generated Stata script to retrieve the data used here is available from us upon request. Any data or computational errors in this paper are our own.

[4]Formal guest workers programmes were implemented in West Germany during the 1950s and 1960s. Foreign workers were recruited from Southern Europe first (bilateral agreements with Italy and Greece were signed in 1955 and 1960, respectively), but soon from Turkey and former Yugoslavia as well. Immigrants who entered the SOEP in the 1980s indicated Yugoslavia as their home country. Aggregate data have been calculated as mean values for the group of current countries that were once enclosed in the Federal Republic.

[5]Merkle and Zimmermann (1992) look at the way migrants' remittance and saving behaviour is influenced by return intentions. Holst et al. (2008, 2010, 2011) investigate the links between gender, transnational networks, legal status and the remittance patterns while Bollard et al. (2011) include SOEP data in their cross-country study and investigate how remittance patterns change according to migrants' different educational levels. Bauer and Sinning (2011) analyse immigrants' savings behaviour while Sinning (2011) focuses on the differences in remitting strategy between permanent and temporary migrants. Similarly to Merkle and Zimmermann (1992), Dustmann and Mestres (2010) look at the way return plans affect the amount remitted but they also exploit the longitudinal nature of the survey in a dynamic panel setting. Bettin and Lucchetti (2016) investigate the issue of time persistence in the decision to remit by means of discrete choice dynamic

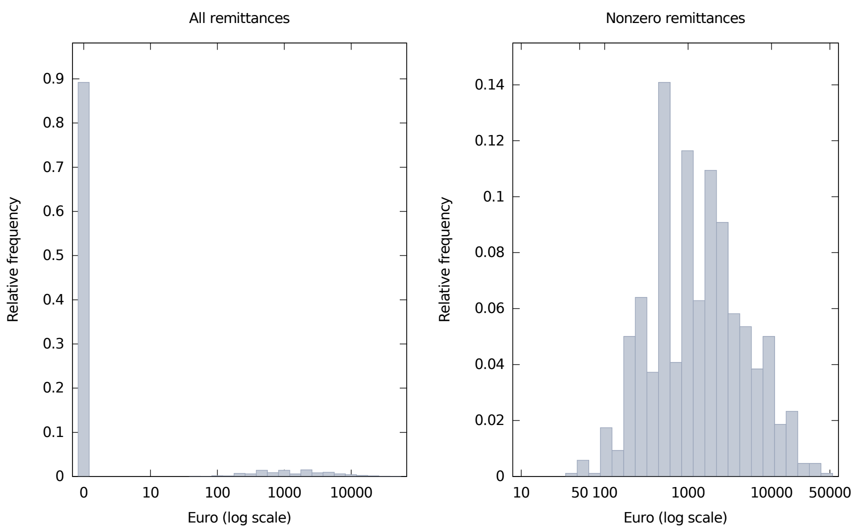
adult immigrant population. Immigrants are defined as foreign-born individuals who moved to Germany after 1948 and therefore include individuals who became German citizens after immigration while excluding second-generation immigrants (see also Bauer and Sinning (2011)).

All waves before 1996 were excluded due to the inconsistency in the questions on remittance behaviour before and after that date.

## 4.1 The amount remitted: definition and some descriptive figures

Information on remittances are collected in the individual questionnaire by asking the following question: "In the last year, that is, in ...., have you personally given money or financial support to relatives or other people outside this household? How much in the year as a whole? ". Specifically, individuals are asked about transfers to parents/parents–in–law, children/son–in–law/daughter–in–law, spouse/ divorced spouse, other relatives and non–relatives. In the definition of the amount remitted, our dependent variable in the main equation, we consider all remittances towards close and distant relatives and express them in natural logarithm[6]. In the selection equation, the dependent variable is equal to 1 when migrants send a positive amount $R$ in year $t$ and is equal to 0 when there are no transfers to any relative back home.

Figure 1: Distribution of amounts remitted

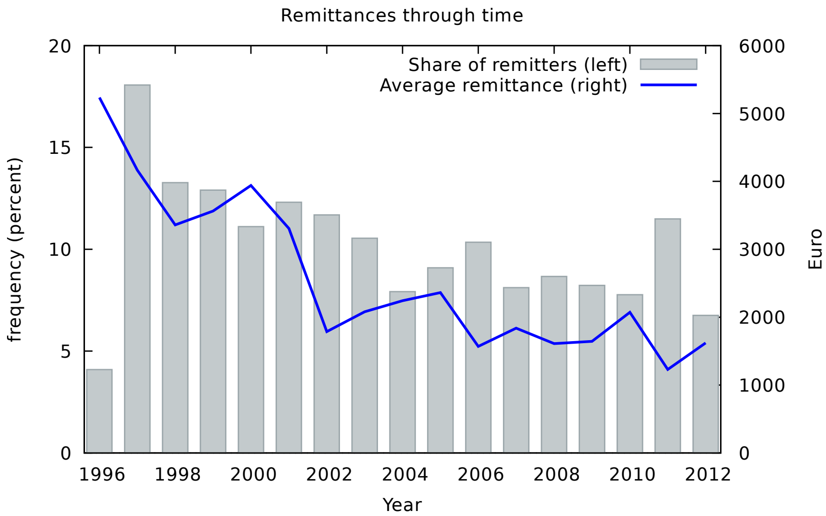



models.

[6]All financial variables (remittances, but also income) before 2002 have been expressed in Euro before taking natural logs.

On average, 11% of migrants in our sample remit (Figure 1). However, it is interesting to note that most non-zero remittances are relatively distant from zero. While of course the distribution displayed in the figure is a marginal, and not a conditional, one, it seems difficult to justify empirically the idea of non-remitting behaviour as the observable outcome of a corner solution at 0 in the migrant's optimisation process.

Figure 2: Share of remitters and average amount remitted by year, 1996-2012




The share of remitting migrants decreases over time, especially in the last years covered in our sample (see Figure 2) . This trend might be related to the consequences of the sovereign debt crisis that spread throughout Europe between 2010 and 2011, although we do not observe significant variations in migrants' income levels and employment outcomes over time. If we focus on the sample of remitters, the mean amount sent home is not constant over time; even adjusting our data for the Euro adoption from 2002 onwards, there seems to be a sensible gap in the size of transfers before and after that date. As a matter of fact, the highest value is registered in 1996, 4.276 Euros, while the lowest in 2011, 1269 Euros.

When looking at remitting behaviour by country of origin (Table 1), sizeable differences emerge. In general, the average share of remitters is higher among migrants from Asia and the Pacific region (33.63%) and from the Balcan region (21.13%). The lowest values (6-7%) are associated to Southern Europeans (Italians, Greeks, Spaniard, the traditional immigration groups in 1960s and 1970s) and other EU-15 or OECD citizens. It is worth noting, however, that these immigrant groups are the ones who send the larger amount of money, with an yearly

Table 1: Average share of remitters and amount remitted by country of origin

| Country | Share of remitters (%) | Mean amount (Euros) | Std. Dev. (Euros) |
|---|---|---|---|
| Turkey | 10.33 | 2581 | 3779 |
| Ita-Gre-Spa | 6.44 | 5129 | 7393 |
| Ex Yugoslavia | 21.13 | 2944 | 3507 |
| Other EU - OECD | 7.44 | 4360 | 9410 |
| New EU members | 13.67 | 1547 | 2464 |
| Ex USSR | 11.17 | 1304 | 2304 |
| Africa | 8.68 | 1991 | 2851 |
| Latin America | 12.18 | 2079 | 2318 |
| Asia-Pacific | 33.63 | 2585 | 2693 |

average remittance above 5100 Euros for individual from Southern Europe and around 4400 Euros for other EU-15 or OECD citizens. Migrants from the ex USSR countries send the lowest amounts (1300 Euros).

### 4.2 The explanatory variables

We include a common set of explanatory variables in both the main and the selection equation. This set includes those immigrants' personal characteristics usually considered in the literature as observable determinants of the decision to remit: gender (1 if male), age and time since migration[7], migrant household composition (number of adult members and number of children), educational level (years of education), intention to stay in Germany (1 for staying, 0 for going back to the home country), German citizenship (1 if acquired), migrants' individual yearly labour income and household net yearly income (both in natural logarithm) and their square terms, a time trend.

We then consider some additional country-level variables in both equations to proxy for the socio-economic conditions of the origin household in the home country that could affect remittance behaviour but are not covered in the SOEP survey. The ratio between per capita GDP[8] in the home country[9] and in Germany (in logs) is included to proxy for the living conditions of those left behind.

[7]Both variables enter the two equations with their value at the first sampling year.

[8]Data are drawn from the World Development Indicators database. GDP per capita is expressed in constant 2005 international dollars.

[9]During the interview, the home country was not chosen from a predefined list, but rather declared freely. For this reason, a non negligible share of individuals list as their home country a territorial entity that is not recognised as a sovereign state per se or no longer exists as such. As a consequence, data for Benelux are calculated as means between those for Belgium and the Netherlands. For Kurdistan and Ex-Yugoslavia we make use of data for Iraq and Serbia, respectively.

Its square is also added to control for possible nonlinear effects. In addition, we also include a set of "pseudo-country" dummies[10] to control for time-invariant factors, such as distance, which might exert an influence on the strength of the relationship with the family at home and therefore affect the decision to remit.

In order to identify the two decision mechanisms correctly, we need to define some exclusion restrictions. Such variables will enter the selection equation, thus affecting the choice whether to remit or not, but are supposed to have no direct effect on the amount remitted. Most of the exclusion restrictions previously employed in the literature relate to either information on recipient households that we are not able to exploit here[11] or to factors which cannot be disregarded *a priori* as determinants of the amount remitted[12].

Following Czaika and Spray (2013), we employ a dummy that takes value 1 if the individual is currently employed in the German labour market on the assumption that being employed (economically active) may affect the decision to remit, but not the size of individual remittances once we control for the migrant's individual and household income. In addition, we exploit the available information on the structure of the receiving household by including four dummy variables respectively for the presence of parents, children, partner and siblings in the home country.

# 5 Estimation results

The three most appropriate estimation alternatives that can be considered for remittance modelling must take into account the censored/truncated nature of the dependent variable.

## 5.1 Pooled models

In table 2 we show the results for the Tobit, sample selection and double hurdle models on the pooled dataset.[13]

[10]Countries of origin are grouped as follows: Turkey, Southern Europe (Italy, Greece, Spain), Ex Yugoslavia, other EU-OECD countries, new EU members, ex URSS, Africa, Latin America, Asia-Pacific.

[11]Hoddinott (1992); Gubert (2002); Amuedo-Dorantes and Pozo (2006) all use proxies for the location of the recipient family that provide an indirect measure of the fixed transaction costs associated with remitting fund.

[12]Hoddinott (1992); Aggarwal and Horowitz (2002) consider the years of absence from the home country, which are likely to have an impact also on the size of transfers as predicted by Poirine (1997).

[13]Estimation results for the linear models are not presented here but are available upon request.

Table 2: Pooled models

| | Tobit | | | Sample selection | | | Double Hurdle | | |
|---|---|---|---|---|---|---|---|---|---|
| | coeff | stderr | | coeff | stderr | | coeff | stderr | |
| Main eq. | | | | | | | | | |
| sex | 0.348 | (0.188) | * | 0.069 | (0.064) | | 0.053 | (0.067) | |
| age | 0.062 | (0.007) | *** | 0.011 | (0.003) | *** | 0.013 | (0.003) | *** |
| education yrs | 0.138 | (0.036) | *** | 0.007 | (0.014) | | 0.008 | (0.014) | |
| yrs since migration | -0.068 | (0.011) | *** | 0.003 | (0.004) | | 0.003 | (0.004) | |
| stay | -0.993 | (0.164) | *** | -0.063 | (0.058) | | -0.075 | (0.059) | |
| germ_nat | -0.355 | (0.266) | | -0.236 | (0.077) | *** | -0.262 | (0.077) | *** |
| n_adults | -0.783 | (0.106) | *** | -0.138 | (0.034) | *** | -0.148 | (0.036) | *** |
| n_children | -0.409 | (0.074) | *** | -0.048 | (0.027) | * | -0.049 | (0.029) | * |
| individual income | 1.543 | (0.189) | *** | -0.009 | (0.060) | | -0.021 | (0.062) | |
| individual income$^2$ | -0.246 | (0.055) | *** | 0.009 | (0.016) | | 0.013 | (0.017) | |
| household income | 3.833 | (0.690) | *** | 0.627 | (0.168) | *** | 0.654 | (0.171) | *** |
| household income$^2$ | -0.630 | (0.242) | *** | -0.032 | (0.038) | | -0.042 | (0.038) | |
| Per capita GDP differential | -2.066 | (0.926) | ** | 0.024 | (0.238) | | 0.028 | (0.247) | |
| Per capita GDP differential$^2$ | -0.344 | (0.228) | | -0.001 | (0.057) | | -0.004 | (0.060) | |
| time trend | -0.151 | (0.019) | *** | -0.022 | (0.007) | *** | -0.019 | (0.008) | ** |
| Selection eq. | | | | | | | | | |
| sex | | | | 0.069 | (0.048) | | 0.072 | (0.048) | |
| age | | | | 0.006 | (0.002) | *** | 0.006 | (0.002) | *** |
| education yrs | | | | 0.008 | (0.010) | | 0.008 | (0.010) | |
| yrs since migration | | | | 0.005 | (0.003) | * | 0.006 | (0.003) | * |
| stay | | | | -0.212 | (0.042) | *** | -0.210 | (0.042) | *** |
| germ_nat | | | | 0.197 | (0.065) | *** | 0.206 | (0.065) | *** |
| n_adults | | | | -0.086 | (0.025) | *** | -0.082 | (0.026) | *** |
| n_children | | | | -0.085 | (0.019) | *** | -0.085 | (0.019) | *** |
| individual income | | | | 0.200 | (0.070) | *** | 0.209 | (0.070) | *** |
| individual income$^2$ | | | | -0.023 | (0.017) | | -0.025 | (0.017) | |
| household income | | | | 0.653 | (0.177) | *** | 0.634 | (0.175) | *** |
| household income$^2$ | | | | -0.126 | (0.062) | ** | -0.121 | (0.061) | ** |
| Per capita GDP differential | | | | -0.525 | (0.238) | ** | -0.534 | (0.241) | ** |
| Per capita GDP differential$^2$ | | | | -0.098 | (0.058) | * | -0.099 | (0.059) | * |
| time trend | | | | -0.046 | (0.005) | *** | -0.047 | (0.005) | *** |
| partner_hc | | | | 0.842 | (0.138) | *** | 0.843 | (0.139) | *** |
| children_hc | | | | 1.753 | (0.102) | *** | 1.755 | (0.103) | *** |
| parents_hc | | | | 1.424 | (0.055) | *** | 1.432 | (0.056) | *** |
| employed | | | | 0.069 | (0.071) | | 0.062 | (0.071) | |
| $\rho$ | | | | -0.507 | | | -0.501 | (0.041) | *** |
| $\sigma_\epsilon$ | 4.586 | | | 1.111 | | | 1.111 | (0.023) | *** |
| Log-lik | | -16853.46 | | | -12686.96 | | | -12466.73 | |
| N. obs. | | 4149 | | | 4149 | | | 4149 | |

Models specifications include an intercept term and country fixed-effects as defined in section 4.

Standard errors are panel and heteroskedasticity robust.

Significance level: * 10%, ** 5%, *** 1%.

The three models considered follow from different assumption on the remittance behaviour. In the Tobit model, it is assumed that the probability mass on zero is due to a corner solution to the individual's optimisation problem: in this case, migrants do not transfer money back home because their income is too low to do so. There is only one decision mechanism that determines at the same time the choice to remit or not and the amount remitted.

Alternatively, we can consider the two decisions as separated from one other. By using the ordinary sample selection model, one implicitly assumes that migrants may attach non-zero utility to sending remittances, or not; those who do always remit, however small is the amount at their disposal. In the double hurdle model, conversely, one allows for a double form of censoring and consider that the zeros might derive either from a budget constraint (possibly including opportunity and transaction costs, usually unobservable) or from the absence of any utility gain related to remittances.

The estimation results reported in table 2 show that the correlation coefficient $\rho$ between the two error terms in equations (1)–(2) is negative and statistically significant. This result would suggest that a lower propensity to remit is associated with larger amounts. An even more relevant implication of this result would be that migrants' behaviour is better described by two non-independent processes, the first one governing the choice as to whether to remit and the second one the decision on how much to transfer, that are characterised by different determinants. Note that this finding does not entail, *per se*, any particular implication on the question whether migrants may have heterogeneous preferences or not.

It may well be that individuals who were observed not to remit were uninterested in remitting at the particular time when they were observed; this shortcoming is clearly attenuated the longer the period of observation is, and constitutes a major reason for preferring longitudinal datasets to cross-sectional one, when available.

As is well known, the literature on sample selection models has long recognised the necessity of having some exclusion restrictions between the selection equation and the main equation to strengthen identification of the model, which otherwise would rely on non-linearity only. In our case, the variables related to the household structure in the country of origin are all extremely significant and positively affect the probability to remit. Migrants' employment status, on the other hand, apparently fails to exert any effect in the selection equation in the pooled specification.

It is important to note that a major adverse consequence of adopting the Tobit model in a case like the present one lies not only in making an incorrect choice as to the decision process of the migrant. Failing to split the remittance decision between extensive and intensive margins also leads to some inevitable confusion in the interpretation of the results. In fact, it can be noted that some explanatory variables only affect the decision whether to remit or not, and that the set of covariates that seems to play a significant role in the amount remitted is much smaller, often with a contrasting sign.

Regarding the other determinants of remitting behaviour, the indications we got from either the sample selection or the double hurdle model are substantially similar. The size of the transfer seems to depend mainly on family-related variables: the larger the household in Germany (that is, the larger the number of adults and children), the lower the amount remitted. On the other hand, a higher household income is associated to larger remittances. Moving to individual characteristics, the only factor exerting a significant (and positive) effect in the main equation is the age of the migrant at the entrance in the sample, which may capture unobservable characteristics, such as individual ability, or migrants' working experience that might be associated to a higher capacity to remit. The time trend variable enters the main equation with a negative sign, suggesting that the size of the transfer decreases over time.

When moving the selection equation, the determinants previously discussed are all significant and with the same sign. In addition, the intention to stay in Germany exerts a strong negative effect on the extensive margin while individual income a positive one. There is also evidence of a non linear relationship between household income and the probability to remit, with a negative coefficient on the squared term. Non linearity characterises also the effect of the GDP differential between the home country and Germany, which is negative both in its linear and its squared term.

The effect of holding German citizenship is a bit more difficult to interpret since the sign of the associated coefficient changes between the two equations. Migrants who already became German citizens seem more likely to remit, but when they do so, they send lower amounts.

### 5.2 Static RE models

In Table 3 we present estimates that exploit the longitudinal nature of the data and model the individual time-invariant effect via a random-effect approach.

This is important because allowing for individual heterogeneity is crucial if we imagine that for some migrants the gains from remittances (either in terms of utility or in terms of future discounted income if remittances are a form of investment) is negligible. This is a conjecture that has been put forward in the literature to explain the high number of migrants for whom observed remittances are zero.

Table 3: Random-effects models

| | Sample selection | | | Double Hurdle | | |
|---|---|---|---|---|---|---|
| | coeff | stderr | | coeff | stderr | |
| Main eq. | | | | | | |
| sex | 0.070 | (0.058) | | 0.053 | (0.059) | |
| age | 0.007 | (0.003) | *** | 0.008 | (0.003) | *** |
| education yrs | 0.006 | (0.011) | | 0.007 | (0.012) | |
| yrs since migration | 0.005 | (0.004) | | 0.005 | (0.004) | |
| stay | -0.040 | (0.045) | | -0.050 | (0.045) | |
| germ_nat | -0.200 | (0.061) | *** | -0.211 | (0.063) | *** |
| n_adults | -0.090 | (0.027) | *** | -0.098 | (0.028) | *** |
| n_children | -0.061 | (0.023) | *** | -0.062 | (0.024) | ** |
| individual income | 0.050 | (0.052) | | 0.048 | (0.053) | |
| individual income$^2$ | 0.000 | (0.014) | | 0.001 | (0.015) | |
| household income | 0.427 | (0.122) | *** | 0.461 | (0.126) | *** |
| household income$^2$ | -0.013 | (0.026) | | -0.021 | (0.026) | |
| Per capita GDP differential | 0.103 | (0.205) | | 0.169 | (0.246) | |
| Per capita GDP differential$^2$ | 0.016 | (0.052) | | 0.027 | (0.061) | |
| time trend | -0.021 | (0.006) | *** | -0.020 | (0.006) | *** |
| Selection eq. | | | | | | |
| sex | 0.148 | (0.097) | | 0.159 | (0.093) | * |
| age | 0.010 | (0.004) | *** | 0.009 | (0.003) | *** |
| education yrs | 0.019 | (0.018) | | 0.021 | (0.018) | |
| yrs since migration | 0.015 | (0.005) | *** | 0.015 | (0.005) | *** |
| stay | -0.272 | (0.055) | *** | -0.269 | (0.055) | *** |
| germ_nat | 0.361 | (0.090) | *** | 0.364 | (0.092) | *** |
| n_adults | -0.066 | (0.033) | ** | -0.061 | (0.033) | * |
| n_children | -0.095 | (0.026) | *** | -0.097 | (0.027) | *** |
| individual income | 0.153 | (0.092) | * | 0.157 | (0.093) | * |
| individual income$^2$ | -0.013 | (0.021) | | -0.013 | (0.022) | |
| household income | 0.611 | (0.139) | *** | 0.576 | (0.140) | *** |
| household income$^2$ | -0.126 | (0.030) | *** | -0.119 | (0.030) | *** |
| Per capita GDP differential | -1.339 | (0.297) | *** | -1.407 | (0.307) | *** |
| Per capita GDP differential$^2$ | -0.326 | (0.070) | *** | -0.341 | (0.073) | *** |
| time trend | -0.042 | (0.007) | *** | -0.042 | (0.007) | *** |
| partner_hc | 1.617 | (0.699) | ** | 1.619 | (0.772) | ** |
| children_hc | 3.610 | (0.264) | *** | 3.654 | (0.250) | *** |
| parents_hc | 2.906 | (0.117) | *** | 2.928 | (0.120) | *** |
| employed | 0.198 | (0.103) | * | 0.183 | (0.104) | * |
| $\kappa$ | -0.216 | (0.060) | *** | -0.221 | (0.053) | *** |
| $\rho$ | -0.494 | (0.040) | *** | -0.488 | (0.040) | *** |
| $\sigma_\alpha$ | 0.738 | (0.023) | *** | 0.745 | (0.025) | *** |
| $\sigma_\eta$ | 1.433 | (0.083) | *** | 1.442 | (0.081) | *** |
| $\sigma_\epsilon$ | 0.776 | (0.017) | *** | 0.772 | (0.017) | *** |
| Log-lik | -10705.86 | | | -10520.41 | | |
| N. obs. | 4149 | | | 4149 | | |

Models specifications include an intercept term and country fixed-effects as defined in section 4.

Standard errors are panel and heteroskedasticity robust.

Significance level: * 10%, ** 5%, *** 1%.

Numerical integration by Gauss-Hermite quadrature method uses 18 grid points.

The results for static random-effects models confirm that the sample selection and the double hurdle models describe migrants' remittance behaviour appropriately, as suggested by the estimated values of the correlation coefficients, which are still significant. Here, $\kappa$ is the correlation between the time-invariant unobserved heterogeneity of the main and selection equations, whereas $\rho$ is the correlation between the time-varying error terms.

Evidence of strong, individual unobserved heterogeneity also emerges from the estimates of the standard deviations of the individual random effects in the two models, and in both the main and selection equations, corresponding to $\sigma_\alpha$ and $\sigma_\eta$, respectively. Even in the absence of a proper "poolability" test, as the reported $p$-value refers to the rejection of a null hypothesis on the frontier of the parameter space, the values of the estimated coefficients and standard errors are such that random-effects models can safely be preferred to specifications based on pooled cross-sections.

A quick comparison between Tables 2 and 3 reveals that estimated parameters for the amounts equation change to a lesser extent than those for the selection equations. This result is coherent with the well-known fact that neglecting to model unobserved heterogeneity provokes estimator inconsistency in binary dependent variable models, while the same does not necessarily apply to linear models, at least as long as the regressors are exogenous with respect to the individual effects $\alpha_i$.

With respect to table 2, the determinants of the amount remitted are substantially unchanged. Some differences instead are noticeable in the selection equation. Migrants' individual income and the length of stay in Germany gain explanatory power in the static random-effects models. Gender as well exerts an effect, albeit weak, on the probability to remit, but only in the double hurdle model.

When considering exclusion restrictions, all the variables related to the origin household are still positive and strongly significant. Moreover, migrants' employment status as well affects the probability to remit in the random-effects models.

### 5.3 Dynamic RE models

In analysing the results in Table 4, the first important result to take note of is that the coefficients for the lagged variables are highly significant; this holds for both equations in both specifications, although in the selection equations significance occurs only jointly but not individually. This confirms and extends the findings by Bettin and Lucchetti (2016) and strongly supports the hypothesis of intertemporal planning of the remittance strategy by migrants, and consequently, the im-

Table 4: Dynamic random-effects models

| | Sample selection | | | Double Hurdle | | |
|---|---|---|---|---|---|---|
| | coeff | stderr | | coeff | stderr | |
| Main eq. | | | | | | |
| $y_{t-1}$ | 0.073 | (0.027) | *** | 0.077 | 0.031 | ** |
| $d_{t-1}$ | -0.515 | (0.189) | *** | -0.239 | 0.101 | ** |
| sex | 0.109 | (0.073) | | 0.092 | (0.074) | |
| age | 0.011 | (0.003) | *** | 0.012 | (0.004) | *** |
| education yrs | 0.014 | (0.016) | | 0.013 | (0.016) | |
| yrs since migration | -0.003 | (0.005) | | -0.003 | (0.005) | |
| stay | -0.036 | (0.053) | | -0.051 | (0.053) | |
| germ_nat | -0.211 | (0.077) | *** | -0.224 | (0.077) | *** |
| n_adults | -0.069 | (0.033) | ** | -0.072 | (0.035) | ** |
| n_children | -0.088 | (0.029) | *** | -0.085 | (0.030) | *** |
| individual income | 0.056 | (0.064) | | 0.045 | (0.065) | |
| individual income$^2$ | -0.006 | (0.018) | | -0.002 | (0.018) | |
| household income | 0.515 | (0.148) | *** | 0.528 | (0.157) | *** |
| household income$^2$ | -0.043 | (0.030) | | -0.046 | (0.032) | |
| Per capita GDP differential | 0.222 | (0.222) | | 0.268 | (0.241) | |
| Per capita GDP differential$^2$ | 0.036 | (0.054) | | 0.043 | (0.058) | |
| time trend | -0.028 | (0.008) | *** | -0.026 | (0.008) | *** |
| Wald test for state dependence $\chi^2(2)$ | | | | | | |
| | 28.172 | | *** | 26.199 | | *** |
| Selection eq. | | | | | | |
| $y_{t-1}$ | 0.054 | (0.044) | | 0.072 | (0.048) | |
| $d_{t-1}$ | -0.046 | (0.286) | | 0.107 | (0.138) | |
| sex | 0.143 | (0.105) | | 0.162 | (0.112) | |
| age | 0.003 | (0.005) | | 0.0031 | (0.006) | |
| education yrs | 0.003 | (0.017) | | 0.007 | (0.017) | |
| yrs since migration | 0.024 | (0.007) | *** | 0.024 | (0.007) | *** |
| stay | -0.300 | (0.067) | *** | -0.301 | (0.068) | *** |
| germ_nat | 0.363 | (0.112) | *** | 0.374 | (0.113) | *** |
| n_adults | -0.028 | (0.052) | | -0.030 | (0.054) | |
| n_children | -0.112 | (0.037) | *** | -0.111 | (0.038) | *** |
| individual income | 0.124 | (0.117) | | 0.139 | (0.118) | |
| individual income$^2$ | -0.011 | (0.026) | | -0.014 | (0.026) | |
| household income | 0.234 | (0.182) | | 0.180 | (0.186) | |
| household income$^2$ | -0.035 | (0.034) | | -0.028 | (0.035) | |
| Per capita GDP differential | -1.313 | (0.356) | *** | -1.441 | (0.385) | *** |
| Per capita GDP differential$^2$ | -0.232 | (0.083) | *** | -0.258 | (0.090) | *** |
| time trend | -0.040 | (0.010) | *** | -0.038 | (0.011) | *** |
| partner_hc | 1.169 | (0.350) | *** | 1.154 | (0.365) | *** |
| children_hc | 3.725 | (0.255) | *** | 3.760 | (0.259) | *** |
| parents_hc | 3.127 | (0.177) | *** | 3.158 | (0.190) | *** |
| employed | 0.297 | (0.136) | ** | 0.269 | (0.140) | * |
| Wald test for state dependence $\chi^2(2)$ | | | | | | |
| | 7.448 | | ** | 6.348 | | ** |
| Wald test for state dependence $\chi^2(4)$ | | | | | | |
| | 37.273 | | *** | 34.563 | | *** |
| $\kappa$ | -0.206 | (0.057) | *** | -0.224 | (0.061) | *** |
| $\rho$ | -0.481 | (0.050) | *** | -0.478 | (0.050) | *** |
| $\sigma_\alpha$ | 0.740 | (0.035) | *** | 0.737 | (0.042) | *** |
| $\sigma_\eta$ | 1.592 | (0.117) | *** | 1.611 | (0.122) | *** |
| $\sigma_\epsilon$ | 0.732 | (0.018) | *** | 0.732 | (0.021) | *** |
| Log-lik | -8061.189 | | | -7877.978 | | |
| N. obs. | 4149 | | | 4149 | | |

Models specifications include an intercept term and country fixed-effects as defined in section 4. Standard errors are panel and heteroskedasticity robust. Significance level: * 10%, ** 5%, *** 1%. Numerical integration by Gauss-Hermite quadrature method uses 18 grid points.

portance of using longitudinal data sets to shed light on the actual mechanism of remittance behaviour.

As for the size of these coefficients, the issue is made considerably complex by the fact that the model described by equations (1)–(2) jointly with (8)–(9) can be thought of as a highly nonlinear VAR akin to the one examined by Koop et al. (1996) in its most general form. In this setting, it may be of interest to analyse the result of a thought experiment not unlike the reference above, or, alternatively, the sequence of dynamic multipliers, that is the effect over time of changes in the explanatory variables in the model. However, this is made very difficult by the intrinsically non-linear and non-differentiable nature of this particular dynamic system. Research is ongoing on this point; suffice it to say that, in contrast to usual dynamic multipliers analysis in linear models, (a) the sequence of multipliers depends on the assumed initial conditions and (b) it is doubtful whether the "effect" of an exogenous variable on the quantity of interest can be safely identified with the conditional mean.

Nevertheless, a qualitative comparison can be drawn with the results obtained by Bettin and Lucchetti (2016), who analysed a dynamic Probit specification for the remittance decision. In that paper it was observed that the sequence over time of remittance decisions by an individual displayed positive autocorrelation, as long as unobservable persistence factors were left unaccounted for. Once these were proxied by a first-order autocorrelation mechanism in the disturbance term $u_{it}$, state dependence turned out to be actually negative, with the positive persistence being explained by other latent factors. Our interpretation of the persistence pattern displayed in Table 4 is that at least some of those factors are captured by the inclusion of the intensive margin equation (1). However, we believe that more research is needed on this particular point before definitive statements can be made.

# 6 Conclusions

We discussed three main methodological issues in the empirical modelling of migrants' remittance behaviour. First, the migration literature has largely emphasised the need to account for the selection bias that arises from the two-sided nature of international transfers: the decision of whether or not to remit and the choice of the amount to send home. This makes sample selection and double hurdle the preferred econometric models in this literature. Secondly, the use of panel data to explore the time dimension requires the proper modelling of permanent unobserved individual effects. Finally, as a complement to the migrants' saving vs. consumption attitudes, remittance choices may be influenced by intertemporal strategies which, therefore, need to be translated into a dynamic panel data econometric model.

In order to perform a comprehensive analysis of the determinants of remittance behaviour, we develop dynamic sample selection and double hurdle models based on a general random-effects formulation that accounts for selection bias, unobserved heterogeneity, and state dependence. We argue that our proposed models offer several advantages in the field of remittance modelling with respect to other available approaches: all the information on the history of highly persistent remittance decisions is retained and we are able to strongly identify the impacts and, ultimately, the partial effects of migrants' characteristics with little time variation, that are often the focus of empirical analyses on remittance determinants.

The estimation results of our proposed models on the GSOEP also offer novel empirical evidence on the analysis of remittances with panel data. We find that neglecting unobserved heterogeneity alters substantially the magnitude of the effect of individual characteristics on the probability to send remittances. Moreover, migrants' remittances are strongly influenced by the past decisions and amounts. Interestingly, we find that the amount sent home in a given year is negatively influenced by the past decision to remit, whereas it exhibits a positive correlation with past remittances.

The formulation of our proposed models is such that they can be easily extended to embed a more detailed description of the migrants behaviour. For instance the assumption of exogeneity of explanatory variables can be relaxed and, following Bettin et al. (2012), we may allow for reverse causation between remittance amounts, income and consumption. In such a case, the modelling framework can easily accommodate additional first-step equations; alternatively, the extension to a correlated random-effects approach is straightforward. This further analysis is, however, left for future research.

## References


Aggarwal R, Horowitz AW. 2002. Are international remittances altruism or insurance? Evidence from Guyana using multiple-migrant households. *World Development* **30**: 2033–2044.

Ahn H, Powell JL. 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* **58**: 3–29.

Akay A. 2012. Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics* **27**: 1189–1204.

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

Arellano M, Bover O. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68**: 29–51.

Arellano M, Bover O, Labeaga JM. 1999. Autoregressive models with sample selectivity for panel data. In Hsiao C LL Lahiri K, MH P (eds.) *Analysis of Panels and Limited Dependent Variable Models*. Cambridge, UK: Cambridge University Press, 23–48.

Banerjee B. 1984. The probability, size and uses of remittances from urban to rural areas in india. *Journal of Development Economics* **16**: 293–311.

Bauer T, Sinning M. 2011. The savings behavior of temporary and permanent migrants in Germany. *Journal of Population Economics* **24**: 421–449.

Bernheim BD, Shleifer A, Summers LH. 1985. The strategic bequest motive. *Journal of Political Economy* **93**: 1045–76.

Bettin G, Lucchetti R. 2016. Steady streams and sudden bursts: persistence patterns in remittance decisions. *Journal of Population Economics* **29**: 263–292.

Bettin G, Lucchetti R, Zazzaro A. 2012. Endogeneity and sample selection in a model for remittances. *Journal of Development Economics* **99**: 370–384.

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

Bouyiour J, Miftah A. 2015. Why do migrants remit? testing hypotheses for the case of morocco. *Journal of Migration* **4**.

Bover O, Arellano M. 1997. Estimating dynamic limited dependent variable models from panel data. *investigaciones economicas* **21**: 141–165.

Brown R, Carling J, Fransen S, Siegel M. 2014a. Measuring remittances through surveys. *Demographic Research* **31**: 1243–1274.

Brown RP. 1997. Estimating remittance functions for Pacific Island migrants. *World Development* **25**: 613–626.

Brown RP, Jimenez-Soto E. 2015. Chapter 20 - migration and remittances. In Chiswick BR, Miller PW (eds.) *Handbook of the Economics of International Migration*, volume 1 of *Handbook of the Economics of International Migration*. North-Holland, 1077 – 1140.

Brown RP, Leeves G, Prayaga P. 2014b. Sharing norm pressures and community remittances: Evidence from a natural disaster in the pacific islands. *Journal of Development Studies* **50**: 383–398.

Butler JS, Moffitt R. 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50**: 761–64.

Chamberlain G. 1980. Analysis of covariance with qualitative data. *The Review of Economic Studies* : 225–238.

Chamberlain G. 1984. Panel data. In Z G, MD I (eds.) *Handbook of Econometrics*, volume 2. Amsterdam: North-Holland, 1248–1318.

Cox D. 1987. Motives for private income transfers. *Journal of Political Economy* **95**: 508–46.

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

Cragg JG. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39**: 829–844.

Czaika M, Spray J. 2013. Drivers and dynamics of internal and international remittances. *Journal of Development Studies* **49**: 1299–1315.

Davidson R, MacKinnon JG. 2004. *Econometric theory and methods*, volume 5. Oxford University Press New York.

de la Briere B, Sadoulet E, de Janvry A, Lambert S. 2002. The roles of destination, gender, and household composition in explaining remittances: an analysis for the dominican sierra. *Journal of Development Economics* **68**: 309–328.

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

Duval L, Wolff FC. 2010. Remittances matter: longitudinal evidence from Albania. *Post-Communist Economies* **22**: 73–97.

Duval L, Wolff FC. 2012. Longitudinal evidence on financial expectations in Albania: Do remittances matter? *Economics of Transition* **20**: 137–161.

Escanciano JC, Jacho-Chvez DT, Lewbel A. 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics* **178, Part 3**: 426 – 443.

Funkhouser E. 1995. Remittances from international migration: a comparison of El Salvador and Nicaragua. *The Review of Economics and Statistics* **77**: 137–146.

Gayle GL, Viauroux C. 2007. Root-n consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics* **141**: 179–212.

Gubert F. 2002. Do Migrants Insure Those who Stay Behind? Evidence from the Kayes Area (Western Mali). *Oxford Development Studies* **30**: 267–287.

Haisken-DeNew JP, Hahn MH. 2010. Panelwhiz: Efficient data extraction of complex panel data sets - an example using the German soep. *Journal of Applied Social Science Studies* **130**: 643–654.

Heckman JJ. 1974. Shadow prices, market wages, and labor supply. *Econometrica* **42**: 679–694.

Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* **47**: 153–161.

Heckman JJ. 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discreted data stochastic process. *Structural Analysis of Discrete Data with Econometric Applications* **MIT Press: Cambridge MA**. Manski CF, McFadden (eds).

Hoddinott J. 1992. Modelling remittance flows in Kenya. *Journal of African Economies* **1**: 206–232.

Hoddinott J. 1994. A model of migration and remittances applied to Western Kenya. *Oxford Economic Papers* **46**: 459–476.

Holst E, Schäfer A, Schrooten M. 2008. Gender, migration, remittances: Evidence from Germany. Discussion Papers of DIW Berlin 800, DIW Berlin, German Institute for Economic Research.

Holst E, Schäfer A, Schrooten M. 2010. Gender, transnational networks and remittances: Evidence from Germany. Discussion Papers of DIW Berlin 1005, DIW Berlin, German Institute for Economic Research.

Holst E, Schäfer A, Schrooten M. 2011. Remittances and gender: Theoretical considerations and empirical evidence. IZA Discussion Papers 5472, Institute for the Study of Labor (IZA).

Holst E, Schäfer A, Schrooten M. 2012. Gender and Remittances: Evidence from Germany. *Feminist Economics* **18**: 201–229.

Koop G, Pesaran M, Potter S. 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* **74**: 119–147.

Kyriazidou E. 1997. Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society* : 1335–1364.

Kyriazidou E. 2001. Estimation of dynamic panel data sample selection models. *The Review of Economic Studies* **68**: 543–572.

Labeaga JM. 1999. A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics* **93**: 49 – 72.

Lucas RE, Stark O. 1985. Motivations to remit: evidence from Botswana. *Journal of Political Economy* **93**: 901–918.

Merkle L, Zimmermann KF. 1992. Savings, remittances, and return migration. *Economics Letters* **38**: 77–81.

Newey WK. 2009. Two-step series estimation of sample selection models. *Econometrics Journal* **12**: 217–229.

Pigini C. 2015. Bivariate non-normality in the sample selection model. *Journal of Econometric Methods* **4**: 123–144.

Poirine B. 1997. A theory of remittances as an implicit family loan arrangement. *World Development* **25**: 589–611.

Powell JL. 1987. Semiparametric estimation of bivariate latent variable models. Working Paper 8704, Social Systems Research Institute, University of Wisconsin-Madison.

Rapoport H, Docquier F. 2006. The economics of migrants' remittances. In Kolm S, Mercier Ythier J (eds.) *Handbook on the Economics of Giving, Altruism and Reciprocity*, volume 2. Elsevier, 1135–1198.

Raymond W, Mohnen P, Palm F, Van Der Loeff SS. 2010. Persistence of innovation in dutch manufacturing: Is it spurious? *The Review of Economics and Statistics* **92**: 495–504.

Rosenzweig MR. 1988. Risk, implicit contracts and the family in rural areas of low-income countries. *Economic Journal* **98**: 1148–70.

Schwiebert J. 2015. Evidence on copula-based double-hurdle models with flexible margins. *Empirical Economics* : 1–45.

Semykina A, Wooldridge JM. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* **157**: 375–380.

Semykina A, Wooldridge JM. 2013. Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics* **28**: 47–61.

Sinning M. 2011. Determinants of savings and remittances: empirical evidence from immigrants to Germany. *Review of Economics of the Household* **9**: 45–67.

Tobin J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* **26**: 24–36.

Vella F, Verbeek M. 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics* **90**: 239–263.

Wooldridge JM. 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of econometrics* **68**: 115–132.

Wooldridge JM. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics* **20**: 39–54.

Yang D. 2008. Coping with disaster: the impact of hurricanes on international financial flows, 1970-2002. *The B.E. Journal of Economic Analysis & Policy* **8**: Article 13.

Yang D. 2011. Migrant Remittances. *Journal of Economic Perspectives* **25**: 129–52.

Yang D, Choi H. 2007. Are remittances insurance? Evidence from rainfall shocks in the Philippines. *The World Bank Economic Review* **21**: 219–248.