IZA

DISCUSSION PAPER SERIES

IZA DP No. 4527

Illegal Migration, Wages, and Remittances: Semi-Parametric Estimation of Illegality Effects

Christian Schluter
Jackline Wahba

October 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Illegal Migration, Wages, and Remittances: Semi-Parametric Estimation of Illegality Effects

Christian Schluter
University of Southampton

Jackline Wahba
University of Southampton and IZA

Discussion Paper No. 4527
October 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4527
October 2009

ABSTRACT

Illegal Migration, Wages, and Remittances: Semi-Parametric Estimation of Illegality Effects*

We consider the issue of illegal migration from Mexico to the US, and examine whether the lack of legal status causally impacts on outcomes, specifically wages and remitting behavior. These outcomes are of particular interest given the extent of legal and illegal migration, and the resulting financial flows. We formalize this question and highlight the principal empirical problem using a potential outcome framework with endogenous selection. The selection bias is captured by a control function, which is estimated non-parametrically. The framework for remitting is extended to allow for endogenous regressors (e.g. wages). We propose a new re-parametrisation of the control function, which is linear in case of a normal error structure, and test linearity. Using Mexican Migration project data, we find considerable and robust illegality effects on wages, the penalty being about 12% in the 1980s and 22% in the 1990s. For the latter period, the selection bias is not created by a normal error structure; wrongly imposing normality overestimates the illegality effect on wages by 50%, while wrongly ignoring selection leads to a 50% underestimate. In contrast to these wage penalties, legal status appears to have mixed effects on remitting behavior.

JEL Classification: J61, J30, J40

Keywords: illegal migration, illegality effects, counterfactuals, selection, control functions, non-parametric estimation, intermediate outcomes, Mexican Migration Project

Corresponding author:

Christian Schluter
Department of Economics
University of Southampton
Highfield
Southampton, SO17 1BJ
United Kingdom
E-mail: C.Schluter@soton.ac.uk

* We are grateful for comments from David Card, Deborah Cobb-Clark, Christian Dustmann, and Jean-Yves Pitarakis as well as seminar participants at EEA/ESEM 2006 and Target 2006, as well as Manchester, Southampton, and St. Andrews.

1 Introduction

Almost 3 percent of the world population are labor migrants. A substantial proportion of this worldwide labor migration is illegal, defined as unauthorized border crossing, or entry by tourist visa and subsequent illegal work. For example[1], in 2005 between 10-15 percent of all migrants are thought to be illegal. In terms of absolute numbers, the USA hosts an estimated 9 million irregular migrants, and in the UK around 0.5 out of 4.3 million migrants are deemed illegal. The annual inflow of illegal migrants into the EU is estimated to equal about half a million. This scale has resulted in huge public concern, and governments spending vast resources on seeking to stem the tide of illegal migration.

This extent of illegal migration presents a challenge for both policy makers and researchers. Despite the size of illegal migration little is known about illegal migrants, their characteristics and behavior. This is perhaps not surprising since illegal migration is an unlawful, criminal activity. Standard data sets simply do not record legal status.

This paper is about the effect of the lack of legal status on migrations' outcomes in the host country. Using a unique data set which records the legal status of migrants we focus on illegal migrants directly and consider, in particular, wages earned in the host country and remitting behavior. The importance of wages is obvious. The relevance of remitting behavior follows from the observation that the huge flow of labour migration has lead to huge remittance flows (World Bank, 2006, see also Rapoport and Docquier, 2006). In 2003 recorded remittances amounted to almost US$ 116 billion or 1.7 percent of the recipient developing countries' GDP, the largest share (29 percent) being remitted from the US. For many LDCs remittances now surpass overseas official development aid, are more stable than foreign direct investment, and instrumental in reducing poverty. Some commentators suggest that informal remittances amount to about 35–75 percent of official remittances to developing countries. Thus the impact of illegal migration on remittances, which has not been studied before, is important for policymakers. We examine these general issues in the particular context of Mexican migration to the US.

More specifically, we quantify the causal impact of the lack of legal status on the host country outcomes wages and remitting, referred to below as "illegality effects". We make both substantive empirical contributions as well as some methodological contributions. Empirically, we are the first to define formally illegality effects using a framework which highlights the principal empirical problem, viz. identification of the causal effect in the presence of endogenous selection. Work which has considered illegality effects on wages has either ignored selection altogether, or has not addressed this issue convincingly. No work has considered illegality effects on remitting behavior, despite the huge importance of remittances.

Methodologically, we apply the potential outcome framework with endogenous selection in which the selection bias is captured using a non-parametrically

[1] For estimates see e.g. IOM (2005), ILO (2004), and Woodbridge (2005).

estimated control function. While this framework has been examined before (Pinkse, 2000, Blundell and Powell, 2004 and Heckman and Navarro, 2004), we propose a new re-parametrisation of the control function. This shape of this control function is of considerable interest for empirical work. The leading empirical parametric modelling of endogenous selection assumes normality of error terms which leads to a linear control function. Hence we test for linearity formally. It turns out that in our empirical application for the 1990s data we find considerable non-linearities. It is therefore of interest to quantify the misspecification bias of the normal model, and we find the illegality effect on wages is overestimated by 50%, while wrongly ignoring selection leads to a 50% underestimate. The model is estimated using penalized regression splines, and inference methods take into account the generated regressor problems as well as potential heteroscedasticity. To improve asymptotic efficiency we also propose an alternative estimator which uses iteratively a weighting scheme based on a variance regression. In our analysis of illegality effects on remitting behavior we extend the potential outcome framework with endogenous selection to allow for endogenous regressors since remitting behavior is a function of wages, which in turn are affected by legal status.

Whether causal illegality effects are present is an empirical question as economic theory does not give a unique prediction. Consider wages earned in the host country. In a wage bargaining setting, illegal migrants typically have lower bargaining power than legal migrants. Legal migrants can search freely the labour market, whereas illegal migrants cannot. In addition, illegal migrants could have lower reservation wages than legal migrants, in which case standard search or Roy models predict lower average wages. Illegal and legal migrants might face segmented labour markets, with illegal migrants taking on different jobs. On the other hand, the main driver of differences in outcomes between legal and illegal migrants could be differences in observable characteristics such as human capital rather than illegality effects. As regards remitting behavior, consider both the incidence of remitting and the amount remitted conditional on remitting. Illegality might have no direct effect on the amount remitted, apart from the indirect effect induced by the wage penalty caused by the lack of legal status. We control for this potentially confounding factor. On the other hand, illegality might give rise to a new "exchange motive" for remitting, in that illegal migrants might have a larger preference for remitting. Alternatively, illegality might reflect a greater need of the source household (and we control for selection on such unobservables). Finally, the illegality effect might be negative if illegal migrants retain more money to reflect the increased precariousness of their situation as a result of intensified efforts by host countries (such as the US) to detect illegal immigrants.

Our specific empirical setting is Mexican migration to the US. Using data from the Mexican Migration project (MMP), we find considerable and robust illegality effects on wages, the penalty being about 12% in the 1980s and 22 % in the 1990s. By contrast legal status appears to have little effect on the incidence of remitting, but does affect negatively the amounts remitted in the 1990s. These results support the hypothesis that wage penalties arise from limited job search and weaker bargaining power. As the US government increased its effort in detecting illegal migrants this bargaining power eroded further, while reduced remittances reflect increased precariousness of the illegal migrant's situation.

The plan of the paper is as follows. In Section 2 we consider illegality effects on wages earned in the host country. We present the potential outcome model with endogenous selection, and our new re-parametrisation of the control function which captures the selection bias. Formal consideration of remitting behavior is postponed to Section 5 since we need to extend the estimation framework to allow for endogenous regressors, as remitting depends on wages. Estimation is carried out using penalized regression splines, and we briefly summarize the estimation method, distributional theory, and methods of inference. Details are collected in Appendix A. Section 3 describes the data, and the empirical analysis of illegality effects on wages is carried out in Section 4. We focus first on the shape of the control function, the estimated illegality effects, and the extent of the misspecification bias of illegality effects were one to ignore selection or impose joint normality of errors. Section 4.4 carries out two robustness checks. Remitting behavior is examined in Section 5, which contains both estimation framework and empirics. The final section concludes.

1.1 Illegality Effects, Mexican Migration, and the Literature

The leading case in the empirical illegality literature is Mexican migration to the US, the reported statistics being truly impressive. More than 1.6 million illegal border crossings have been intercepted by border guards in 2000 (INS), one third of Mexican residents in the USA are believed to be unauthorized, and Passel (2002) estimates that 80% of all Mexican immigrants who arrived in the 1990s were unauthorized. Illustrating the importance of remittances, the World Bank estimates that Mexican workers abroad remitted about 2% of the country's GNI in 2003. Our empirical analysis is also placed in this geographical setting.

While much is known about legal Mexican migration to the US, the empirical literature on illegality effects is thin. Rivera-Batiz (1999) seeks to identify illegality effects through the amnesty of 1986 (the Immigration Reform and Control Act). Using the 1990 Legalized Population Survey, he finds a 52% wage gap unexplainable by differences in observable characteristics. The analysis is based on Blinder-Oaxaca wage decompositions and ignores potential selection effects. In our analysis we find important selection biases for our 1990s data. By contrast, Kossoudji and Cobb-Clark (2002) address the selection issue using a difference estimator. They also use the Legalized Population Survey to consider illegal migrants, and construct a comparison group of Latino men from NLSY data. Hanson (2006) criticizes this strategy and observes that the comparison group of Latino men (i.e. mainly US citizens of Hispanic descent who were born and educated in the US) does not capture the required counterfactual, i.e. the average wage outcomes of migrants who migrated illegally had they migrated legally instead. The illegality effect is identified by a difference-in-difference if the systematic difference between the two treatment and comparison group, the selection bias, is time invariant and thus differences out. It is open to debate whether this parallel trends assumption holds at a time of many changes in migration policy. The reported wage penalty ranges from 14% to 24%. We obtain similar estimates, but employ a completely different and much less restrictive identification strategy. Neither Rivera-Batiz (1999) nor Kossoudji and Cobb-Clark (2002) consider the role of networks, whereas we do. To the best of our knowledge, no one has examined illegality effects on remittances. See also Hanson (2006) for other aspects of illegal Mexican migration.

We estimate illegality effects based on the Mexican Migration project (MMP). This is one of the few well-tested data sets which records the legal status of migrants. Work by economists based on MMP data has considered, among other issues, measuring network effects (e.g. Munshi, 2003), the effects of border enforcement and people smuggler prices on demand (e.g. Angelucci, 2005), and the estimation of hazard rate models of illegal migration (Orrenius and Zavodny, 2005).

2 Estimating Illegality Effects

Does the mere lack of legal documents have a causal effect on outcomes? For expositional clarity we consider first only final outcomes such as wages earned in the host country. The case of remitting behavior is more involved since it is an outcome which depends on an intermediate outcome (wages), both of which are potentially affected by legal status. We postpone the discussion of this case to Section 5 where we extend the framework to include endogenous regressors.

We use a potential outcome framework to rigorously define illegality effects. We then add more structure to the potential outcomes, formalize the endogenous selection problem, and capture the selection bias using a non-parametrically estimated control function.[2] Similar frameworks have been considered by e.g. Pinkse (2000), Blundell and Powell (2004) and Heckman and Navarro (2004). We add to these insights by proposing an insightful re-parametrisation of the control function. The empirical literature typically either ignores selection or imposes linearity on our control function via a normality assumption. Hence we will focus on testing the shape of our control function, and in the empirical application we will quantify the misspecification biases in terms of the associated estimated illegality effects.

[2] We follow the literature and ignore additional selection into work in the host country. This is justifiable in our empirical setting of Mexican migration to the US, since almost all Mexican labor migrants work in the US. As in the literature we also ignore selection into migration. However, by conditioning on illegal status (I) we do capture common unobservables that also affect the migration decision. We believe that the most likely barrier to costly migration is credit constraints, and that its incidence should be uncorrelated with the error terms in the wage equations conditional on I.

2.1 Potential Outcomes and Endogenous Selection

We formalize the illegality effect using a potential outcome model. Let $I_i = 1\,(i \text{ is illegal})$ denote the indicator of whether migrant i is illegal (equal to 1 if i is illegal and zero otherwise).[3] For outcomes, denote by Y_{1i} the *potential* outcome of migrant i if he migrates illegally, and by Y_{0i} if he migrates legally. Y_{0i} is the counterfactual outcome for $I_i = 1$, and the *observed* outcome is $Y_i = Y_{1i}I_i + Y_{0i}(1 - I_i)$. The idiosyncratic loss from migrating illegally is $Y_{1i} - Y_{0i}$, and the average illegality effect, our principal object of interest, is

$$\delta \equiv E\{Y_{1i} - Y_{0i}|I_i = 1\}. \tag{1}$$

The potential outcomes (Y_{0i}, Y_{1i}) and I_i might not be independent, so selection might be endogenous, which gives rise to the principal estimation problem. Such dependence could stem from unobserved ability or motivation and willingness to endure hardship as illegal migration is typically physically and mentally taxing, or from unobserved network effects.

We add more structure to the potential outcomes and the selection indicator I. In particular, let

$$\begin{aligned} Y_1 &= \mu_1 + W_1 \\ Y_0 &= \mu_0 + W_0. \end{aligned}$$

We have dropped the migrant subscript i for notational simplicity. W_I with $I \in \{0, 1\}$ denotes idiosyncratic deviations from the population averages with $E\{W_I\} = 0$. These deviations are allowed to vary with observable characteristics X, and we assume that $W_I = g(X) + V_I$ for some smooth scalar-valued function g with $E\{V_I\} = 0$, $Var\{V_I\} = \sigma_I^2$ and $E\{g(X)\} = 0$. $g(X)$ is a common effect.

Next, consider the event $I = 1$. Determination of illegal status is modelled by the reduced form[4] $I = 1\,(\mu_U(Z) + U > 0)$ where $\mu_U(Z)$ is some smooth scalar-valued function of observables Z. U is the individual's variation about the population value $\mu_U(Z)$ with distribution function F, $E\{U\} = 0$ and $Var\{U\}$ normalized to 1 for identification. Let $\overline{F}(x) = 1 - F(x)$. The propensity score of the event $I = 1$ is denoted by

$$p(Z) \equiv \Pr\{I = 1|Z\} = \overline{F}(-\mu_U(Z)). \tag{2}$$

[3] In our data, illegality is self-reported. Potential misclassification error is more a theoretical than practical concern because data collectors are believed to have established good relationships with interview subjects which has resulted in high quality data. Interviewees have no incentives to lie about their legal status. In particular, the incidence of illegality in our data is consistent with external estimates (e.g. Passel, 2002); other covariates have been similarly validated (see Section 3). We have therefore not incorporated potential misclassification error into our estimation framework.

[4] We are agnostic about the underlying structural model giving rise to this reduced form. One possibility is a choice framework based on a Roy model with individual legal status specific migration costs. Alternatively, the reduced form could describe outcomes of institutional processes.

Z is assumed to contain some observables, say Z_2, which are excluded from the outcome equations and thus are not part of X, $Z \equiv [X, Z_2]$, and the propensity score is assumed to be a non-trivial function of these Z_2. We discuss the specific exclusion restrictions in the empirical part of the paper. We also assume that (U, V_0, V_1) are independent (or at least mean-independent) of X and Z, and that $0 < p(z) < 1$.

(V_0, V_1) and U might be dependent, inducing the selection problem. We capture the resulting biases explicitly by control functions. Define

$$\begin{aligned} K_1(Z) &\equiv E\{V_1|Z, I=1\} = E\{V_1|U > -\mu_U(Z)\} \\ K_0(Z) &\equiv E\{V_0|Z, I=0\} = E\{V_0|U < -\mu_U(Z)\}. \end{aligned} \tag{3}$$

The illegality effect in this extended framework becomes

$$\begin{aligned} \delta &= [\mu_1 - \mu_0] + E\{V_1 - V_0|I=1\} \\ &= [\mu_1 - \mu_0] + E_{Z|I=1}\left[K_1(Z) + \frac{1-p(Z)}{p(Z)}K_0(Z)\right]. \end{aligned} \tag{4}$$

The observable outcomes are $Y = Y_0 + (Y_1 - Y_0) I$, therefore

$$\begin{aligned} Y &= \mu_0 + g(X) + V_0 + [\mu_1 - \mu_0 + V_1 - V_0] I \\ &= \mu_0 + g(X) + [\mu_1 - \mu_0] I + K(I, Z) + \varepsilon, \end{aligned} \tag{5}$$

with composite control function $K(I, Z) = K_0(Z) + [K_1(Z) - K_0(Z)] I$ and $E\{\varepsilon|X, Z, I\} = 0$ where $\varepsilon = (1-I)[V_0 - K_0(Z)] + I[V_1 - K_1(Z)]$. This is a non-parametric version of Heckman's (1979) classic control function approach.

2.1.1 Identification

Do the observable data identify this illegality effect ? The identification issues in models of this type are well know (see e.g. Imbens and Angrist, 1994). The control function is identified up to a constant because of the assumed exclusion restrictions concerning Z_2. The illegality effect is identified in the special cases when the error terms (U, V_0, V_1) are independent, or when $V_0 = V_1$ so that all individual effects are constant, $Y_1 - Y_0 = \mu_1 - \mu_0$. The identification of $\mu_1 - \mu_0$ and thus of δ in the general case requires a further restriction. We follow common practice and assume 'identification at infinity', i.e. the existence of two limits sets for the covariates such that the propensity score converges to 0 on one set and to 1 on the other. In addition, our re-parametrisation of the control function, described next, requires that the distribution function of U is strictly monotonic.

2.2 Re-Parametrising the Control Function

We propose a new insightful re-parametrisation of the control function. The leading case in the applied literature which does seek to control for endogenous selection assumes joint normality of errors. In this case our re-parametrised control function is shown below to be linear.

Following Das et al. (2003), we impose a restriction on the joint distribution of (U, V_0, V_1) such that the individual control functions are smooth functions only of the propensity score: $K_0(Z) = g_0(p(Z))$ and $K_1(Z) = g_1(p(Z))$ for some smooth functions g_I. Given the independence assumption above, this property follows if the distribution function of U, F, is strictly monotonic, since then $K_1(Z) = E\{V_1 | U > -\mu_U(Z)\} = E\{V_1 | W < p(Z)\}$ with $W = \overline{F}(U)$ and $K_0(Z) = E\{V_0 | W > p(Z)\}$.

Next, we seek a further transformation which relates to the Normal model. If (U, V_I) follows a joint Normal distribution with correlation coefficient ρ_I and $Var\{V_I\} = \sigma_I^2$, the resulting control functions are $K_1(Z) = \rho_1 \sigma_1^2 \lambda(-\mu_U(Z))$, and $K_0(Z) = \rho_0 \sigma_0^2 \lambda(-\mu_U(Z)) \times (-p(Z)) / (1 - p(Z))$ where λ denotes the inverse Mill's ratio $\lambda(-x) = \phi(x) / \Phi(x)$, and ϕ and Φ denote the Gaussian density and distribution. The control functions are linear in λ, the propensity score is $p = \Phi(\mu_U(Z))$, and we have $\lambda(-\mu_U(Z)) = p^{-1}\phi\left(\Phi^{-1}(p)\right)$.

Returning to the general case, define the strictly decreasing transformation $t(x) = \phi\left(\Phi^{-1}(x)\right)/x$ with $t(p) = \lambda$. The event $W < p$ is equivalent to the event $t(W) > \lambda$, and the control function K_1 can therefore be written as smooth function of λ: $K_1(Z) = f_1(\lambda)$. In order to re-parametrise the control function K_0, we follow application of the transform t by multiplying by the strictly decreasing transform $-p/(1-p)$, to arrive at $K_0(Z) = f_0\left(-\frac{p}{1-p}\lambda\right)$. The composite control function is thus $K(I, Z) = (1 - I) f_0\left(-\frac{p}{1-p}\lambda\right) + I f_1(\lambda)$.

A final re-parametrisation internalizes the switching between f_0 and f_1, and thus achieves a compact formulation as follows. Define the hazard $h(I, Z)$ by

$$h(I, Z) = \begin{cases} \lambda(-\mu_U(Z)) & \text{if } I = 1 \\ -\frac{p(Z)}{1-p(Z)}\lambda(-\mu_U(Z)) & \text{if } I = 0 \end{cases} \tag{6}$$

with $h > 0$ for $I = 1$ and $h < 0$ for $I = 0$. For notational convenience we suppress the dependence of h on I and Z, as we did for the dependence of p and λ on Z. The composite control function $K(I, Z)$ can then be written as a function of h,

$$K(I, Z) = f(h) \tag{7}$$

with $h = h(I, Z)$. The assumption that $E\{V_0\} = E\{V_1\} = 0$ implies that f is continuous at 0 with $f(0) = 0$.[5] We finally assume that f is a smooth function everywhere, i.e. twice continuously differentiable, except for a possible kink at 0.

[5] For $I = 1$ we have $h = \lambda \to 0^+$, i.e. $p \to 1$ and the covariates $Z \to Z_{L,1}$ are such that $K_1(Z_{L,1}) = E\{V_1 | W < 1\} = E\{V_1\} = 0$. Similarly, for $I = 0$, $h \to 0^-$ is equivalent to $p \to 0$ the covariates $Z \to Z_{L,0}$ are such that $K_0(Z_{L,0}) = E\{V_0 | W > 0\} = E\{V_0\} = 0$.

2.2.1 A Special Case: Joint Normality of Errors

We show that f is linear if the errors follow a normal distribution.[6] Assume then first that (U, V_0, V_1) are jointly normal, and further that $V_0 = V_1$. Then f is linear throughout with

$$f(h) = \rho_{U,V}\sigma_V \times h, \tag{8}$$

where $\rho_{A,B}$ is the correlation coefficient between A and B, and σ_j^2 the variance of V_j . This is the leading case in the applied literature (e.g. Maddala, 1983). With $V_0 = V_1$ it follows then that the illegality effect satisfies $\delta = \mu_1 - \mu_0$.

This setting can be generalized by allowing the error terms in the outcome equation to differ, $V_0 \neq V_1$. We have

$$f(h) = \left[\rho_{U,V_1}\sigma_1 1(h \geq 0) + \rho_{U,V_0}\sigma_0 1(h < 0)\right] \times h \tag{9}$$

The unobservable idiosyncratic expected losses then are $E\{V_1 - V_0|Z, I = 1\} = p(Z)^{-1} \times \phi\left(\Phi^{-1}(p(Z))\right)[\rho_{U,V_1}\sigma_1 - \rho_{U,V_0}\sigma_0]$.

2.3 The Estimation Equation

Returning to the estimation equation (5), writing $K(I, Z) = f(h)$, the principal estimation equation is

$$Y = \mu_0 + g(X) + [\mu_1 - \mu_0] I + f\left(\widehat{h}\right) + \theta, \tag{10}$$

with $\theta = \varepsilon + f(h) - f\left(\widehat{h}\right)$ satisfying $E\{\theta|X, Z, I\} = 0$.[7] We have taken explicitly into account that h will have to be generated from a first-stage estimation. $\mu_1 - \mu_0$ is estimated without bias by the coefficient of I. In order to estimate the illegality effect δ given by (4), we recover the individual control functions from our estimate of f, then compute for each illegal migrant the bracketed term in (4), and average this over the entire group of illegal migrants.

2.4 Estimation via Penalized Regression Splines, Inference, and Testing: A brief Sketch

We estimate equation (10) and the preliminary propensity score equation (2) non-parametrically using penalized regression splines. In view of the curse of

[6] Note that we argue that if (not iff) (U, V_0, V_1) are joint normal, then a linear control function results. Hence we test for normality by testing for linearity. It is conceivable that linearity also obtains for distributions other than the Gaussian. We control for this situation in the empirical application by comparing the point estimates of the illegality coefficient (the principal object of interest) for the unrestricted and the normal model.

[7] The equality holds to first order. More specifically $E(\varepsilon|X, Z, I) = 0$. Second order expanding $f\left(\widehat{h}\right)$ about $-\mu_U(Z)$ shows that the error induced by the generated regressor, $f(h) - f\left(\widehat{h}\right)$, is proportional to $(\widehat{\mu}_U(Z) - \mu_U(Z))$. Hence the rate of convergence of the estimator is not negatively affected by the generated regressor problem.

dimensionality and the sample sizes available to us, we follow common practice and impose additively separability of covariates, so that $g(X) = \sum g_k(x_k)$ with $X = [x_1, .., x_K]$. Hence our estimation method is semi-parametric.[8 9]

Details and formal definitions are given in Appendix A. Our estimators are standard, and we follow the non-parametric literature in using the generalized cross-validation (GCV) score both to estimate the smoothing parameter (in our context the roughness penalty), and to select between competing models. In particular, we want to compare estimates of equation (10), and specifically of $[\mu_1 - \mu_0]$, when the control function f is unrestricted, to estimates when f is restricted to be identically zero (so endogenous selection is ignored) or linear (the two normal models of equations (8) and (9)). For the estimator see Appendix Section A.1, and for the GCV score see equation (16).

The distributional theory for the estimator (see Appendix Section A.2) is standard, and the estimators are asymptotically normally distributed under standard regularity conditions. Moreover, the estimator achieve the optimal nonparametric convergence rate.

Inference is more involved than in a standard setting. Estimation of the covariance matrix of the estimators needs to take into account the generated regressor problem introduced by the first step estimation of h and a heteroscedasticity issue. Appendix A.3 sketches the details of our covariance estimation, where we extend standard methods. Efficiency gains in estimating, say, the control function f are conceivable were one to consider the error structure in the estimation step. We therefore propose an alternative estimator, in which the error structure is directly taken into account via a variance regression, which in turn induces a weighted point estimation (see Appendix Section A.3.2). The standard penalized regression spline estimate of the control function is denoted by $\widehat{f}$, whereas the weighted estimate is denoted by $\widehat{f}_w$.

Hypothesis testing is carried out using Wald tests. Of particular interest is the linearity of the control function f since the normal model (8) is often used in the literature. The Wald test statistic is given by equation (17) in Appendix Section A.4.1. We complement this formal procedure, of course, by visual inspection of $\widehat{f}$. Additional information is provided and reported below by the estimated rank of the part of the smoother matrix pertaining to f, given by the number of positive eigenvalues, since using a full rank penalized cubic regression spline to estimate a linear function results in an estimated rank of one.

[8] For the data at our disposal this is not too restrictive, since many of our regressors are dummies. We have also experimented with interactions.

[9] Other semi-parametric estimators are available. Das, Newey, and Vella (2003) use unpenalized series estimation, i.e. an estimator which uses a polynomial basis. Our estimator uses cubic spline basis instead and imposes a roughness penalty, because of its more attractive approximation-theoretic properties (see e.g. Wahba, 1990, and Green and Silverman, 1994).

2.5 Summary and Empirical Agenda

Our empirical analysis will focus on two issues, namely first the shape of the control function f, and second the estimate of the illegality effect δ. The principal estimation equation is (10) and we estimate the unknown functions under the assumption of additive separability using penalized regression splines. Methods for inferences are adapted to the particular error structure of (10).

The illegality effect δ in the econometric potential outcome model with endogenous selection is given by equation (4). Its first term, $\mu_1 - \mu_0$, is estimated directly as the illegality coefficient in the estimation equation (10), the second term (the selection bias) is obtained as an averaged prediction for the group of illegal migrants.

Our interest in the shape of the control function f arises from the observation that the literature either ignores endogenous selection ($f(h) \equiv 0$), or imposes linearity by assuming joint normality of the errors (U, V_0, V_1) and $V_0 = V_1$, see equation (8). We therefore formally test linearity of f using both the unweighted and the asymptotically more efficient weighted estimator $\widehat{f}$ and $\widehat{f}_w$.

Finally, we use a standard model selection criterion, the GCV score, to select between the model with the unrestricted control function f and models with the parametric restriction of the literature imposed ($f \equiv 0$, equation (8), or its generalization in equation (9)), and quantify the extent of the selection bias by comparing the illegality coefficients $[\mu_1 - \mu_0]$ using the unrestricted f and the potentially misspecified parametric f.

3 The Data

We use data from the Mexican Migration Project (MMP93), a collaborative research project based at the Princeton University and the University of Guadalajara. We briefly describe some general features of the data.[10] Each year since 1982, the MMP interviews a random sample of 200 households in two to five Mexican communities. Following completion of the Mexican surveys, interviewers travel to destination areas in the United States to administer identical questionnaires to migrants from the same communities sampled in Mexico who have settled north of the border. However, we have excluded the US sample because of questionable survey methods as it is not clear what the population sampled from represents. The survey is cross-sectional. The complete Mexican sample includes ca. 16,000 households in 93 communities in 17 Mexican states. These communities represent a wide range of regions, ethnic compositions and economic conditions, covering isolated rural towns, larger farming communities, as well as metropolitan areas.

The principal focus of the survey is the household head. Crucially for our purposes, the survey elicits information about the mode of entry to the US. We defined as illegal migrants those migrants who have entered the US 'without documents', or who have entered the US on a tourist visa and have subsequently

[10] For more details see the project's website at http://mmp.opr.princeton.edu.

worked. This defines our illegal indicator I_i which is equal to 1 if migrant i is illegal, and zero otherwise. Detailed information about the household head's illegal migration history include the number of border apprehensions, and the price paid to a people smuggler ('coyote'). In order to be included in the Mexican survey, the household head must have completed or temporarily interrupted the migration spell or the main respondent reports on his behalf. Note, however, that the typical Mexican migration to the US tends to be circular and of 'short' duration.

The survey records standard labour market performance measures for the time before and during the last migration spell, as well as remitting behavior. Also included are standard demographic and economic information about the individual migrant and his household (except for household income), as well as community and municipality characteristics based on census data. Events at the household level which might have led to the migration, such as crop failure or other adverse events, are not surveyed.

Massey and Zenteno (2000) report a systematic comparison between the MMP and a nationally representative survey (Encuesta National de la Dinamica Demografica implemented by Mexico's national statistical institute), and conclude a close correspondence between *migrant* characteristics in the two sources (we do not require the MMP data to be representative of the entire Mexican population). The MMP is a well-tested and popular data set.

3.1 The Sample

We focus on labor migrants. We therefore consider prime aged male heads of households (aged below 55) who have migrated for work. Given the temporary nature of Mexican migration we restrict our analysis to recent migrants at survey date whose migration duration is less than 5 years, in order to eliminate permanent migrants. We also use retrospective data on migration history and wages, and limit the recall period to 10 years. Our sample is comprised of individuals who have migrated in the 1980s and 1990s, and we consider the two decades separately. This split by decade is motivated by the concerns about changing characteristics of Mexican immigrants in the literature, and the changing laws and border enforcement strategies. The samples are of size 595 for the 1980s and 872 for the 1990s. Finer control of time effects is not practicable given these sample sizes.

In the Data Appendix, we describe in some detail the variables and measures used in the subsequent analysis. Selected summary statistics for some of the covariates used in the analysis are reported in Table 1. The table reveals that the incidence of illegal migration in our sample is substantial: 81 percent of migrations embarked on in the 1980s and 63.5 percent in the 1990s are illegal. Since our sample is confined to heads of households, most of the migrants in our sample are married. Illegals are on average four years younger than legals, more likely to be on their first trip, the duration of their migration spell (measured in months) is slightly longer and does not exceed two years (but recall the exclusion of spells above five years). Illegal migration from urban areas is substantial, as

	1980s				1990s			
	Legal migrants		Illegal migrants		Legal migrants		Illegal migrants	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Incidence [%]	19		81		36.5		63.5	
Age at migration	37.3	(10.0)	32.0	(9.5)	36.6	(7.9)	32.3	(8.73)
Married	0.98	(0.13)	0.93	(0.26)	0.96	(0.2)	0.91	(0.28)
Years of education	5.97	(4.64)	5.34	(3.93)	5.59	(3.23)	6.22	(3.46)
Primary school	0.17	(0.39)	0.24	(0.43)	0.36	(0.48)	0.30	(0.46)
Secondary school	0.16	(0.36)	0.15	(0.36)	0.15	(0.36)	0.25	(0.43)
> Secondary school	0.14	(0.35)	0.09	(0.29)	0.07	(0.26)	0.08	(0.28)
English proficiency	0.14	(0.35)	0.03	(0.16)	0.06	(0.24)	0.03	(0.16)
Skilled	0.20	(0.4)	0.26	(0.44)	0.13	(0.34)	0.26	(0.44)
Agriculture in Mx	0.38	(0.49)	0.36	(0.48)	0.53	(0.50)	0.34	(0.47)
Migration duration	11.6	(10.7)	12.7	(13.0)	9.66	(9.90)	12.02	(11.45)
First trip	0.07	(0.25)	0.38	(0.49)	0.04	(0.19)	0.37	(0.48)
California	0.62	(0.49)	0.66	(0.47)	0.42	(0.49)	0.38	(0.49)

Table 1: Summary Statistics: Covariates

only around one-third of illegal migrants have worked in agriculture in Mexico.

As regards the human capital variables, in line with the literature, the quality of legal migrants has deteriorated over time while the proportion of legal migrants who worked in agriculture in Mexico has increased. The evidence on whether illegal migrants are typically worse educated or trained than legal one is mixed. In the 1980s, they had sightly less but in the 1990s slightly more education. The share of highest level of (self-reported) English proficiency is smaller for illegals. On the other hand, average skills levels are slightly higher for illegal migrants. A detailed skill tabulation (not reported here) reveals that most migrants report not to have changed occupation after migrating to the US. As in Card (2005) we observe a move away from California in the 1990s, and this applies to both groups of migrants.

In the Data Appendix, we also consider extensively measures of social networks. We find that both 'weak ties' and 'strong ties' are more pronounced for legal migrants than for illegals. For both groups weak ties appear to have become weaker over the two periods, whereas strong ties have become stronger. At the same time illegal migrants rely more on relatives to find jobs than legal ones, which is consistent with the hypothesis that they cannot freely search the labor market because of the lack of legal status.

3.1.1 Outcomes: Wages and Remitting Behavior

The principal outcomes of interest are real hourly log wages (measured in constant 1998 prices) earned in the US and the incidence and intensity of remitting. Table 2 reports the means and standard deviation.

	1980s				1990s			
	Legal migrants		Illegal migrants		Legal migrants		Illegal migrants	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Log US wages	2.03	(0.44)	1.79	(0.45)	1.93	(0.40)	1.78	(0.45)
Remittances	321.2	(329.6)	327.1	(408.1)	396.4	(327.0)	330.1	(356.7)

Table 2: Summary Statistics: Outcomes

First consider wages. In the 1980s the means are 1.79 and 2.03 respectively, and the difference of -0.23 translates into a substantial wage gap of 21 percent. For the 1990s data, mean wages for illegal migrants remain stable at 1.78, but those for legal migrants fall substantially to 1.93 in line with the observed drop in human capital. The resulting observed wage gap falls to 13 percent. These wage gaps are consistent with the literature using different datasets (e.g. Massey, 1987, Rivera-Batiz, 1999, and Kossoudji and Cobb-Clark, 2002). Rather than limiting attention to the first moment, we proceed to consider the entire wage distribution. Figure 1 depicts simple kernel density estimates of the unconditional log wage distributions for legal and illegal migrants. These distributions are fairly symmetric, concentrated, and have single modes. Compared to legal migrants, the wage densities for legal migrants are shifted to the left. Inspecting the plots of the respective empirical distribution functions shows that the wage distribution for illegal migrants first order stochastically dominates that for legal migrants. However, the gap is less pronounced for the 1990s than for the 1980s.

Figure 1: approximately here

The second outcome to be studied is remitting behavior. In the 1980s legal and illegal migrants remitted about the same on average, in the 1990s legal migrants remitted on average more. The incidence of remitting is high. In the 1990s ca. 85% of both legal and illegal migrants remitted. For remitting migrants, Figure 2 depicts simple kernel density estimates of log remittances, and after filtering out log wages (since we are interested in a direct illegality effect, and remittances are positively correlated with wages). For the 1980s data, the densities for illegal and legal migrants are very similar. For the 1990s data a small shift of the density appears to be present, even after filtering out log wages.

Figure 2: approximately here

In summary, we find, first, that our samples exhibit characteristics broadly comparable to other studies using MMP and other data despite different sample selection rules and objectives. Second, we observe differences in average characteristics between the groups of illegal and legal migrants, as well as over time, but these are moderate. Hence there is scope for examining causal effects of legal status conditional on similar characteristics. At the same time the observed differences in outcomes allow scope for the potential presence of illegality effects.

4 Empirical Results I: The Control Function, and Illegality Effects on Wages

We proceed to implement the empirical agenda outlined in Section 2.5, discussing the estimates of the control function before considering the estimates of the illegality effects on wages.

The implementation of the control function approach requires a prior estimate of the propensity score. Since the specific results are of no direct interest, we have collected these in Appendix B. Relevant for the what follows are the conclusions that (i) the covariates collected in Z_2 and included in the propensity score equation but excluded from the wage equation are relevant, so the propensity score is indeed a non-trivial function of Z_2; these encompass some demographics measures (marital status, numbers of workers and number of children in the source household in Mexico), and the wealth and deprivation measures relating to the financeability of costly migration (assets and livestock owned by source household prior to migration, dwelling descriptors, illiteracy and migration measures at community level, an indicator for traditional migration states); (ii) we conclude the existence of limits sets for the covariates necessary for the identification at infinity.

For our estimation of equation (10), the list of regressors collected in X include the human capital variables, age, and the network measures. The marginal effects are as expected, and we concentrate on the two principal features of the estimation equation. [11]

4.1 The shape of the Control Function

We depict the estimates of the control function $\widehat{f}$ and pointwise confidence bands in Figure 3, while in Figure 4 we juxtapose this estimate with the alternative weighted estimate $\widehat{f}_w$.

[11] All our final semi-parametric specifications are arrived at by backward selection. We have started with a fairly large set of regressors, including interaction terms. Backward selection has resulted in fairly parsimonious specifications with good explanatory power. As a final check on our specifications, we have verified the good behavior of normal QQ plots of the residuals. Detailed results are available from the authors.

Figures 3 and 4: approximately here

Consider first $\widehat{f}$ for the 1980s. The estimated function is linear over its entire support and visually indistinguishable from $f(h) = 0$. The formal Wald test supports the null of linearity (the p-value is 1). In addition, both estimated degrees of freedom and estimated rank are equal to 1, and the estimate is statistically insignificant. Using instead the alternative weighted estimator yields a linear function with a positive, but statistically insignificant gradient. We conclude that for the 1980s the error terms are approximately independent. By contrast the control function for the 1990s data is strongly significant and non-linear. The visual conclusion is supported formally by the Wald test which clearly rejects the null of linearity (the p-value is 0.00). The estimated degrees of freedom are 4.3, and the estimated rank is 8. The alternative weighted estimate exhibits a slightly larger amplitude.

4.2 Wage Penalties in the 1980s and 1990s

control function	1a. $\widehat{f}(h)$	1b. $\widehat{f_w}(h)$	2a. linear (normal, $V_1 = V_0$)	2b. linear (normal, $V_1 \neq V_0$)	3. $f(h) \equiv 0$ ($U \perp V_0, V_1$)
			A. **1980s**		
$\mu_1 - \mu_0$	-0.12	-0.12	-0.12	-0.12	-0.13
SE	0.071	0.080	0.071 [0.080]	0.073 [0.078]	0.042
GCV	0.1309	0.1312	0.1309	0.1313	0.1304*
			B. **1990s**		
$\mu_1 - \mu_0$	-0.20	-0.19	-0.30	-0.30	-0.09
SE	0.116	0.118	0.111 [0.106]	0.111 [0.103]	0.042
GCV	0.1809*	0.1809*	0.1822	0.1826	0.1828

Table 3: Illegality coefficients in the log wages regression (10). Notes: $\mu_1 - \mu_0$ is the coefficient of the illegality indicator, SE is based on a sandwich estimator and includes a correction for the generated regressor problem, GCV is the generalised cross-validation score. GCV: * indicates specification with lowest GCV score. Columns 2a and 2b: The control functions for the normal models are given by (8) and (9), [SE] is computed using the parametric assumptions. The 1980s raw difference between illegal and legal migrants is -.23, for the 1990s data we have -.14.

Table 3 Columns 1a and 1b report the estimates of the illegality coefficient $\mu_1 - \mu_0$ when the control function is non-parametrically estimated. Using either unweighted or weighted estimator for f makes only a negligible difference, as the estimates of the illegality coefficient closely agree as do the GCV scores. In the 1980s the estimated illegality coefficient is -.12. The estimate of the control function in this period is insignificant, hence the estimate of the illegality effect δ equals the estimate of the illegality coefficient $\mu_1 - \mu_0$. For the 1990s the estimate of the illegality coefficient $\mu_1 - \mu_0$ is -.2 Based on the estimate of the control function $\widehat{f}$ we estimate $E_{Z|I=1}\left[K_1(Z) + \frac{1-p(Z)}{p(Z)}K_0(Z)\right]$ to equal -0.024. Hence, by equation (4) the estimated illegality effect $\widehat{\delta}$ for the 1990s data equals -0.22.

Overall, our estimates indicate substantial wage penalties for illegal migrants, which have increased from -.12 in the 1980s to -.22 in the 1990s. Our sensitivity analyses conducted in Section 4.4 suggest that these findings are robust. The increased wage penalty correlates with the measures taken by the US government designed to curb illegal migration, including harsher penalties. In this more hostile environment we observe that illegal migrants search the labour market less for fear of detection,[12] and, more importantly, we infer that their bargaining power has diminished further.

4.3 Misspecification Biases

Given the prevalence in the empirical literature to either ignore selection or to assume joint normality of errors, it is of interest to quantify the misspecification bias of these naive approaches focussing on the illegality coefficient $[\mu_1 - \mu_0]$. It is also of interest which model for the control function is selected by the GCV criterion. Recall that in the absence of selection we have $f(h) \equiv 0$, and in the normal case depending on whether $V_0 = V_1$ the control function is given by (8) or (9). Table 3 Columns 2 and 3 reports the associate estimates.

For the 1980s our estimate of the control function is linear and indistinguishable from zero. Hence it is no surprise that the restricted models yield essentially the same estimates of $[\mu_1 - \mu_0]$. Imposing normality yields insignificant estimates of the correlation coefficients of the error terms.[13] Imposing the inferred restriction directly should and does yield efficiency gains. Moreover, the lowest GCV score is achieved by this model.

Turning to the 1990s data, our estimate of the control function is highly nonlinear. Imposing a different shape therefore should result in noticeably different estimates of the illegality coefficient. Compared to our preferred estimate of -.2 (column 1a), ignoring selection (column 3) leads to an estimate which is half this magnitude, and using either of the two normal models (columns 2a and 2b) yields an estimate which is 50% larger in magnitude. We also observe that the GCV score is minimized by the unrestricted model.

The naive approaches thus enjoy a lucky escape for the 1980s data. By contrast, the misspecification biases for the 1990s data are substantial.

[12] In a separate experiment we also include a dummy for whether the job was found through the migrant's own search. The effect on wages is positive, small, and borderline significant for the 1990s with no effect on the estimate of $\mu_1 - \mu_0$, and insignificant for the 1980s.

[13] Specifically, in the normal model with $V_0 = V_1$ the point estimate of $(\rho_{U,V}, \sigma_V)$ is (-.02,.35). For the normal model with $V_0 \neq V_1$, the point estimates of $(\rho_{U,V_0}, \rho_{U,V_1})$ are (.02,-.06) and of (σ_0, σ_1) are (.32,.36).

4.4 Wage Penalties: Robustness Analyses

We verify the robustness of our estimated illegality effects in two experiments.

4.4.1 Non-Parametric Matching Estimates for the 1980s

For the 1980s data we concluded that the error terms U and (V_0, V_1) are independent. Given this independence fully non-parametric estimators become available and we re-estimate the illegality effect using matching. This is of interest since additive separability, imposed in our estimation of equation (10) might not hold. This exercise is not repeated for the 1990s data since the control function for this case is significant and non-linear, the result of dependent errors.

If U and (V_0, V_1) are independent, the illegality effect is then identified non-parametrically by the observable difference $E_{Z|I=1}\{E\{Y|Z, I=1\} - E\{Y|Z, I=0\}\}$. Matching proceeds by estimating for each illegal migrant the counterfactual outcome Y_0 from a group of legal migrants deemed similar. Denoting the imputed counterfactual by $\widehat{Y}_0$, the illegality effect is then estimated by the sample analogue of $E_{Z|I=1}\left\{E\left\{Y_1 - \widehat{Y}_0|Z, I=1\right\}\right\}$. The literature (see e.g. Imbens, 2004, for a survey) contains many procedures for non-exact matching, with no best procedure. The usual approach is to use a distance functions which measures the distance between observational units in some metric, and to match the k nearest neighbors. A common procedure is to measure distance by the Euclidean distance between the propensity scores, or alternatively by a vector norm metric of the form $||v||_W^2 = v^\intercal W^{-1} v$, where W is the covariance matrix of v and v is taken to be the vector of differences in observables Z (Abadie and Imbens, 2006).

log wages in the 1980s	nearest neighbors			
Matching	1	2	3	4
vector norm	-0.12	-0.11	-0.11	-0.11
SE	0.06	0.05	0.05	0.05
propensity score	-0.13	-0.13	-0.23	-0.23
SE	0.16	0.13	0.12	0.11

Table 4: Matching estimates of causal illegality effects on log wages in the 1980s.

Table 4 reports the results. The smoothing parameter (number of nearest neighbors) is varied from 1 to 4. The matching estimates using the vector norm are very close to the results of Table 3, statistically significant, and stable across the numbers of nearest neighbors considered. Matching instead on the propensity score, the estimates vary across the number of nearest neighbors and exhibit greater variability, but are of the same order of magnitude.

4.4.2 Difference-in-Difference Estimates for Wage Penalties in the 1990s

For the 1990s data we concluded that the error terms U and (V_0, V_1) are dependent. Part of this dependence could have been induced by a person-specific but time-invariant effect, say η. We can address this possibility for a subset of our sample since some migrants in the 1990s not only have experienced a previous migration spell, but also have changed legal status and have become legal migrants. This data structure enables us to difference out η and to identify the causal illegality effect by a difference-in-difference.

We briefly sketch how the model can be extended to this data structure by consider distinct spells indexed by $s \in \{1, 2\}$. The potential outcomes are thus indexed by (I, s), and we have

$$\begin{array}{rcr} Y_{0,1} = \mu_0 + g(X_1) + V_{0,1} & \text{and} & Y_{1,1} = \mu_1 + g(X_1) + V_{1,1}, \\ Y_{0,2} = \mu_0 + g(X_2) + V_{0,2} + \tau e & \text{and} & Y_{1,2} = \mu_1 + g(X_2) + V_{1,2} + \tau e. \end{array}$$

The passage of time is assumed to affect migrants equally (this is a standard parallel trends assumption), and the growth effect is measured by τe where e is the elapsed time between the spells. We consider 'switchers', who become legal migrants, and 'stayer', who remain illegal migrants throughout. Letting dI be the indicator for switching, we assume that the event $dI = 1$ is given by

$$dI = 1(\mu_{dI}(Z) + U_{dI} > 0). \tag{11}$$

The source of dependence between the error terms U_{dI} and $V_{I,s}$ is assumed to be a person-specific random effect η, in particular we assume that $U_{dI} = \eta + \varepsilon_{dI}$, and $V_{I,s} = \eta + \varepsilon_{I,s}$, where ε_{dI} and $\varepsilon_{I,s}$ are assumed to be independent. Then the expected conditional change in log wages over the two spells for switcher is

$$E\{Y_{0,2} - Y_{1,1} | X_1, X_2, e, dI = 1\} = \tau e + \mu_0 - \mu_1 + g(X_2) - g(X_1).$$

For stayers we have

$$E\{Y_{1,2} - Y_{1,1} | X_1, X_2, e, dI = 0\} = \tau e + g(X_2) - g(X_1).$$

The observable difference in difference between stayers and switchers thus equals $\mu_1 - \mu_0$, which in turn equals δ in this model.

For our robustness analysis for the 1990s data we consider those migrants whose first spells falls into the 1980s, and whose last spell falls into the 1990s. This produces a sample of 237 observations, of which 102 are switchers. Mean log wages for switchers and stayers in the 1980s and 1990s are comparable to the full sample means. In particular, for the migration spell in the 1980s, both stayers and switchers are illegal migrants, and the mean log wages are 1.81 and 1.83 respectively. In the 1990s stayers' mean (real) log wages fall to 1.74, whereas switchers' mean log wages are 1.84. The unconditional difference in difference thus equals -.072. Conditioning, however, on observables and time effects, our estimate of the causal wage penalty $\mu_1 - \mu_0$ is -.12 with (robust) standard error

.06.[14] The estimate is thus significant and substantial, and comparable to the result reported in Table 3 Panel B. The difference-in-difference estimate for the subsample of repeated migrants is smaller in magnitude than for the full sample. This is to be expected since the full sample include migrants on the first trip, and being on the first trip impacts negatively on wages.

5 Remitting Behavior

Estimating illegality effects on remitting behavior requires a generalized approach since remitting behavior might be a function of wages which are themselves affected by legal status. We thus extend the potential outcome framework with endogenous selection to allow for endogenous regressors.

5.1 Estimating Illegality Effects in the Presence of Intermediate Outcomes

Consider potential outcome models for both wages E and remittances R, allowing wages to affect remittances. We have

$$\begin{aligned} E_1 &= \mu_{E,1} + g_E(X_E) + V_{E,1} \\ E_0 &= \mu_{E,0} + g_E(X_E) + V_{E,0}, \end{aligned} \tag{12}$$

and

$$\begin{aligned} R_1 &= \mu_{R,1} + g_R(X_R) + \psi(E_1) + V_{R,1} \\ R_0 &= \mu_{R,0} + g_R(X_R) + \psi(E_0) + V_{R,0.} \end{aligned} \tag{13}$$

The notation X_E and X_R allows different regressors to affect earnings and remitting. Let Z collect all relevant regressors.

The illegality effect on remitting is

$$\begin{aligned} \delta_R &= E\{R_1 - R_0 | I = 1\} \\ &= E\{R | I = 1\} - E_{Z|I=1}\{E\{R_0 | Z, I = 1\}\}. \end{aligned} \tag{14}$$

Estimating this faces two problems. The first empirical problem is the endogenous selection problem addressed before. The new problem is that remitting might depend on wages, which are also correlated with the other error terms. This endogenous regressor confounds the illegality effect on remitting.

These problems necessitate a new estimating strategy. We estimate the constituent parts of (14) in a multi-stage procedure, noting that the first term $E\{R|I = 1\}$ is directly estimable from the data, whereas the second term is a counterfactual.

[14] Log wages are regressed on the indicators for the last spell and illegality, the interaction between the latter two, and covariates. The difference-in-difference estimate is given by the coefficient of the interaction term.

As regards the second term, we have $E\{R_0|Z, I=1\} = \mu_{R,0} + g_R(X_R) + E\{\psi(E_0)|Z, I=1\} + E\{V_{R,0}|Z, I=1\}$. The last term satisfies $E\{V_{R,0}|Z, I=1\} = -\frac{1-p(Z)}{p(Z)} E\{V_{R,0}|Z, I=0\}$ which is estimable. Consider a first order Taylor series approximation of the penultimate term yielding $E\{\psi(E_0)|Z, I=1\} \simeq \psi\left(\mu_{E,0} + g_E(X_E)\right) + \psi'\left(\mu_{E,0} + g_E(X_E)\right) E\{V_{E,0}|Z, I=1\}$. The last term satisfies $E\{V_{E,0}|Z, I=1\} = -\frac{1-p(Z)}{p(Z)} E\{V_{E,0}|Z, I=0\}$. Putting these expression together we have

$$\begin{aligned} E\{R_0|Z, I=1\} &= \mu_{R,0} + g_R(X_R) + \psi\left(\mu_{E,0} + g_E(X_E)\right) \\ &\quad - \frac{1-p(Z)}{p(Z)} \psi'\left(\mu_{E,0} + g_E(X_E)\right) E\{V_{E,0}|Z, I=0\} \\ &\quad - \frac{1-p(Z)}{p(Z)} E\{V_{R,0}|Z, I=0\}. \end{aligned} \tag{15}$$

We seek to estimate the various functions involved in (15). This regression cannot be carried out directly, since the argument of the functions ψ and ψ' needs to be estimated first. But an estimate of $\mu_{E,0} + g_E(X_E)$ as well as $E\{V_{E,0}|Z, I=0\}$ can be obtained from the separate control function augmented regression using (12). This step 1 regression is

$$\begin{aligned} E\{E_0|Z, I=0\} &= \mu_{E,0} + g_E(X_E) + E\{V_{E,0}|Z, I=0\} \\ &= \mu_{E,0} + g_E(X_E) + f_E(h) \end{aligned}$$

where $f_E(h)$ denotes the appropriate control function (with $h < 0$). Next, to estimate the functions $\mu_{R,0} + g_R(X_R)$ and ψ and hence ψ' in (15) consider the remitting equation (13) for R_0 for the sample of legal migrants. The endogeneity problem implies that E_0, $V_{R,0}$ and I are potentially correlated. We overcome this endogeneity problem in step 2 by conditioning on an estimate of the error term $V_{E,0}$ as in Das et al. (2003). Thus $E\{R_0|Z, I=0, E_0, V_{E,0}\} = \mu_{R,0} + g_R(X_R) + \psi(E_0) + E\{V_{R,0}|Z, I=0, V_{E,0}\}$. The error term $V_{E,0}$ conditioned on can be estimated by the residual $E_0 - \widehat{\mu}_{E,0} - \widehat{g}_E(X_E)$ in step 1. We estimate the bias $E\{V_{R,0}|Z, I=0, V_{E,0}\}$ in practice non-parametrically by a bivariate (non-isotropic tensor product) spline defined over h and the estimate of $V_{E,0}$.

Finally, we assemble the estimate of $E\{R_0|Z, I=1\}$ by computing the predicted sample analogue of the RHS of (15). This is then averaged over observables Z of illegal migrants to obtain the estimate of $E_{Z|I=1}\{E\{R_0|Z, I=1\}\}$. The illegality effect δ_R is then estimated by the difference between the estimates of $E\{R|I=1\}$ and $E_{Z|I=1}\{E\{R_0|Z, I=1\}\}$.

To summarize, we estimate the unknown functions in (15) by estimating separately the control function augmented earnings regression (step 1) and the control function augmented remitting regression (step 2) for legal migrants . The latter's control function includes conditioning on the residual of the earnings regression. We then predict $E\{R_0|Z, I=1\}$ for illegal migrants, average across covariates Z, and finally difference from $E\{R|I=1\}$ to arrive at the estimate of the direct illegality effect on remitting δ_R. Given the complexity of the procedure, inference is carried out using a bootstrap procedure.

5.2 Empirical Results II: Incidence and Intensity of Remitting.

We consider the incidence of remitting, and the log (average monthly) remittances conditional on remitting. Recall Figure 2, which depicts kernel density estimates of log remittances conditional on remitting, and of residual remittances after filtering out wages. For the 1980s the densities appear to be the same for the two groups of migrants, whereas for the 1990s a small difference appears to persist.

	1980		1990	
1980	1. incidence	2. log remittances	3. incidence	4. log remittances
raw difference	0.07	-0.21	-.02	-0.23
illegality effect (δ)	0.004	-0.07	-0.04	-0.35
bMedian	0.07	-0.08	-0.05	-0.34
bQuantiles	-0.37,0.55	-1.09,0.85	-0.11,0.01	-0.47,-0.20

Table 5: Illegality effects on remitting behaviour. Notes: bMedian is the median of the illegality effect in the bootstrap distribution, and bQuantiles are the .1 and .9 quantiles of the bootstrap distribution; R=999 bootstrap replications.

Table 5 reports the raw differences in means, as well as our estimates of the illegality effects. Consider first the raw differences. Compared to legal migrants, in the 1980s illegal migrants exhibit a slightly greater incidence of remitting, but this is reversed in the 1990s. Legal migrants remit larger amounts. However, wages could be a confounding factor, and we have shown in Section 4.2 that illegal migrants suffer a wage penalty.

We therefore turn to the results of our estimation procedure. We briefly comment on the estimated control function. The step 1 control function is the same as discussed in Section 4.1 with $h < 0$. For the 1980s data, it is insignificant, but for the 1990s it is significant and non-linear. The step 2 control function mirrors this in that it is insignificant for the first period but significant for the second period.[15]

Turning to the estimate of δ_R, we find that in the 1980s, legal status has no statistical impact on either the incidence or the intensity of remitting. The estimated effects are somewhat smaller in magnitude than the raw differences, imprecisely estimated, and a basic percentile (90%) bootstrap confidence interval for δ_R include zero both for incidence and log remittances.

For the 1990s the evidence is mixed. As regards the incidence of remitting, the bootstrap confidence interval for δ_R includes zero. However, conditional on remitting, the illegality effect on the amount remitted is larger in magnitude than the raw difference, and statistically significant, since the bootstrap confi-

[15] The regressors in the first step wage regression are those also used in Section 4. The regressors used in the second stage regression include human capital variables, person level characteristics (e.g. married) and wages, household level charactersitics (e.g. number of children), and deprivation indicators.

dence interval does not include 0. This evidence suggests that illegal remitters in the 1990s sent 30 percent less because of the lack of legal status. The estimates of the illegality effect on log remittances is also consistent with the estimates of the wage-filtered log remittance density of Figure 2. The difference between the results for the 1980s and the 1990s correlates again with the more hostile environment for illegal migrants in the 1990s. Presumably, it has become more difficult for illegal migrants to transfer money from within the US and more important to retain money to cope with the increased insecurity in the US.

5.2.1 Robustness Check: Remittances in the 1980s

As in the case of wages, we can investigate the robustness of our results for the 1980s data using matching, since we infer that the error terms in the selection equation and the outcome equations are independent. Again, this conclusion does not apply to the 1990s data since the control functions are significant. Unfortunately, a difference-in-difference investigation for this period is not feasible since remittances are only reported for the last spell.

Table 6 presents the matching results for the 1980s. Note that matching does not take into account the intermediate wage outcome. These constitute a confounding factor as remittances are an increasing function of wages. Given the negative illegality effect on wages, matching yields a lower bound on the effect on remittances. Using propensity score matching, illegality effects for both incidence and amounts remitted are never statistically different from zero. The instability of the estimates is fairly pronounced across the smoothing parameters. By contrast, using the vector norm for matching yields much more stable results. For the amount remitted, the illegality effects is again insignificant. For the incidence, the estimated illegality effect is significant but small. We conclude that these results in line with the estimate of our model.

	nearest neighbors				nearest neighbors			
Matching	1	2	3	4	1	2	3	4
	incidence				log remittances			
vector norm	-0.12	-0.10	-0.10	-0.09	-0.05	-0.12	-0.08	-0.06
SE	0.05	0.05	0.05	0.05	0.16	0.14	0.14	0.14
propensity score	-0.15	-0.15	0.02	-0.02	0.30	0.23	-0.05	0.18
SE	0.17	0.13	0.14	0.13	0.42	0.32	0.30	0.29

Table 6: Matching estimates of illegality effects on remitting behavior in the 1980s.

6 Conclusions

This paper contains a number of conceptual, empirical, and econometric contributions. We have defined rigorously the concept of illegality effects, using the potential outcome framework, as causal effects of the lack of legal status on outcomes. Endogenous selection, if ignored, gives rise to a confounding selection bias. We capture the selection bias using a control function. We propose a new re-parametrisation of the control function, which is linear in case of a normal error structure, estimate it non-parametrically, and test linearity in the empirical application. To consider remitting behavior, the framework is extended to allow for endogenous regressors.

Despite the substantial extent of illegal migration (about 3 percent of the world population are migrants, of which 10-15% are estimated to be illegal; about 80% of all Mexican immigrants who arrived in the US in the 1990s are estimated to be unauthorized), the empirical literature on illegality effects is small. The leading papers consider wage effects, either ignore selection or address the issue not fully convincingly. In contrast to this, the selection bias is an important focus of our analysis. Moreover, the literature has ignored altogether the issue of illegality effects on remitting behavior despite the extent and importance of remittances.

We find, for the case of Mexican migration to the US based on MMP data, considerable and robust illegality effects on wages, the penalty being about 12% in the 1980s and 22 % in the 1990s. For the latter period, the selection bias is not created by a normal error structure; wrongly imposing normality overestimates the illegality effect on wages by 50%, while wrongly ignoring selection leads to a 50% underestimate. While we obtain similar estimates of the wage penalty as Kossoudji and Cobb-Clark (2002), we have employed a completely different and much less restrictive identification strategy. Since search behavior appears to have changed only marginally (see Table 7), we surmise that the increase in the wage penalty is due to the further erosion of the bargaining power of illegal migrants resulting from the more aggressive measures taken by Immigration and Naturalization Service designed to detect and curb illegal migration.

In contrast to these wage penalties, legal status appears to have small effects on remitting behavior. Hence lack of legal status does not give rise to a new "exchange motive" for remitting, but illegal migrants seem to retain more money to reflect the increased precariousness of their situation. These results are relevant to migration policy as they quantify the substantial losses to the migrant arising from the lack of legal status; turning this around, these illegality effects quantify the substantial gains that would arise from amnesty-style policies or the losses to the migrant from more restrictive migration policies.

References

Abadie, A. and G. Imbens (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, vol. 74(1), 235-267, 2006.

Angelucci, M. (2005), "U.S. Border Enforcement and the Net Flow of Mexican Illegal Migration", IZA DP1642.

Blundell, R. and J.L. Powell (2004), " Endogeneity in Semiparametric Binary Response Models Endogeneity in Semiparametric Binary Response Models", *The Review of Economic Studies*, Vol. 71, No. 3, pp. 655-679

Card, D. (2005), "Is the New Immigration Really So Bad?", NBER Working Paper No. 11547.

Das, M, Newey, W. and F. Vella (2003), "Nonparametric estimation of sample selection models", *The Review of Economic Studies*, 33-58.

Granovetter, M. S. (1973), "The Strength of Weak Ties." *American Journal of Sociology*, 78, 1360-80.

Green, P.J. and B. Silverman (1994), *Nonparametric regression and generalized linear models*, Chapman and Hall.

Hall. P. and J. Opsomer (2005), "Theory for penalised spline regression", *Biometrika*, 92, 105-118.

Hanson, G. H. (2006), "Illegal Migration from Mexico to the United States", Journal of Economic Literature, 44, Issue 4, 869–924.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica,* 47, 153–161.

Heckman, J.J. and S. Navarro (2004), "Using matching, instrumental variables, and control functions to estimate economic choice models", The Review of Economics and Statistics, 86(1), 30–57

Hirano, K., G. Imbens and G. Ridder (2003), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score", *Econometrica*, 71, 1161-1189.

International Labor Organization (2004) "Towards a fair deal for migrant workers in the global economy: Report VI" Conference Report, ILO: Geneva.

International Organization for Migration (2005) World Migration 2005: Costs and Benefits of International Migration.

Kossoudji, S. and D. A. Cobb-Clark (2002), "Coming out of the Shadows: Learning about legal status and wages from the legalized population", *Journal of Labour Economics*, 20, 3, 598-628.

Massey, D.S. and R. Zenteno (2000), "A Validation of the Ethnosurvey: The Case of Mexico-U.S. Migration", *International Migration Review*, Vol. 34, No. 3, 766-793.

Munshi, K. (2003), "Networks in the Modern Economy: Mexican Migrants in the US Labor Market", *The Quarterly Journal of Economics*, 549-599.

Orrenius, P.M and M. Zavondy (2005), "Self-selection among undocumented immigrants from Mexico,"*Journal of Development Economics*, 78, 215-240.

Passel, J. (2002), "New Estimates of the the Undocumented Population in the United States", Migration Information Source. Washington DC: Migration Policy Institute.

Pinske, J. (2000), "Nonparametric two–step regression estimation with regressors and error are dependent," *Canadian Journal of Statistics,* 28, 289–300.

Rapoport, H. and F. Docquier (2006), "The Economics of Migrants Remittances", in the *Handbook on the Economics of Reciprocity, Giving and Altruism*, Vol. 2, Ch. 17, Elsevier-North Holland, (Editors: S.-C. Kolm and J. Mercier-Ythier).

Rivera-Batiz, F. (1999), "Undocumented Workers in the Labor Market: An Analysis of the Earnings of Legal and Illegal Mexican Immigrants in the United States", *Journal of Population Economics*, 12, 91-116.

Ruppert, D., M.P. Wand, and R.J. Carroll (2003), *Semiparametric Regression*, Cambridge University Press.

US Commision On Immigration Reform (1998), *Migration Between Mexico and the United States: Binational Study*, Washington DC.

Wahba, G. (1990), *Spline models for observational data*, CBMS-NSF regional conference series in applied mathematics 59, Philadelphia, Pa. : Society for Industrial and Applied Mathematics.

Wood, S. (2006), *Generalized additive models : an introduction with R*, Chapman & Hall/CRC.

World Bank (2006), "Global Economic Perspectives - Economic Implications of Remittances and Migration", The International Bank for Reconstruction and Development, Washington, D.C.

A Estimation Details

First, we briefly describe our estimator, which is standard (see e.g. Green and Silverman, 1994, or Wood, 2006), and its distributional theory. Second, we explain our methods for inference which extend standard methods. The need for these arise since $\widehat{h}$ is a generated regressor, and the error term θ is possibly heteroscedastic of an unknown form. Finally, we state our test of the linearity of the control function.

For expositional simplicity we make a notational digression and consider only the generic univariate case : Assume that response y_i is generated by the DGP $y_i = f(x_i) + \theta_i$ with $E\{\theta_i\} = 0$ and $i = 1, ..n$. The function $f : R \longrightarrow R$ is assumed to be smooth. Given the additively separable structure of our model above, the multivariate case is immediate, requiring only column binding of model matrix components and row binding of the associated coefficients. In an additive setting involving more than one function note that these are identified up to an additive constant. For identification we therefore impose centering constraints of the type $\sum_i \widehat{f}(x_i) = 0$.

A.1 Estimation: Penalized Regression Splines

We assume that $f(x)$ can be represented in terms of known basis functions as $f(x) = \sum_j^{K_f} b_j(x)\gamma_j$ where K_f is the number of basis functions. Given our estimate of $\widehat{\gamma}_j$, our estimate of f is $\widehat{f}(x) = \sum_j^{K_f} b_j(x)\widehat{\gamma}_j$.

The coefficient vector γ is estimated by solving a trade-off between fidelity to the data and penalizing the estimated variability of the function f. The smoothness of the estimate of the function can be controlled by applying a penalty π to a measure of the roughness of the function. The particular measure used here is $\int [f''(x)]^2 dx$ which can always be equivalently stated, using the above representation of f, as $\gamma^\intercal P\gamma$ where P is the penalty matrix consisting of known elements.

Write the DGP more compactly in matrix form $Y = f + \varepsilon$ with $f = X\gamma$ where the i-th row of X is $\left[b_1(x_i), .., b_{K_f}(x_i)\right]$. Then $\widehat{\gamma}(\pi) = \arg\min$ $||Y - X\gamma||^2 + \pi\gamma^\intercal P\gamma$ where $||Y||^2 = Y^\intercal Y$ denotes the squared length of a vector. The solution is $\widehat{\gamma}(\pi) = (X^\intercal X + \pi P)^{-1} X^\intercal Y$, and we have $\widehat{f} = S_\pi Y$ where $S_\pi = X(X^\intercal X + \pi S)^{-1} X^\intercal$ is the smoother matrix. Note that linear functions are unpenalized, and that $\pi \to \infty$ results in a linear estimate.

The smoothing parameter π is itself estimated by minimizing the generalized cross-validation score

$$GCV(\pi) = n\sum_i^n \left[y_i - \widehat{f}_i(\pi)\right]^2 / \left[tr(I_n - S_\pi)\right]^2 \tag{16}$$

where tr denotes trace (as ordinary cross validation suffers from a lack of invariance problem). The remaining problem is the choice of the basis functions. We use cubic splines in view of the approximation-theoretic optimality properties, using the so-called value-second derivative representation (see e.g. Green and Silverman, 1994). Such splines require selecting knot points at which the basis function is to be evaluated. This selection is arbitrary, and we use knots placed at equi-distant probabilities of the quantile function.[16] The GCV criterion is also used to select between models with different numbers of knots.

Finally, we briefly comment on the first-step estimation of the propensity score $p = E\{I|Z\}$. Let l be a known link function such that $l(E\{I|Z\}) = \mu_U(Z)$. Three link functions are usually considered: the identity link which results in a linear probability model for the propensity score (Das et al., 2003); the logit link $l(p) = p/(1-p)$ (Hirano et al., 2003); the probit link $l = \Phi^{-1}$. In this paper we use the probit link because we seek to nest the normal case.[17] The link function introduces a non-linearity, so the above estimation scheme needs to be generalized. In practice this is done using iterated re-weighted least squares.

Our computations are carried out in the statistical program language `R`, and use Wood's (2006) library `mgcv`.

A.2 Distributional Theory and Rates of Convergence

The distribution theory for the estimators conditional on the vector of smoothing parameters follows from the fact that the estimators are linear since the estimation problem can be transformed into a least squares problem. In particular, $\widehat{Y} = S_\pi Y$ where $Y = [Y_1, .., Y_n]^\intercal$ and S_π is the smoother matrix which depends on the vector of smoothing parameters π. The estimators are therefore asymptotically normally distributed under standard regularity conditions. Hypothesis testing can then be carried out using Wald tests.

Convergence rates are studied in Hall and Opsomer (2005) who show that, in a simplified setting, penalized spline regression estimators achieve the optimal nonparametric convergence rate.

A.3 Inference: Non-Constant Variance

We pursue two approaches, namely a two-stage sandwich estimator, and, alternatively, a variance regression for greater asymptotic efficiency.

A.3.1 Sandwich Variance Estimator

First, as regards the two-stage sandwich estimator, we have
$Var\{\widehat{\gamma}\} = (X^\intercal X + \pi S)^{-1} X^\intercal Var\{\theta\} X (X^\intercal X + \pi S)^{-1}$. In the first stage, the parameter vector is estimated ignoring the potential heteroscedasticity, and we generated the vector of the squared residuals, which is denoted by r^2. In the second stage, $Var\{\widehat{\gamma}\}$ is estimated by $(X^\intercal X + \pi S)^{-1} X^\intercal diag(r) X (X^\intercal X + \pi S)^{-1}$.

[16] We have verified the robustness of our results to the choice of basis functions and knots.

[17] In our data, it turns out that logit and probit specifications yield very similar estimates, whereas the normal probability model generates a significant number of fitted probabilities exceeding one.

A.3.2 Variance Regression and the Weighted Point Estimator

Second, we propose to estimate a variance regression, and to compute iteratively a weighted estimate of the parameter vector. The details of the procedure are as follows (see also Ruppert et al, 2003). Denote by $g(x) = Var\{y|x\}$ the conditional variance. Let $g(x) = \exp\{f_{var}(x)\}$, the exponential being chosen to ensure a positive variance estimate. The object is to estimate the function f_{var} and thus g non-parametrically. To this end we assume that the vector $(y - f)$ has a normal distribution with zero mean and possibly non-constant variance $diag(g)$, $N(0, diag(g))$ (we verify this normality assumption using e.g. QQ plots). Then the vector $(y - f)^2$ follows a Gamma distribution $G\left(0.5, [2g]^{-1}\right)$ with mean g. We use this result for an iterative estimation procedure using the squared fitted residuals. (i) Denote by $\widehat{f}$ the first stage estimate of f ignoring heteroscedasticity. (ii) Obtain the vector of squared residuals $\widehat{r}^2 = \left(y - \widehat{f}\right)^2$ and fit to it the Gamma model. This generates $\widehat{g}(x) = \exp\left\{\widehat{f}_{var}(x)\right\}$. (iii) Re-estimate the heteroscedastic model $y \sim N(f, diag(\widehat{g}))$ to obtain the new $\widehat{f}$. Return to step (ii) and iterate until convergence.

In the main text we use the notation $\widehat{f_w}$ in order to distinguish this weighted estimate from the unweighted estimate $\widehat{f}$.

In our application to equation (10), we consider a semi-parametric specification of the variance function g. Specifically, we consider g as a function of the propensity score p: $g = \exp\{f_{var}(p)\}$. Note that this nests the normal model in which $Var\{Y_i|Z, I\}$ is a quadratic in the hazard h, $\gamma_0 + \gamma_1 h_i \Phi^{-1}(p_i) + \gamma_2 h_i^2$ with $h_i = h(p_i)$, and thus a function of the propensity score p_i.

A.3.3 The Generated Regressor Problem

Note also that in our application the error term is a composite one, which includes the generated regressor problem. The latter can be isolated by the delta method: Asymptotically, $Var\left\{\widehat{f}(h_i)\right\} = [\partial f/\partial h_i]^2 Var\left\{\widehat{h}_i\right\}$. If $\widehat{f}(h) = \sum_j^{K_f} b_j(h)\widehat{\gamma}_j$, then $\partial f/\partial h_i$ can be estimated by $\sum_j^{K_f} [\partial b_j(h_i)/\partial h_i]\widehat{\gamma}_j$. With $\widehat{\mu_U}(Z_i) = \sum_j^{K_{\mu_U}} c_j(Z_i)\widehat{\omega}_j \equiv C_i\widehat{\omega}$, we also have $Var\left\{\widehat{h}_i\right\} = [h_i(h_i + \widehat{\mu_U}(Z_i))]^2 C_i\widehat{\omega}C_i^{\intercal}$. This expression nests, of course, the parametric normal case, since the control function f is then linear.

A.4 Hypothesis Testing

Testing nested restrictions on the non-parametric model is usual implemented using an approximate F-tests, where the degrees of freedom of the non-parametric model are given by the trace of the smoother matrix. The test statistics has approximately an asymptotic F-distribution (Hasties and Tibshriani, 1990), one reason for the approximation being that the smoothing parameter is estimated but assumed fixed for the asymptotic theory. We also compare the GCV scores of various specifications.

A.4.1 Testing the Linearity of the Control Function

Of particular interest is the linearity of the control f in equation (10). As regards the formal test, note that a linear function (i) implies linear restrictions on the cubic spline, and (ii) is not penalized. Since the estimator has an asymptotic normal distribution, the null of linearity can then be tested using a Wald test.

The details of the Wald test depend on the parametrisation of the spline, in our case the value-second derivative representation subject to an identifiability constraint.[18] Let $f(h) = X_{B,f}(h)\gamma_f$ where $X_{B,f}$ is the part of the basis matrix pertaining to f and evaluated at the vector h, and γ_f is the associated vector of parameters. The estimate is of course $\widehat{f}(h) = X_{B,f}(h)\widehat{\gamma}_f$. Let t_f denote the vector of knots of the spline, and n_f the numbers of knots. Then the null of linearity implies $R\gamma_f - r = 0$ with $R = X_{B,f}(t_f)$ and $r = f(t_f) = at_f + b1_{n_f}$ with scalars a and b. Let V_f be the estimated covariance matrix of γ_f. This matrix is likely to be rank deficient when the smoothing (penalty) parameter is large; this situation occurs under the null of linearity since the cubic spline of the unrestricted model needs to be heavily penalized. Denote by V^{g-} the Moore-Penrose generalized inverse of a matrix V. Then the Wald test statistic

$$\left[R\widehat{\gamma}_f - r\right]^{\intercal}\left[RV_fR^{\intercal}\right]^{g-}\left[R\widehat{\gamma}_f - r\right] \tag{17}$$

is a random variable which follows asymptotically a chi-squared distribution with degrees of freedom equal to the number of elements of γ_f (i.e. $n_f - 1$). This result is only approximate, however, since uncertainty induced by the estimation of the smoothing parameter is ignored. The values of a and b are obtained in practice by regressing $\widehat{f}(t_h)$ on t_h.

We complement this formal test of course by the visual inspection of the estimate $\widehat{f}$. Additional information is provided by the estimated rank of the smoother matrix, given by the number of positive eigenvalues, since using a full rank ($K_f \gg 1$) penalized cubic regression spline to estimate a linear function results in an estimated rank of one (as well as a large estimate of the smoothing parameter).

B The Propensity Score Equation

The regressors for the propensity score equation can be grouped into measures of demographics, of individual human capital, of household and locality specific networks, and finally of household and locality specific wealth and deprivation.[19]

[18] The identifiability constraint is $C_f\gamma_f = 0$ with $C_f = 1_n^{\intercal}X_{B,f}(h)$ where 1_n is a vector of ones of length equal to the sample size.

[19] Specifically, the regressors include human capital variables (education and English), person level characteristics (age, married), source household characteristics in Mexico (numbers of workers and children), network measures (greencard, first trip, total trips, traditional, mi-

Significant time effects relate to changes in legislation and border enforcement (1987 onwards, 1991, 1996). All the regressors have the expected signs, and the estimated equations have good explanatory power, explaining 47% of the deviance for the 1980s data, and 42% for the 1990s data.

Important for the empirical implementation is the inclusion of covariates Z_2 which induce variation in the propensity score, but which are excluded from the wage equation. In our particular case, we include variables which relate to the financeability of a costly migration trip. In particular, illegal migration is more costly than legal migration since people smugglers are often involved, and is typically beyond the means of the poorest households. We thus include some demographics measures (marital status, numbers of workers and number of children in the source household in Mexico) which relate to the needs and incomes of the household in Mexico, and direct wealth and deprivation measures for the source household in Mexico prior to migration. For instance, we include livestock which can be considered as a liquid asset. We verify the signifiance of these covariates. Moreover, these covariates should play no direct role in determining wages in the US.

Finally, Figure 5 depicts simple kernel density estimates of the propensity score for illegal and legal migrants for the two periods. The main feature of the plots are: (i) the propensity scores overlap: for illegal migrants with a given propensity score there are legal migrants with similar propensity scores; (ii) for both groups we common support is $(0, 1)$, and we conclude the existence of limits sets for the covariates necessary for the identification at infinity.

Figure 5: approximately here

C Data Appendix

This brief appendix collects brief definitions of some selected variables.

The illegality indicator: We define as `illegal migrants` those migrants who have entered the US 'without documents', or who have entered the US on a tourist visa and have subsequently worked.

Human capital variables: Education is measured in both years, and the following ordered categories based on years: Less than Primary ($<$ 6 years), `Primary` (6 years), `Secondary` ($>$ 6 years and $<$ 12 years), `Higher than Secondary` ($>$12 years). `English` refers to the migrant self-reporting a high level of English proficiency. Skill refers to the skill level of the last job in Mexico before migration, and is obtained from reported occupational classifications, and `Skilled`

gration prevalence), wealth and deprivation measures of the source household in Mexico prior to migration (livestock, refrigerator, phone, car, dirt floor, and similar indicators, as well as adult illiteracy rate), indicators for work in agriculture and living in a small village, and time effects.

refers to skilled occupations. `Agriculture` indicates individuals working in agriculture in Mexico before migration. Also reported are skill levels occupations in the US and `US agriculture` indicates those working in the agricultural sector in the US.

Individual characteristics: `Age (at migration)` refers to age at the time of migration and `Married` to marital status at the time of survey.

Migration trip characteristics: `First trip` is a dummy equals to one if the migrant is on his first trip to the US. `Migration duration` is measured in months. We restrict our sample to recent migrants whose migration duration is less than 5 years at the time of survey. `California` captures the destination of migrants in the US.

Networks measures distinguish between weak ties and strong ties. The former is measured by `total trips` which refer to the total number of trips to the US by household members, and the latter by `migration prevalence` which refers to the proportion of adults in the community with migration experience based on census data. Given the long thin tail of `Total trips`, we also define the indicator, set to one if the total number of trips exceeds 6 (the modal number). We also consider an indicator for whether the migrant has friends in the US (`Friends in US`), and the number of family members in the US (`Fam in US`). `Traditional` refers to traditional sending states (Aguascalientes, Colima, Durango, Guanajuato, Jalisco, Michoacán, Nayarit, San Luis Potosí and Zacatecas). `Greencard` indicates whether the person has a parent or sibling in possession of a greencard.

Demographics: `Household size` is the number of household members. We also consider the number of children, and the number of workers in the household.

Wealth indicators. We use several variables to capture household wealth indicators: `Car` is a dummy equals to one if the household owns a car not bought using US Dollars. `Land` refers to land owned by the Household if not bought using US Dollars. `Livestock` refers to any livestock owned by the household but not those bought using US Dollars. `Dirtfloor` refers to the flooring of the dwelling of the household.

Municipality level covariates are all based on census data**.** Municipalities. `Less than minimum wage` refers to the proportion of workers earning less than the official minimum wage and was supplied by MMP. Note that the Mexican minimum wage system was first introduced in 1917, decentralised to the level of the municipality. A major reform took place in 1986 from when levels were set centrally. We also consider other deprivation indicators, e.g. the `Adult illiteracy rate` of individuals aged at least 15 years.

Outcome variables: `Log US Wages` refer to log hourly US wages measured in 1998 prices. `Remittances` is the average monthly amount (in constant pesos) remitted back from the US to Mexico.

	1980s				1990s			
	Legal migrants		Illegal migrants		Legal migrants		Illegal migrants	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total US trips	13.68	(12.57)	3.78	(4.34)	11.13	(7.74)	3.29	(3.24)
Migration prevalence	0.35	(0.15)	0.26	(0.14)	0.33	(0.12)	0.24	(0.12)
Friends in US	0.74	(0.44)	0.67	(0.47)	0.69	(0.46)	0.57	(0.49)
Family in US	1.85	(2.04)	1.00	(1.54)	1.70	(2.08)	1.25	(1.72)
Green Card	0.01	(0.09)	.03	(0.16)	0.19	(0.39)	0.15	(0.36)
Traditional	0.99	(0.09)	0.91	(0.27)	0.92	(0.26)	0.71	(0.45)
How US job was obtained:								
Own search	0.30	(0.46)	0.25	(0.43)	0.30	(0.46)	0.23	(0.42)
Recom. by relative	0.29	(0.45)	0.37	(0.48)	0.30	(0.46)	0.38	(0.49)
Recom. by friend	0.25	(0.44)	0.31	(0.46)	0.31	(0.46)	0.30	(0.48)

Table 7: Summary Statistics: Network Variables

C.1 Measuring Networks

Table 7 considers the networks measures, which distinguish between weak and strong ties. The former is measured by the proportion of households in the community with migration experience. We also include an indicator for whether the migrant reports to have friends in the US, and whether he comes from a traditional migration state in Mexico. Strong ties are measured by the total number of trips to the US by household members. In addition we consider whether anyone in the family is reported to be in possession of a green card, and the number of family members in the US. Both weak ties and strong ties are more pronounced for legal migrants than for illegals. Legal migrants have more members of their households with US migration experience than illegal migrants. The proportion of legal migrants in the 1990s with a relation in possession of a greencard is slightly more than that among illegals, in the 1980s the incidence is negligible. For both groups weak ties appear to have become weaker over the two periods, whereas strong ties have become stronger.

Migrants also report how they obtained their job in the US. For both groups own search and recommendations by relatives and friends account for more than 90 percent of job matches. Legal migrants exhibit somewhat more successful individual search than illegal migrants (30 percent vs. 23 percent in the 1990s), and show less dependence on family members (30 percent vs. 38 percent in the 1990s). This is another facet of the importance of strong ties to illegal migrants. The pattern remains stable over time. Illegal migrants are expected to search the labor market less because of their criminal status, and this is borne out by the data. However, successful own search has fallen only slightly over time despite a more hostile environment for illegal migrants in the 1990s.

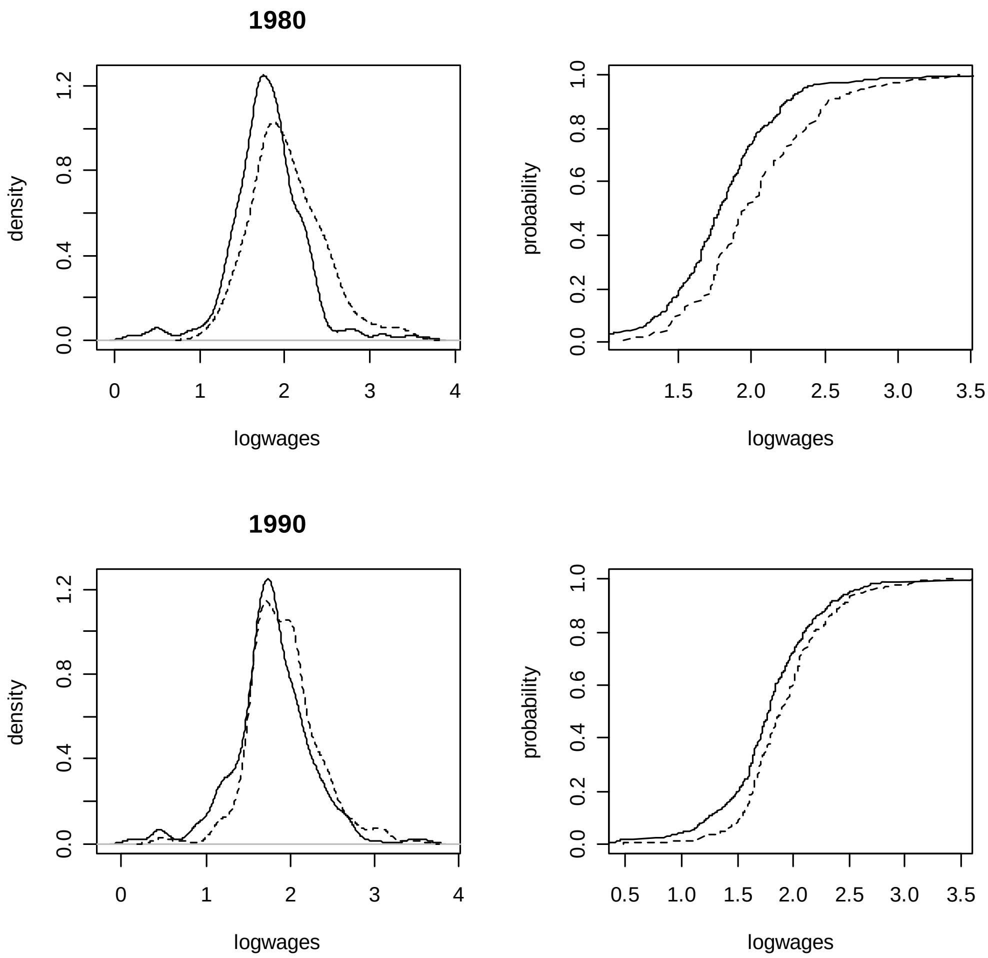

Figure 1: The distributions of hourly log wages for legal and illegal migrants. Notes: Solid lines refer to illegal, and dashed lines to legal migrants. The left panels depict kernel density estimates, right panels depict the empirical distribution functions.

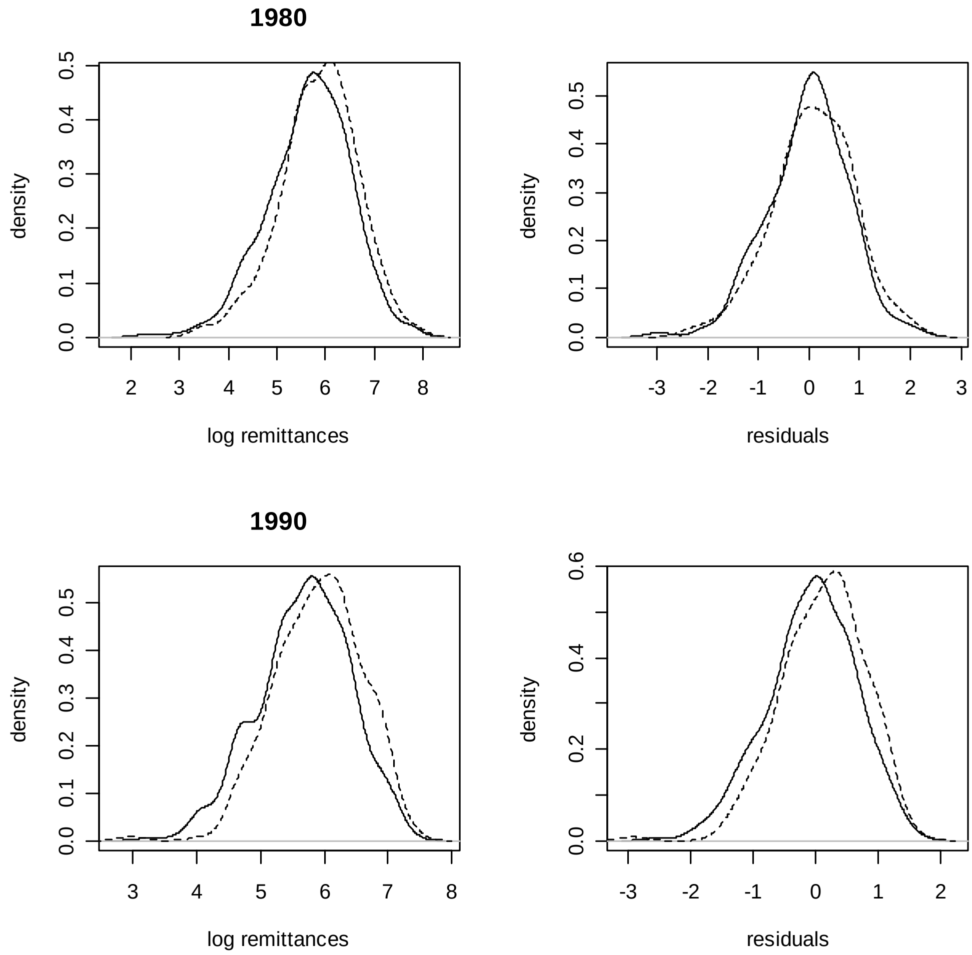

Figure 2: The density of log remittances, and of the wage-filtered residuals. Notes: Solid lines refer to illegal, and dashed lines to legal migrants. Kernel density estimates, conditional on migrants remitting. The residulas are obtained from a linear regression of log remittances on log wages.

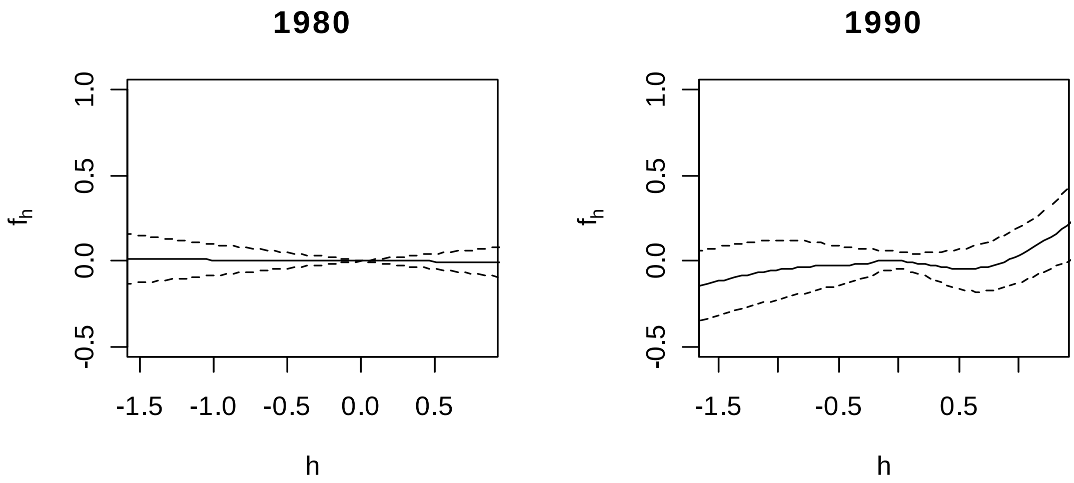

Figure 3: Non-parametric estimates of the control functions. Notes: the dashed lines represent a pointwise 95% confidence interval.

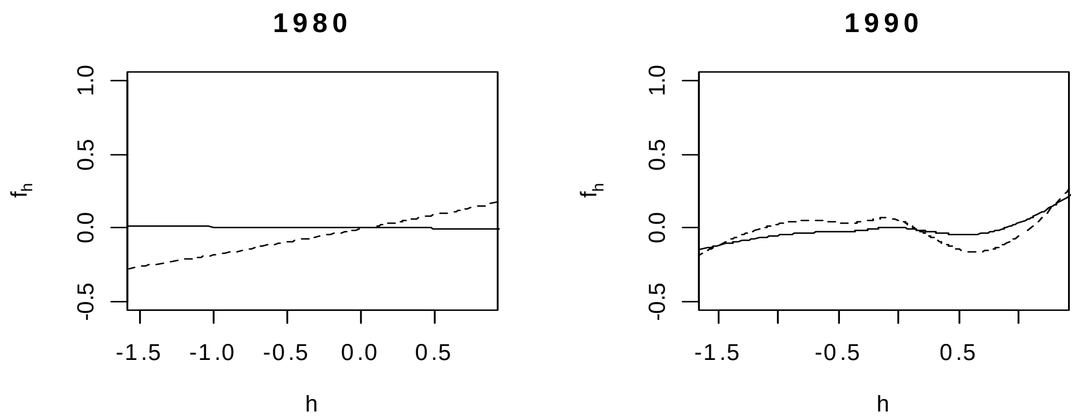

Figure 4: Unweighted v. weighted estimate of the control function by period. Notes: the solid line is $\widehat{f}$, the dashed line $\widehat{f}_w$.

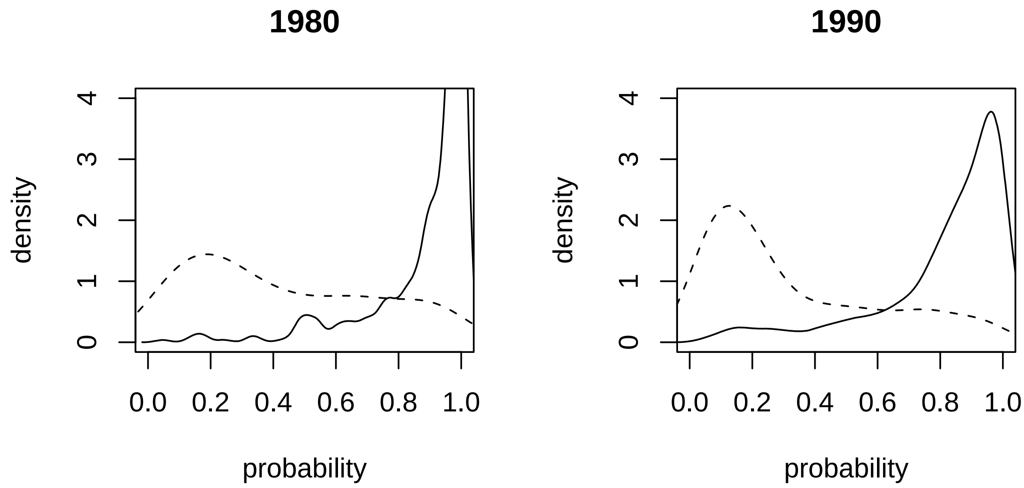

Figure 5: Kernel density estimates of the illegality propensity scores. Notes: The solid line refers to illegal migrants, the dashed line to legal migrants.