University of Wisconsin-Madison
Department of Agricultural & Applied Economics

Staff Paper No. 507, Rev. 2nd Revision, August 2008

Interprovincial Migration and Inequality During Vietnam's Transition

By

Diep Phan and Ian Coxhead

AGRICULTURAL &

APPLIED ECONOMICS

STAFF PAPER SERIES

Interprovincial Migration and Inequality During Vietnam's Transition*

Diep Phan[†] and Ian Coxhead[‡]

August 4, 2008

Staff Paper # 507

Department of Agricultural and Applied Economics

University of Wisconsin – Madison

Abstract

Vietnam's economic boom during the transition to a market economy has centered on very rapid growth in some sectors and provinces, yet poverty has diminished across the entire country. With capital investments highly concentrated by province and sector, geographic labor mobility may be critical in spreading the gains from growth. Conversely, rising income inequality may be attributable in part to impediments to migration. We first use census data to investigate migration patterns and determinants. We then examine the role of migration as an influence on cross-province income differentials. The former analysis robustly confirms economic motives for migration but also suggests the existence of poverty-related labor immobility at the provincial level. Examination of income differentials between pairs of provinces reveals that the impact of migration on inequality can be either negative or positive. A robust inequality-reducing impact of migration is found for migration flows into provinces where most of Vietnam's trade-oriented industrial investments are located.

1 Introduction

Economic growth in diversified economies is inherently unbalanced, because sectors have different factor intensities and factor endowments grow at differing rates. In the developing world, where capital endowments are typically highly concentrated by location, any new investments, price or

*Funding for this research comes from the Ford Foundation Office for Vietnam and Thailand. We are grateful to Paul Glewwe and to participants in the University of Wisconsin Development Workshop for helpful comments on an earlier draft.

[†]dnphan@wisc.edu

[‡]coxhead@wisc.edu

productivity shocks, or policy interventions that alter the value marginal products of factor inputs at a sectoral level thus induce a spatial reallocation of the relatively mobile factor, labor. In this way, economic growth and internal migration are complements: growth stimulates migration, and migration facilitates growth.

The growth-migration relationship is a source of many empirical questions with normative and policy implications. In this paper we address two such questions, using data on interprovincial migration from a rapidly-growing low-income economy, Vietnam. First, we inquire into the determinants of inter-provincial migration during an era of rapid growth, testing the extent to which labor flows between provinces can be explained by distance, income of sending and receiving provinces, and past migration. These variables have been found to be significant in other studies, but equally important, an investigation of this type can also yield information about impediments to migration. This is important to our second inquiry, into the links between migration and inter-provincial income inequality.

Since the 1986 adoption of the economic reform package known as doi moi, Vietnam's economy has experienced rapid growth, averaging 7% per year, accompanied by impressive poverty reduction. But growth and poverty reduction have been unevenly distributed, with regions such as the Northern Mountains, North Central Coast, and Central Highlands falling behind (please see figure 1 for Vietnam's regional and provincial map). A large share of industrial capital in Vietnam is concentrated in a very small number of urban centers, mainly Ho Chi Minh City (HCMC) and surrounding provinces (Ba Ria-Vung Tau, Dong Nai, and Binh Duong) in the South, and Hanoi, Hai Duong, Hai Phong and Quang Ninh in the North.[1] As a result, labor market adjustments and migration have become vital in spreading the benefits of growth from urban centers to hinterland areas.

[Figure 1 about here]

Despite the apparent importance of the topic, the literature on labor mobility and migration in Vietnam only started to receive attention very recently. Earlier studies (Guest 1998, Dang 2003, Nguyen 2003) are mostly descriptive and are useful only in giving migrants' characteristics and correlates of migration. More recent studies, including Hoang et al (forthcoming), de Brauw and Harigaya (2007), ADB (2007), and Phan (2008) study the impact of migration on the migrant's family and the source community, and begin to unravel the link between migration and poverty and inequality. These studies generally indicate that internal migration benefits not only migrants but also their families and source communities, and that internal migration tends to be pro-poor although it might increase inequality within the origin.

Our paper's contribution to the literature is twofold. First, while we employ the popular empirical gravity model, which hypothesizes that the flow of migrants between locations is a

[1] In 2002, the four provinces in the HCMC cluster made up 50 percent of the country's industrial output (HCMC alone accounted for 22 percent), while the northern cluster made up 14 percent. According to Mekong Economics (2002), the key areas in the South attract as much as 60 percent of all licensed foreign direct investment projects, and 53 percent of total registered capital.

function of population, distance, wage/income differentials, differences in unemployment rate, and other variables (Greenwood 1997, Fan 2005, Dhar 1984, Adrienko and Guriev 2004, Mueser 1989), we construct a theoretical framework that gives the model a solid structural interpretation. Furthermore, the framework incorporates a subsistence constraint and thus yields hypotheses regarding the impact of a sending province's income on migration. This allows us to examine the implications of liquidity constraints on poverty-related labor immobility.

The second contribution is to unravel the link between internal migration and regional income inequality. The empirical literature typically addresses this question with a standard convergence/growth regression, but this approach has also been subjected to methodological and empirical critiques (see section 4). We examine the migration-inequality relationship using a simple yet novel approach: we relate the impact of the outmigration rate from province i to province j in one period to the change in income differential between the two provinces in the next period.

This second contribution also relates to the existing literature on Vietnam's migration and spatial inequality. Increasing regional inequality is becoming a concern in Vietnam, as evidenced by a steady increase in the standard deviation of per capita GDP across provinces (see figure 2). To study the impact of migration on regional or provincial inequality, micro data sets with national coverage are needed. Yet there is a serious lack of such data sets, making it difficult, if not impossible, to relate macroeconomic changes at the regional and sectoral levels with economic decision-making and welfare changes at the micro level. For an empirical study, then, there is a trade-off between national-level coverage at the provincial level and much more limited coverage at the level of the household. Most of the Vietnam migration studies cited above forgo the former, and hence cannot draw economy-wide inferences on migration issues. Our study uses aggregate data on interprovincial migration. The limitations of this data set for the purpose of obtaining insights into household-level migration decisions and outcomes are obvious. Unlike household level data, however, it does yield economy-wide inferences on migration. In this way our approach is complementary to the more usual household level analysis.

[Figure 2 about here]

Our econometric analysis robustly confirms economic motives for migration, but also suggests the existence of poverty-related labor immobility at the provincial level. This in turn may imply persistence of poverty in certain regions if labor mobility is indeed a major channel through which the benefits of growth are distributed. We find that the impact of migration on income differentials between pairs of provinces can be positive, negative or zero, depending on the destination. A particularly interesting result is the robust and negative impact on income differentials of migration flows going to HCMC and surrounding provinces, where most of Vietnam'slabor-intensive manufacturing growth and investments are concentrated. This lends support to the importance of the labor market and migration in distributing the benefits of trade-driven manufacturing growth.

The rest of the paper proceeds as follows. Section 2 provides a theoretical framework for the migration decision of a representative household, and links migration with inter-provincial income inequality. Section 3 investigates determinants of inter-provincial migration flows, and tests the hypothesis that poorer people/provinces have low migration propensity because of their inability to finance migration costs. Section 4 examines the relationship between migration and inter-provincial income differentials. Section 5 offers concluding thoughts and directions for future research.

2 Theoretical Framework

2.1 Literature review

Lucas (1997) provides an excellent review of the literature on inter-provincial migration in developing countries. There are two main theoretical approaches on inter-regional migration, the disequilibrium approach and the equilibrium approach (Greenwood 1997). Each is built around a core model of wage/income differences in origin and destination as predictors of migration flows, but each provides a different interpretation of key variables–notably of income in the origin. The disequilibrium approach is formulated in the context of individual utility maximization. Migration is driven by the existence of a set of non-market clearing regional wages or incomes. Differences in wages/incomes, which are caused by exogenous shocks, reflect opportunities for utility gains through migration. Hence migration acts as an equilibrating force in bringing the wage/income differential back to zero, and is independent of location-specific amenities. This framework provides the rationale for including variables on income, earning, wage, unemployment rate, degree of urbanization, etc. in equations estimating the determinants of migration flows, as most studies indeed do. Some models also include demographic and education variables of the origin, such as median age or median number of years of schooling, to proxy for the average characteristics of the population from which the migrants are drawn. This approach predicts that wage/income variable in the origin should take a negative sign, while the destination wage/income variable should take a positive sign. According to Greenwood (1997), empirical studies employing this theoretical perspective do not yield uniform results, but the weight of available evidence favors the importance of wages/incomes in determining migration flows. Some notable examples include Fan (2005), Andrienko and Guriev (2003), Dhar (1984), and Greenwood (1969).

The equilibrium approach is also formulated in the context of individual utility maximization, and also assumes that migration is motivated by spatial variations. But it differs from the former approach in the source and persistence of these variations (Greenwood 1997). This approach assumes that the system is already in equilibrium even with the existence of wage/income differential, because these differentials simply reflect the differences in location-specific amenities (both social and natural). Both migration and changes in wages act as the equilibrating forces in response to exogenous changes in amenity demand. Using this framework, proponents of the equilibrium hypothesis include a wide variety of regional amenities in their econometric specifications, such as average humidity, number of hot days, national forest lands, presence of absence of a sea coast, public services, health facilities, schools, etc.

Is either the equilibrium or disequilibrium model to be preferred ex ante in a developing-country context? This decision comes down to a question of the importance of regional amenities vs. the importance of income/wage differentials in the migration decision. Most studies in the migration literature for developed countries adopt one approach and include variables suggested by that approach only–with the exception of a few that have nested both approaches into one model, with mixed results (Graves 1983; Greenwood and Hunt 1989). Graves (1983) finds evidence supporting the equilibrium hypothesis since his amenity variables are all significant, while income variables are only significant and have expected signs when amenity variables are present. On the other hand, Greenwood and Hunt (1979) find that amenities are less important determinants of migration. The equilibrium approach has not been used much in the literature for developing countries, probably because at lower income levels, amenities are less important as arguments in the utility function, so that demand for them play a smaller role in the migration decision. For this reason, the model in this section follows the equilibrium approach. It extends existing models (in particular the model in Yang 2004) by allowing the possibility of a liquidity constraint to study non-linearity in the migration decision.

2.2 Basic model - motivations for migration

Consider a pair of provinces, denoted s and r. Assume that in the initial period t=0, province r (the receiving province) is richer, or has higher per capita income $Y_{r,0}$ than per capita income $Y_{s,0}$ of province s (the sending province). At the end of $t = 0$, a representative household in province s must decide whether or not to allocate some labor to migration in order to maximize its expected income in period $t = 1$:

$$Y_{s,1} = MAX_{m_{sr,1}} \{(1 - m_{sr,1})Y_{s,0} + m_{sr,1}Y_{r,0} - C(m_{sr,1}, m_{sr,0}, D_{sr})\}$$

where $C(m_{sr,1}, m_{sr,0}, D_{sr}) = \overline{C}m_{sr,1}^{\sigma}m_{sr,0}^{\theta}D_{sr}^{\gamma}$ is the assumed structure of the migration cost function, $m_{sr,0}$ is the past migration rate, and D_{sr} is the geographical distance between the provinces. Total labor supply of this representative household is normalized to one, so $m_{sr,1}$ is also the out-migration rate from s to r.

It is assumed in this model that $\sigma > 1$, $\theta < 0$, and $\gamma > 0$. $\sigma > 1$ implies that the migration cost is convex in migrant labor supply; that is, migration costs increase at an increasing rate as the number of migrants increases. This assumption is widely used in the regional economics literature (Yang 2004). Convexity can arise from sources such as increasing urban rents or increasing unwillingness to migrate for remaining family members (who tend to be older and more attached to the home area than the first migrants). As we will see shortly, this assumption on the value of σ is important, because it may determine the relationship between migration rate and sending province's per capita income. The literature recognizes past migration rates and geographical distance between provinces as important determinants of migration costs. In particular, distance serves as a proxy for the cost of transportation, cost of job search and information acquisition, as well as the psychological cost of migration. Greater distance should increase these costs and hence deter migration, i.e., $\gamma > 0$. The past migration rate, on the other hand, measures the stock of existing migrants and proxies for a migration network, which can significantly reduce these costs and thus encourage migration, i.e., $\theta < 0$.

The first-order condition of the migration function yields the optimal migration rate:

$$m^*_{sr,1} = \left(\frac{Y_{r,0} - Y_{s,0}}{\sigma \overline{C} m^{\theta}_{sr,0} D^{\gamma}_{sr}} \right)^{\frac{1}{\sigma - 1}} \tag{1}$$

Taking logs on both sides, we get:

$$ln(m^*_{sr,1}) = \frac{-ln(\sigma\overline{C})}{\sigma - 1} + \frac{ln(Y_{r,0} - Y_{s,0})}{\sigma - 1} - \frac{\theta}{\sigma - 1} ln(m_{sr,0}) - \frac{\gamma}{\sigma - 1} ln(D_{sr})$$

or

$$ln(m^*_{sr,1}) = \beta_0 ln(\sigma C) + \beta_y ln(Y_{r,0} - Y_{s,0}) + \beta_m ln(m_{sr,0}) + \beta_d ln(D_{sr}) \tag{2}$$

where

$$\beta_y = \frac{1}{\sigma - 1} \quad , \quad \beta_m = \frac{-\theta}{\sigma - 1} \quad , \quad \beta_d = \frac{-\gamma}{\sigma - 1}$$

are the elasticities of migration rate with respect to income differential, distance, and past migration rate, respectively. Equation (2) has the form of a typical modified gravity migration equation, in which the migration rate is a function of the income differential, distance, and past migration (Greenwood 1997, Fan 2005, Dhar 1984, Andrienko 2004, Mueser 1989).

2.3 Liquidity constraint and non-linearity in migration decisions

As mentioned earlier, most empirical work in the migration literature finds a negative impact of origin income on migration. However, a few studies, either using micro or macro data, find a positive effect in some income ranges, suggesting a non-linear relationship between origin income and migration propensity (Connell et al. 1976, Barnerjee and Kanbur 1981, Burda et al. 1998; Andrienko and Guriev 2003). This contradicts classical theory, which predicts a monotonic negative relationship between the two variables.

In reconciling this difference, Banerjee and Kanbur (1981) were among the first to incorporate migration costs to investigate specifically the role of origin's income in the migration process. Their model is based on the idea that when migration is costly and when the capital market operates imperfectly, increased income would raise the capacity to finance migration. In what follows, we use a similar idea but model the migration cost function explicitly; the resulting formulation exhibits a non-linear relationship between migration and income in the origin.

Suppose the household faces a subsistence constraint S, such that any savings from period 0's income to finance migration at the start of period 1 must satisfy the constraint $Y_{s,0} - C \geq S$. By assumption, there is no borrowing due to capital market imperfections. For simplicity, and without changing key results, we drop past migration rate and distance from the migration cost function. Several subscripts are also dropped for notational ease. The household's optimization problem becomes:

$$\begin{aligned} &MAX_m \left\{(1-m)Y_0 + mY_r - \overline{C}m^{\sigma}\right\} \\ &s.t. \quad Y_s - \overline{C}m^{\sigma} \geq S \end{aligned}$$

The first-order condition yields:[2]

$$m^* = \left(\frac{Y_r - Y_s}{\sigma\overline{C}(\lambda+1)}\right) \tag{3}$$

where λ is the Lagrange multiplier. If the subsistence constraint binds ($\lambda > 0$), then the optimal migration rate, using the subsistence constraint at equality, is $m^* = \left(\frac{Y_s - S}{\overline{C}}\right)^{1/\sigma}$. If not ($\lambda = 0$), then the optimal migration rate, using the first order condition, is $m^* = \left(\frac{Y_r - Y_s}{\sigma\overline{C}}\right)^{1/(\sigma-1)}$. Let $\overline{Y}_s$ be the solution to $\left(\frac{Y_r - Y_s}{\sigma\overline{C}}\right)^{1/(\sigma-1)} = \left(\frac{Y_s - S}{\overline{C}}\right)^{1/\sigma}$.[3] This is the level of income above which the household is no longer bound by the subsistence constraint. So we have:

$$m^* = \begin{cases} \left(\frac{Y_s - S}{\overline{C}}\right)^{1/\sigma} & if \;\; Y_s \leq \overline{Y}_s \\ \left(\frac{Y_r - Y_s}{\sigma\overline{C}}\right)^{1/(\sigma-1)} & if \;\; Y_s > \overline{Y}_s \end{cases}$$

and

$$\frac{\partial m^*}{\partial Y_s} = \begin{cases} \frac{1}{\sigma}\left(Y_s - 1\right)^{(1-\sigma)/\sigma}\overline{C}^{-1/\sigma} & > 0 \;\; if \;\; Y_s \leq \overline{Y}_s \\ \frac{-1}{\sigma-1}\left(Y_r - Y_s\right)^{(2-\sigma)/(\sigma-1)}\left(\sigma\overline{C}\right)^{-1/(\sigma-1)} & < 0 \;\; if \;\; Y_s > \overline{Y}_s \end{cases}$$

Thus, the marginal impact of the sending province's per capita income on the out-migration rate

[2]Note that when $\varsigma > 1$, the objective function is globally concave, and the solution is as presented in the paper. When $\varsigma = 1$, we clearly have a corner solution: the household compares the expected income from not migrating at all to that from sending all labor away, and choose the option that yields higher income. When $\varsigma >$ 1, the objective function is globally convex, so no global minimum exists. However, since the choice variable m is constrained to be in the interval [0,1], we again have a corner solution as in the case of $\varsigma = 1$.

[3]Note that $\left(\frac{Y_s - S}{\overline{C}}\right)$ is monotonically increasing in Y_s, while $\left(\frac{Y_r - Y_s}{\sigma\overline{C}}\right)^{1/(\sigma-1)}$ is monotonically decreasing in Y_s, so a unique solution exists

is always positive at low income levels (when $Y_s \leq \overline{Y_s}$) as suggested by the liquidity constraint hypothesis, which maintains that the poorer a province is, the lower the capacity to finance migration cost, and hence the smaller the out-migration rate. But at sufficiently high level of income (when $Y_s > \overline{Y_s}$), the marginal impact of the sending province's per capita income on migration is negative (provided $\sigma > 1$ as assumed here), as suggested by the push effect hypothesis, which maintains that poorer people want to migrate more. This result suggests that a "labor mobility trap" is possible, at least theoretically, as long as the migration cost is sufficiently high and the capital market is imperfect.

2.4 Migration and income inequality

Let us go back to the notation in section 2.2. Following Yang (2004), the realized net income gain from migration in period 1 for the whole household is:

$$\begin{aligned} G &= (1-m^*_{sr,1})Y_{s,1} + m^*_{sr,1}Y_{r,1} - \overline{C}(m^*_{sr,1})^{\sigma} m^{\theta}_{sr,0} D^{\gamma}_{sr} - Y_{s,1} \\ &= (Y_{r,1} - Y_{s,1})m^*_{sr,1} - \overline{C}(m^*_{sr,1})^{\sigma} m^{\theta}_{sr,0} D^{\gamma}_{sr} \end{aligned} \tag{4}$$

Substituting (1) into (4) yields:

$$G = m^*_{sr,1}\left((Y_{r,1} - Y_{s,1}) - \frac{1}{\sigma}(Y_{r,0} - Y_{s,0})\right)$$

Define $I_{sr,0} \equiv Y_{r,0}/Y_{s,0}$ as the relative per capita income inequality between s and r in period $t = 0$. For simplicity, assume that $Y_{r,0}$ and $Y_{s,0}$ are exogenously given and that they are the same as per capita GDP, i.e., there is no remittance flow between the two provinces in the initial period. Let $g_r(m_{sr}, Z_r)$ and $g_s(m_{sr}, Z_s)$ be the gross growth rates[4] in per capita GDP from $t = 0$ to $t = 1$ of the two provinces. Z_r and Z_s are (exogenously given) province-specific characteristics that determine its long-run per capita GDP growth rate. Let τ be the measure of the propensity to remit. The relative income inequality between s and r in period $t = 1$ is then:[5]

$$\begin{aligned} I_{sr,1} &= \frac{Y_{r,1}}{Y_{s,1} + \tau m_{sr,1}\left((Y_{r,1} - Y_{s,1}) - \frac{1}{\sigma}(Y_{r,0} - Y_{s,0})\right)} \\ &= \frac{g_r(m_{sr,1}, Z_r)Y_{r,0}}{g_s(m_{sr,1}, Z_s)Y_{s,0} + \tau m_{sr,1}\left[Y_{r,0}\left(g_r(m_{sr,1}, Z_r) - \frac{1}{\sigma}\right) + Y_{s,0}\left(g_s(m_{sr,1}, Z_s) - \frac{1}{\sigma}\right)\right]} \end{aligned} \tag{5}$$

The expression shows that the current-period income differential is a function of initial per capita incomes, migration, and all the province-specific variables that determine the growth rates of per

[4] Defined as g = 1 + r, where r is the net growth rate

[5] It is implicitly assumed that migrants are accounted as one separate group, so the defined relative income inequality only measures the income differential between natives of province r and stayers of province s.

capita GDP in the two provinces. This theoretical relationship forms the foundation for our empirical analysis of the impact of migration on inter-provincial income differences in section 4

3 Determinants of inter-provincial migration 1984-1989 and 1994-1999

3.1 Methodology review

There are two main empirical approaches to estimate the determinants of inter-provincial or inter-regional migration using aggregate data: the gravity model approach and the polytomous logistic model approach.

Gravity model approach

The basic formulation of this approach is:

$$M_{ij} = P_i^\alpha P_j^\beta / D_{ij}^{-\gamma}$$

The hypothesis is that the gross migration flow from province i to province j, M_{ij}, is directly proportional to the size of the origin and destination provinces' populations, P_i and P_j respectively, and inversely proportional to the distance between them, D_{ij}. This basic model is usually modified by adding the variables $G_{ij} = \{X_i, X_j\}$, as suggested by either the equilibrium or the disequilibrium approach, or both. These variables are theorized to affect the gravity force, or attraction between the two provinces:

$$M_{ij} = G_{ij} P_i^\alpha P_j^\beta / D_{ij}^{-\gamma} = X_i^\theta X_j^\eta P_i^\alpha P_j^\beta / D_{ij}^{-\gamma} \tag{6}$$

In the modified gravity model, distance is interpreted more generally to encompass not only geographical distance but also any variables that influence the cost of migration between the two provinces, such as migrant's network . Empirically, (6) is often estimated in double-log form, by taking the logarithms of both sides and adding a multiplicative error term and a constant term:

$$ln(M_{ij}) = C + \theta ln(X_i) + \eta ln(X_j) + \alpha ln(P_i) + \beta ln(P_j) + \gamma ln(D_{ij}) + \epsilon_{ij} \tag{7}$$

In many cases (e.g., Fields 1982), migration flows are first normalized by dividing (6) by origin population, P_i, before taking logarithms. This yields the following alternative specification:

$$ln(m_{ij}) = C + \theta ln(X_i) + \eta ln(X_j) + (\alpha - 1) ln(P_i) + \beta ln(P_j) + \gamma ln(D_{ij}) + \epsilon_{ij} \tag{8}$$

where $m_{ij} = M_{ij}/P_i$ is the migration rate.

There are two ways in which origin and destination characteristics can enter the estimation equation. In symmetric models, the independent variables are the difference or ratio of the origin and destination characteristics. Such models assume that the impact of these origin and destination characteristics are equal but of opposite signs. The asymmetric models, on the other hand, take origin and destination characteristics as separate explanatory variables, thus allowing the impacts of origin and destination conditions to be different from one another. Not surprisingly, asymmetric models are often the preferred choice, because it is a more generalized model; the possibility of symmetric impact can be tested as restrictions on the estimated parameters of asymmetric models.

Logit model approach

The gravity model and its double log form have been the more popular approach in the literature, probably because it yields good fits, and because the coefficients are elasticities and hence easy to interpret. However, Schultz (1977) argues that the gravity model does not recognize that the migration decision is inherently a choice between a finite number of mutually exclusive discrete alternatives, including non-migration. Moreover, in the gravity model, more non-migration will appear to exist for regions that are larger in population and land size simply because a large share of all the moves will occur within boundaries of the large regions. This causes non-migration to be spuriously correlated with origin population size and land area. He then proposes a polytomous logistic model[6]:

$$m_{ij} = \frac{exp(Z_{ij})}{\sum_{j'} exp(Z_{ij'})}\ ,\ \sum_{j} m_{ij} = 1\ \ for\ i = 1, 2, ...n \tag{9}$$

As above, m_{ij} is the gross migration rate, and also the probability that an individual locating in province i decides to move to province j. $i = j$ means that a person decides not to move, or only move within the origin's boundary. Z_{ij} is a function of a set of origin and destination conditions (including both $G_{ij} = \{X_i, X_j\}$ and D_{ij}) that affect the migration decision. Schultz provided several possible specifications for Z_{ij}, but only one is presented here:

$$Z_{ij} = \alpha + \beta_i ln(X_i) + \beta_j ln(X_j) + \gamma ln(D_{ij}) \tag{10}$$

Maximum likelihood can be used to estimate (9) and (10). Further extension to the model can be made to take into account of the complications that might arise with non-migration (namely the occurrence of m_{ii}) and to allow for the asymmetric impact of origin and destination characteristics. This is done by viewing the decision process as a two-stage decision process: first the potential migrant decides whether to migrate, then he chooses where to migrate. This might be interpreted as the distinction between the response parameters for the case $i = j$ and those for the case $i \neq j$. That is, when $i = j$:

[6] Like other logit models, the independence of irrelevant alternatives (IIA) condition (i.e., the odds ratio of any two probabilities is independent of the characteristics of other locations) is required.

$$Z_{ii} = \alpha^* + \beta_i^* ln(X_i) + \beta_j^* ln(X_j)$$
$$= \alpha^* + (\beta_i^* + \beta_j^*) ln(X_i) \quad (11)$$

where the asterisks mean that the coefficient values differ from those in (11). Taking the logarithms of the odd ratios, one obtain the estimated model:

$$ln(m_{ij}/m_{ii}) = \alpha' + \beta_i' ln(X_i) + \beta_j ln(X_j) + \gamma ln(D_{ij}) + \epsilon_{ij} \quad (12)$$

where $\alpha' = \alpha - \alpha^*$, $\beta_i' = \beta_i - (\beta_i^* + \beta_j^*)$. OLS can be used to estimate (12), in addition to maximum likelihood logit.

Comparing (12) with (8), the difference is that in the logit model the dependent variable is the relative migration rate, while in the gravity model it is the absolute migration rate. As the interval of time over which migration is measured diminishes, m_{ii} approaches unity, and $ln(m_{ij}/m_{ii})$ approaches $ln(m_{ij})$, so that (12) and (8) become equivalent. In applying the model to Venezuelan census data, Schultz found that the maximum likelihood logit estimator outperforms the logit OLS, which outperforms the gravity model in its predictive power .

3.2 Empirical strategy

The empirical exercise in this section employs the extended gravity model approach, using the model developed in section **??** as its theoretical foundation. Both the gravity OLS model and the logit model are estimated. As will be seen shortly, there are hardly any differences in the coefficient estimates, although these estimates are to be interpreted differently in the two models. For the most part, we will focus on the results from the gravity model for ease of interpretation.

From (2), the system of structural equations to be estimated is:

$$\begin{aligned} ln(m_{sr}^{84-89}) &= \beta_0^{84} C + \beta_y^{84} ln(y_r^{84}/y_s^{84}) + \beta_d^{84} ln(d_{sr}) + \eta^{84} X + \epsilon_{sr}^{84} \\ ln(m_{sr}^{94-99}) &= \beta_0^{94} C + \beta_y^{94} ln(y_r^{94}/y_s^{94}) + \beta_d^{94} ln(d_{sr}) \\ &\quad + \beta_m ln(m_{sr}^{84-89}) + \eta^{94} X + \epsilon_{sr}^{94} \end{aligned}$$

where d_{sr} is the bus distance (in kilometers) between the capital cities of the two provinces, and X is a vector of other control variables, such as regional dummies for sending and receiving provinces, or province fixed effects. As written, these equations also incorporate the implicit restriction that the impacts on migration of sending and receiving provinces' incomes are equal but of opposite signs. In line with the literature, however, we instead estimate a more general specification in which these are allowed to differ:

$$
\begin{aligned}
ln(m_{sr}^{84-89}) &= \beta_0^{84}C + \beta_r^{84}ln(y_r^{84}) + \beta_s^{84}ln(y_s^{84}) + \beta_d^{84}ln(d_{sr}) + \eta^{84}X + \epsilon_{sr}^{84} \qquad (13)\\
ln(m_{sr}^{94-99}) &= \beta_0^{94}C + \beta_r^{94}ln(y_r^{94}) + \beta_s^{94}ln(y_s^{94}) + \beta_d^{94}ln(d_{sr}) \qquad (14)\\
&\quad + \beta_m ln(m_{sr}^{84-89}) + \eta^{94}X + \epsilon_{sr}^{94}
\end{aligned}
$$

Equation (13) includes only pre-determined variables, so ordinary least square (OLS) provides a consistent estimator. Equation (14) includes pre-determined variables plus the past migration rate, which is endogenous in (13). If the assumption that ϵ_{sr}^{84} and ϵ_{sr}^{94} are uncorrelated is imposed, the system becomes fully recursive; then OLS applied to each structural equation will yield unbiased and consistent estimates of the direct effects of the covariates on migration in both periods. The indirect effect of a change in a time-invariant predetermined variable, for example distance, on the left-hand-side through the endogenous right-hand-side variable (past migration) can be computed from these estimates by $\beta_d^{94} * \beta_m$. The total effects, or the sum of direct and indirect effects, can be computed as follows:

$$
\begin{aligned}
\beta_0 &= \beta_m \beta_0^{84} + \beta_0^{94}\\
\beta_d &= \beta_m \beta_d^{84} + \beta_d^{94}\\
\eta &= \beta_m \eta^{84} + \eta^{94}
\end{aligned}
$$

However, there may be unobserved factors that affect migration flows in both periods, causing ϵ_{sr}^{84} and ϵ_{sr}^{94} to be correlated, and OLS would yield inefficient estimates of (13) and inconsistent estimates of (14). In this case, a system estimator such as three-stage least squares (3SLS) is needed for both consistency and efficiency. In the empirical work below, we implement both OLS and 3SLS. But as will be seen shortly, the two estimation methods do not yield qualitatively different results.

Throughout the empirical analysis, we will run the above regressions first on the full sample, then on a sub-sample in which the only receiving provinces included are HCMC and its three neighboring provinces (Ba Ria-Vung Tau, Dong Nai, and Binh Duong). As mentioned earlier, this is a special urban cluster where most manufacturing capital and foreign direct investment are located, and where most of recent growth, especially export-oriented manufacturing growth, is concentrated. As a result, it is of particular interest to examine the determinants and consequences of migration flows into these provinces.

3.3 Data

Table 1 gives summary statistics of variables used in the regressions in this chapter. Most data on province characteristics, such as per capita income, population, etc., are collected from different issues of the Statistical Year Book published by the General Statistics Office (GSO) of Vietnam. Migration data come from the 1989 and 1999 Population and Housing Censuses, also conducted by the GSO. The censuses cover the entire nation, and gather information on demographic and socio-economic characteristics of the population. In particular, they ask questions on place of birth, duration of residence, place of last residence, and place of residence at a fixed prior date. A respondent is identified as a migrant if he/she was at least five years of age at the time of census, and changed place of residence within the past five years. This allows estimation of inter- and intra-provincial migration flows during the prior five-year period. As noted, a major drawback of the approach is that it excludes temporary/seasonal and return migrants, as well as those who were born during the five-year interval. Besides, the exact timing of any reported move is unknown. Thus the census data must underestimate actual migration, and are more likely to reflect permanent than temporary moves. Lastly, since we do not have data on provincial per capita incomes in 1984, per capita industrial output in 1986 is used instead.

[Table 1 about here]

From 1984 to 1989, inter-provincial migration in Vietnam closely followed the resettlement program discussed earlier. From 1994 to 1999, rural-rural migration flows continued to be correlated with those in the 1980s: they were large and involved long-distance moves from the Red River Delta and North Central Coast regions to the Central Highlands (see table 2). But unlike the 1980s, there also emerged new short-distance rural-urban moves from the Red River Delta into Hanoi and from the Southeast and Mekong River Delta into HCMC.

[Table 2 about here]

Such migration patterns accord with theoretical predictions that people tend to move from low income to high income areas, and also from land-scarce to land-abundant regions. The former is borne out by figure 3, which shows a strong and statistically significant relationship between per capita income and net in-migration, confirming that internal migration in Vietnam is motivated to a large extent by income differences. However a particular case against this trend is the low level of labor mobility either into or out of the North West region – the poorest and also most remote region of Vietnam. The persistence of poverty in this region might be attributable to a combination of high migration costs and household-level liquidity constraints (i.e., inability to finance migration cost due to low income). These in turn are correlated with ethnicity; the more remote regions are also those with the highest proportion of ethnic minorities, whose language and cultural barriers drive migration costs especially high.

[Figure 3 about here]

3.4 Results

Tables 3a through 3c give the estimates for equations 13 and 14 with regional fixed effects, using OLS, logit, and 3SLS estimation, respectively. For the 1990s OLS regression, both total and indirect effects of OLS and logit regressions are also shown. As can be seen, OLS and logit models hardly yield any differences in the results. There are also no major qualitative differences between OLS and 3SLS estimates, suggesting that any endogeneity caused by correlation of the error terms does not have serious effects on the estimates. This may be because regional dummies already capture much of the unobserved influence on migration flows in both periods. So to save space, we will discuss only the OLS estimates in table 3a.

[Tables 3a, 3b, and 3c about here]

Table 3a shows that all explanatory variables have coefficients of expected signs that are statistically significant at either 1% or 5% level. Provinces that are further apart send fewer migrants to each other, as the coefficient of distance variable is negative. The direct effect of distance in the 1990s is much smaller than in the earlier period. In the 1980s, the distance elasticity of migration is about -1.1, meaning that a 1% increase in distance between province s and province r leads to a 1.1% decrease in migration between s and r, other things equal. In the 1990s, the direct effect of distance has fallen to only 0.437%. This is expected, given that the cost of transportation must have reduced considerably over the decade. After adding the indirect effect, nevertheless, the total effect is approximately the same as in the 1980s.

Provinces with high per capita income attract more migrants. In the 1990s, a 1% increase in the per capita income of receiving province leads to a 1.5% increase in the migration rate, implying a rather high responsiveness of migration to income. By taking the absolute value of the ratio of the estimated distance elasticity to the estimated elasticity on destination income, one can compute the income-distance trade-off, a rough indicator of the cost of moving a given distance further (Greenwood 1997). That value for the 1990s regression is 0.73 (= |-.1.1/1.5|), meaning that a .73% increase in destination income is needed to offset a 1% increase in distance. Computed at the sample averages, this means that a move 7km further away can be offset by an increase in monthly income of 11,800 VND (approximately $1.00 at 1999 prices). Although comparisons across studies are hard to make, it is still worth mentioning that such figures for Canada were found to be in the range of 0.146 to 0.439, with variations across age and education groups and across different time periods (Courchene 1970). For China, the figure is 0.35 (Fan 2005). Thus the income-distance trade-off of 0.73% found in this study seems relatively high by comparison with other studies. This suggests that in Vietnam the part of migration cost that is correlated with distance is higher than in other countries.

In the 1980s, the regional dummies show that the major receiving regions were the Central Highlands and the South East, even though per capita income in the Central Highlands was among the lowest in the country during that time. Such flows should thus be interpreted as capturing migration policy effects. Controlling for income differences, the dummy for urban areas in the 1980s did not show any statistically significant impact, which is perhaps due to the policy bias against rural-urban migration during this period. In the 1990s, the Central Highlands and the South East continued to be major receiving regions; however, their dummies should no longer be interpreted as measuring policy effects. Rather they should now be interpreted as capturing the impact of structural shocks such as Vietnam's agricultural export boom, especially coffee in the Central Highlands. Also, the dummy for urban areas is positive and significant in the 1990s, indicating the pulling effect of both higher income and rapid manufacturing growth in urban centers. In general, the signs of regional dummies are in accord with the theoretical prediction that the mobile factor (labor) flows to regions with high concentrations of immobile factors (land and capital).

The estimated coefficient of per capita income in the sending province is positive. There are two potential explanations for this. First, although past migration and distance control for much of the migration cost, there is no reason to expect that they capture all of it. There are perhaps omitted variables or unobserved province effects that influence access to information in the sending province, which should reduce the cost of migration and increase migration flows. Omitting these variables biases the estimated coefficient β_s^{94} upward, since the omitted variables are likely to be positively correlated with per capita income. To partially solve this issue, we estimate equations (13) and (14) with dummies for sending and receiving provinces, to control for some of the omitted variables bias caused by provinces' heterogeneity. Table 3d reveals no major qualitative changes in the results, especially for the 1990s regression.

[Table 3d about here]

Another explanation for the positive sign of sending province's per capita income is that the liquidity constraint effect might be stronger than the push effect, causing migration flows to be larger when the sending province is richer. We will test this hypothesis in section 3.6.

3.5 Alternative specifications and models

In the previous section, we follow closely the theoretical framework in section **??** and include only income, distance, past migration rate, and regional/provincial dummies as regressors determining inter-provincial migration. But the migration literature usually includes a number of other provincial/regional characteristics, such as population, unemployment rate, educational levels, per capita land, percentage of ethnic minorities, etc.[7] Furthermore, the theoretical model in section **??** may not be the only model that can explain inter-provincial migration in Vietnam. It has already been mentioned that migration policies in the 1980s directed migration toward

[7]Regional amenities such as average temperatures or number of hot days, etc. are not included because it was explained earlier that for countries with lower income levels such as Vietnam, amenities are less important arguments in the utility function. Consequently, demand for amenities play a smaller role in the migration decisions, whereas wage or income differentials are thought to be more influential.

provinces in the Central Highlands region, which had lower per capita income but higher per capita agricultural land. In other words, there might be two distinct migration flows motivated by two different reasons. One flow is motivated strongly by income differential, and perhaps mostly involves rural-urban migration. The other flow is motivated more by per capita land differential, and probably involves more of rural-rural migration.

In this sub-section, we consider alternative specifications by adding other regressors and by repeating the regressions for different sub-samples. For regressors whose data are not available, we argue how this might affect our estimates. The goal is to study how robust earlier estimates are. This will demonstrate how well the proposed theoretical model in section **??** explains reality in competition with other potential models. we will focus on examining alternative specifications only for 1990s migration flows, because migration in this decade is more recent and is of higher interest.

The current study does not include unemployment rate as a regressor because these data are not available. Still, we feel reasonably comfortable with excluding this variable for two reasons. First, a significant portion of migration in Vietnam are from rural to rural areas which tends to involve the whole family and lead to no occupational change (i.e., the migrants continue to be self-employed farmers in the destination). Second, anecdotal evidence, as well as migration surveys from Guess (1998) reveal that unemployment rates are particularly low among migrants, and that most migrants have jobs arranged before migration or find jobs within a relatively short time after arrival. This is probably because in Vietnam the job search associated with migration is usually done in the origin before the move, which explains the low unemployment rate among migrants. This rural-based job search strategy is very different from an important assumption in the famous Harris-Todaro model, which holds that migrants come to the cities without a pre-arranged job and with an anticipated employment rate (Harris and Todaro 1970). Other than Guest (1998), there has not been much research to rigorously study the issue of rural-based vs. urban-based job search for migration in Vietnam. But based on the available evidence, it seems including the unemployment rate will not affect the current estimates much.

Data on the other provincial characteristics are available, and they indeed deserve to be included in the regression. The role of education in the migration decision have been confirmed by many micro level studies, which find evidence that migrants tend to have higher level of education (Connell et al 1976). As said above, a large portion of migration in Vietnam are rural-rural, so the differences in per capita agricultural land between the origin and the destination should affect the migration decision, especially for rural-rural migration. In Vietnam, many studies have found that ethnic minorities are much less mobile than the Kinh majority (Turk and Swinkels 2002 and 2006). So we would expect that provinces with high ethnic minorities percentages would have lower out migration rate. These provinces may or may not have high in-migration rate. On the one hand, they tend to have lower per capita income, implying that in-migration rate should be low. On the other hand, all provinces in the Central Highlands have high percentage of ethnic minorities, and this region is also a major receiving region because of the government sponsored program in the 1980s, as this region has high per capita land.

Unfortunately, there is one practical issue regarding the addition of provincial characteristics other than income. As can be seen in table 4, these variables tend to be highly correlated with each other, with income (especially within certain regions), and also with regional dummies. The correlation coefficients suggest that the regional dummies might have captured some impact of provincial characteristics other than income, and the addition of these characteristics might cause multi-collinearity problem, making the parameter estimates sensitive to alternative specifications. The multi-collinearity issue is the most serious when we want to test the impact of income differential vs. the impact of per capita land differential. The hypothesis is that for rural-rural migration, especially migration into the Central Highlands, per capita land differential would be more important, while for rural-urban migration, per capita income differential would be more important. But table 2.4 reveals that for the Central Highlands region and urban areas, all provincial characteristics are highly correlated with each other, if not indistinguishable in some cases.

[Table 4 about here]

With these caveats in mind, we try adding these provincial variables in the regression because they are important for policy implications. The estimates are presented in table 5a. In model I, the regression is run with regional fixed effects. In model II, these regional fixed effects are excluded. The first important result to note is that parameter estimates of distance, past migration rate, and income are highly robust to the addition of these other provincial characteristics or to the exclusion of regional dummies. This strengthens the results found in the earlier regressions, and we feel confident that income is among the most important and robust factors, if not the most, in explaining inter-provincial migration in Vietnam.

[Table 5a about here]

The second result to note is that parameter estimates as well as t-statistics of several newly added regressors change considerably from model I to model II. This must be due to the high correlation between regional dummies and provincial characteristics. Acknowledging this observation, let us discuss the results. In both models, sending province's percentage of ethnic minorities has a negative impact on migration rate as expected, again confirming limited labor mobility of ethnic minorities in Vietnam. Receiving province's percentage of ethnic minorities has a positive impact despite their lower per capita income. This might reflect the legacy of past migration policy in the 1980s, as well as the pull effect of higher per capita land. Indeed, when we take out the regional dummies (which presumably capture migration policy impact), the coefficient of this variable doubles. In model I, upper secondary school enrollment rates have insignificant impact on migration. But in model II, they now have positive significant impact.

This is in accordance with the migration literature which usually finds that migrants tend to be more educated. Finally, in both models per capita agricultural land of sending province has a negative effect on migration, as we would expect, even though it is much larger and more significant in model I. The coefficient of receiving province's per capita land change dramatically from model I to model II. With regional dummies controlled for, it is associated with less migration, while without regional fixed effects, it is associated with more migration.

We next run the regressions separately for rural-rural migration and rural-urban migration, and the results are in table 5b (We present the results of regressions without regional dummies only; interestingly, for the sub-samples, there are no major qualitative differences in the results between including or excluding sending provinces' dummies). Only major differences between the two regressions are now discussed. Distance has a more negative and more significant impact on rural-rural migration, reflecting the fact that most rural-rural migration involves longer moves, while most rural-urban migration are from the surrounding areas of the urban areas into the urban areas. In both regressions, receiving province's ethnic minority percentages are positive, but their magnitudes differ significantly, and they are also to be interpreted differently. Ethnic minority variable for rural-rural migration may proxy for migration policy and higher per capita land, as said earlier. Ethnic minority variable for rural-urban migration, on the other hand, may proxy for the high percentage of Chinese minority who mostly reside in urban areas and are relatively more affluent than the rest of the population including the Kinh majority. For rural-rural migration, per capita agricultural land of receiving provinces has a strong and positive impact on in-migration. For rural-urban migration, this variable has a negative impact. Such results support the hypothesis that agricultural availability of land is more important for rural-rural migration, while income differential is more important for rural-urban migration.

[Table 5b about here]

Finally, in table 5c, we present the regression estimates when the dependent variable is gross migration flow as opposed to migration rate, and the covariates include origin's and destination's population. This is because population might be proxying for various factors that may influence migration flows, including labor market conditions. Moreover, as explained in the logit model approach in section 3.1, non-migration might be spuriously correlated with origin population size. The estimates in table 5c maintain results similar to those in earlier regressions, although the colinearity issue remains unresolved as in table 5a or 5b (1994 population has a high correlation with 1994 per capita income, shown in table 4).

[Table 5c about here]

In summary, people may be migrating for different reasons, and several competing theories can explain inter-provincial migration flows. But the empirical results presented so far indicate that income differential is among the most important and most robust factors motivating migration, especially for rural-urban migration. This supports the validity of the simple theoretical model in section **??**. For rural-rural migration, availability of agricultural land is also important. Ethnic minorities are found to have limited labor mobility, and this trend is especially clear for out-migration. For in-migration, the result is more mixed because of two competing effects: provinces with more ethnic minorities are poorer and so should attract less migrants, but these provinces also have more land and so should attract more migrants. The impact of education on migration is generally positive, although it is not robust to alternative specifications.

3.6 Testing the liquidity constraint hypothesis

It was shown earlier that per capita income of the sending province can proxy for two opposite forces. It may represent the push (or income differential) effect, meaning that the poorer a province is, the more people want to migrate; thus β_s should have a negative sign in (13) and (14). But it might also represent the liquidity constraint effect, meaning that the poorer a province is, the lower is the capacity to finance migration costs, in which case β_s should have a positive sign. Which effect dominates depends on the income level of sending province itself. At low levels of income, the liquidity constraint effect should dominate. At higher levels, the push effect should dominate.

Methodology:

To test this hypothesis, we need to first compute the income threshold for each inter-provincial flow from s to r. Remember that it is a function of many variables, such as distance, sending province's income, etc., and so there exists one distinct threshold for each flow/observation. Then, we could run two separate OLS regressions, one for the observations in which each sending province's income is below its province-specific threshold, and another for the observations in which it is above the threshold. Comparing the coefficient estimates of these two OLS regressions would provide a test of the liquidity constraint hypothesis. The difficulty is that we do not know what the income thresholds are, because we do not have all the variables needed to compute them, nor do we have specific functional form for the computation. Thus a direct test of this hypothesis is not possible.

Still, there are a number of ways to indirectly test the hypothesis. In this chapter, we propose three different methods to do so. The first one is the simplest and based on the intuition that the liquidity constraint effect varies with the level of migration cost, becoming material only when migration cost is material. We therefore break the sample into percentiles of distance and past migration rate (which affect migration costs), then examine how β_s varies across the sub-samples.

Since we are concerned about non-linearity in the relationship between income and migration, another way is to add the square and/or cube of log of sending province's per capita income to test for higher-order impacts. However, rather than searching for such refinements of the parametric specification, we can also estimate the following semi-parametric model, which is the second proposed method:

$$ln(m_{sr}^{94-99}) = f(ln(y_s^{94})) + \beta Z_{sr} + \epsilon \quad (15)$$

where the vector Z_{sr} includes all explanatory variables other than sending province's income. Equation (16) falls into the class of partially linear models, which consist of a linear part, βZ_{sr}, and a non-parametric part, $f(.)$. In this model, no parametric assumption is imposed upon $f(.)$, except that it is a smooth function, while for the rest of the variables the usual parametric assumptions are used. The model thus allows the data to freely determine the shape of the influence of origin's income on migration[8].

Finally, the third method is to employ maximum likelihood to estimate the following regime-switching model with unknown and varying thresholds:

$$\begin{aligned} ln(m_{sr}^{94-99}) &= \beta_s^1 ln(y_s^{94}) + \beta Z_{sr} + \epsilon_{sr}^1 \quad if \quad y_s \leq \overline{y}_s \equiv \xi T_{sr} + u_{sr} \\ ln(m_{sr}^{94-99}) &= \beta_s^2 ln(y_s^{94}) + \beta Z_{sr} + \epsilon_{sr}^2 \quad if \quad y_s > \overline{y}_s \equiv \xi T_{sr} + u_{sr} \end{aligned}$$

where T_{sr} is the vector of variables that determine the unobserved and stochastic income threshold $\overline{y}_s$. The hypothesis is that $\beta_s^1 > 0$ and $\beta_s^2 < 0$. The likelihood function is:

$$\mathcal{L} = \prod_{sr} \left\{ Pr\left(u_{sr} \geq y_s - \xi T_{sr} | \epsilon_{sr}^1\right) * f(\epsilon_{sr}^1) + Pr\left(u_{sr} \leq y_s - \xi | \epsilon_{sr}^2\right) * f(\epsilon_{sr}^2) \right\}$$

where $f(.)$ is the probability density function of the error terms ϵ_{sr}^1 and ϵ_{sr}^2. Assume that ϵ_{sr}^1, ϵ_{sr}^2, and u_{sr} are normally distributed with mean zero and variance-covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \sigma_{1u}^2 \\ \sigma_{12}^2 & \sigma_2^2 & \sigma_{2u}^2 \\ \sigma_{1u}^2 & \sigma_{2u}^2 & \sigma_u^2 \end{bmatrix}$$

We then have (Dickens and Lang 1985):

$$\begin{aligned} Pr\left(u_{sr} \geq y_s - \xi T_{sr} | \epsilon_{sr}^1\right) &= 1 - Pr\left(u_{sr} \leq y_s - \xi T_{sr} | \epsilon_{sr}^1\right) \\ &= 1 - \Phi\left(\frac{y_s - \xi T_{sr} - \frac{\sigma_{1u}}{\sigma_1}\epsilon_{sr}^1}{\sqrt{1 - \frac{\sigma_{1u}^2}{\sigma_1}}}\right) \\ Pr\left(u_{sr} \leq y_s - \xi T_{sr} | \epsilon_{sr}^2\right) &= \Phi\left(\frac{y_s - \xi T_{sr} - \frac{\sigma_{2u}}{\sigma_2}\epsilon_{sr}^2}{\sqrt{1 - \frac{\sigma_{2u}^2}{\sigma_2}}}\right) \end{aligned}$$

The log likelihood function is thus:

[8] Detailed discussions of such a model and its estimation can be found in Hardle et al. (2004, chapter 7)

$$log\mathcal{L} = \sum_{sr} ln \left\{ \left[1 - \Phi\left(\frac{y_s - \xi T_{sr} - \frac{\sigma_{1u}}{\sigma_1}\epsilon^1_{sr}}{\sqrt{1-\frac{\sigma^2_{1u}}{\sigma_1}}} \right) \right] * \phi\left(\epsilon^1_{sr}, \sigma_1\right) + \Phi\left(\frac{y_s - \xi T_{sr} - \frac{\sigma_{2u}}{\sigma_2}\epsilon^2_{sr}}{\sqrt{1-\frac{\sigma^2_{2u}}{\sigma_2}}} \right) * \phi\left(\epsilon^2_{sr}, \sigma_2\right) \right\}$$

Results:

We first break the sample into five separate subsamples by quintiles of distance and estimate the 1990s regression separately for each subsample. The results are in table 6a. A Chow test (F=54) rejects the null hypothesis that the coefficients of the five piecewise regressions are equal. A pattern seems to emerge: the impact of sending province's per capita income at first increases in both magnitude and statistical significance as distance becomes greater, then falls at the highest distance quintile. This unexpected pattern at the highest distance quintile might be explained by the fact that there is high mobility among the major urban centers in Vietnam, and the average distance among these urban centers are relatively high, because they are spaced out across the country's long geographical border. Such mobility by the richest urban centers might offset the trend that at highest distance quintile the impact of sending province's income matters more. In brief, these results provide partial evidence that at relatively long distances (but not the longest distances), out-migration is more likely to occur from higher-income provinces.

[Table 6a about here]

We then break the sample into five sub-samples by quintiles of past migration rate and again run piecewise regressions (see table 6b). A Chow test (F=50.6) again rejects the null hypothesis of constant coefficients. At lower quintiles of past migration rate (i.e., when migration cost is higher), the coefficient on per capita income of the sending province is positive, high, and more statistically significant. At higher quintiles (i.e., lower migration cost) this coefficient becomes less statistically significant and also decreases in magnitude. This suggests that the liquidity constraint effect is indeed stronger when migration costs are higher.

[Table 6b about here]

The second test of the liquidity constraint hypothesis involves semi-parametric regressions. Figures 3a and 3b present semi-parametric estimates for the impact of income in the sending province on migration, as specified in equation (16), using the full sample and using the sub-sample consisting of just the HCMC cluster as receiving provinces.[9] For the full sample (figure 3a), there seems to be a non-monotonic impact of origin's income on the out-migration rate, but this impact is insignificant (the 95% confidence limits cover the zero axis; also, for the test of the null hypothesis that this impact is zero, p = 0.16). For the poorer two-thirds of the provinces,

[9]Estimates for the parametric part are qualitatively similar to those in parametric regressions and are available from the authors upon request.

the relationship follows that predicted by the liquidity constraint hypothesis: sending province's income's impact on out-migration rate first increases, then decreases with income. For the richest one third, the impact again increases as income rises, but the confidence limits widen, casting doubt on the robustness of the estimated relationship. It is therefore difficult to either reject or accept the liquidity constraint hypothesis. Nevertheless, in figure 3b, in which the only receiving provinces included are HCMC, Ba Ria-Vung Tau, Dong Nai, and Binh Duong, the null hypothesis of zero impact of origin's income can be rejected (p-value = 0.012). Furthermore, there is a clearer inverse U-shape relationship, as predicted by the liquidity constraint hypothesis: the impact of origin's income on out-migration rate at first increases, then decreases with income, once income passes the level of around 137,000 VND (about 12-13 USD) per month (mean monthly per capita income in the sample is 162,000 VND, about 14-15 USD).

[Figures 3a and 3b about here]

The third and final test of the liquidity constraint hypothesis involves maximum likelihood estimation of a regime-switching model with unknown and varying thresholds, as specified earlier in this subsection. While the econometric model is straightforward theoretically, its implementation runs into numerous difficulties. First, we do not know or have all the variables in the vector of variables T that determine income threshold (the minimum income required to be free of liquidity constraint). We do have two potential candidates, distance and past migration rate, because these variables both affect the migration cost. Second, the model is putting a tremendous demand on the data, because the regression must estimate not only the coefficients for the two separate regimes, but also the coefficients of the regime-determining equation. Third, given the complicated likelihood function, numerical maximization must be used. A further practical issue associated with numerical maximization is the choice of starting values. We must have a reasonably good initial guess of which regime each observation might fall into, and a reasonably good initial guess of all the parameters given the regime. This is difficult as we do not have any other substitutable models whose estimates can be used as starting values.

Because of these issues, there is no guarantee a priori that the available data will be able to estimate the model. In the context of numerical maximization, this means that there is no guarantee a priori that convergence will be achieved. It turned out to be quite difficult to achieve convergence for the sample used in this chapter. Convergence could not be achieved at all when we tried to include too many regressors in the threshold-determining equation. So we left with no choice but to include them one at a time. The results are given in table 7. For model I in which distance is the threshold-determining variable, we sorted the data in ascending values of the distance variable, divided the sample into two halves, ran OLS on each half of the sample, then used OLS estimates from the two sub-samples as starting parameter values for the two regimes. A similar procedure was carried out for model II in which past migration is the threshold-determining variable, but the data were sorted in ascending values of the past migration[10].

[Table 7 about here]

The results in table 7 reveal very minor differences compared with OLS results in table 3. As expected, distance has a positive impact on the income threshold (model I) while past migration has a negative impact (model II). The coefficients of sending province's income below or above threshold turn out to be very consistent for whichever threshold-determining regressor used. This provides us with some confidence in the results, given that the two models are given with different starting parameter values. For all provinces whose incomes are below their province-specific income thresholds, the sending province's income is positively associated with migration. For provinces above their province-specific income thresholds, we expected the coefficient for this variable to be negative, but it turns out to be still positive, even though it is much smaller than that for the below-threshold provinces. We may interpret these result as follows. For the full sample that we have, the liquidity constraint effect might be dominating the push effect everywhere, so the impact of sending province's income is positive everywhere. At the same time, there are still two regimes: for the group of provinces below their income thresholds, this liquidity constraint effect is much stronger than that for the group of provinces above the thresholds–this in a way is still consistent with the liquidity constraint hypothesis.

These results are to be interpreted with much caution however, for several reasons. The regime-switching regression rests on a critical assumption on the distribution of the error terms: they must be trivariately normally distributed. This assumption is perhaps too strong for most developing country data sets. Furthermore, there were many difficulties with achieving convergence, and when the model is fed with wildly different starting parameter values, convergence might still be attained but yielding quite different estimates. Finally, as said earlier, we are still in the dark about which variables actually determine the income threshold. The two variables used, distance and past migration, are potentially good candidates, but there might be many other omitted variables in the threshold equation.

4 Inter-provincial migration and income inequality

Consider equation 5 in section 2.4. If we have specific functional forms for the growth rates $g_r(m_{sr}, Z_r)$ and $g_s(m_{sr}, Z_s)$, then a non-linear least squares technique can be employed to estimate the relationship between current period's income differential and migration, provinces growth rate determinants, and previous period's incomes. For this section, we employ a log-linear least squares model:

[10]We also ran the regressions with sending and province dummies, which yield almost the same result when distance is the threshold-determining variable. But for the case of past migration as the threshold-determining variable, convergence could not be achieved with the addition of these dummies.

$$ln(I_{sr,01}) \equiv ln(Y_{r,02}/Y_{s,02}) = \rho_0 + \rho_m ln(m_{sr}^{94-99}) + \rho_{yr} ln(Y_{r,99}) + \rho_{ys} ln(Y_{s,99}) + \rho_{gr} ln(g_{r,99-02}) + \rho_{gs} ln(g_{s,99-02}) + \epsilon_{sr} \quad (16)$$

In equation 16, provincial growth rates enter directly as independent variables, acknowledging that part of the impact of migration on income differential might go through these (although it is theoretically ambiguous whether this impact is negative or positive). The results (model I, table 8) show that a lower growth rate and initial income of the sending (poorer) province leads to a higher inter-provincial income differential, while a lower growth rate and initial income of the receiving (richer) province has the opposite effect, as expected. Not surprisingly, most of the income differential in 2002 comes from the income differential in 1999; the coefficients on 1999 per capita income of both sending and receiving provinces are very large, and also have very high t-statistics. Migration seems to have a mitigating effect on inter-provincial inequality, but the elasticity of this impact is very small: in model I, a 1% increase in out-migration rate from s to r is associated with a decrease in income differential of approximately .01%.

[Table 8 about here]

Given that a specific form for the theoretical relationship in 5 cannot be derived, we also tried non-parametric estimation. The disadvantage is the so-called "curse of dimensionality": non-parametric estimation becomes difficult when there is more than one predictor, because the potential approximation error grows very fast with the number of predictors. Accordingly, we regress the change in the relative income inequality on out-migration rate only (as opposed to a full non-parametric version of equation 16):

$$D \equiv ln\left(\frac{Y_{r,02}}{Y_{s,02}} - \frac{Y_{r,99}}{Y_{s,99}}\right) = f(m_{sr}^{94-99}) + \epsilon_{sr}$$

Figure 4 graphs the result. Again, as with the parametric analysis, there is overall a slight negative relationship between migration and relative income inequality. Note, however, that this relationship fluctuates widely as out-migration rate changes, which leads us to the next step.

[Figure 4 about here]

Several robustness checks and regression diagnostics were carried out for model I. In model II, migration is instrumented using the independent variables in the earlier regressions in section 3. The results do not change much compared with those from model I. In model III, two receiving provinces, Dak Lak and Lam Dong, are excluded from the sample. The results indicate that that these two provinces are very influential, as the coefficient on migration changes substantially once these are removed from the sample. These coffee-growing provinces in the Central Highlands region are among the largest migrant-receiving provinces; however, their per capita incomes decreased between 1999 and 2002 as the 1990s coffee boom ended.[11] HCMC as a receiving province is likewise found to be an outlier. This is not surprising, as HCMC is the largest and fastest-growing urban center in Vietnam and hence may behave differently compared with the rest of the country. Such information points to the varying impact of migration depending on the receiving province, which is in accordance with our prior knowledge.

We therefore estimate equation 16 separately for selected receiving provinces, or groups of them. Models I and II in table 9 indicate a negative and statistically significant relationship between migration and relative income inequality between the HCMC cluster and the rest of the country. As emphasized earlier, provinces in this cluster form the largest manufacturing center of Vietnam, making up 50% of the country's industrial output in 2002, and attracting 60% of licensed FDI projects and 53% of registered capital. This suggests that the labor market might play an important role in spreading the benefits of growth from this manufacturing cluster and in reducing regional inequality. In contrast, the opposite relationship is found for major industrial centers in the North (Hanoi, Hai Duong, Hai Phong, and Quang Ninh) and the rest of the country (models III and IV): migration and inequality have a positive association for these receiving provinces.

[Table 9 about here]

Contrasting results are also found for migration into rural areas. Model V shows no statistically significant impact of migration on income differentials between the receiving provinces in the Central Highlands– the largest rural receiving region– and other provinces. However, model VI shows astatistically significant but small negative impact of migration on income differentials between provinces in the Mekong River Delta– the "rice basket" of Vietnam– and other provinces.

Summing up, both parametric and non-parametric analyses provide evidence of a small impact of migration on income inequality between pairs of provinces. This impact varies widely, depending on the receiving provinces, and appears to be related to industrial structures in those provinces. In particular, a negative impact of migration on income differential is found on flows going to the export-oriented manufacturing cluster in HCMC and surrounding areas, while a positive impact is found on flows going to urban centers in the North. No statistically significant impact is found on flows going to the coffee-growing Central Highlands, while a small negative impact is found on flows going into the Mekong River Delta where the rice sector agricultural productivity is high. That the impact of migration on income inequality varies by receiving provinces is not a surprising result. It was discussed earlier that there is large variation in the economic characteristics of regions and provinces in Vietnam. As a result, population movements to different regions might represent entirely different types of migration, which in turn represent entirely different livelihood strategies at the household level.

[11] In Dak Lak and Lam Dong, where the coffee boom occurred, per capita incomes increased dramatically from about 215,000VND/person/month in 1994 to about 400,000VND in 1999 and then dropped back to about 230,000 and 238,000 in 2002, respectively. The increase in income in the 1990s could easily be attributed to the coffee boom, while the decrease in early 2000s could be the result of the coffee bust at the end of the 1990s.

It is worth re-emphasizing that in our analysis of the impact of migration on income inequality, we do not attempt to study inequality at the household level, nor inequality within each sending or receiving province, nor inequality for Vietnam as a whole. Given the available data and the aggregate level of analysis, we simply examine how "changes" in the income ratio of pairs of provinces vary with migration flows between them. This sheds light on inequality at the province level, and is meant to be complementary to household-level analysis.[12]

5 Conclusion

This paper has examined two questions: that of the determinants of inter-provincial migration flows in Vietnam, and that of the impact of these flows on inter-provincial inequality. Regarding the first question, we find that migration flows follow patterns predicted by theory: people move from low-income to high-income provinces. There is also evidence of a liquidity constraint effect which leads to poverty-related labor immobility at the provincial level.

Regarding the second question, some support is found for the income inequality-reducing effect of migration for flows into the manufacturing centers in the South and to the agriculturally rich Mekong Delta. Meanwhile, flows going into urban centers in the North or to other regions have no measurable effects. These results suggest that the impact of migration on income inequality between pairs of provinces is contingent on the impetus for the flow. Migration has a significant association with reduced inequality when the destination provinces are those hosting the greatest concentration of labor-intensive, export-oriented manufacturing industries. Other forms of growth, such as have occurred in the other net in-migration regions of Vietnam, do not. A deeper investigation of the reasons for this difference requires microeconomic data on migrants, jobs, and household incomes. This could permit testing of the relative magnitudes of 'brain drain' effects versus the gains from remittance income.

Everything considered, the evidence confirms that economic growth and internal migration are complements. On the one hand, migrants respond to increases in the marginal productivity of labor in sectors where Vietnam is exploiting its comparative advantage and acquiring new investments; that is, migration fuels growth. On the other hand, regression analysis indicates a robust and negative impact of migration on income inequality for those migration flows going to the largest manufacturing centers. This implies that migration helps offset some of the increase in spatial inequality caused by location- and sector-specific growth. At the same time, however, impediments to migration, such as liquidity constraints due to low incomes and imperfect capital markets, may imply persistence of poverty for populations in disadvantaged locations. Such evidence suggests that policies facilitating internal migration will be good both for economic growth and inequality reduction. It is important to note, however, that such policies must ensure broad-based access to migration across households and regions, so as to avoid increasing

[12] A recent addition to this literature is Hoang et al (forthcoming), who find that internal migration is pro-poor but also increases inequality within the sending region.

inequality within the sending areas.

The most serious constraint to the type of analysis employed in this paper is the restriction posed by aggregate data. Such data fail to account for different types of move, and specifically, do not do a good job of capturing return and circular migration. With these data, moreover, we observe only average characteristics and trends at provincial level; differences across individuals or among subgroups of the population are not taken into account in evaluating the decision to migrate. Finally, the gross migration rate msr for aggregates is an unbiased estimator for the underlying individual probability of migrating from s to r only if the characteristics affecting migration decisions are identically and independently distributed across households/individuals, such that behavior of a representative agent mimics that of the aggregate. This assumption is likely to be violated; conclusions on individual migration behavior using aggregate data should thus be drawn with care. The advantage of using these data, for all their limitations, is that they alone are nationally representative at a provincial scale. By gathering migration, employment and remittance data for nationally representative samples of households and individuals, future rounds of the national living standards survey could open the door to considerably richer modes of analysis.

References

[1] Andrienko, Yuri and Sergei Guriev (2004) "Determinants of inter-regional mobility in Russia" Economics of Transition Vol. 12(1), pp. 1-27

[2] Asian Development Bank (2007) "Rural labor markets and migration" Making Markets Work Better for the Poor Discussion Paper #18, Asian Development Bank, Hanoi

[3] Barro, Robert J. and Xavier Sala-i-Martin (2004) Economic Growth, The MIT Press, Cambridge Massachusetts

[4] Banerjee, Biswajit and S. M. Ravi Kanbur (1981) "On the Specification and Estimation of Macro Rural-Urban Migration Functions: With an Application to Indian Data" Oxford Bulletin of Economics and Statistics 43: 7-29

[5] de Brauw, Alan and Tomoko Harigaya (2007) "Seasonal Migration and Improving Living Standards in Vietnam" American Journal of Agricultural Economics, Vol. 89(2), pp. 430-447

[6] Connell, J.B., B. Dasgupta, R. Laishley, and M. Lipton (1976) Migration from rural areas: the evidence from village studies, Oxford University Press, Delhi

[7] Courchene, Thomas (1970) "Interprovincial migration and economic adjustment" The Canadian Journal of Economics Vol. 3(4), pp. 550-576

[8] Dang, Nguyen Anh (2003) "Migration in Vietnam: Review of Information on Current Trends and Patterns, and Their Policy Implications" Paper presented at the Regional Conference on Migration, Development, and Pro-Poor Policy Choices in Asia, Dahka, Bangladesh, 22-24 June 2003

[9] Dhar, Sanjay (1984) "Inter-state and within-state migration in India," Contractual Arrangements, Employment and Wages in Rural Labor Markets in Asia, eds. H.P. Binswanger and M. R. Rosenzweig, Economic Growth Center, Yale University, pp. 281-304

[10] Dickens, William T & Lang, Kevin (1985). "A Test of Dual Labor Market Theory," *American Economic Review*, 75(4), 792-805, September.

[11] Fan, C. Cindy (2005) "Modeling interprovincial migration in China 1985-2000" Eurasian Geography and Economics Vol. 46(3), pp. 165-184

[12] Greenwood, Michael J. (1997) "Internal Migration in Developed Countries," Handbook of Population and Family Economics, eds. M. R. Rosenzweig and O. Stark, Elsevier Science B.V.

[13] General Statistics Office (2001) Population and Housing Census Vietnam 1999: Completed Census Results, General Statistics Office, Hanoi: Statistical Publishing House

[14] General Statistics Office (1991) Population and Housing Census Vietnam 1989: Completed Census Results, General Statistics Office, Hanoi: Statistical Publishing House

[15] General Statistics Office, various years, Statistical Year Books, Hanoi, Vietnam

[16] Guest, Philip (1998) "The dynamics of internal migration in Vietnam," UNDP Discussion Paper 1, Hanoi, Vietnam

[17] Hardle, W., M. Muller, S. Sperlich, and A. Werwatz (2004) Nonparametric and Semiparametric Models, Springer Series in Statistics, Springer-Verlag Berlin Heidelberg

[18] Hoang, Thanh Ha, Nguyen Thi Nguyet, Nguyen Thu Phuong, Remco Oostendorp, Tran Ngo Thi Minh Tam "Determinants and Impacts of Migration in Vietnam," forthcoming, Market, Policy and Poverty Reduction in Vietnam, National Political Publisher (Henrik Hansen and Nguyen Thang, eds.), Hanoi (Vietnam)

[19] Lucas, Robert E.B. (1997) "Internal Migration in Developing Countries," Handbook of Population and Family Economics, Chapter 13, page 721-798

[20] Mekong Economics (2002) "A Study of Trade, Foreign Direct Investment, and Labour in Vietnam." An input to the DFID, ESCOR funded projected on Globalisation, Production, and Poverty: Macro, Meso, and Micro Level Studies, Hanoi

[21] Mueser, Peter (1989) “The spatial structure of migration: an analysis of flows between states in the USA over three decades” Regional Studies Vol. 23.3, pp. 185-200

[22] Nguyen, Duc Vinh (2003) "Urban savings and remittance in Vietnam" Migration in Vietnam: Theoretical Approaches and Evidence from a Survey, ed. Dang, Nguyen Anh

[23] Phan, Diep (2008) “Migration and credit constraints: evidence from four provinces in Vietnam” Manuscript, Department of Agricultural and Applied Economics, University of Wisconsin-Madison

[24] Schultz, T. Paul (1977) “A conditional logit model of internal migration: Venezuelan lifetime migration within education strata” Center Discussion Paper #266, Economic Growth Center, Yale University, September 1977

[25] Yang, Liu (2004) “Unequal Provinces but Equal Families? An Analysis of Migration and Inequality in Thailand” Essays on the Determinants and Consequences of Internal Migration, PhD dissertation, University of Chicago 2004

A Tables and Figures

Table 1: regional characteristics

	1994 per capita agricultural land (ha)	1994 average annual rice yield	1996-1997 net enrollment rate for upper secondary school	1994 population	1994-1999 net in-migration	1994-1999 net in-migration rate
Red River Delta	0.063	38.41	35.5	12,127,950	-285,570	-2.35%
North East Moutain	0.09	28.14	21.1	10,860,740	-105,031	-0.97%
North West Moutain	0.156	23.03	11.8	2,227,686	-2,929	-0.13%
North Central Coast	0.069	25.7	24.1	10,007,215	-234,113	-2.34%
South Central Coast	0.079	35.28	23.4	6,525,841	-93,548	-1.43%
Central Highland	0.293	24.15	16.7	3,062,293	265,763	8.68%
South East	0.229	26.97	16.3	7,671,727	115,979	1.51%
Mekong River Delta	0.179	39.54	13.6	16,132,024	-159,775	-0.99%
Hanoi	-	-	49.4	2,672,122	141,307	5.29%
HCMC	-	-	32.5	5,037,155	336,263	6.68%
Vietnam	0.103	31.99	22.4	76,324,753	-	-

Source: authors' calculation using data from different issues of the Statistical Year Book and the 1999 Census

Table 2: summary statistics

Variables	unit	Mean	Std. Dev.	Min	Max
94-99 migration flow	people	577	1,770	0	33,590
84-89 migration flow	people	395	1,091	0	21,353
distance	km	1,018	719	20	2,597
1986 pc industrial output	thousand VND, 1982 price	1.50	1.60	0.15	9.77
1994 monthly per capita income	thousand VND, 1994 price	161.9	48.7	99.2	413.3
1999 monthly per capita income	thousand VND, 1999 price	283.3	103.5	160.7	828.2
2002 monthly per capita income	thousand VND, 2002 price	318.3	117.4	173.1	904.1

Table 3a: OLS regression of migration flows with regional fixed effects

	1980s			1990s			
	Coeff.	t-stat		Coeff. (direct effect)	t-stat	total effects	indirect effects
ln(1984-1989 migration rate from s to r)	-	-		0.672	27.2	-	-
ln(bus distance)	-1.131	-48.01		-0.437	-14.2	-1.198	-0.760
ln(receiving province's per capita income, 1994)	-	-		1.560	13.2	-	-
ln(sending province's per capita income, 1994)	-	-		0.858	8.160	-	-
ln(receiving province's pc industrial output, 1986)	0.798	14.0		-	-	-	-
ln(sending province's pc industrial output, 1986)	0.093	1.84		-	-	-	-
sending region = North East	-0.212	-2.6		-0.687	-11.6	-0.829	-0.142
sending region = North West	-0.438	-3.5		-0.590	-5.3	-0.884	-0.294
sending region = North Central Coast	-0.172	-1.9		-0.361	-5.5	-0.476	-0.115
sending region = South Central Coast	-1.092	-11.1		-0.704	-7.7	-1.438	-0.734
sending region = Central Highland	-0.101	-0.9		-0.471	-4.9	-0.539	-0.068
sending region = South East	-0.858	-9.2		-0.401	-5.5	-0.978	-0.577
sending region = Mekong River Delta	-1.899	-20.2		-0.876	-10.7	-2.152	-1.276
sending region = urban areas	-0.480	-4.2		-0.618	-7.8	-0.940	-0.322
receiving region = North East	-1.283	-15.0		-0.160	-2.1	-1.022	-0.863
receiving region = North West	-0.968	-5.8		-0.340	-2.2	-0.991	-0.651
receiving region = North Central Coast	0.660	7.5		-0.303	-4.0	0.141	0.444
receiving region = South Central Coast	-0.339	-4.2		0.162	2.2	-0.066	-0.228
receiving region = Central Highland	0.991	7.6		0.931	11.0	1.597	0.666
receiving region = South East	0.813	9.2		0.266	3.3	0.812	0.546
receiving region = Mekong River Delta	-0.648	-8.5		-0.463	-6.2	-0.898	-0.436
receiving region = urban areas	-0.022	-0.2		0.212	2.5	0.197	-0.015
Constant	-1.223	-6.9		-12.066	-15.2	-12.888	-0.822
Adj. R2	0.60			0.73			
N	3613			3613			

Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r for the relevant period

b) Direct effects come from structural equation estimation

c) t-statistics are for structural equation estimates, and based on White heteroskedasticity consistent covariance matrix

d) The omitted sending and receiving region is Red River Delta

Table 3b: Logit regression of migration flows with regional fixed effects							
	1980s			1990s			
	Coeff.	t-stat		Coeff. (direct effect)	t-stat	total effects	indirect effects
ln(1984-1989 migration rate from s to r)	-	-		0.672	27.2	-	-
ln(bus distance)	-1.131	-47.940		-0.437	-14.1	-1.198	-0.760
ln(receiving province's per capita income, 1994)	-	-		1.559	13.2	-	-
ln(sending province's per capita income, 1994)	-	-		0.860	8.170	-	-
ln(receiving province's pc industrial output, 1986)	0.798	14.0		-	-	-	-
ln(sending province's pc industrial output, 1986)	0.092	1.830		-	-	-	-
sending region = North East	-0.214	-2.6		-0.699	-11.8	-0.843	-0.144
sending region = North West	-0.452	-3.6		-0.617	-5.5	-0.921	-0.304
sending region = North Central Coast	-0.177	-2.0		-0.371	-5.6	-0.490	-0.119
sending region = South Central Coast	-1.109	-11.2		-0.718	-7.8	-1.463	-0.745
sending region = Central Highland	-0.103	-0.9		-0.491	-5.1	-0.560	-0.069
sending region = South East	-0.869	-9.3		-0.412	-5.7	-0.996	-0.584
sending region = Mekong River Delta	-1.922	-20.5		-0.893	-10.9	-2.186	-1.292
sending region = urban areas	-0.495	-4.3		-0.636	-8.0	-0.969	-0.333
receiving region = North East	-1.283	-15.0		-0.160	-2.1	-1.022	-0.863
receiving region = North West	-0.968	-5.8		-0.340	-2.2	-0.991	-0.651
receiving region = North Central Coast	0.660	7.5		-0.303	-3.9	0.141	0.444
receiving region = South Central Coast	-0.339	-4.2		0.162	2.2	-0.066	-0.228
receiving region = Central Highland	0.990	7.6		0.931	10.9	1.596	0.666
receiving region = South East	0.813	9.2		0.266	3.3	0.812	0.547
receiving region = Mekong River Delta	-0.648	-8.5		-0.463	-6.2	-0.898	-0.436
receiving region = urban areas	-0.022	-0.2		0.212	2.5	0.197	-0.015
Constant	-1.187	-6.7		-12.030	-15.1	-12.828	-0.798
Adj. R2	0.60			0.73			
N	3613			3613			

Notes: a) dependent variable = log (gross migration rate from s to r divided by non-migration rate)

b) Direct effects come from structural equation estimation

c) t-statistics are for structural equation estimates, and based on White heteroskedasticity consistent covariance matrix

d) The omitted sending and receiving region is Red River Delta

Table 3c: 3SLS regression of migration flows with regional fixed effects					
	1980s			1990s	
	Coeff.	t-stat		Coeff. (direct effect)	t-stat
ln(1984-1989 migration rate from s to r)	-	-		0.599	3.6
ln(bus distance)	-1.131	-48.070		-0.517	-2.7
ln(receiving province's per capita income, 1994)	-	-		2.345	4.9
ln(sending province's per capita income, 1994)	-	-		0.981	7.110
ln(receiving province's pc industrial output, 1986)	0.798	14.9		-	-
ln(sending province's pc industrial output, 1986)	0.093	1.800		-	-
sending region = North East	-0.212	-2.6		-0.700	-8.1
sending region = North West	-0.438	-3.4		-0.599	-4.3
sending region = North Central Coast	-0.172	-1.8		-0.368	-3.7
sending region = South Central Coast	-1.092	-10.3		-0.785	-3.9
sending region = Central Highland	-0.101	-0.9		-0.496	-4.5
sending region = South East	-0.858	-8.9		-0.502	-2.7
sending region = Mekong River Delta	-1.899	-22.6		-1.044	-3.1
sending region = urban areas	-0.480	-3.6		-0.703	-4.6
receiving region = North East	-1.283	-15.5		-0.235	-1.1
receiving region = North West	-0.968	-7.6		-0.277	-1.4
receiving region = North Central Coast	0.660	6.7		-0.222	-1.5
receiving region = South Central Coast	-0.339	-3.2		0.144	1.4
receiving region = Central Highland	0.991	8.7		0.914	4.5
receiving region = South East	0.813	8.4		0.094	0.9
receiving region = Mekong River Delta	-0.648	-7.7		-0.693	-3.4
receiving region = urban areas	-0.022	-0.2		-0.090	-0.7
Constant	-1.223	-7.3		-16.665	-5.6
Adj. R2	0.60			0.72	
N	3613			3613	

Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r for the relevant period

b) The omitted sending and receiving region is Red River Delta

c) t-statistics are based on White heteroskedasticity consistent covariance matrix

Table 3d: OLS regression of migration flows with provincial fixed effects						
	1980s		1990s			
	Coeff.	t-stat	Coeff. (direct effect)	t-stat	total effects	indirect effects
ln(1984-1989 migration rate from s to r)	-	-	0.721	28.3	-	-
ln(bus distance)	-1.137	-51.950	-0.350	-11.8	-1.170	-0.820
ln(receiving province's per capita income, 1994)	-	-	3.034	10.3	-	-
ln(sending province's per capita income, 1994)	-	-	0.793	3.350	-	-
ln(receiving province's pc industrial output, 1986)	1.118	12.7	-	-	-	-
ln(sending province's pc industrial output, 1986)	0.529	6.8	-	-	-	-
Constant	-2.991	-11.3	-20.719	-9.2	-22.875	-2.156
Adj. R2	0.73		0.79			
N	3613		3613			

Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r for the relevant period

b) Direct effects come from structural equation estimation

c) t-statistics are based on White heteroskedasticity consistent covariance matrix

d) Fixed effects for sending and receiving provinces were included but not reported

Table 4: correlation among provincial characteristics							
	1994 per capita income	1986 per capita industrial output	1994 per capita agricultural land	1986 per capita land	% of ethnic minorities	upper secondary school enrollment rate	1994 population
All provinces							
1994 per capita income	1.00						
1986 per capita industrial output	0.75	1.00					
1994 per capita agricultural land	0.03	-0.19	1.00				
1986 per capita land	-0.22	-0.07	0.60	1.00			
% of ethnic minorities	-0.32	-0.24	0.25	0.52	1.00		
upper secondary school enrollment rate	0.12	0.36	-0.60	-0.37	-0.45	1.00	
1994 population	0.53	0.50	-0.36	-0.37	-0.39	0.24	1.00
Central Highlands							
1994 per capita income	1.00						
1986 per capita industrial output	-0.99	1.00					
1994 per capita agricultural land	-0.84	0.78	1.00				
1986 per capita land	-0.90	0.85	0.99	1.00			
% of ethnic minorities	-0.96	0.92	0.92	0.97	1.00		
upper secondary school enrollment rate	0.93	-0.89	-0.89	-0.95	-0.99	1.00	
1994 population	0.76	-0.75	-0.86	-0.83	-0.72	0.64	1.00
Urban areas							
1994 per capita income	1.00						
1986 per capita industrial output	0.98	1.00					
1994 per capita agricultural land	-0.82	-0.67	1.00				
1986 per capita land	-0.70	-0.55	0.96	1.00			
% of ethnic minorities	0.92	0.98	-0.52	-0.37	1.00		
upper secondary school enrollment rate	-0.68	-0.72	0.46	0.53	-0.69	1.00	
1994 population	0.94	0.92	-0.77	-0.74	0.85	-0.88	1.00

Table 5a: OLS regression of migration flows with additional provincial characteristics

	(I)			(II)	
	Coeff.	t-stat		Coeff	t-stat
ln(1984-1989 migration rate from s to r)	0.698	28.3		0.762	35.3
ln(bus distance)	-0.388	-13.1		-0.328	-12.7
ln(receiving province's per capita income, 1994)	1.362	11.2		1.484	16.6
ln(sending province's per capita income, 1994)	0.497	4.2		0.578	7.3
ln(receiving province's % of ethnic minorities 1994)	0.034	2.0		0.059	4.8
ln(sending province's % of ethnic minorities 1994)	-0.051	-2.9		-0.060	-5.5
ln(receiving province's upper secondary school enrollment rate 1994)	0.054	0.7		0.282	3.9
ln(sending province's upper secondary school enrollment rate 1994)	0.084	1.3		0.133	2.1
ln(receiving province's per capita agricultural land 1994)	-0.123	-1.6		0.100	2.2
ln(sending province's per capita agricultural land 1994)	-0.523	-6.5		-0.093	-1.9
sending region = North East	-0.199	-2.1		-	-
sending region = North West	0.299	1.9		-	-
sending region = North Central Coast	-0.127	-1.6		-	-
sending region = South Central Coast	-0.428	-4.1		-	-
sending region = Central Highland	0.609	3.9		-	-
sending region = South East	0.483	3.8		-	-
sending region = Mekong River Delta	0.005	0.0		-	-
sending region = urban areas	-0.726	-8.6		-	-
receiving region = North East	-0.235	-2.4		-	-
receiving region = North West	-0.347	-2.0		-	-
receiving region = North Central Coast	-0.387	-4.5		-	-
receiving region = South Central Coast	0.081	0.9		-	-
receiving region = Central Highland	0.945	6.6		-	-
receiving region = South East	0.335	2.7		-	-
receiving region = Mekong River Delta	-0.322	-2.8		-	-
receiving region = urban areas	0.137	1.6		-	-
Constant	-5.768	-4.1		-12.022	-13.3
Adj. R2	0.74			0.71	
N	3613			3613	

Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r
b) The omitted sending and receiving region is Red River Delta
c) t-statistics are based on White heteroskedasticity consistent covariance matrix

Table 5b: OLS regression of migration flows with additional provincial characteristics

	(III)			(IV)	
	Rural-rural migration			Rural-urban migration	
	Coeff.	t-stat		Coeff	t-stat
ln(1984-1989 migration rate from s to r)	0.761	34.7		0.801	17.0
ln(bus distance)	-0.334	-12.4		-0.204	-3.1
ln(receiving province's per capita income, 1994)	1.089	8.4		(dropped)	
ln(sending province's per capita income, 1994)	0.599	7.2		0.236	1.4
ln(receiving province's % of ethnic minorities 1994)	0.047	3.7		0.334	7.7
ln(sending province's % of ethnic minorities 1994)	-0.060	-5.3		-0.060	-3.0
ln(receiving province's upper secondary school enrollment rate 1994)	0.355	4.7		-0.450	-1.9
ln(sending province's upper secondary school enrollment rate 1994)	0.135	2.0		0.117	1.4
ln(receiving province's per capita agricultural land 1994)	0.278	4.4		-0.662	-6.3
ln(sending province's per capita agricultural land 1994)	-0.069	-1.3		-0.442	-4.6
Constant	-11.733	-12.5		7.110	4.9
Adj. R2	0.69			0.93	
N	3374			239	

Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r
b) The omitted sending region is Red River Delta
c) Regional fixed effects could not be included because of colinearity
d) t-statistics are based on White heteroskedasticity consistent covariance matrix

Table 5c: OLS regression of migration flows with additional provincial characteristics

	(V)			(VI)	
	Coeff.	t-stat		Coeff	t-stat
ln(1984-1989 migration rate from s to r)	0.659	26.2		0.661	26.4
ln(bus distance)	-0.436	-14		-0.425	-14
ln(receiving province's per capita income, 1994)	1.264	9.93		1.127	8.99
ln(sending province's per capita income, 1994)	0.488	4.01		0.401	3.19
ln(receiving province's population, 1994)	0.322	5.74		(dropped)	
ln(sending province's population, 1994)	0.686	12.3		1.031	11
ln(receiving province's % of ethnic minorities 1994)	-	-		0.054	3.11
ln(sending province's % of ethnic minorities 1994)	-	-		-0.045	-2.6
ln(receiving province's upper secondary school enrollment rate 1994)	-	-		0.136	1.65
ln(sending province's upper secondary school enrollment rate 1994)	-	-		0.254	4.03
ln(receiving province's per capita agricultural land 1994)	-	-		-0.344	-5.8
ln(sending province's per capita agricultural land 1994)	-	-		0.388	5.62
Constant	-18.152	-17		-18.410	-17
Adj. R2	0.75			0.71	
N	3613			3613	

Note: a) dependent variable is log of gross migration *flow*
b) Regional dummies were included but not reported
c) Receiving province's per capita income in 1994 was dropped in model (IV) because of colinearity
d) t-statistics are based on White heteroskedasticity consistent covariance matrix

Table 6a: OLS regressions of migration flows by quintiles of distance					
	1990s				
	1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile
ln(1984-1999 migration rate from s to r)	0.694	0.740	0.743	0.508	0.309
	(20.22)	*(24.52)*	*(15.82)*	*(8.18)*	*(6.55)*
ln(bus distance)	-0.572	-0.843	0.189	0.207	-9.854
	(-9.99)	*(-5.76)*	*(1.07)*	*(0.38)*	*(-7.87)*
ln(receiving province's per capita income, 1994)	0.657	1.031	1.847	2.295	1.995
	(3.49)	*(6.22)*	*(7.26)*	*(6.31)*	*(4.67)*
ln(sending province's per capita income, 1994)	**0.152**	**0.608**	**1.214**	**1.715**	**0.806**
	(0.85)	***(3.62)***	***(5.6)***	***(5.4)***	***(1.89)***
Adj. R2	0.72	0.71	0.73	0.71	0.62
N	688	730	732	732	731
Notes: regional dummies and constant terms were included but not reported					
1st quintile = shortest distance (meaning lower migration cost)					
t statistics in parentheses					

Table 6b: OLS regressions of migration flows by quintiles of past migration rate					
	1990s				
	1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile
ln(1984-1999 migration rate from s to r)	0.186	0.259	0.165	0.249	0.507
	(5.00)	*(6.09)*	*(4.46)*	*(4.46)*	*(6.24)*
ln(bus distance)	-0.612	-0.396	-0.568	-0.242	-0.298
	(-2.37)	*(-5.02)*	*(-9.56)*	*(-6.35)*	*(-7.66)*
ln(receiving province's per capita income, 1994)	1.898	1.962	1.674	1.001	1.370
	(4.59)	*(7.71)*	*(7.31)*	*(4.92)*	*(9.01)*
ln(sending province's per capita income, 1994)	**1.967**	**1.036**	**0.489**	**0.274**	**-0.081**
	(6.11)	***(3.89)***	***(2.15)***	***(1.45)***	***(-0.62)***
Adj. R2	0.37	0.41	0.39	0.38	0.68
N	723	723	722	723	722
Notes: regional dummies and constant terms were included but not reported					
1st quintile = lowest past migration rate (meaning higher migration cost)					
t statistics in parentheses					

Table 7: Regime-switching regression with unknown thresholds				
	(I)		(II)	
	Coeff.	S.E.	Coeff.	S.E.
Variables determining the threshold income				
ln(distance)	0.453***	0.009	-	-
ln(1984-1989 migration rate from s to r)	-	-	-0.29***	0.007
Variables determining migration flows				
ln(sending province's per capita income, 1994) below threshold	2.515***	0.094	2.358***	0.112
ln(sending province's per capita income, 1994) above threshold	0.64***	0.060	0.543***	0.060
ln(1984-1989 migration rate from s to r)	0.803***	0.011	0.84***	0.011
ln(bus distance)	-0.256***	0.019	-0.185***	0.019
ln(receiving province's per capita income, 1994)	1.246***	0.067	1.227***	0.065
Constant	-9.845***	0.453	-9.326***	0.454
# of observations	3613		3613	
# of iterations	24		113	
Mean log-likelihood	-1.33718		-1.34372	
Notes: a) dependent variable = ln(m_sr) = log of gross migration rate from s to r				
b) regional fixed effects were included but not reported				

Table 8: impact of migration on income differentials						
	(I)		(II)		(III)	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
ln(out-migration rate 1994-1999)	-0.011	-5.9	-0.015	-5.3	-0.002	-1.4
ln(pc income of richer province) 1999	0.927	55.8	0.936	56.4	0.974	69
ln(pc income of poorer province) 1999	-0.92	-27	-0.91	-26	-0.9	-34
ln(99-02 pc GDP growth rate of richer province)	0.035	3.2	0.034	2.99	0.018	2.6
ln(99-02 pc GDP growth rate of poorer province)	-0.07	-3.2	-0.07	-3.3	-0.06	-3.4
constant	-0.13	-0.8	-0.23	-1.3	-0.43	-3
N	1830		1801		1723	
Adj. R2	0.6		0.6		0.74	
a) dependent variable is log of ratio of per capita income of richer province to that of poorer province						
b) In model I & II, the whole sample was used						
c) In model II, the instruments are the same independent variables in table 2.3a						
d) In models III, Dak Lak and Lam Dong receiving provinces were taken out						

Table 9: impact of migration on income differentials for selected receiving provinces

	(I)	(II)	(III)	(IV)	(V)	(VI)
	HCMC	HCMC & surrounding	Hanoi	Northern industrial provinces	Dak Lak & Lam Dong	Mekong River Delta
ln(out-migration rate 1994-1999)	-0.049**	-0.036**	0.002	0.011*	-0.008	-0.017***
ln(pc income of richer province) 1999	-	0.994***	-	1.031***	0.062	0.618***
ln(pc income of poorer province) 1999	-0.759***	-0.783***	-0.828***	-0.857***	-0.933***	-0.813***
ln(99-02 pc GDP growth rate of richer province)	-	-0.011	-	-0.065	-0.089***	0.061**
ln(99-02 pc GDP growth rate of poorer province)	-0.057	-0.075*	-0.043	-0.061	-0.022	-0.095**
constant	5.341	-1.261	5.309***	-0.987	4.49	1.052*
N	60	230	59	175	140	522
Adj. R2	0.76	0.83	0.7	0.77	0.75	0.7

a) Dependent variable is log of ratio of per capita income of richer province to that of poorer province

b) * = 10% significance, ** = 5% significance, *** = 1% significance (White's heteroskedasticity robust covariance matrix)

c) Model (II) includes the following adjacent southern receiving provinces: HCMC, Ba Ria Vung Tau, Dong Nai, Binh Duong

In 2002 these four provinces make up 50% of total industrial output in Vietnam

d) Model (IV) includes the following northern receiving provinces: Hanoi, Hai Phong, Quang Ninh, Hai Duong

In 2002, these four provinces make up 14% of total industrial output of Vietnam

e) Dak Lak & Lam Dong are in the Central Highland where the coffee boom and then bust occurred; their per capita incomes increased dramatically from 1994 to 1999 but then dropped sharply from 1999 to 2002

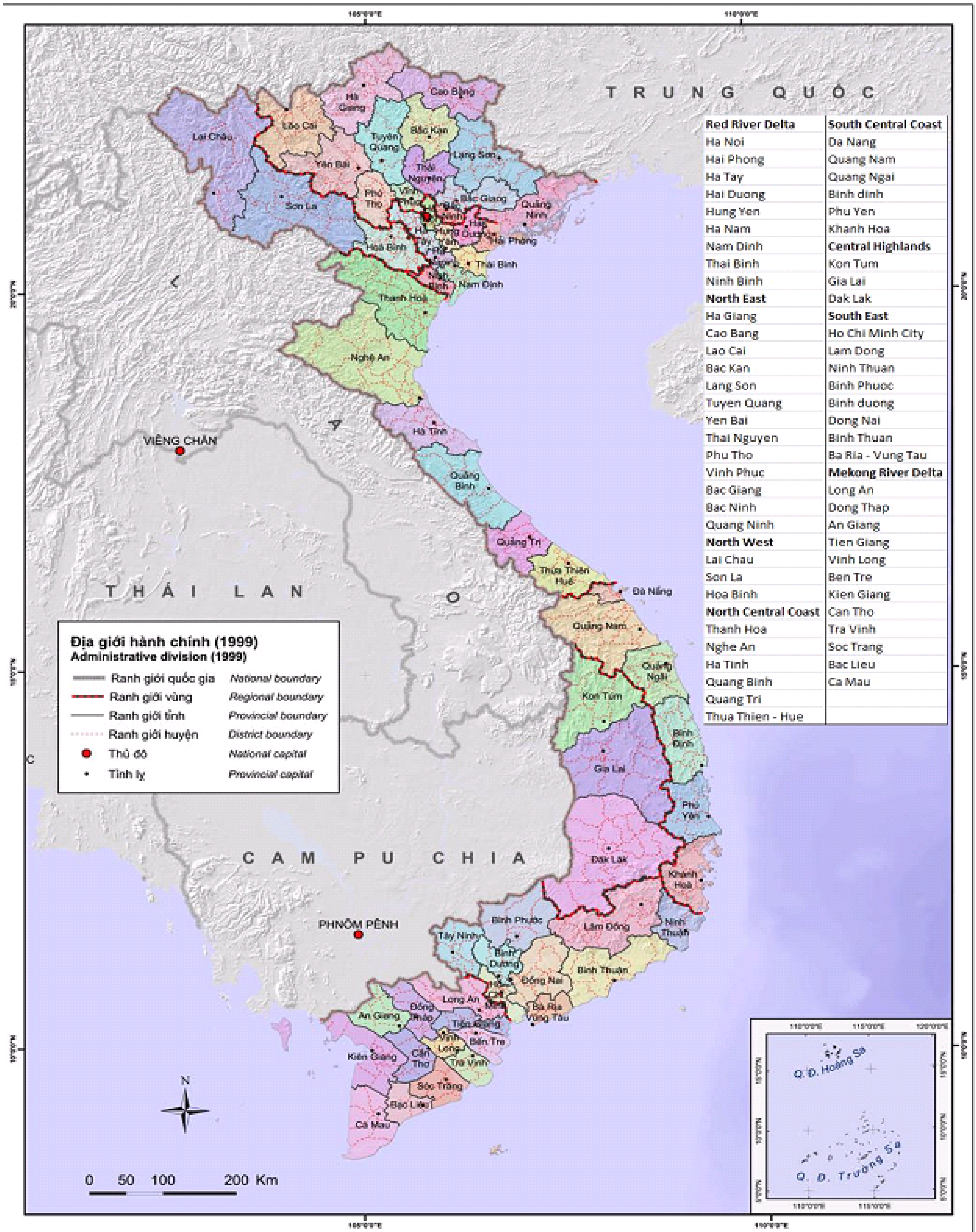

Figure 1: Map of Vietnam's provinces and regions

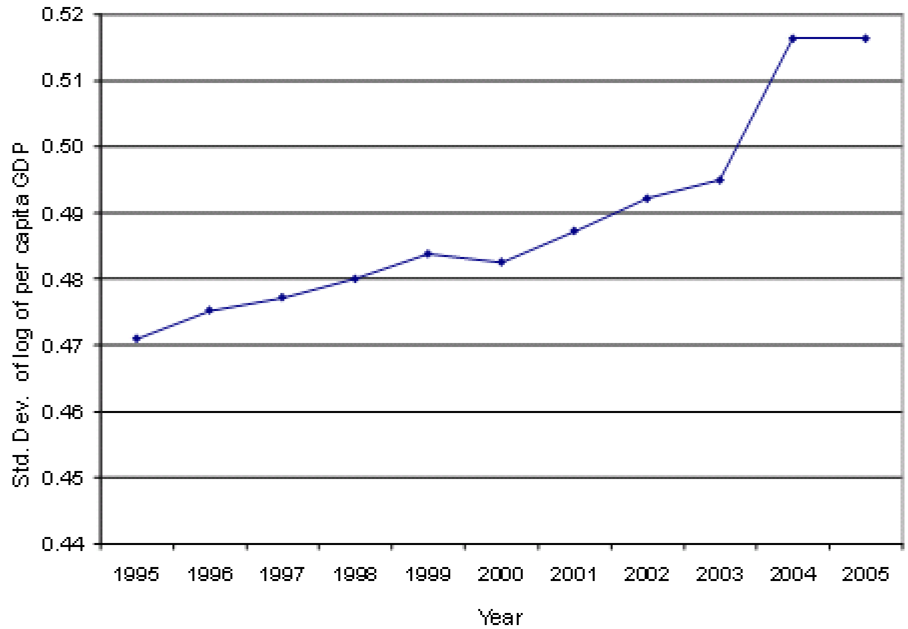

Figure 2: Trend in provincial income disparity in Vietnam 1995-2005

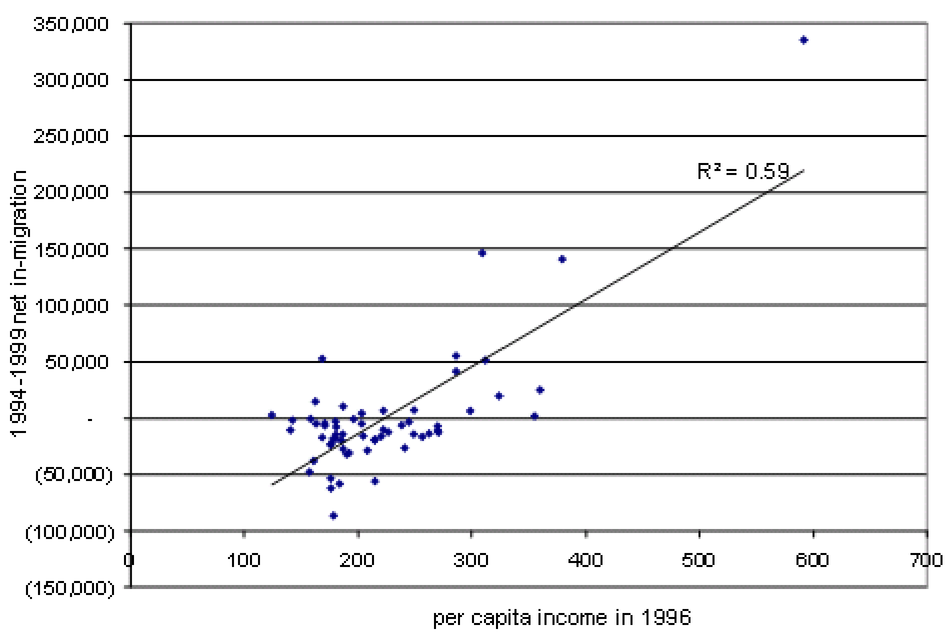

Figure 3: Per capita income vs. net in migration

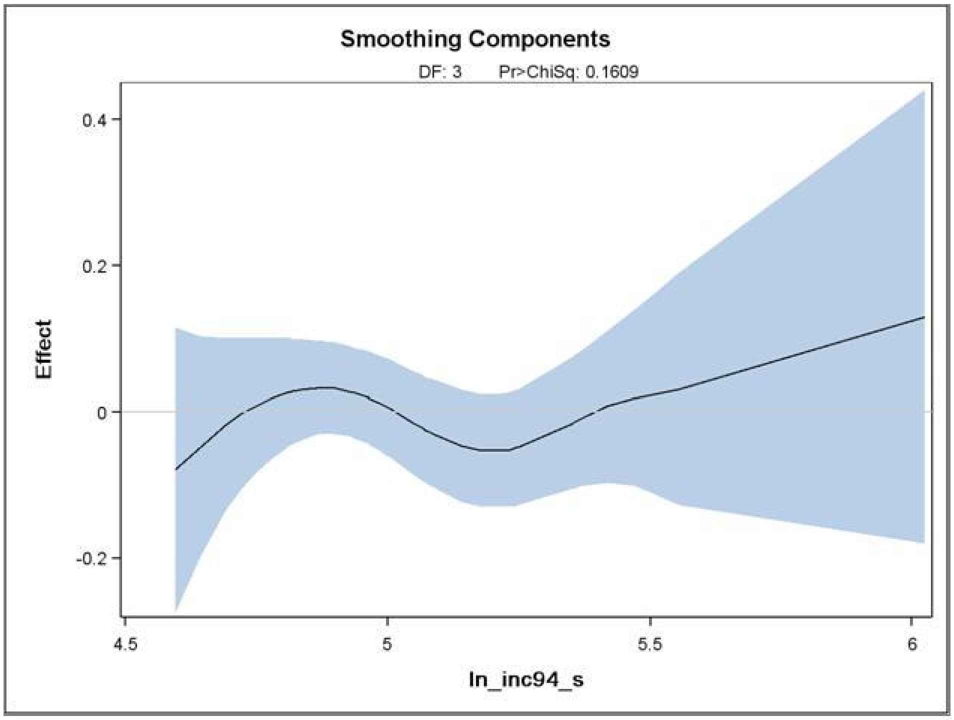

Figure 4a: Semi-parametric estimate (full sample result)

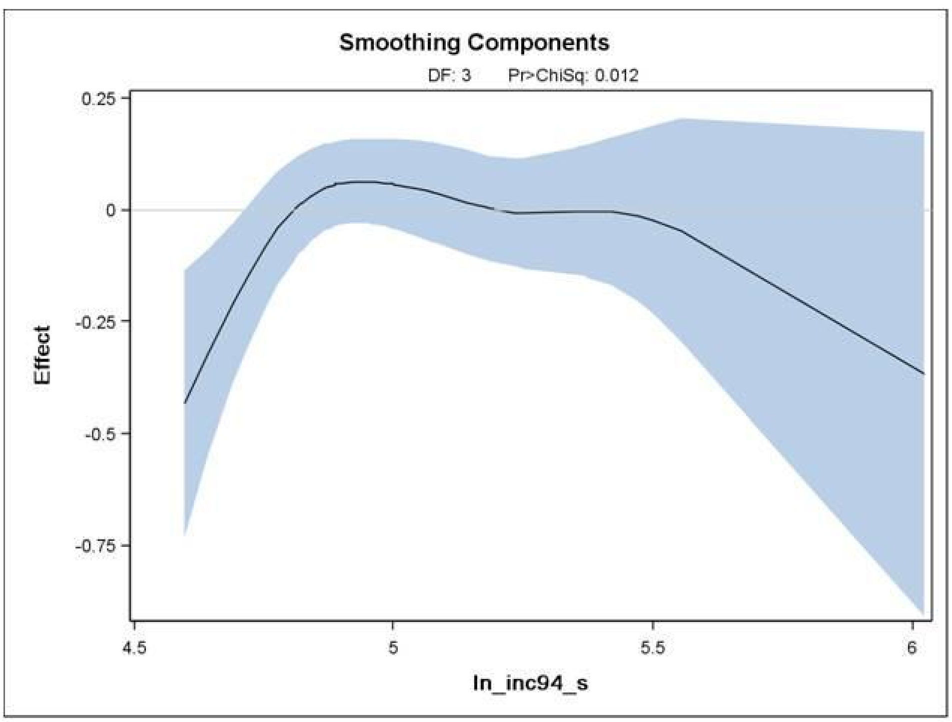

Figure 4b: Semi-parametric estimate
(sub-sample includes HCMC, Binh Duong, Dong Nai, Ba Ria - Vung Tau)

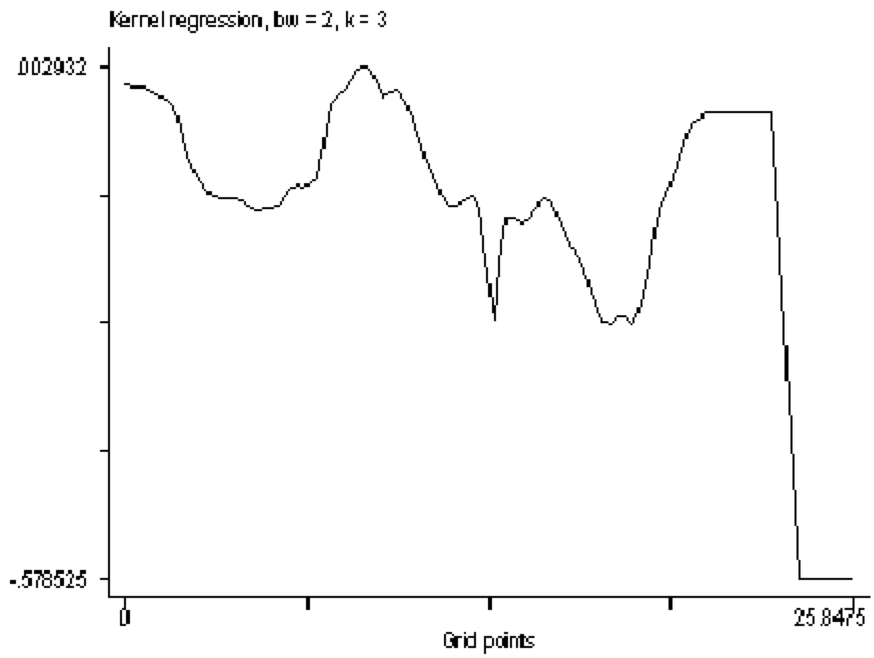

Figure 5: Conditional kernel density

Notes: bandwidth = 2, weight function = Epanechnikov; Independent variable: 94-99 out-migration rate from s to r (*1000); Dependent variable: Change in relative income inequality between s and r from 1999 to 2002