

# International Migration and the Propagation of HIV in Sub-Saharan Africa\*

F. Docquier<sup>a,b</sup>, Ch. Vasilakis<sup>a</sup>, D. Tamfutu Mumsi<sup>a</sup>

<sup>a</sup> IRES, Université Catholique de Louvain, Belgium

<sup>b</sup> FNRS, National Fund for Scientific Research, Belgium

December 2011

## Abstract

In this paper, we identify and quantify the role of international migration in the propagation of HIV across sub-Saharan African countries. We use a panel database on bilateral migration flows and HIV prevalence rates covering 44 countries over the nineties. Controlling for unobserved heterogeneity, spatial autocorrelation, reverse causality and reflection issues, and incorrect treatment of country fixed effects, we regress the log-change of HIV prevalence rates on the average levels of prevalence at destination and origin of migrants. We find evidence of a very robust emigration-induced propagation mechanism. On the contrary, immigration has no significant effect. Numerical experiments reveal that the long-run effect of emigration accounts for more than 5 percent of HIV prevalence rates in 18 countries (resp. 20 percent in 9 countries).

**Keywords:** international migration, labor mobility, HIV/AIDS, pandemics, propagation of diseases. **JEL codes:** F22, I12, J61

## 1 Introduction

It has long been recognized that international migration is a powerful force that shapes the distribution of human populations across the globe. It affects economic inequality between nations and contributes to propagate economic shocks (see Hatton and Williamson, 2008; Massey, 1988; Docquier and Rapoport, 2011; Gibson and

---

\*We thank Michel Beine, J. Paul Elhorst, Catherine Gurkinger, Fatemeh Shadman-Metha, and Vincenzo Verardi for helpful comments. The first author is grateful for the financial support from the Belgian French-speaking Community (convention ARC 09/14-019 on “Geographical Mobility of Workers and Firms”). The usual disclaimers apply. Corresponding author: Frédéric Docquier (frederic.docquier@uclouvain)

McKenzie, 2011). A recent strand of literature gives support to the view that migration also induces important transfers of political, cultural, sociological or behavioral norms and values between countries. Spilimbergo (2009) shows that foreign-educated individuals promote democracy in their home country, but only if foreign education is acquired in democratic countries. Lodigiani and Salomone (2011) demonstrate that emigration to democratic countries improves female political empowerment in the origin country. Fargues (2007) and Beine et al. (2008) provide evidence of migration-induced transfers of behavioural fertility norms, i.e. fertility behaviour at origin is affected by fertility rates at destination.

While the literature has mainly focused on transfers of positive norms, it is pretty obvious that movements of people can also propagate negative shocks across countries. In particular, migration can be a source of propagation of pandemic diseases within and across regions. History shows that colonization served to propagate germs across countries and continents. Migration contributed to spread bubonic plague within Europe in the 14th century. It propagated the Spanish flu from East Asia to Russia, Europe and North America in the beginning of the 20th century (Diamond, 1995).

Labor mobility is also perceived to be a major factor explaining the spreading of HIV/AIDS. Africa is the most infected continent with average HIV prevalence rates as high as 25 percent in Southern and Eastern Africa. The HIV virus causes AIDS, which is expected to induce the death of about 100 million people per year by 2025. Many case studies have highlighted the mechanism through which workers' mobility contributes to propagate the disease (see among others Anarfi, 1993; Decosas et al, 1995; Hope, 2001; Ateka, 2001; Brummer, 2002). Although many migrants have regular sexual partners, some have relations with casual partners and face a higher risk to be infected. This is especially the case of male workers migrating or commuting to find jobs on plantations or in mines, where prostitutes are brought in. The circular nature of migration and the maintenance of links with home through frequent visits puts people at risk at both ends of the migratory movement.<sup>1</sup>

Due to lack of comparable data on international migration, existing studies focus on internal migration or build on anecdotal evidence. To the best of our knowledge, our paper is the first to quantify the effect of international migration, relying on standard albeit rigorous econometric techniques. In this paper, we use aggregate data despite their limits (lack of infra-geographical information, imperfect measurement of migration flows, difficulty to interpret the mechanisms at work, etc.). An alternative strategy would be to collect micro and macro data at the local level (household, village) in order to identify how local changes in exits and entries affect the household's probability to be infected, and study the channels through which the virus is transmitted. With macro data, unobserved heterogeneity is key, causation is harder to establish, and it is uneasy to identify the channels. However, macro-data have also some advantages: they are homogeneous, constructed by the same authors or insti-

---

<sup>1</sup>Another factor relates to the migration of unhealthy widows away from their deceased spouse (Ntozi, 1997).

tutions for different periods and countries, they cover longer horizons. Beyond the mere advantage of using more observations, availability of panel data allows solving some of the problems listed above and limits the risk of misspecification and endogeneity biases (Islam, 1995 and 2003; Caselli et al., 1996; Roodman, 2009; Bazzi and Clemens, 2010).

Using a panel regression model, we identify and quantify the effect of international migration on HIV spreading across sub-Saharan African countries. We take advantage of a new database on bilateral migration between sub-Saharan African countries and combine it with annual panel data on HIV prevalence rates. Our data cover 44 sub-Saharan African countries over the nineties. Controlling for omitted variables, spatial correlation and endogeneity problems, we estimate the effect of immigration and emigration on the dynamics of HIV prevalence rates. Our analysis reveals that emigration to high-prevalence destination countries increases infection rates at origin. On the contrary, immigration does not generate significant effects. Although other mechanisms are plausible, these results are consistent with the widespread view that migrants have unprotected relations with prostitutes who were already infected in the host country. Hence, immigration does not induce significant changes in prevalence rates at destination. However new infected migrants propagate the virus to their origin countries through circulation, visits and/or return migration.

Over the nineties, data show that average levels of HIV at destination decreased in 20 sub-Saharan African countries, and increased in 24 countries. These variations can be due to changes in emigration flows and/or emigrants' location choices. Numerical experiments based on our estimated parameters reveal that the effect of emigration is rather low in about half of the countries included in the sample. However the long-run effect of emigration accounts for more than 5 percent of HIV prevalence rates in 18 countries, and more than 20 percent in 9 countries. In particular, HIV prevalence rates in the year 2000 would have been at least 20 percent larger without decreasing emigration in countries such as Mauritius, Lesotho, Swaziland, Botswana or Namibia. On the contrary, prevalence rates would have been at least 20 percent lower without increasing emigration in countries such as Burkina Faso, Comoros, Liberia or Equatorial Guinea.

The remainder of this paper is organized as following. Section 2 describes the empirical model and discusses econometric issues. Data are presented in Section 3. Section 4 provides empirical results. Finally, Section 5 concludes.

## 2 Model

Our goal is to analyze the determinants of HIV prevalence rates, defined as the percentage of people aged 15-49 who are infected with HIV. The HIV prevalence of country  $i$  ( $i = 1, \dots, N$ ) at year  $t$  ( $t = 1, \dots, T$ ) is denoted by  $H_{i,t}$ . Our model combines the time series dimension and the cross section variation of the data. Given its stock nature, HIV prevalence rates exhibit some inertia and we need a dynamic regression model to

explain their evolution. We use a standard convergence specification which features the annual log-change in HIV prevalence as the dependent variable. The explanatory variables are: past level of HIV prevalence, average level of HIV prevalence in destination countries of native emigrants from country  $i$  (denoted by  $Z_{i,t}^e$ ), average level of HIV prevalence in origin countries of foreign immigrants to country  $i$  (denoted by  $Z_{i,t}^i$ ). In this section, we present the specification used in our empirical analysis and then discuss some econometric issues.

## 2.1 Benchmark specification

The basic specification writes as follows:

$$\begin{aligned} \Delta \ln(1 + H_{i,t}) = & \alpha + \beta \ln(1 + H_{i,t-1}) + \gamma \ln(1 + Z_{i,t-1}^e) \\ & + \delta \ln(1 + Z_{i,t-1}^i) + \eta X_{i,t-1} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

where  $\Delta \ln(1 + H_{i,t}) \equiv \ln(1 + H_{i,t}) - \ln(1 + H_{i,t-1})$ ,  $Z_{i,t-1}^e$  and  $Z_{i,t-1}^i$  are the average levels of HIV prevalence in emigration and immigration countries (defined below),  $X_{i,t-1}$  is a set of other determinants of HIV,  $(\alpha, \beta, \gamma, \delta, \eta)'$  is the vector of parameters to be estimated, and  $\varepsilon_{i,t}$  is the error term. We use a specification with  $\ln(1 + x)$  to avoid losing observations with  $x = 0$ , i.e. to be consistent with countries where HIV prevalence rates (domestic, at destination or at origin) are null or small.

In equation (1),  $Z_{i,t-1}^e$  and  $Z_{i,t-1}^i$ , are constructed in line with previous studies on migration-induced transfers of norms and values (see Spilimbergo, 2009; Beine et al, 2008; Lodigiani and Salomone, 2011). For  $Z_{i,t-1}^e$ , we add up HIV prevalence rates in destination countries of native emigrants from country  $i$ , weighted by bilateral emigration rates. The latter is defined as the ratio of emigration flow from  $i$  to  $j$  to the native population in country  $i$ . We consider migration flows (rather than stocks) to eliminate earlier migrants who settled in the destination country a long time ago (possibly before the rise of HIV) or who migrated as children. Using migration flows, we focus on recent migrants who are likely to keep strong ties with their home country. This gives:

$$Z_{i,t}^e \equiv \frac{1}{N_{i,t}} \sum_j M_{ij,t} H_{j,t} \quad (2)$$

where  $M_{ij,t}$  stands for the emigration flow from country  $i$  to country  $j$  at time  $t$ , and  $N_{i,t}$  is the resident population in country  $i$ . Our rationale is that emigrants maintain ties with their home country through frequent visits, especially when they migrate for seeking jobs abroad. This puts people at risk in the origin country, in line with the literature described above.

Similarly,  $Z_{i,t}^i$  is the sum of HIV prevalence rates in origin countries of foreign immigrants to country  $i$ , weighted by bilateral immigration rates (defined as the

ratio of bilateral immigration flow to native population):

$$Z_{i,t}^i \equiv \frac{1}{N_{i,t}} \sum_j M_{j,i,t} H_{j,t} \quad (3)$$

Our coefficients of interest in (1) are  $\beta$ ,  $\gamma$  and  $\delta$ . They can be interpreted as following. Coefficient  $\beta$  determines the speed of convergence of HIV prevalence rates to their steady state level, the latter being impacted by country-specific characteristics ( $Z_{i,t}^e$ ,  $Z_{i,t}^i$ ,  $X_{i,t}$ ). Note that a positive value for  $\beta$  would imply that the growth rate of HIV prevalence increases with the current rate, i.e. a pattern of explosive dynamics. However a negative value for  $\beta$  implies that the growth rate of HIV prevalence tends to zero as the prevalence rate increases. It means that the prevalence rate progressively converges to a stationary level in the long-run. We expect our estimate of  $\beta$  to be comprised between 0 and -1, which reflects a process of monotonic and conditional convergence of HIV prevalence rates. Coefficient  $\gamma$  captures the short-run effect of emigration on HIV prevalence. If  $\gamma$  is positive and significant, it means that emigration to countries with high HIV prevalence rates increases the prevalence rate at origin. Coefficient  $\delta$  captures the short-run effect of immigration on HIV prevalence. If  $\delta$  is positive and significant, it means that immigration to HIV infected countries increases the prevalence rate at destination.

Because our estimates below strongly support  $\beta \in [-1; 0]$ , we now focus on the conditional-convergence interpretation. Model (1) predicts that the HIV prevalence rate of country  $i$  will converge towards a long-run equilibrium level defined as

$$\ln(1 + H_{i,ss}) = \frac{\alpha + \gamma \ln(1 + Z_{i,ss}^e) + \delta \ln(1 + Z_{i,ss}^i) + \eta X_{i,ss}}{-\beta} \quad (4)$$

where subscript  $ss$  stands for steady state (i.e. long-run level). Hence, the long-run effects of emigration and immigration are given by  $-\gamma/\beta$  and  $-\delta/\beta$ .

## 2.2 Econometric issues

The estimation of (1) entails several econometric issues that might lead the OLS technique to generate inconsistent estimates. Four main problems are considered here: omitted variables, spatial correlation, reverse causality and reflection problem, and correlated individual effects.

**Omitted variables.** The dynamics of HIV prevalence is a clearly an endogenous process affected by a large number of determinants of a varied nature, captured by the vector of controls  $X_{i,t}$  in (1). Demographic variables (age and gender structures, density, urbanization, etc.), economic variables (level of development, education, gender inequality, unemployment, structure of industry, etc.), quality of institutions (information about HIV risk, medical staffing and infrastructure, etc.) and cultural characteristics (religion and beliefs, sex practices, ethnic fractionalization, etc.) are

among the main determinants. It is difficult to control for all these characteristics given the lack of data on sub-Saharan African countries. Hence, omitted variables can cause biased parameter estimates (Islam, 1995). We take advantage of the panel dimension of our problem and use country and time fixed effects. The introduction of fixed effects accounts for the time-invariant unobservable factors and common trends. Although some determinants can vary across years and countries, using fixed effects is much less limitative that it seems at first glance. First, a lot of factors such as ethnic diversity or degree of urbanization are stable over time. Second, other factors such as education or the quality of institutions exhibit a lot of inertia. It is thus unclear whether their explicit inclusion (should we have observations for these factors) in the regression model would significantly improve the quality of fit and would reduce the degree of misspecification bias. We hypothesize that fixed effects are informative enough to account for unobserved heterogeneity. We write

$$\eta X_{i,t-1} = \alpha_i + \alpha_t$$

where  $\alpha_i$  is the fixed effect for country  $i$  and  $\alpha_t$  is the fixed effect for year  $t$ . Hence, we will first estimate the model using ordinary-least-squares method with fixed effects (referred to as OLS-FE).

**Spatial correlation.** International migration might not be the only spreading channel of HIV across nations. Commuting, tourism, visits abroad, unrecorded movements of people can also propagate the virus across countries. To control for this, we depart from standard OLS-FE and estimate two alternative models, the spatial error model (SEM) and the dynamic spatial autoregressive model (SAR).

In the SEM case, the spatial influence operates through the error term<sup>2</sup>. We estimate the SEM model assuming that the error term  $\varepsilon_{i,t}$  in (1) exhibits spatial correlation. Like migration, the magnitude of alternative propagation channels is likely to vary with the geographic distance between countries. Hence, our  $N \times N$  weight matrix  $W$  includes geographic distances between the 44 countries in our sample. Removing the country index  $i$  and using vectorial notations, we rewrite the error terms in (1) in the following fashion:

$$\varepsilon_t = \rho_1 W u_t + v_t \tag{5}$$

where  $W$  is the weight matrix (with zeroes on the diagonal),  $\rho_1$  is the spatial autoregressive parameter to be estimated,  $u$  and  $v$  are assumed to be normal and independently distributed with  $E(v) = 0$ ,  $E(vv') = \sigma^2 I$ .

In the SAR case, we follow the dynamic spatial autoregressive specification described in Lee and Yu (2010). We add two terms to (1). First, we allow the variation of HIV prevalence rate of a given country to depend on the lagged HIV prevalence rates in the other countries in addition to its own HIV prevalence rate. The difference

---

<sup>2</sup>For the estimation of SEM model, we use the Matlab routines for spatial panel data described in Elhorst (2003,2010a)

with  $Z_{t-1}^e$  and  $Z_{t-1}^i$  is that these HIV prevalence rates abroad are weighted by the distance matrix  $W$  (rather than the emigration and immigration rates). Coefficient  $\lambda$  captures the existence of alternative dynamic propagation mechanisms related to geographic distance. Second, we allow for contemporaneous spatial and social interactions between countries. We multiply the vector of current prevalence rates by the same distance matrix. Coefficient  $\rho_2$  captures contemporaneous interactions. We estimate the model using the Quasi-Maximum Likelihood (QML) estimator as in Lee and Yu (2010).<sup>3</sup>

Using vectoral notations (with  $\alpha_n$  standing for the vector of country fixed effects), our general SAR specification writes as

$$\Delta \ln(1 + H_t) = \alpha + \rho_2 W \ln(1 + H_t) + \lambda W \ln(1 + H_{t-1}) + \beta \ln(1 + H_{t-1}) \quad (6) \\ + \gamma \ln(1 + Z_{t-1}^e) + \delta \ln(1 + Z_{t-1}^i) + \alpha_n + \alpha_t + \varepsilon_t.$$

Under the latter specification, we allow spillover effects to operate with distance (on top of migration flows to high-infection countries). This model becomes non linear in the parameters. We need to express its reduced form to obtain the spillover effects as follows:

$$(I - \rho_2 W) \ln(1 + H_t) = \alpha + [(1 + \beta)I_n + \lambda W] \ln(1 + H_{t-1}) \\ + \gamma \ln(1 + Z_{t-1}^e) + \delta \ln(1 + Z_{t-1}^i) + \alpha_n + \alpha_t + \varepsilon_t$$

where  $I_n$  is the identity matrix. This model allows quantifying the effect of a change in the average HIV prevalence levels in emigration and immigration of a given country on its own HIV prevalence rate but also the HIV prevalence rate of all other countries in the sample. We will multiply the left and right-hand sides of the latter equation by  $(I - \rho_2 W)^{-1}$ , and then estimate the non linear model. We can compute the matrix of partial short-run effects of emigration and immigration as<sup>4</sup>:

$$\gamma^+ = \frac{\partial \ln(1 + H_t)}{\partial \ln(1 + Z_{t-1}^e)} = (I - \rho_2 W)^{-1} \gamma$$

and

$$\delta^+ = \frac{\partial \ln(1 + H_t)}{\partial \ln(1 + Z_{t-1}^i)} = (I - \rho_2 W)^{-1} \delta$$

Off-diagonal elements of the  $W$  matrix represent indirect effects. The parameter  $\rho_2$  measures the strength of contemporaneous spatial interdependencies between countries. If there is no contemporaneous spatial correlation,  $\rho_2 = 0$ , the direct effects of emigration and immigration are simply captured by  $\gamma$  and  $\delta$ . This does not prevent the existence of dynamic contagion effects if  $\lambda$  is positive and significant.

---

<sup>3</sup>MATLAB codes are available upon request.

<sup>4</sup>The computation of partial long-run effects is straightforward.

**Reverse causality and reflection.** The OLS regression model assumes that all covariates are independent of the error term. Although fixed effects control for possible misspecifications caused by unobserved characteristics, it does account for other possible sources of endogeneity of the regressors. Endogeneity problems may arise for several reasons.

First, equation (1) is dynamic because the presence of  $\ln H_{i,t-1}$ . The use of fixed effects and AR terms leads to inconsistency of estimates, especially when the number of periods is increasing (Nickell, 1981). Although the ratio of the cross-section dimension to the time dimension suggests that the Nickell bias should be limited in our regressions, it is interesting to look at alternative approaches.

Second, a positive effect of emigration or immigration on HIV prevalence could be explained by reverse causality if migration rates and/or the choice of destination country are endogenous. For instance, people originating from a risky country could be willing to emigrate more.

A third endogeneity source comes from the reflection problem (Manski, 1993). If country-specific equations were written as a system, the HIV prevalence rate in country  $i$  would depend on that in country  $j$ , which itself depends on that of country  $i$ . The reflection problem leads to two separate issues: the first problem is the possibility that the model is not identified, and the second is the simultaneity issue implied by the social interaction term. Concerning the identification problem we are in line with Calvo, Patacchi and Zenou (2009) who show that network models are identified if and only if networks of individuals are not similar. The weights in the social connections in the network context play the same role as the migration structure in our context. However, the simultaneity issue remains important.

In sum, there are two separate econometric problems related to equation (1). The first one is related to its dynamic structure and is well discussed in the recent econometric literature (Islam, 2003). The second one is related to the possible endogeneity of average levels of HIV at destination of emigrants and at origin of immigrants. As discussed by Islam (2003), there is no optimal estimation method for convergence equations in a panel data set-up. We use the most intuitive and transparent solution proposed in the literature, which consists in considering  $\ln H_{i,t-1}$  and average HIV prevalence rates at origin and destination as endogenous variables. As in Beine et al. (2008), we rely on two-stage least squares methods (2SLS). We will use the  $t-2$  levels to instrument  $\ln(1 + H_{i,t-1})$ ,  $\ln(1 + Z_{t-1}^e)$  and  $\ln(1 + Z_{t-1}^i)$ . We will also consider a specification with two lags ( $t-2$  and  $t-3$ ) per variable.

**IV with differences.** As argued by Caselli et al. (1996), the overwhelming majority of empirical studies on convergence are plagued by the incorrect treatment of country fixed effects. It is usually assumed that those effects are uncorrelated with the other right-hand-side variables. The fixed effect  $\alpha_i$  in (1) is used as a determinant of the log-change in the HIV prevalence rate, or equivalently of  $\ln(1 + H_{i,t})$ . By construction, it is also a determinant of  $\ln(1 + H_{i,t-1})$ , which is a regressor in (1). Hence, the assumption of uncorrelated fixed effects is violated in panel dynamic regressions.



Although fixed effects are used as control variables, it is desirable to correct for this collinearity bias.

To solve this problem, Caselli et al. (1996) suggest to estimate the model in differences to eliminate country fixed effects. Abstracting from spatial autocorrelation terms, equation (1) can be rewritten as

$$\ln H_{i,t} = \alpha + (1 + \beta) \ln H_{i,t-1} + \gamma \ln Z_{i,t-1}^e + \delta \ln Z_{i,t-1}^i + \alpha_n + \alpha_t + \varepsilon_{i,t}$$

Differentiating yields

$$\Delta \ln H_{i,t} = (1 + \beta) \Delta \ln H_{i,t-1} + \gamma \Delta \ln Z_{i,t-1}^e + \delta \Delta \ln Z_{i,t-1}^i + \tilde{\alpha}_t + \tilde{\varepsilon}_{i,t} \quad (7)$$

where country fixed effects are eliminated,  $\tilde{\alpha}_{i,t} \equiv \alpha_t - \alpha_{t-1}$  is the new time fixed effect, and  $\tilde{\varepsilon}_{i,t} \equiv \varepsilon_{i,t} - \varepsilon_{i,t-1}$  is the transformed error term.

As above, the model in differences can be estimated after instrumenting right-hand side variables using their lagged values. This eliminates the inconsistency due to endogeneity. For robustness check, we also adopt the GMM approach following Bond (2002). The method follows Bond (2002): We start by differencing each variable and then instrumenting  $\Delta \ln (1 + H_{i,t-1})$ ,  $\Delta \ln (1 + Z_{t-1}^e)$  and  $\Delta \ln (1 + Z_{t-1}^i)$  by the levels observed in  $t - 3$  and/or  $t - 4$ .

### 3 Data

Our sample is restricted to the 44 sub-Saharan African countries. We choose these countries because Africa is the most infected continent and HIV prevalence rates drastically increased in the nineties. We focus on the period 1990-2000 given the availability of migration data.

Comprehensive longitudinal data on HIV prevalence rates ( $H_{i,t}$ ), defined as the percentage of people aged 15-49 who are infected with HIV, were recently released by UNAIDS for 1990-2004 (UNAIDS, 2006). They reveal increasing levels of HIV prevalence rates in many countries between 1991 and 2000, and large difference across countries. The same data were used in Bhargava and Docquier (2008) who study the links between HIV prevalence, medical brain drain and number of deaths due to AIDS in Africa.

Data on international migration are taken from Ozden et al. (2010). They collected bilateral data on migration stocks ( $S_{ij,t}$ ) for more than 200 countries from 1960 to 2000, with one observation every ten years. Following the United Nations definition, they define a migrant as "any person that changes his or her country of usual residence" (United Nations, 1998) and classify migrants by country of birth. They use various sources to record migrants, mainly census and population register records collected in destination countries. These data reveal a striking fact: quantitatively, South-South migration dominates the global migrant stock, and explains one half of the world migration stock. It is worth noticing that South-North migration is the

fastest growing component of international migration, and North-South migration is negligible as all OECD countries send most of their migrants to other OECD countries. However, the intensity of South-South migration varies considerably across country pairs and over time. This is key since these variations will be used to identify and quantify the propagation mechanism of HIV.

We use the 1990 and 2000 migration matrices. As migration stocks vary slowly and smoothly over time, we interpolate the annual migration stock data assuming a constant annual growth rate over the nineties. Then, migration flows are proxied by differentiating stocks:  $M_{ij,t} \equiv S_{ij,t} - S_{ij,t-1}$ . These proxied flows will be used to weight annual prevalence rates at destination and origin as explained in (2) and (3).<sup>5</sup> Combining the HIV prevalence rate data with data on bilateral migration flows, we compute the average (or weighted) HIV prevalence levels at destination of emigrants ( $Z_{i,t}^e$ ) and the average HIV prevalence levels at origin of immigrants ( $Z_{i,t}^i$ ).

Table 1 summarizes the descriptive statistics for our main variables. We provide the sample means and standard errors calculated for the full sample of 44 sub-Saharan African countries in 1991 and 2000. We also provide data for the 25 most infected countries in 2000. On average, the HIV prevalence rate has been multiplied by 2.5 between 1991 and 2000, increasing from 2.83 to 7.00 percent. In 2000, prevalence rates range from 0.2 percent in Mauritius and Comoros to 28.6 percent in Botswana. Important changes were observed in Southern Africa (Swaziland, Lesotho, South Africa and Mozambique).

Average prevalence rates in emigration and immigration countries (expressed per 100,000 native people in Table 1 for clarity) have increased in absolute value, due to the global trend in HIV. However bilateral migration flows can be negative or positive and the sign of these average rates varies across countries. In 2000, the emigration-induced rate were large in Botswana, Gabon, Côte d'Ivoire, Burkina Faso\* or Swaziland, and low in South Africa, Republic of Congo or Zimbabwe. As far as the immigration-induced rate is concerned, large values were reported for Liberia, Equatorial Guinea, Burkina Faso\*, Mali\*, Uganda, Togo\*, while low levels were observed in Lesotho, Botswana, Swaziland, Namibia, Rwanda, Malawi or Zimbabwe.<sup>6</sup> We will take advantage of the high heterogeneity in HIV growth and prevalence rate at origin and destination to identify the migration-induced propagation mechanism.

---

<sup>5</sup>In unreported tables, we have also considered an alternative sample with extrapolated levels of migration flows from 2001 to 2004. We did so by assuming that the annual rate of growth of migration stocks between 2000 and 2004 equals the growth rate observed between 1990 and 2000. Regression results were similar (available upon request).

<sup>6</sup>A superscript \* indicates that countries are not reported in Table 1, due to lower levels of HIV prevalence in 2000.

**Table 1. Summary statistics for the years 1991 and 2000**

	$H_{i,91}^{(a)}$	$Z_{i,91}^e{}^{(b)}$	$Z_{i,91}^i{}^{(b)}$	$H_{i,00}^{(a)}$	$Z_{i,00}^e{}^{(b)}$	$Z_{i,00}^i{}^{(b)}$
Sample mean	2.83	0.61	0.90	7.00	1.31	-5.35
St. error	3.66	4.34	9.15	7.79	9.18	26.98
Botswana	6.99	4.93	-4.01	28.59	28.83	-51.87
Swaziland	2.40	0.29	-3.73	28.05	8.58	-55.77
Zimbabwe	13.89	-1.41	-1.25	26.08	-4.09	-12.67
Lesotho	1.86	0.16	-9.24	23.83	1.04	-147.00
Namibia	2.57	-0.74	-3.28	18.52	-2.78	-41.82
Zambia	14.34	-0.87	-1.84	17.28	-1.99	-8.10
South Africa	1.00	-10.55	-0.16	16.49	-37.46	0.23
Malawi	6.81	-0.26	-2.49	14.72	-0.91	.16-01
Mozambique	0.68	0.43	-2.40	13.99	2.61	-5.34
Central Afr. Rep.	4.91	-1.83	-1.65	11.14	-3.92	-1.03
Kenya	4.86	14.27	-0.03	8.01	9.89	-0.43
Uganda	12.01	-6.90	6.87	7.80	-4.51	8.60
Cote d'Ivoire	5.89	10.83	1.39	7.14	17.82	1.74
Gabon	0.91	8.50	-0.45	7.10	20.88	-0.09
Tanzania	5.79	-2.12	-0.65	7.09	-3.35	-0.90
Congo, Rep.	6.68	-11.16	-0.45	5.85	-14.38	-1.92
Cameroon	1.45	-0.30	0.13	5.74	-0.86	0.62
Rwanda	8.73	0.04	-27.25	5.31	0.01	-16.52
Angola	0.99	0.10	-0.71	3.82	0.21	-2.23
Burundi	4.18	-1.50	-3.46	3.49	-1.12	-1.97
Liberia	2.55	-0.14	23.37	3.46	-0.39	23.32
Guinea-Bissau	0.46	-0.01	0.49	3.42	-0.05	1.30
Equatorial Guinea	1.43	0.50	3.98	3.31	1.70	23.02
Nigeria	0.53	0.21	0.14	3.30	0.68	0.19
Cong, Dem. Rep.	3.56	-0.26	-0.18	3.18	-0.29	-0.67

Our sample includes the 44 sub-Saharan African countries. Table 1 only reports data for countries exhibiting the 25 highest HIV prevalence rates in 2000. <sup>(a)</sup> Prevalence rate per 100 people (source: UNAIDS, 2006). <sup>(b)</sup> Prevalence rate at destination/origin per 100,000 native residents (own calculations).

## 4 Results

In this section, we provide four sets of results. We first provide the results of the OLS-FE model. Then, we correct for spatial correlation and describe the results of the SEM and SAR models. Third, we correct for endogeneity biases and report the results obtained with the IV method. Finally, we estimate the model in differences using IV methods. All regressions include the full set of country fixed effects and year dummies

(except the model in differences). Moreover, we control for heteroskedasticity and only report robust standard errors in parentheses. We use the fully robust variance-covariance matrix described in Wooldridge (2002).

**OLS-FE estimates.** Table 2 reports OLS regressions for the main variables of interest. The columns contain different versions of model (1) covering the period 1990-2000.

**Table 2. OLS-FE regressions - Dependent =  $\Delta \ln(1 + H_{i,t})$**

	(1)	(2)	(3)	(4)
$\ln(1 + H_{i,t-1})$	-0.082*** (0.012)	-0.079*** (0.012)	-0.103*** (0.011)	-0.079*** (0.0114)
$\ln(1 + Z_{i,t-1}^e)$		1.334*** (0.244)		1.329*** (0.243)
$\ln(1 + Z_{i,t-1}^i)$			-1.571 (1.883)	-0.092 (1.845)
Const.	0.250*** (0.026)	0.246*** (0.027)	0.269*** (0.025)	0.246*** (0.027)
# Obs.	440	396	396	396
R <sup>2</sup>	0.789	0.871	0.862	0.871
Country FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$  and \* $p < 0.1$ . OLS regressions with a full set of year and country fixed effects. Robust stand errors in parentheses.

In all variants, we find out a significant and negative impact of the HIV prevalence rate of the previous period,  $\ln(1 + H_{i,t-1})$ . Hence, higher HIV prevalence rate of previous period tend to reduce the growth of HIV prevalence rate. The absolute value for  $\beta$  is lower than one. This implies that the model characterizes a stable and monotonic dynamic process through which HIV prevalence rates converge towards a country-specific long-run equilibrium. The speed of convergence is around 0.08, which means that it takes about 12.5 years to reach the long-run equilibrium.

As far as propagation mechanisms are concerned, all regressions show a positive and significant impact of the average prevalence rate at destination,  $\ln(1 + Z_{i,t-1}^e)$ . The short-run elasticity ( $\gamma$ ) is around 1.3 while the long-run one ( $\gamma/\beta$ ) is around 16. This effect is significant at the one percent level. On the contrary, we find no evidence of immigration-induced propagation. In all regressions, the average HIV prevalence rate at origin,  $\ln(1 + Z_{i,t-1}^i)$ , turns out to be insignificant.

Although data do not allow us to disentangle the mechanism, this is in line with the existing literature emphasizing the role of labour migration and circulation of people in the propagation of diseases across countries. How can we reconcile the fact that the emigration impact is significant while the effect of immigration is not? There could be several explanations. In particular, our macro analysis supports the widespread view

that migrants have unprotected relations with sexual partners abroad. These partners include prostitutes in the host country, among whom infection rates are high. Hence, immigration does not induce significant changes in prevalence rates at destination. However, new infected migrants propagate the virus to their origin countries through circulation, visits and/or return migration. This explains the positive and highly significant effect of HIV at destination.

Finally, it is worth emphasising that the fixed effects capturing unobserved heterogeneity play a key role. They enable us to explain more than 80 percent of the variability in HIV prevalence rate. Our results strongly support the hypothesis of conditional convergence, with long-run level depending on country characteristics and emigration patterns.

**Spatial regressions.** As stated above, migration might not be the only spreading channel of HIV. To evaluate the robustness of our results, we now correct for possible spatial correlation using the spatial error (SEM) and spatial autoregressive (SAR) models. The weighting matrix is based on latitude and longitude data collected for the 44 sub-Saharan African countries. Columns (1) and (2) in Table 3 provide the results of the SEM model, assuming that spatial correlation operates through the residual term, as formalized in (5). In columns (3) and (4), we provide results for the SAR model depicted in (6).

Qualitatively and quantitatively, results obtained with the SEM model are very similar to the OLS-FE ones. Both regressions point to a conditional convergence process, a highly significant impact of average HIV prevalence at destination of emigrants,  $\ln(1 + Z_{i,t-1}^e)$ , and an insignificant effect of average HIV prevalence at origin of immigrants,  $\ln(1 + Z_{i,t-1}^i)$ . It is worth noticing that the spatial correlation coefficient  $\rho_1$  is not significant, and the LM test of no spatial error is not significant. This means that we cannot reject the null hypothesis that we do not need to account for spatial correlation in the residuals.

The same conclusions emerge with the dynamic SAR model. The estimated spatial correlation parameters,  $\lambda$  and  $\rho_2$ , are never significantly different from zero. We found no evidence of dynamic or contemporaneous interactions besides the effect of emigration to countries with high infection rates. Again, it means that OLS-FE should be preferred to spatial correlation techniques for consistency and efficiency reasons. Emigration appears to be the main channel of transmission of HIV/AIDS between countries.

**Table 3. Spatial regressions**

	(1)	(2)	(3)	(4)
	SEM	SEM	SAR	SAR
Dependent	$\Delta \ln(1 + H_{i,t})$	$\Delta \ln(1 + H_{i,t})$	$\ln(1 + H_{i,t})$	$\ln(1 + H_{i,t})$
$\ln(1 + H_{i,t-1})$	-0.078*** (0.009)	-0.078*** (0.010)	0.996*** (0.0146)	0.997*** (0.0147)
$\ln(1 + Z_{i,t-1}^e)$	1.380*** (0.255)	1.330*** (0.265)	2.15*** (0.315)	2.2*** (0.369)
$\ln(1 + Z_{i,t-1}^i)$		0.072 (1.132)		0.9117 (1.8311)
Const.	0.190 (0.221)	0.189 (0.220)	0.185 (0.025)	0.185 (0.221)
# Obs.	396	396	308	308
$\rho_1, \rho_2$	0.120 (0.079)	0.118 (0.079)	0.1467 (0.097)	0.1386 (0.0979)
$\lambda$			-0.1516 (0.092)	-0.1433 (0.093)
LM test	0.106	0.108		
State FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$  and \* $p < 0.1$ . Spatial regressions with full set of year and country fixed effects. Columns 1-2 show results obtained with the SEM model, where  $\rho_1$  is the spatial correlation coefficient. Columns 3-4 show regressions results obtained with the SAR model, where  $\lambda$  and  $\rho_2$  are the spatial correlation coefficients. In line 'LM test', we provide the p-value of the test of no spatial error.

**IV regressions.** We now investigate whether our results could be driven by reverse causality and reflection problems. Tables 4 presents the results obtained with a 2SLS estimation technique, with lags of  $\ln H_{i,t-1}$ ,  $\ln Z_{i,t-1}^e$  and  $\ln Z_{i,t-1}^i$  used as instruments for their current values. In column (1), we only instrument  $\ln(1 + H_{i,t-1})$  with the  $t - 2$  lag. In column (2), we instrument all explanatory variables with  $t - 2$  lags. In column (3), we drop the non significant variable and add on lag to the number of instrument, i.e. we use the  $t - 2$  and  $t - 3$  lags.

The two necessary conditions for instrumentation are fulfilled in our regressions. Since Cragg-Donald and Stock and Yogo tests are not strictly valid in the presence of heteroskedasticity, we use the "rule of thumb" of a F-stat above 10 to test for the presence of weak instruments. In all first-stage regressions, F-stats are always far above 10 so that our instruments are not weak. They also pass the Kleinbergen-Paap's test of weak-identification test in the presence of heteroskedasticity. Moreover, our specification is robust to the Sargan-Hansen test of joint validity of instruments of Column(3).

The main findings of the IV estimations are broadly similar to those of the OLS estimations. We confirm a strong and significant propagation effect through emigration

to infected countries, while immigration remains insignificant. Although significant at the one percent level, the values obtained for the elasticity to emigration ( $\gamma$ ) and the speed of convergence ( $\beta$ ) are slightly lower than those obtained with OLS-FE. This could reflect the existence of a reverse causal link: HIV prevalence acts as a push factor for emigration.

**Table 4. 2SLS regressions - Dependent =  $\Delta \ln(1 + H_{i,t})$**

	(1)	(2)	(3)
$\ln(1 + H_{i,t-1})$	-0.124*** (0.013)	-0.131*** (0.013)	-0.117*** (0.010)
$\ln(1 + Z_{i,t-1}^e)$	0.699** (0.306)	0.994*** (0.225)	1.295*** (0.161)
$\ln(1 + Z_{i,t-1}^i)$	-1.012 (2.124)	-0.044 (1.948)	
Const.	0.288*** (0.027)	0.234*** (0.027)	0.199*** (0.020)
# Obs.	396	352	308
R <sup>2</sup>	0.863	0.906	0.936
Country FE	yes	yes	yes
Year FE	yes	yes	yes
F-stat of 1st Regression	4891	8373	23221
Cragg-Donald Wald F-stat	5132	1769	1.4e+04
Kleibergen-Paap Wald F-stat	4110	1349	6805
Hansen test overidentification test p-value			0.16

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$  and \* $p < 0.1$ . 2SLS regressions with full set of year and country fixed effects. Robust standard errors in parentheses. In Column (1), we use as instrument the second lag of HIV only. In Column (2), we use the second lags of variables  $\ln(1 + H)$ ,  $\ln(1 + Z^e)$ ,  $\ln(1 + Z^i)$ . In Column (3), we use the second and third lags of  $\ln(1 + H)$  and  $\ln(1 + Z^e)$  as instruments.

**IV in difference and GMM.** In Table 5, we estimate the model in difference described in equation (7). Although country fixed effects are eliminated by differentiating, we still include time fixed effects. All explanatory variables are instrumented using two lags, the preferred specification in IV in levels or in difference. The coefficient of the lagged term,  $\Delta \ln(1 + H_{i,t-1})$ , is around 0.70 and highly significant. This means that the speed of convergence now increases to about 30 percent a year. It takes 3 to 4 years to reach the country-specific steady state once explanatory variables are kept constant. Previous results about migration-induced propagation effects are also comforted. The effect of immigration,  $\Delta \ln(1 + Z_{i,t-1}^i)$ , is never significant, and the effect of emigration,  $\Delta \ln(1 + Z_{e,t-1}^e)$ , remains positive and significant at the one percent level. The elasticity increases to 1.5 in the preferred specification of column (1). Similar results are obtained when we instrument first-differenced variables by

levels in  $t - 3$  and  $t - 4$  and use the GMM approach following Bond (2002). All the tests support the fact that there is no evidence of weak instruments problem. All the specifications are robust to the Sargan-Hansen test of joint validity of instruments.

**Table 5. First difference IV regressions- Dependent =  $\Delta \ln(1 + H_{i,t})$**

	(1)	(2)	(3)	(4)
$\Delta \ln(1 + H_{i,t-1})$	0.71*** (0.026)	0.72*** (0.025)	0.72*** (0.0259)	0.71*** (0.025)
$\Delta \ln(1 + Z_{i,t-1}^e)$	1.5*** (0.44)	1.4*** (0.47)	1.5*** (0.4365)	1.21**
$\Delta \ln(1 + Z_{i,t-1}^i)$		-1.61 (4.22)		-6.1 (4.11)
Const.	-0.012*** (0.003)	-0.012*** (0.003)	-0.012*** (0.0021)	-0.012***
# Obs.	264	264	264	264
R <sup>2</sup>	0.86	0.86	0.86	0.86
Country FE	no	no	no	no
Year FE	yes	yes	yes	yes
Cragg-Donald Wald F-stat	994	639	994.2	639.2
Kleibergen-Paap Wald F-stat	757.1	539	757.7	538
Hansen test overidentification test p-value	0.3	0.35	0.4	0.43

Notes: \*\*\* $p < 0.01$ ; \*\* $p < 0.05$  and \* $p < 0.1$ . First-difference (columns 1-2) regressions with a full set of year fixed effects. First-difference with gmm (columns 3-4) regressions with a full set of year fixed effects. Robust standard errors in parentheses. Dependent variable is the change in HIV prevalence. We use the third and fourth lags of  $\ln(1 + H_{i,t-1})$ ,  $\ln(1 + Z_{i,t-1}^e)$  and/or  $\ln(1 + Z_{i,t-1}^i)$  as instruments.

**To what extent does emigration affect HIV prevalence rate?** Our regressions identify a significant and very robust effect of emigration on the growth rate and level of HIV prevalence rates between 1990 and 2000. To illustrate the extent of this effect, we use a simple numerical experiment. We start from the distribution of HIV prevalence rate observed in 2000. We then cut emigration flows in all countries (setting  $\ln(1 + Z^e) = Z^e = 0$ ) and simulate changes in HIV prevalence obtained after one year and in the long-run. Our simulation is based on parameter values obtained in Table 5, i.e.  $\gamma = 1.5$  and  $\beta = 0.29$ . Results are presented in Table 6. We only focus on sub-Saharan African countries where emigration induces a long-run change in HIV prevalence greater than 5 percent (i.e. 18 countries out of 44). Column (1) gives the HIV prevalence rate observed in 2000 (UNAIDS, 2006). Columns (2) and (3) give the simulated prevalence rate observed after one year with  $Z^e = 0$ , and the percentage deviation from the observed level. Columns (4) and (5) give the long-run effects.

We identify 20 countries in the sample where average levels of HIV at destination is decreasing. This can be due to decreasing emigration flows or changes in emigrants'



location. Nine of these countries are reported in Table 6 (top of the table). Most of them have high prevalence rate, except Mauritius and Rwanda. From our experiments, it appears that observed decreases in emigration flows and/or HIV levels at destination have reduced HIV prevalence rates in countries such as Lesotho, Swaziland, Botswana or Namibia. Without such decreasing emigration flows and in relative terms, long-run HIV prevalence rates would be at least 26 percent larger in these five countries.

On the contrary, we identify 24 countries where average levels of HIV at destination is increasing. Nine of these countries are reported in Table 6 (bottom of the table). Observed increases in emigration and/or HIV levels at destination have deteriorate HIV prevalence rates in countries such as Burkina Faso, Comoros, Liberia or Equatorial Guinea. Without such increasing emigration flows and in relative terms, HIV prevalence rates would be at least 23 percent lower in these countries. The emigration-induced propagation mechanism is important in these countries.

**Table 6. No-emigration counterfactual HIV prevalence rates (short-run and long-run)**

	Obs 2000	Short-run		Long-run	
	HIV rate	HIV rate <sup>a</sup>	Perc. dev <sup>b</sup>	HIV rate <sup>a</sup>	Perc. dev <sup>b</sup>
Mauritius	0.0	0.0	+77.3%	0.1	+93.3%
Lesotho	23.8	31.5	+24.4%	72.2	+67.0%
Swaziland	28.1	31.0	+9.6%	41.9	+33.1%
Botswana	28.6	31.4	+8.9%	41.5	+31.1%
Namibia	18.5	20.0	+7.4%	25.1	+26.2%
Rwanda	5.3	5.5	+3.3%	6.1	+12.5%
Malawi	14.7	15.2	+2.9%	16.5	+11.0%
Zimbabwe	26.1	26.7	+2.2%	28.5	+8.6%
Zambia	17.3	17.5	+1.5%	18.3	+5.7%
Benin	2.5	2.5	-1.5%	2.4	-6.1%
Somalia	0.9	0.9	-1.5%	0.8	-6.3%
Uganda	7.8	7.7	-1.7%	7.3	-6.8%
Togo	3.1	3.0	-1.9%	2.8	-7.8%
Mali	1.9	1.9	-2.4%	1.7	-10.1%
Eq. Guinea	3.3	3.1	-5.2%	2.7	-23.0%
Liberia	3.5	3.3	-5.2%	2.8	-23.0%
Comoros	0.0	0.0	-6.9%	0.0	-34.5%
Burkina Faso	2.3	2.1	-11.0%	1.5	-55.7%

<sup>a</sup> Counterfactual HIV prevalence rate obtained after one year (short-run) or in the long-run (long-run) when the average HIV level at destination equals zero. <sup>b</sup> Percentage of deviation from the observed level in 2000.

## 5 Conclusion

This paper shows that emigration to high-prevalence destination countries tend to increase HIV prevalence rates in sub-Saharan African countries. This comforts the widespread view that migrants have sexual relations in their host country and propagate the virus to their origin countries through circulation, visits and/or return migration. Consequently, changes in the size and/or structure of emigration flows affect growth rates and levels of HIV prevalence. The effect is very robust to the specification and the choice of the estimation method. Its magnitude varies a lot across countries. In 24 African countries (out of 44), decreasing emigration flows contributed to lower HIV risks over the nineties. In the remaining 24 countries, increasing emigration flows contributed to raise HIV prevalence. The effect was particularly strong in 9 countries, where changes in recent emigration flows are responsible for more than 20 percent of the number of HIV cases.

## References

- [1] Anarfi, J.K. (1993). Sexuality, migration and AIDS in Ghana - A socio-behavioural study. *Health Transition Review* 3, 1-22.
- [2] Arellano, M., and S. Bond (1991). Some tests of specification for panel data: Monte-Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-297.
- [3] Ateka, G.K (2001). Factors in HIV/AIDS transmission in sub-Saharan Africa. *Bulletin of the World Health Organization* 79 (12), 1168.
- [4] Bazzi, S., and M. Clemens (2010). Blunt Instruments: On Establishing the Causes of Economic Growth. Center for GLobal Development, Working Paper 171.
- [5] Beine, M., F. Docquier and M. Schiff, (2008). International Migration, Transfers of Norms and Home Country Fertility. UCL Discussion Paper 2008-43.
- [6] Barghava, A., and F. Docquier (2008). HIV prevalence and migration of health-care staff in Africa. *World Bank Economic Review* 22, 345-366.
- [7] Bond,S.(2002). Dynamic panel data models: a guide to micro data methods and practice, CEMMAP Working paper CWP09/02.
- [8] Brummer, D. (2002). Labour migration and HIV/AIDS in Southern Africa. IOM internal report : IOM's Regional Office for Southern Africa.
- [9] Calvo-Armengol, A.E. Patacchini and Y. Zenou (2009). Peer Effects and Social Networks in Education.*Review of Economic Studies* 76(4),1239-1267.

- [10] Caselli, F., G. Esquivel and F. Lefort (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth*, 1, 363-389.
- [11] Decosas, J., F. Kane, J.K. Anarfi, K.D. Sodji and H.U. Wagner (1995), Migration and AIDS. *Lancet* 346 (8978), 826-828.
- [12] Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. New York: W.W. Norton.
- [13] Docquier, F., and H. Rapoport (2011). Globalization, brain drain and development. *Journal of Economic Literature* (forthcoming).
- [14] Elhorst, J.P. (2003). Specification and Estimation of spatial panel data models. *International Regional Science Review* 26, 826-828.
- [15] Elhorst, J.P. (2010a). Spatial panel data Models. In *Handbook of applied spatial analysis*, eds. M.M. Fischer and A. Getis, 377-407, Berlin: Springer.
- [16] Elhorst, J.P. (2010a). Applied Spatial econometrics: raising the bar. *Spatial Economic Analysis* 5, 9-28.
- [17] Fargues, Ph. (2007). The demographic benefit of international migration: a hypothesis and its application to Middle Eastern and North African countries. In: Ozden, C. and M. Schiff (eds), *International migration, economic development and policy*, World Bank and Palgrave Macmillan: Washington DC.
- [18] Hatton, T., and J. Williamson (2008). *Global migration and the world economy: two centuries of policy and performance*. Cambridge (MA) and London: the MIT Press.
- [19] Holtz-Eakin, D., W. Newey, and H.S. Rosen. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56, 1371-1395.
- [20] Hope, K.R. (2001). Population mobility and multi-partner sex in Botswana : implications for a spread of HIV/AIDS. *Journal of Reproductive Health* 5 (3), 73-93.
- [21] Islam, N. (1995). Growth Empirics: a Panel Data Approach. *Quarterly Journal of Economics* 110 (4), 1127-1170.
- [22] Islam, N. (2003). What Have we Learnt From the Convergence Debate? *Journal of Economic Surveys* 17 (3), 309-362.
- [23] Lee, L.F., and J. Yu (2010). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26, 564-597.

- [24] Lodigiani, E., and S. Salomone (2011). International migration and female political empowerment. IRES Discussion paper.
- [25] Lurie, M. (2000). Migration and AIDS in southern Africa : a review. South African Journal of Science 96 (6), 343-347.
- [26] Nickell (1981). Biases in Dynamic Models with Fixed Effects. Econometrica 49 (6), 1417-1426.
- [27] Ntozi, J.P.M. (1997). Widowhood, remarriage and migration during the HIV/AIDS epidemic in Uganda. Health Transition Review 7, 125-144.
- [28] Massey, D.S. (1988). Economic development and international migration in comparative perspectives). Population and Development Review 14 (3), 383-413.
- [29] Gibson, J., and D. McKenzie (2011). Eight questions about brain drain. Journal of Economic Perspectives (forthcoming).
- [30] Ozden, C., C. Parsons, M. Schiff and T. Walmsley (2011). The Evolution of Global Bilateral Migration 1960-2000. World Bank Economic Review (forthcoming).
- [31] Roodman, D. (2009). A note on the theme of too many instruments. Oxford Bulletin of Economics and Statistics, 71, 135-158.
- [32] Spilimbergo, A., (2009). Democracy and Foreign Education. American Economic Review, 99(1), 528-43.
- [33] UNAIDS (2006). HIV Data. Geneva ([www.unaids.org/en/hiv\\_data/default.asp](http://www.unaids.org/en/hiv_data/default.asp)).
- [34] Wooldridge (2002). Econometric Analysis of cross section and panel data. The MIT press.