Modeling migration flows in the Mekong River Delta region of Vietnam: an augmented gravity approach

Huynh Truong HUY
School of Economics and Business, Can Tho University
Street 3/2, Ninh Kieu District, Can Tho city, Vietnam
Email: hthuy@ctu.edu.vn
Tel: (0084) 939409555
Fax: (0084) 7103839168

Walter NONNEMAN
Faculty of Applied Economics, Antwerp University
Email: walter.nonneman@ua.ac.be

Abstract

This article aims at modeling inter-provincial migration flows between provinces of the Mekong River Delta (MRD) region and 3 major urban cities in Vietnam. The key feature of the model is that it departs from the time proofed gravity model, which is expected to verify whether hypothesis on determinants of migration suggested by the literature hold or not in the case of the MRD region. The result of estimations indicates that migration flows between the MRD provinces and 3 major urban cities vary with the square root of the product of province populations and the ratio of income at destination over income at source, but inversely relate with distance. In addition, the forecast shows that the MRD region remains an important out-flow region with out-flows from provinces increasing by 0.4 million in the next five years, among Ca Mau, Kien Giang, Dong Thap and An Giang will see the largest increases in out flows.

Keywords: migration flows, distance, income ratio, poverty rate.

JEL classification: J61, C10, C31, C53.

1. Context

The Mekong River Delta (MRD) region is home to 17.3 million people (2010) – about 20 percent of the population of Vietnam. The region has 13 provinces and cities and with a density of 426 people per square kilometer is one of the most populated areas of the Southeast Asia basin. The population growth rate is a steady pace of 1.8 to 2 percent since the 1990s. Approximately 85% of the MRD population lives from agriculture. The region produces about 90% of national rice exports and 60% of Vietnam's fishery product exports. Despite being the largest granary in South East Asia and increasing household standards of living, poverty is still a major policy concern, as well as other welfare issues such as education, health and environmental issues.

It is not surprising that this rural area is the main migrant sending region of Vietnam. Over the period 2004-2009 slightly more than 250,000 entered the MRD region from provinces out of the region, whereas more than 900,000 people left the MRD region for other provinces in the country.

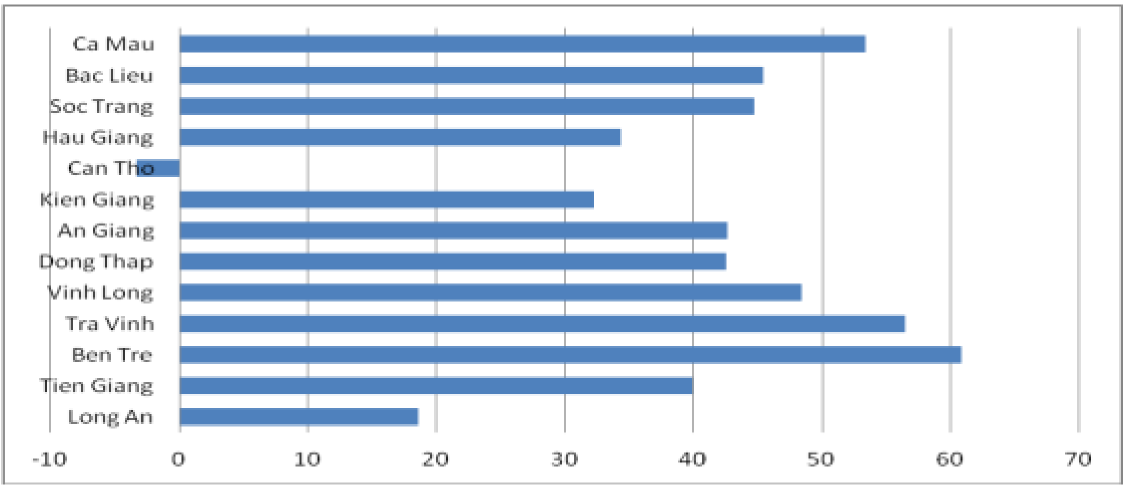
The most important destinations for these MRD out migrants are the urban provinces of Ho Chi Minh City (45.9% of all MRD out migration) and Binh Duong (20.8%). The others are going to provinces within the MRD region (20.4% of all MRD out migration) of which 25.5% are destined for the main urban area of the MRD region namely Can Tho. The rest of MRD out migrants (12.0% of all MRD out migration) moved to other areas in Vietnam.

Based on descriptive statistics, many typical stylized facts on migration in developing countries are valid for Vietnam and the MRD region: migration from rural to urban areas, feminization of migration, migrants are predominantly young people and on average with more human capital (VGSO, 2010b, 99-101).

Figure 1 gives an overview of net out migration of MRD provinces over the period 2004-2009. All provinces are net-sending areas, except for the urban province of Can Tho. However, net in migration of Can Tho (3.3 per 1000 population over the 5 year period) is very small compared with other urban areas of attraction such as Binh Duong (448.6 per 1000), Ho Chi Minh City (149.1 per 1000) and Ha Noi (94.4 per 1000).

Figure 1

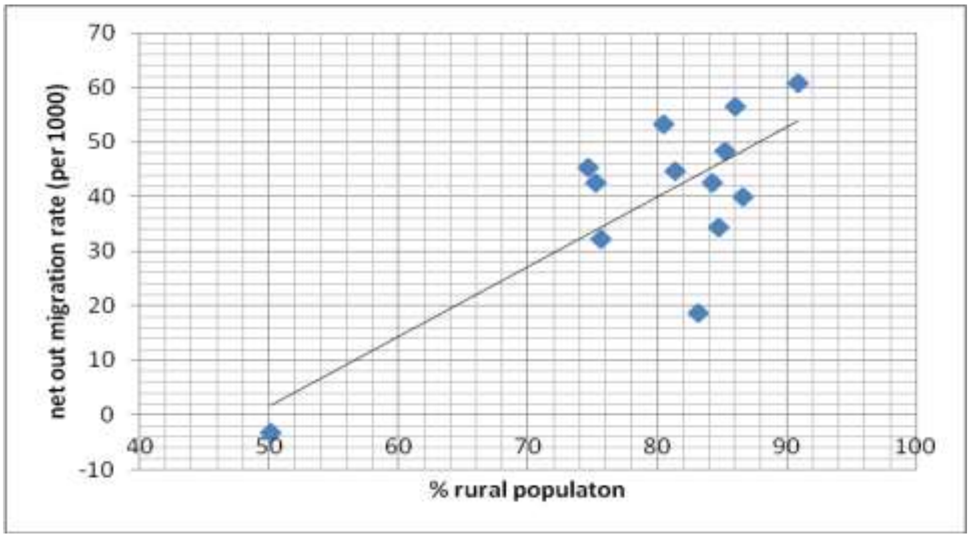
Net Out Migration MRD Provinces (2004-2009, Net out per 1000 Population)



The scatter diagram of Figure 2 illustrates the rural-urban migration phenomenon within the MRD region.

Figure 2

Net Out Migration in MRD Provinces and Urbanization



Modeling migration between provinces of the MRD and the rest of the country goes beyond description but it attempts to explain these stylized facts, identifying and estimating the relative importance of possible determinants of migratory flows. Such knowledge may be useful to predict the course of future migration flows.

The purpose of this article is to model migration flows between the provinces of the MRD and 3 major urban cities and the rest of Vietnam using the time proofed gravity model. The aim is to explain migration flows, to verify whether hypothesis on determinants of migration suggested by the literature hold or not in the case of the MRD region and finally, to forecast migration flows. The next section (2) discusses theory and hypothesis related to gravity models of migration and econometric issues involved in estimating parameters. The section 3 explains the data used, the main descriptive statistics and some bi-variate analysis between migration flows and key explanatory variables are shown. Section 4 is devoted to multivariate analysis, verifying various hypotheses ventured in the migration literature. A suitable model is selected for forecasting and forecasts for the period 2009-2014 are presented in section 5. Finally, conclusions and caveats are presented.

2. Gravity models of migration: theory and hypothesis

Over time, different approaches have been developed in the literature to model migration flows and to structure economics of migration (Greenwood & Hunt, 2003). Gravity models were popular in the 1950s and 60s. They are still often used to structure explanations and to forecast of migration flows (Lewer & Van den Berg, 2008).

Most early studies – for example (Zipf, 1946) – framed the gravity model in Newtonian terms i.e. flows were proportional to the population" masses" of source and destination area and inversely related to "distance" to some positive exponent or

$$M_{ij} = k\frac{P_i P_j}{d_{ij}^{\alpha}} \qquad (1)$$

During the 60s "modified gravity type" models were developed. These models featured the standard proportionality of migration flows to size of origin and destination population and an inverse proportional relation with distance, but added – based on ad hoc reasoning of what could attract or repel migrants – several additional variables. Most frequent additional variables used are income, tax rates, unemployment rates, degree of urbanization and amenity variables such as climate, access to public services, etc.

Modified gravity models do not have a strong or explicit choice-theoretic foundation, except for some naïve efforts. For example, Niedercorn *et al* have argued that equation (2) is the outcome of a utility maximizing decision by assuming that migration yields utility directly (Niedercorn & Bechdolt Jr, 1969). However, it is generally accepted that migration does not generate utility in a direct way but only indirectly as an investment in human capital, involving costs that are hopefully covered by future benefits (Sjaastad, 1962).

Despite the lack of an explicit choice-theoretic framework – with migrant behavior as the outcome of a constrained utility maximization model – the extensive literature on migration and development[1] – suggests several key variables to include as independent variables.

[1] For an excellent survey on migration and development from a broad perspective, see de Haas, 2010.

The “classic” rural-urban migration model (Harris & Todaro, 1970) stresses the difference in expected labor income between the rural source and the urban destination as the key determinant. This justifies the inclusion of income and employment opportunities or unemployment as independent variables.

As migration is an investment requiring sufficient capital funds to overcome the initial cost of migration, financing migration in the absence of proper capital markets may be a problem for the poorest of families (Lucas, 1997, 746-747). Hence, migration may not be an option for the poorest of families and poverty may be associated with less rather than more migration.

The “new economics of labor migration” adds migration as a means of risk diversification (Stark, 1991, 55). As agriculture is a high risk activity with nature playing havoc with farm output and income, one way to alleviate family risk is by urban migration of a dependable family member. When insurance schemes against adversity in agricultural output are lacking, rural to urban migration may occur even if urban expected incomes are lower than the rural income. This line of thought justifies using some measure of urbanization in source and destination as independent variables.

Another class of models suggests that “relative deprivation” is a major driving force of migration (Stark 1991, 87-101) (Stark, 1984). If a person compares himself to his peers and finds himself well off - or “relatively deprived” - and sees an opportunity to improve his and rank order by migration, he will have a strong incentive to do so. This effect may be captured by including a variable that measures relative deprivation in the context of the local community.

In sum, if the Harris-Todaro model holds, then differentials in expected income per capita should perform better as an explanatory variable than the differential in average income. If low income or high poverty implies a liquidity trap for potential migrants, then the deterrent effect of distance should be higher. If urbanization of the destination region has an independent significant impact on migration, then Stark’s argument on risk diversification is empirically supported. Finally, if Stark’s hypothesis on relative deprivation holds, then a variable capturing inequity in the source income distribution should be significant. These different hypotheses are not mutually exclusive and may hold simultaneously. Several of these hypotheses are tested for in empirical part of the article.

Econometric issues

Modified gravity models are usually estimated in double logarithmic form so that coefficients can be interpreted as elasticities and that linear estimation techniques can be applied. A typical model, including relative income, is for example (Fields, 1979)

$$\ln M_{ij} = a_0 + a_1 \ln D_{ij} + a_2 \ln(P_i P_j) + a_3 \ln(Y_i / Y_j) + \varepsilon_{ij} \quad (2)$$

A more general formulation is

$$\ln M_{ij} = \beta_0 + \beta_1 \ln D_{ij} + \beta_2 \ln P_i + \beta_3 \ln P_j + \sum_n \alpha_n \ln X_{ni} + \sum_m \alpha_m \ln X_{mj} + \varepsilon_{ij} \qquad (3)$$

with X_{ni} are presumed determinants in location i and X_{mj} potential determinants in location j.

A third class of models are so-called "systemic gravity models" (Hunt & Greenwood, 1985). Such models explicitly recognize that the flow of migration from location i to j depends upon the attractiveness of location j but compared to all other possible locations a migrant can choose to go to. These models include features of push, pull and cost, not only for the region of destination but for all potential destinations.

Hence, to include the potential effect of other options a migrant has, equation (3) is further modified to

$$\ln M_{ij} = \beta_0 + \sum_j \beta_{1j} \ln D_{ij} + \beta_2 \ln P_i + \sum_j \beta_{3j} \ln P_j + \sum_n \alpha_n \ln X_{ni} + \sum_j \sum_m \alpha_{jm} \ln X_{jm} + \varepsilon_{ij} \qquad (4)$$

These different gravity models are usually estimated in its linear double logarithmic form as in equation (2), (3) or (4). Several problems are associated with this procedure (Schultz, 1982).

Zero migration flows

As gravity models are usually estimated in double logarithm, zero flows between regions pose a problem. Several options are open to deal with zero flows.

First, observations with zero flows may be omitted but this biases the regression results as the sample is truncated.

Second, an alternative is to estimate a Tobit model or censored regression model, using maximum likelihood (Verbeek, 2008, 230-235). There is some economic rationale to use the censored regression model. People in an origin decide first on whether or not to migrate, and second, if they do so, the decision on the destination on comparing attractions at destinations and repulsions at the origin.

Third, one could add 1 to all migration flows before taking logarithms and estimate the equation with scaled OLS (SOLS). This procedure boils down to multiplying the OLS estimators by the reciprocal of the proportion of non zero migration flows (Lewer & Van den Berg, 2008).

Non-migration and spurious correlation with population size

Usually regions differ substantially in population and size. It is likely that large areas have a larger share of within area migrations. These within area migrations go unobserved. Apparently there will be more non-migration and less migration in these large areas compared to smaller areas. Hence, migration will be spuriously (negatively) correlated with the size of population at the origin.

To also include information on the relative importance of non migration, as well as to recognize that the destination is picked out of range of alternative destinations, a logistic specification is advocated. (Greenwood & Hunt, 2003).

In a logistic formulation, the underlying assumption is that an individual's decision to migrate from i to j is specified as (Fields, 1979)

$$P_{ij} = \frac{e^{z_{ij}}}{\sum_j e^{z_{ij}}} \tag{5.a}$$

where $\sum_j P_{ij} = 1$ (5.b)

The values of z are (log) linear functions of the origin and destination determinants and distance or

$$z_{ij} = \beta_0 + \sum_i \beta_m \ln X_{mi} + \sum_j \gamma_m \ln X_{mj} + \delta \ln D_{ij} \tag{6}$$

By substituting (6) in (5) and rearranging the logistic form of the gravity model is obtained, namely

$$\ln\left(\frac{P_{ij}}{P_{ii}}\right) = \tilde{\beta}_0 + \sum_i \tilde{\beta}_m \ln X_{mi} + \sum_j \tilde{\gamma}_m \ln X_{mj} + \tilde{\delta} \ln D_{ij} \tag{7}$$

Note however that, if the variation in the share of non migrants is small so that P_{ii} is almost constant, then the logistic model will yield similar results to a log-log formulation.

Bilateral variables

Logistic gravity models such as (7) usually contain "bilateral variables" such as distance between regions, relative income differentials, population ratios, etcetera. However, there may be specific influences of one destination region that are common across all source regions or common across all sources of a destination country. Not taking into account such influences implies clustering of standard errors into the coefficients of bilateral variables and this may bias estimates. A dummy for each source and each destination may be added to equation (7) to capture such region specific effects (Redding & Venables, 2004).

Simultaneity bias

Migration is influenced by current economic conditions in source and destination locations. However, migration itself – if substantial - may affect current economic conditions at both locations. Hence, a simultaneity bias is real. The risk of simultaneity may be minimized by choosing all independent values at the base year of the migration flow. Even this precaution may not entirely exclude simultaneity between migration and population. Present population is likely to be influenced by past migrations, itself the results of past economic conditions. As present conditions are strongly correlated with past conditions, there is a risk of simultaneity when including population as an independent variable.

3. Data

3.1. Dependent variable

The dependent variable is observed migration flows (M_{ij}) or the observed flows relative to population of source and destination ($p_{ij}=M_{ij}/(P_i.P_j)$) between 17 locations in Vietnam. As the focus is on migration in and from the MRD the flows cover interprovincial flows in the 13 provinces of the MRD. As most migrants from the MRD region migrating to the rest of the country mainly go to the three major cities (provinces) with more than 250,000 inhabitants - Ho Chi Minh city, Binh Duong and Ha Noi - these three cities (provinces) are also included. The rest of Vietnam is included as a 17^{th} location to cover the complete system of migration flows in Vietnam. Data on migration flows are directly derived from the Population Census 2009, reporting on the population of age 5 and over that changed its usual province of residence between 1/4/2004 and 1/4/2009. [Source: (VGSO, 2010a, 242-277)].

3.2. Independent variables

Distances (in km)

The distances between provinces and cities are based on line distance measurements between the approximate centers of gravity in each of the provinces (using the Google Earth measurement tool). Distances between all MRD provinces and between MRD provinces and the 3 major cities can be directly measured.

The "distance" between an MRD province and "the rest of Vietnam" is calculated as the weighted average distance between the approximate center of gravity of each MRD province and the approximate center of gravity of the different regions of Vietnam (other than MRD provinces and the 3 cities), with the share of each region in total out-migration from the MRD province to the rest of Vietnam as weight or

$$\bar{d}_{ir} = \sum_r \frac{M_{ir}}{\sum_r M_{ir}} d_{ir} \qquad (8)$$

A similar approach is taken for the "distance" between the 3 cities and "the rest of Vietnam".

Other variables

Data on provincial population size, the rate of unemployment and the degree of urbanization are from the Statistical Yearbook 2010 (VGSO, 2010c). The data on provincial average income per capita and the provincial poverty rate data are from the Vietnam Household Living Standard Survey 2006 and 2010 (VGSO, 2010d).

In order to minimize simultaneity population data are from 2004, the start of the period (see Fields (1979) for a similar approach). Data for all other variables are averages for the period 2004-2009 except for the poverty rate where data for 2006 are used as earlier data on this variable are not available.

In order to test Stark's relative deprivation hypothesis, a local inequality measure should be used. In the VHLSS the percentage of households in each province with an income below a national minimum standard (y') is reported (p). Also the average household income in each province (y") is known. One option is to use this reported poverty rate in the multivariate analysis. However, this poverty rate is defined against a national standard and not against a local standard. Relative deprivation typically refers to the rank position in the local income distribution. An alternative is to use a measure of local inequality such as a Gini coefficient. This coefficient is estimated as follows. Assume that the local income distribution follows a Pareto distribution defined by two (unknown) parameters *ym* and *alfa*. The cumulative distribution or the fraction of people F(y) with an income less than y equals

$$F(y) = 1 - \left(\frac{ym}{y}\right)^{\alpha} \qquad (9)$$

If the local income distribution follows a Pareto distribution, then it can be shown that the Gini coefficient equals to

$$G = 1 - \frac{1}{2\alpha - 1} \qquad (10)$$

We know the fraction of people p below the national poverty standard y' and the provincial average income y" in the province. Hence for each province, it holds that

$$F(y') = p = 1 - \left(\frac{ym}{y'}\right)^{\alpha} \qquad (11.a)$$

$$E(y) = \frac{\alpha . ym}{\alpha - 1} = y'' \qquad (11.b)$$

These two equations form a non linear system of equations with two unknown provincial income distribution parameters *alfa* an *ym*. Solving for *alfa* and *ym* specifies the local provincial income distribution. With the parameter *alfa*, the provincial Gini coefficient – a measure of local inequality – can be calculated. Relative deprivation at the level of the province can be approximated by the Gini coefficient for the province as an alternative to the provincial poverty rate.

3.3. Descriptive statistics

Dependent variables - Mij and pij

Table 1 summarizes the descriptive statistics of the dependent variables.

Table 1

Descriptive Statistics Dependent (N=272)

Variable	Mean	Std.dev.	Min	Max
M_{ij}	8973.2	45016.2	4.000	567049
p_{ij}	0.997	0.066	0.955	0.999
p_{ii}	0.003	0.066	0.000	0.045

First, it is important to note that there are no zero migration flows. Hence, there is no immediate need to bias the sample by omitting zero flows or for the use of a corrective procedure such as Tobit or SOLS. However, the distribution of flows is positively skewed (skewness = 9.80). The skewness of this variable is predominantly due to the very large migration flows to the urban areas of Ho Chi Minh City and Binh Duong and flows to the aggregate area grouped as "the rest of Vietnam". This area was added to cover the total of all internal Vietnamese migration flows and avoid sample selection bias. This positive skewness should not necessarily be a problem as an important explanatory variable, namely distance, is also positively skewed (skewness distance = 2.40). However, in view of this skewed dependent variable, it seems especially appropriate to check for normality of error terms in explanatory models.

Second, the share of non-migrants in each province (p_{ii}) shows little variation as the coefficient of variation (standard deviation on mean) is less than 1%. That implies that the bias from not taking into account non-migrants because of possible correlation between size of region and non accounted for internal migration is minimal. Hence, models based on relative flows such as in equation (7) are not explored further here.

Independent variables

In Table 2 the descriptive statistics for the independent variables are listed.

As Vietnam is a large S shaped country, the distribution of distances is positively skewed with distances between provinces ranging from less than 20km to over 2000 km with an average of about 350km.

Relative average income and relative expected income is highly correlated as the variation in unemployment rates is relatively low (ranging from 3.7 to 5.0%). On average the income premium of a destination province over a source country is relatively low (some 8.5-8.6%). However, the variation in relative income is wide, ranging from 0.35 to 2.85.

Also, the population distribution is skewed. Within the MRD region, population size of provinces ranges from about 0.75 million in Hau Giang to 2.1 million in An Giang. Large provinces are Ho Chi Minh City (6.0 million) and Ha Noi (3.0 million). The maximum value of 54.5 million is the population for the aggregate region “rest of Vietnam”.

Table 2

Descriptive Statistics Independent Variables

Variable		Mean	Std.dev.	Min	Max
D_{ij}	Distance source-destination (km)	337.7	563.0	13.7	2070.0
Y_j/Y_i	Relative average income destination/source	1.086	0.466	0.361	2.850
EY_j/EY_i	Relative expected income destination/source	1.085	0.464	0.352	2.838
POP_i	Population (in 1000 units) source (destination)	4925	12406	754	54105
URB_i	Share of urban population (%)	27.49	18.70	9.57	82.57
POV_i	Poverty rate (%)	11.12	5.69	0.40	21.45
$GINI_i$	Gini coefficient	0.485	0.058	0.317	0.572
$UNEMP_i$	Unemployment rate (%)	4.289	0.390	3.763	5.004

The degree of urbanization varies from about 10% (Ben Tre) to over 80% (Can Tho). On average somewhat more than ¼ of the population is urbanized.

The average poverty rate (an absolute standard) is 11% but ranges from less than 1% in the cities of Binh Duong and Ho Chi Minh City to over 20% in the rural area of Tra Vinh. Correspondingly, Gini coefficients are lowest in the cities (around 0.32) but reach over 0.50 in some rural areas (for example Tra Vinh).

3.4. Bi-variate analysis

Bi-variate analysis offers an initial indication of the validity of the different explanatory hypothesis on migration flows.

From Figure 3 it follows that size of origin and destination population clearly matter for the volume of migration flows. The coefficient of determination between the natural log of migration flows and the natural log of the product of origin and destination population (R^2=0.475) is highly significant (better than 1%).

Figure 3

Migration Flows and Population Size (Gravity)

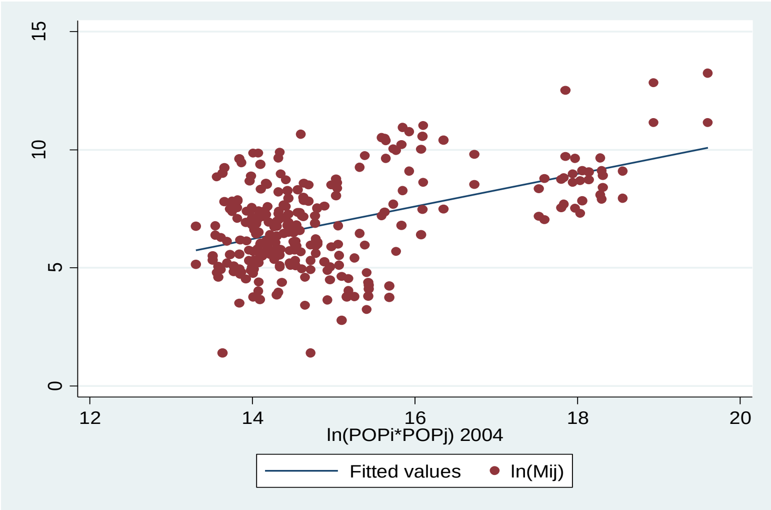
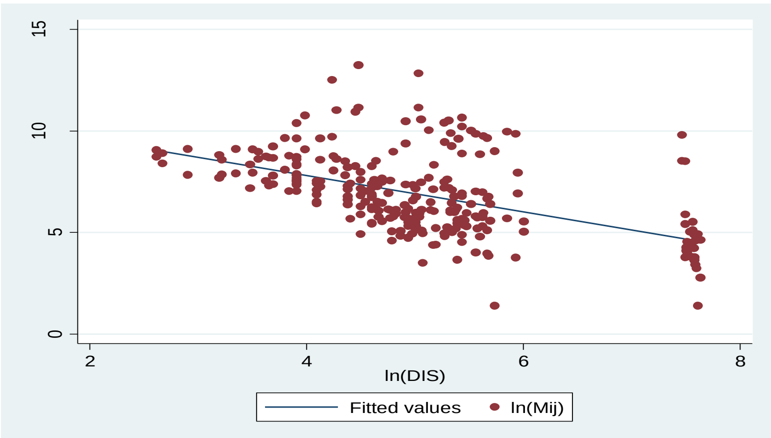


Figure 4 shows the relationship between the natural log of migration flows and the natural log distance – a proxy for the cost of migration. There is a clear and significant (better than 1%) negative relationship (R^2=0.513) between both variables supporting the hypothesis that distance (cost) is a deterrent to flows.

Figure 4

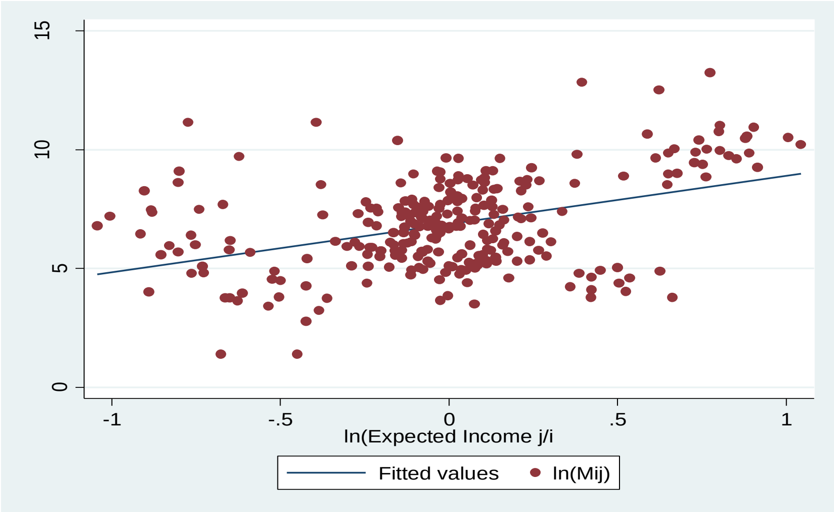
Migration Flows and Distance (Cost)



Expected relative income (or relative income taking into account the probability to get employment) between source and destination also is positively correlated to migration flows, as follows from Figure 5, supporting the Harris-Todaro insight. The correlation is strong (R^2=0.418) and significant (better than 1%). There is no obvious indication from the graph of a "liquidity trap" or a non-linearity at the low end of income. However, this will be checked further in the multivariate analysis in relation with distance (cost).

Figure 5

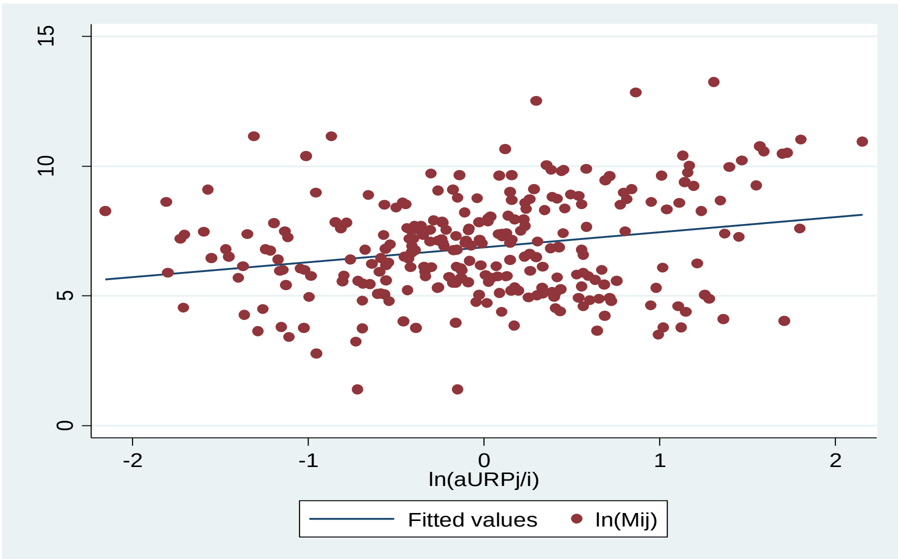
Migration Flows and Relative Expected Income (Harris-Todaro)



The attractiveness of migration of family members to urban areas – even in the absence of better income prospects – as an option to cover family risk was put forward by Stark and others. Figure 6 offers some preliminary and tentative evidence in support of this as there is a positive but weak relationship between relative urbanization and migration flows (R^2=0.233, significance better than 1%). However, this bi-variate analysis may be misleading as higher urbanization is correlated with higher income and its independent effect can only be checked in a multivariate model.

Figure 6

Migration Flows and Urbanization (Stark)



Finally, another hypothesis offered by Stark is that relative deprivation is an explanatory factor for migration. Figure 7 is a scatter between migration flows and the (estimated) Gini coefficient at origin. A positive relationship would be expected if deprivation (or inequality) is conducive to migration.

From the graph, there is no significant relationship (R^2=0.017). However, if one omits the flows associated with more equal areas (coinciding with the urban areas such as Ho Chi Minh City and Binh Duong), then some positive relationship for more rural areas may be discerned.

Figure 7

Migration Flows and Inequality at Origin

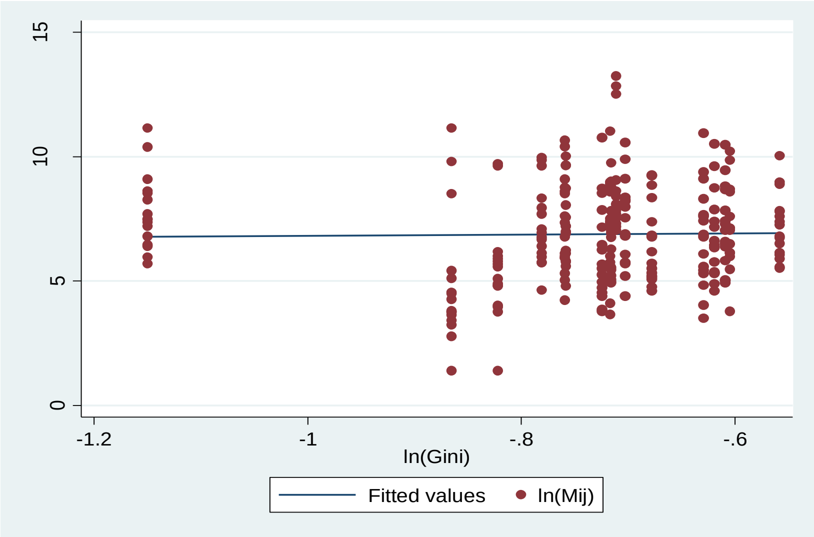
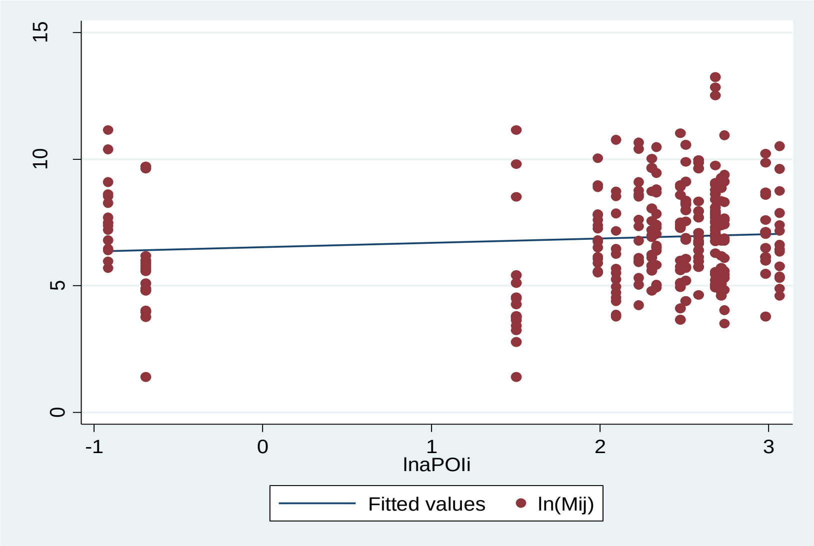


Figure 8

Migration Flows and Poverty Rates



In Figure 8 an alternative measure to capture the effect of deprivation namely the poverty rate is used. High poverty (or a possible large group of relatively deprived persons) should be conducive to migration. However, again no significant relationship is found (R^2=0.098).

4. Multivariate analysis

4.1. Basic gravity model and relative income

In Table 3 regression results for the basic gravity model and two models with relative income added are reported. All models were tested for heteroskedasticity (White test). OLS estimates for models 2 and 3 suffered from heteroskedasticity and robust standard errors were estimated.

All three models show a decrease in migration flows with -0.74% per percent increase in distance. This distance or cost elasticity is statistically significant from zero (and one) and precisely estimated (standard error of 0.09).

The estimates show that migration flows approximately vary in proportion with the square root of population at source and at destiny. The exact elasticity from all three models is 0.541 and is fairly accurately estimated.

Models show that relative income is a very important variable. Including this variable (model 2 and model 3) increases the explanatory power of the basic gravity model to a modified gravity model with more than 20% as the R^2 increases from 0.394 to 0.569.

The effect of an income premium of destination over source is substantial. Migration flows increase with the square of the relative income ratio or a doubling of relative income leads to a fourfold increase in migration flows, etc.

Table 3

Basic Gravity Model and Relative Income - Dependent ln(M_{ij})

	Model 1 *(b/se)*	Model 2 *(b/se)*	Model 3 *(b/se)*
Ln(DIS)	-0.737***	-0.737***	-0.737***
	(0.09)	(0.08)	(0.07)
Ln(POP_i*POP_j)	0.541***	0.541***	0.541***
	(0.07)	(0.08)	(0.06)
Ln(Y_j/Y_i)		2.022***	
		(0.24)	
Ln(EY_j/EY_i)			2.031***
			(0.19)
Constant	2.505*	2.503	2.505*
	(1.24)	(1.31)	(1.05)
R^2	0.394	0.569	0.569
N	272	272	272

* p<0.05, ** p<0.01, *** p<0.001

There is as no difference between model 2 – where relative average income is used – and model 3 – with relative expected income. Both models have the same predictive power and coefficients are practically equal. This could be expected as low unemployment and low variation in unemployment rates over provinces lead to high correlation between average income and expected income. Due to this, the expectancy aspect of the Harris-Todaro model cannot really be verified in this case. However, the empirical evidence supports the general economic theory that migration is strongly determined by the comparison between income prospects at destination with income prospects at source and that flows are deterred by costs (distance).

4.2. Augmented gravity models

In Table 4 estimation results of modified gravity models – i.e. models including population, distance and relative income – augmented with additional variables are reported. These models test for a liquidity trap of restraining migration, an autonomous effect of urbanization (risk sharing by urban migration) or migration out of relative deprivation. Although the present data at the more aggregate level of a province are not ideal to test these micro assumptions at family or individual level, it seems worthwhile to prompt for possible confirmation.

First, the augmented gravity models add some 15 to 19% in explanatory power. In terms of explanatory power and significance of coefficients model 5 seems to dominate model 4. The augmented models yield smaller elasticities for population size (almost half the value in model 5 compared to models 1 to 3) but yield relative income elasticities that are almost double those from the basic models. A possible explanation may be that previous models clustered more influences of different variables with counteracting effects into a single variable namely relative income.

Table 4

Augmented Gravity Model - Dependent $\ln(M_{ij})$

	Model 4 *(b/se)*	Model 5 *(b/se)*
Ln(DIS)	-0.823***	-0828***
	(0.15)	(0.07)
$Ln(POP_i*POP_j)$	0.410***	0.280***
	(0.05)	(0.05)
$Ln(Y_j/Y_i)$	5.094***	5.444***
	(0.32)	(0.30)
$Ln(POV_i)*ln(DIS)$	-0.057	-0.118***
	(0.07)	(0.02)
$ln(URB_j/URB_i)$	-0.782***	-0.757***
	(0.13)	(0.12)
$ln(POV_i)$	-0.672	
	(0.35)	
ln(Gini)		-5.347***
		(0.76)
Constant	6.876***	4.170***
	(1.10)	(0.85)
R^2	0.721	0.761
N	272	272

* p<0.05, ** p<0.01, *** p<0.001

Both models (model 4 and model 5) include a variable to test for a possible "liquidity trap" for poor migrants. Costs may be particularly prohibitive or restrictive for low income migrants, lacking funds or capital to finance the cost of migrating. This is tested by including an interaction term between the poverty rate and distance. If cost is more of a concern for provinces with a high percentage of poor, then the deterrent effect of distance on migration flows would be larger. Hence, a negative interaction term would be indicative of a liquidity trap. The estimated results seem to confirm the hypothesis of a liquidity trap. The coefficients of the interaction term are relatively small and have the correct sign. The coefficient is statistically significantly different from zero and rather precisely estimated in model 5. As (relative) poverty is also included directly in model 4, co-linearity between the interaction term and this variable renders the estimate of the interaction term less accurate. Taking the estimate of model 5, the coefficient implies that an increase in the number of poor in a province with one percent implies that the elasticity of distance with respect to migration flows increases from -0.83 to -0.95. Hence, keeping all other factors constant, poor people will tend to migrate to less distant destinations.

Both models also incorporate the rate of urbanization of the destination relative to the rate of urbanization of the source area. An autonomous effect of relative urbanization may be an indication for risk spreading strategies of agricultural families. The autonomous urbanization effect is large and statistically significant but has the wrong sign! This does not confirm the earlier finding in the bi-variate analysis. This negative effect may be explained as a congestion effect, i.e. that more urbanization – ceteris paribus ultimately leads to a more expensive and less attractive way of life. However, this hypothesis is difficult to test with these date. Also, strong co linearity between urbanization, population and relative income may be a reason for this sign reversal.

Finally, some indicators for relative deprivation are included. In model 4 the absolute poverty rate at source is included and in model 5 the estimated Gini coefficient is put in as an alternative. The estimates are problematic in both models. In model 4 the estimated coefficient is negative, implying that poverty at the source is a deterrent but statistically not significant. This deterrent effect would be on top of the interaction effect with distance. The result on the Gini coefficient in model 5 is puzzling. A larger Gini or more inequality at the source would dampen migration, which is contrary to expectations. One would expect more relatively deprived persons with more inequality and hence more migration if Stark's theory of relative deprivation prevails. However, these aggregate data are not ideal to test this micro level hypothesis.

5. Forecasting migration flows 2009-2014

Gravity models are very informative for policy. For example, the large impact of relative income on migration flow indicates that migration is highly sensitive to unbalanced development of the economy. Growing divergence of income per capita between provinces will have a more than proportional effect on migration and differentially impacting future demands for living space, education, health provisions in the richer areas. Declining poverty reduces the deterrent effect of migration in poor areas as the liquidity trap is less stringent adding to immigration pressures in traditional destination areas.

To put a numerical dimension on such future policy challenges, migration flows forecasts are required. Gravity models are well suited for forecasting. A modified gravity model with n regions and with distance, population and relative income as independent variables requires only 2n forecasts of independent variables to generate forecasts for n(n-1) migration flows (assuming distances and parameters constant over time).

In order to forecast migration flows for the period 2009-2014, a final model was estimated leaving out more problematic parameters such as those on income distribution and degree of urbanization. The following model is selected for forecasting purposes:

Table 5
Augmented Gravity Model For Forecasts - Dependent ln(Mij)

	Model 6 *(b/se)*
Ln(DIS)	-0.578***
	(0.06)
$Ln(POV_i)*ln(DIS)$	-0.168***
	(0.02)
$Ln(POP_i*POP_j)$	0.412***
	(0.05)
$Ln(Y_j/Y_i)$	3.760***
	(0.25)
Constant	5.352***
	(0.96)
R^2	0.677
N	272

* p<0.05, ** p<0.01, *** p<0.001

All coefficients in this model have small standard errors and are statistically different from zero with better than 1% significance. The model explains somewhat more than 2/3 of total variation in migration flows.

Recall that this model is estimated based on the migration flows covering a five year period from 2004 to 2009, using population data of 2004 (to minimize simultaneity problems) and income, poverty and urbanization data based on average values or mid period values for the period 2004-2009.

To construct a forecast of migration flows for the next five year period 2009-2014, consistent with the timing of data inputs used in parameter estimation model, non forecasted data inputs namely interprovincial distances (fixed) and observed population data 2009 are required, but also forecasts for the period averages 2009-2014 of the other independent variables namely income and poverty.

Forecasts of future income for each province are calculated using a simple extrapolation method or

$$Y_{it} = Y_{i0}(1+r_i)^t \tag{12}$$

Assuming that the growth rate of income in a province during 2009-2014 (r_i) is equal to the growth rate observed over 2004-2009.

Forecasts for poverty are based on an inverse relation (as the poverty rated is bounded from below at A%) namely

$$POV_{it} = \frac{A_i}{B_i + t} \tag{13}$$

Observed poverty rates in 2004 and in 2009 are used as reference points to derive the parameters A and B.

Finally, the estimated error term for each observation of the forecasting equation for the period 2004-2009 is added to take into account observation specific factors not taken into account by the independent variables included in the estimated forecasting equation. The observed migration flows 2004-2009 and the forecasted flows 2009-2014 are reported in Appendix.

Table 6 summarizes the row totals (out migration) and column totals (in migration) for all locations.

Table 6

Migration flows from the MRD region and 3 major cities (2004-2009 & 2009-2014)

	Out-migration		**In-migration**	
	2004-2009	2009-2014	2004-2009	2009-2014
Long An	65.331	82.653	39.533	40.990
Tien Giang	89.891	101.006	24.368	30.479
Ben Tre	91.280	88.219	13.569	20.033
Tra Vinh	66.702	83.235	11.042	12.293
Vinh Long	71.107	73.599	21.811	31.518
Dong Thap	88.252	143.596	19.029	16.422
An Giang	108.149	185.865	18.382	20.310
Kien Giang	71.431	117.905	19.907	20.914
Can Tho	52.127	48.397	55.865	84.013
Hau Giang	37.395	57.434	11.675	10.754
Soc Trang	67.358	104.791	11.428	11.149
Bac Lieu	42.673	59.604	6.323	7.964
Ca Mau	70.618	139.774	7.965	6.799
Ha Noi	92.773	94.584	382.832	298.356
Binh Duong	34.732	21.058	500.003	1.189.176
HCM city	137.031	362.090	1.033.028	770.783
Rest of VN	1.253.862	1.220.727	263.952	412.583
Total	**2.440.712**	**2.984.536**	**2.440.712**	**2.984.536**

First, migration will remain a major issue in Vietnam. Flows over the period 2009-2014 are expected to amount to almost 3 million people or an increase with more than 0.5 million people or 22% compared with 2004-2009. Dealing with the consequences of such large flows for land use, housing, education, health care and the job market will be a major policy challenge.

Second, the table shows some major shifts in out-migration to the major cities of Vietnam. Ho Chi Minh city will no longer be the main destination in the coming period with in migration flows declining from 1 million to 0.77 million. Binh Duong will be the main pole of attraction of the future with flows increasing from 0.5 million from 2004-2009 to almost 1.2 million in 2009-2014. Finally, in flows in Ha Noi – previously 0.4 million – will decline to less than 0.3 million.

Third, the MRD region will continue to be a major source of migrants. Total out-migration will increase with almost 40% from 922.000 in 2004-2009 to 1.286.000 in 2009-2014. The growth of in-migration in the region will be much smaller (20%) from 261.000 to 314.000 in-migrants. All provinces – except Can Tho – will remain net sources of migrants. The city of Can Tho – with an almost equal number of in- and out- migrants in 2004-2009 – can expect an excess of 36.000 in-migrants over out-migrants. Net-out migration of all provinces of the MRD will increase except for Can Tho but also for Ben Tre and Vinh Long where a slight decrease in net-out migration can be expected. Provinces with the largest increase in out-migration are Ca Mau – with net out-migration expected to double – but also – all areas quite close to the urban attraction pole of Can Tho.

6. Conclusions

In this article migration flows in the period 2004 to 2009 between the 13 provinces of the Mekong Delta River region, 3 cities (Ha Noi, Binh Duong and Ho Chi Minh City) and the rest of Vietnam were modeled using basic modified and augmented gravity models. These basic modified models include distance as a proxy for cost, population sizes of source and destination and relative income. As there are no zero flows, models were estimated with standard OLS correcting standard errors when heteroskedasticity was detected. To avoid simultaneity problems independent variables base year data for the independent variables were used. The basic modified model explains about 57% of the variation in provincial migration flows over this 5 year period and which range from a low of 4 to a high of over 0.5 million. The basic modified model shows that migration flows between provinces of the MRD (and cities and the rest of Vietnam) approximately vary with the square root of the product of province populations and with the square of the ratio of income at destination over income at source. Migration flows vary inversely with distance and the estimated elasticity between distance and migration is about -3/4.

The basic modified model is augmented with additional variables with the purpose of testing some theories on migration. More specifically, four hypothesis are tested namely whether (i) expected relative income – combining income with job opportunities - is a better predictor of migration flows than simply relative average income, (ii) lack of funds and poverty may inhibit the poor to migrate (iii) urbanization has an independent effect perhaps as the result of a family risk diversification strategy and (iv) feelings of relative deprivation resulting from poverty or income inequality at a source are enhancing migration.

Augmenting the basic modified model with additional variables adds some 15 to 19 percent to explanatory power with more than ¾ of all variation in migration flows explained. From the estimated coefficients it follows that the deterrent from distance is larger in provinces with more poor. Hence, there is some support for a "liquidity trap" at work. Urbanization seems to have a strong independent effect however opposite to what is expected. Poverty or income inequality yields non significant results.

The results broadly confirm standard economic investment theory on explaining migration flows, namely that higher expected returns (relative income) and lower costs (distance) are major explanations for observed flows. Findings do confirm the idea that lack of resources to migrate limits the poorest but not the presumed impact of inequality and urbanization. However, a major caveat of these findings is that the data used here, namely aggregates at the provincial level, are not ideal to test theories that are formulated an individual level or household level. A second caveat is that causal relations are difficult to argue with cross section data and strictly panel data should be used to verify such relationships. Further research is required to test these micro level data preferably by using individual panel data.

Forecasts for the period 2009-2014 show that a substantial increase in migration flows can be expected from some 2.5 million people in 2004-2009 to about 3.0 million people for the next five years. Apparently in flows into Ho Chi Minh city are expected to come down from over 1 million in 2004-2009 to about 0.8 million over the next five years. Binh Duong will see the largest inflows – 1.2 million – up from 0.5 million in 2009-2004. It will be the fastest growing urban area in Vietnam. The MRD region remains an important out-flow region with out-flows from provinces increasing from 0.9 million to 1.3 million in the next five years. All provinces will remain sending areas, except for the urban area of Can Tho. The provinces in the neighborhood of Can Tho such as Ca Mau, Kien Giang, Dong Thap and An Giang will see the largest increases in out flows.

Appendix A. Observed and forecasted migration flows

Observed flows 2004-2009

Ori(down)/Des(right)	Column1	Long an	Tien Giang	Ben Tre	Tra Vinh	Vinh Long	Dong Thap	An Giang	Kien Giang	Can Tho	Hau Giang	Soc Trang
		1	2	3	4	5	6	7	8	9	10	11
Long An	1	0	2568	640	244	291	1279	192	141	511	111	80
Tien Giang	2	5328	0	1548	310	1621	1801	402	276	1426	139	164
Ben Tre	3	2132	4087	0	968	1650	886	440	203	1971	124	230
Tra Vinh	4	1275	577	628	0	2613	314	315	213	1637	206	746
Vinh Long	5	1657	1137	670	2527	0	1466	722	335	5848	593	679
Dong Thap	6	2960	1883	916	432	3693	0	4318	984	3925	296	309
An Giang	7	2006	873	397	370	1528	5382	0	6329	4973	444	370
Kien Giang	8	1340	509	315	267	1079	1321	3159	0	6115	1916	448
Can Tho	9	855	676	356	463	1604	896	2484	1967	0	2446	1397
Hau Giang	10	478	207	161	244	874	115	298	922	10252	0	1606
Soc Trang	11	1236	456	236	1235	1143	395	455	664	5333	1991	0
Bac Lieu	12	1022	232	141	158	536	185	253	1121	1781	859	1889
Ca Mau	13	860	790	608	312	1197	303	385	2820	4112	452	1042
Ha Noi	14	226	71	94	38	88	45	42	25	166	30	43
Binh Duong	15	476	320	401	266	123	163	287	53	131	121	55
HCM city	16	8992	5541	3903	1328	1553	1736	1789	604	2211	643	895
Rest of V	17	8690	4441	2555	1880	2218	2742	2841	3250	5473	1304	1475
Sub-IM by Pro		39533	24368	13569	11042	21811	19029	18382	19907	55865	11675	11428

Bac Lieu	Ca Mau	Ha Noi	Binh Duong	HCM city	Rest of VN	Sub-OM by pro	m)/Des(right)2
12	13	14	15	16	17		
91	47	44	5063	47871	6158	65331	Long An
39	165	61	7895	61377	7339	89891	Tien Giang
33	266	56	11850	57314	9070	91280	Ben Tre
99	203	131	14929	36625	6191	66702	Tra Vinh
136	152	154	12810	35560	6661	71107	Vinh Long
81	181	80	19791	39276	9127	88252	Dong Thap
153	201	69	42993	33123	8938	108149	An Giang
327	1017	121	15569	22295	15633	71431	Kien Giang
258	425	250	7224	22912	7914	52127	Can Tho
171	182	98	6972	10580	4235	37395	Hau Giang
1203	464	44	19168	27372	5963	67358	Soc Trang
0	2476	136	8239	17149	6496	42673	Bac Lieu
2207	0	102	19173	21054	15201	70618	Ca Mau
4	16	0	4902	18124	68859	92773	Ha Noi
4	43	360	0	15347	16582	34732	Binh Duong
386	297	5034	32534	0	69585	137031	HCM city
1131	1830	376092	270891	567049	0	1253862	Rest of V
6323	7965	382832	500003	1033028	263952		Sub-IM by Pro

Forecasted flows 2009-2014

Ori(down)/Des(right)	Column1	Long an	Tien Giang	Ben Tre	Tra Vinh	Vinh Long	Dong Thap	An Giang	Kien Giang	Can Tho	Hau Giang	Soc Trang
		1	2	3	4	5	6	7	8	9	10	11
Long An	1	0	3416	987	330	478	1284	244	175	879	124	92
Tien Giang	2	5076	0	2027	367	2276	1542	451	303	2144	136	165
Ben Tre	3	1681	3810	0	884	1853	612	381	171	2321	94	179
Tra Vinh	4	1136	615	747	0	3328	245	306	203	2181	177	656
Vinh Long	5	1264	1028	681	2230	0	980	607	279	6588	435	514
Dong Thap	6	3362	2567	1394	573	6034	0	5359	1191	6675	323	347
An Giang	7	2026	1056	538	435	2208	4666	0	6708	7417	428	367
Kien Giang	8	1493	681	468	349	1743	1270	3740	0	9876	2029	493
Can Tho	9	618	585	344	391	1651	567	1950	1505	0	1685	995
Hau Giang	10	492	255	222	293	1296	104	334	1014	15883	0	1639
Soc Trang	11	1254	554	322	1469	1671	352	503	718	8141	1948	0
Bac Lieu	12	1040	282	191	185	767	164	284	1220	2668	832	1844
Ca Mau	13	1084	1191	1027	459	2155	334	527	3808	7692	545	1279
Ha Noi	14	117	44	65	23	65	20	23	14	128	15	22
Binh Duong	15	242	195	273	158	90	73	159	29	100	59	28
HCM city	16	13981	10452	8320	2571	3629	2491	3223	1078	5440	1026	1478
Rest of V	17	6125	3749	2428	1578	2272	1718	2218	2498	5883	899	1050
subtotal		40991	30481	20036	12297	31523	16428	20317	20922	84022	10764	11160

Bac Lieu	Ca Mau	Ha Noi	Binh Duong	HCM city	Rest of VN	Sub-OM by pro	m)/Des(right)2
12	13	14	15	16	17		
126	50	56	15575	48922	9914	82653	Long An
48	154	70	21323	54857	10067	101006	Tien Giang
31	191	48	25088	40651	10224	88219	Ben Tre
104	163	123	35690	29519	8042	83235	Tra Vinh
123	106	129	26613	24738	7284	73599	Vinh Long
108	184	93	59910	40162	15316	143596	Dong Thap
185	183	75	116480	30194	12900	185865	An Giang
434	1036	148	46894	22567	24684	117905	Kien Giang
221	278	192	14081	15020	8313	48397	Can Tho
208	168	103	19165	9819	6437	57434	Hau Giang
1451	424	46	51957	25040	8942	104791	Soc Trang
0	2185	151	22611	15836	9343	59604	Bac Lieu
3187	0	138	64974	24030	27347	139774	Ca Mau
2	7	0	6807	8529	78702	94584	Ha Noi
2	20	189	0	7052	12390	21058	Binh Duong
768	454	9572	134930	0	162677	362090	HCM city
965	1194	287224	527077	373848		1220727	Rest of V
7976	6812	298370	1189191	770799	412600	2984689	

References

de Haas, H. 2010. "Migration and Development: A Theoretical Perspective." *International Migration Review* 44:227-264.

Fields, G.S. 1979. "Place-to-place migration: Some new evidence." *The Review of Economics and Statistics* 61:21-32.

Greenwood, M.J. & G.L. Hunt. 2003. "The early history of migration research." *International Regional Science Review* 26:3-37.

Harris, J & M Todaro. 1970. "Migration, unemployment and development: a two sector analysis." *American Economic Review* 60:126-142.

Hunt, G.L. & M.J. Greenwood. 1985. "Econometrically accounting for identities and restrictions in models of interregional migration:: Further Thoughts." *Regional Science and Urban Economics* 15:605-614.

Lewer, J.J. & H. Van den Berg. 2008. "A gravity model of immigration." *Economics Letters* 99:164-167.

Lucas, R E B. 1997. "Internal migration in developing countries." Pp. 721-798 in *Population and family economics*, edited by M. R. Rosenzweig and O. Stark. Amsterdam: Elsevier Science.

Niedercorn, J.H. & B.V. Bechdolt Jr. 1969. "An economic derivation of the "Gravity Law" of spatial interaction." *Journal of Regional Science* 9:273-282.

Redding, S. & A.J. Venables. 2004. "Economic geography and international inequality." *Journal of International Economics* 62:53-82.

Schultz, T.P. 1982. "Lifetime migration within educational strata in Venezuela: estimates of a logistic model." *Economic Development and Cultural Change* 30:559-593.

Sjaastad, L A. 1962. "The costs and returns of human migration." *The Journal of Political Economy* 70:80-93.

Stark, O. 1991. *The migration of labor*. Cambridge: Blackwell.

Stark, O. 1984. "Rural-to-urban migration in LDCs: a relative deprivation approach." *Economic Development and Cultural Change* 32:475-486.

Verbeek, M. 2008. *A guide to modern econometrics*: Chichester: John Wiley & Sons.

VGSO. 2010a. "The 2009 Vietnam Population and Housing census: major findings." in *Part III: Tabulated Tables*. Ha Noi, Vietnam: Statistical Publisher.

—. 2010b. "Migration and urbanization in Vietnam: patterns, trends and differentials." in *The 2009 population census*. Ha Noi, Vietnam: Statistical Publisher.

—. 2010c. "Statistical Yearbook of Vietnam 2010." Ha Noi, Vietnam: Statistical Publisher.

—. 2010d. "A summary of the result of Vietnam household living standard survey in 2010." Ha Noi, Vietnam: Statistical Publisher.

Zipf, G.K. 1946. "The P 1 P 2/D hypothesis: On the intercity movement of persons." *American Sociological Review* 11:677-686.