

Peeters, Ludo

**Conference Paper**

## Heterogeneity and asymmetry in cross-section gravity models of aggregate migration

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Peeters, Ludo (2010) : Heterogeneity and asymmetry in cross-section gravity models of aggregate migration, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118856>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# HETEROGENEITY AND ASYMMETRY IN CROSS-SECTION GRAVITY MODELS OF AGGREGATE MIGRATION

Ludo Peeters\*

**ABSTRACT.** The standard empirical gravity model of aggregate migration typically assumes homogeneity, quasi-symmetry, and (log-) linearity of bilateral relationships. However, estimation is likely to be plagued with severe misspecification bias if no attempt is made to control for unobserved migration selectivity and unmeasured spatial heterogeneity. This paper presents an alternative specification of the cross-section gravity model, controlling for unobserved unilateral (origin- and destination-specific) and, most importantly, bilateral (flow-specific) effects. Specifically, individual slopes account for correlated heterogeneity in the bilateral dimension, where linkage factors such as distance and previous migration (migrant stock) are allowed to interact with unobserved bilateral effects. This generalized gravity model is applied for an exploratory analysis of inter-state migration in Mexico and estimated by applying the Generalized Maximum Entropy method. The empirical importance of relaxing the standard assumptions is demonstrated by comparing the GME estimates with those obtained from OLS. Furthermore, it is found that variations in the unobserved individual (flow-specific) distance and migrant-stock effects account for 30 to 45% of the variation in (log) migration flows, whereas observed socio-economic “push and pull” factors at origins and destinations (unemployment rates, incomes per capita, etc.) have only a small explanatory power of 5 to 10%. Finally, it is shown that the inclusion of bilateral fixed effects in the double-log form of the gravity model is instrumental in keeping with arbitrary heteroskedasticity.

*JEL:* R23; C14

*Key words:* Gravity model; Aggregate migration; Cross section; Fixed effects; Heterogeneity; Asymmetry; Entropy estimation.

\* Hasselt University  
Department of Business and Economics  
KIZOK Research Centre for Entrepreneurship & Innovation  
Agoralaan – Gebouw D, 3590 Diepenbeek, BELGIUM  
[ludo.peeters@uhasselt.be](mailto:ludo.peeters@uhasselt.be)

## 1. INTRODUCTION

Spatial patterns of aggregate migration are the result of a set of complex combinations of populations' motivations and constraints that influence the size and composition of observed migration flows between places. Using aggregate data, however, puts obvious limits to the study of migration, since they do not tell us much about the characteristics of different subgroups of migrants and their distinctive spatial patterns of movement (who is moving to where). Yet, the gravity model of aggregate migration has been widely used as a baseline model for empirical studies of the determinants of migration and for simulating or predicting the impacts of spatial variations in policies and changes in economic, social, and other conditions. However, despite its long track record, the "standard" empirical gravity model suffers from important deficiencies when applied to aggregate migration flows and place characteristics.

The deficiencies can be traced back to (at least) three mutually related assumptions that are typically utilized in empirical implementations of the gravity model, which are not firmly grounded on theory and do not generally match with real phenomena. First, the standard empirical gravity model assumes *homogeneous* migrant populations and *perfect* information. In reality, however, aggregate migration represents more than a simple spatial re-distribution of "homogeneous" and "perfectly informed" people. Reliance on such assumptions makes it hard to explain, for instance, the puzzling phenomenon of counter-streams (or the very large volume of gross migration flows relative to net flows). Evidently, the prevalence of counter-streams can only be explained in a satisfactory way by making allowance for heterogeneity of people and places (e.g., [Greenwood and Hunt, 2003](#)) and imperfect information (e.g., [Allen and Eaton, 2005](#)). Also, migrants may select their destinations endogenously, for reasons largely unobservable to the researcher (and possibly linked with observable determinants of migration). Nevertheless, most studies of aggregate migration would simply "brush aside" any heterogeneity, selectivity, or imperfect information by assuming (quite unrealistically) that these are just randomly distributed across migrant populations and over space and can, thus, simply be "averaged out". Second, the standard empirical gravity model assumes *quasi-symmetry* ([Bavaud, 2002](#)). This essentially means that, say, the distance between two places has an equally impeding effect on migration regardless of the direction of the move. Yet, migrants may be more, or less, sensitive to distance in one direction than in another. For example, moving out of a (desolate) rural area into a (populous) urban area is something quite different from moving in the opposite direction, and moving in either direction likely involves distinct "types" of "qualities" of migrants. Hence, any sensible model of aggregate migration flows should somehow allow for asymmetry, or directional heterogeneity, in the deterring effect of distance, and permit observed migration flows in one or the other direction to reveal (in an ex-post sense) differences in motives, opportunities, and preferences (i.e., differences in objective function) across specific types of migrant groups (e.g., [Gordon and Vickerman, 1982](#); [Gordon and Molho, 1998](#)). [Millington \(2000\)](#), for example, provides a good illustration of this point in his study of the effect of aging on the sensitivity to migration stimuli: "*Taking any pairing of locations, the young may be moving (on net) in one direction while the old are moving in the other way*" (p. 525). Such observations also call for employing a more *generalized* conception of "distance" in studying aggregate migration (e.g., [Mueser, 1989a](#)), in order to acknowledge that physical distance is likely to be perceived differently by migrant sub-populations moving in multiple directions, and that, therefore, aggregate migration flows are likely to be sensitive to distance at varying levels of intensity—contingent upon differences in their age structure, educational attainment, risk profile, etc. (e.g., [Schwartz, 1973](#)). Third, for ease of estimation, "migration costs" in the standard empirical gravity model are generally assumed to be a (log-) *linear* function of observable proxies (e.g., distance, contiguity, and migrant stock). This, however, is likely to impose implausibly strong regularity conditions on the data (constant coefficients). Therefore, a more "flexible" specification of the gravity model should be used to account for possible *nonlinearities* in the ways in which, say, distance (or any other bilateral linkage factor) affects current migration flows (see also, for example, [Henderson and Millimet, 2008](#)).

Several attempts have been made in the empirical literature to "correct" for unobserved heterogeneity in gravity models of migration (or trade), either in a cross-sectional or a panel-data setting, by introducing *fixed effects* (FE).<sup>1</sup> However, there appears to be little agreement as to *how* to actually specify the FE, while most attempts generally fail to account for FE in the *bilateral* dimension. The cross-section gravity model (of trade) proposed by [Egger \(2005\)](#) controls for unobserved effects in the unilateral (origin- and destination-specific) dimensions by using a "two-way" FE estimator and a Hausman-Taylor instrumental-variables (IV) estimator. Most panel-data gravity models also tend to be limited to FE in *either* the unilateral dimensions (e.g., [Zavodny, 1997](#); [Etzo, 2008](#); [Mayda, 2009](#), [Warin and Blakely, 2009](#), [Leblang et al. 2009](#); [Balderas and Greenwood, 2009](#), all on migration) *or* the bilateral dimension (e.g., [Cheng and Wall, 2005](#), on trade; [Mayda, 2009](#), on migration). Two notable exceptions are the "three-way" FE gravity models (of trade) in [Baier and Bergstrand \(2007\)](#) and [Grant and Lambert \(2008\)](#), however at the expense

of “loosing” the parameter estimates associated with time-constant variables. Only two (published) gravity-based studies of aggregate migration can be found in the literature (as far as this author is aware of) that control for unobserved *bilateral* (flow-specific)—besides unilateral—effects, namely, the early study by Mueser (1989a) and the recent study by Pedersen et al. (2008). Both studies, however, suffer from important shortcomings. The first study actually develops a three-way FE gravity (spatial-interaction) model *avant la lettre*, using a first-differencing estimator to remove time-constant unobserved flow effects (what Mueser calls the “spatial structure”), but symmetry of the (generalized) distance effects was assumed throughout (which, in fact, allowed Mueser to factor them out). The second study employs a General Estimating Equations (GEE) random-effects (RE) estimator, hence assuming away—or, rather, disregarding (see their footnote 14 on p. 1175)—any correlated heterogeneity in the bilateral dimension.

Given the apparent shortcomings of these earlier approaches, the objective of this paper is to advance a novel three-way FE procedure as a workable solution to account for unobserved fixed heterogeneity along *all* relevant dimensions in a *cross-section* gravity model<sup>2</sup>—i.e., the unilateral (origin- and/or destination-specific) dimensions *and* the bilateral (flow-specific) dimension. To overcome the usual (identification) limitations of standard cross-sectional methods, the three-way FE approach will be implemented by using the *Generalized Maximum Entropy* (GME) methodology (Golan et al., 1996). GME has several advantages over earlier (traditional) approaches. First, through controlling for unobserved heterogeneity in the three dimensions (*i*, *j*, and *ij*) of the gravity equation, the proposed FE-GME estimator extracts the largest possible amount of information contained in the cross-sectional data and is, therefore, more likely to return consistent coefficient estimates, so that they are less likely to pick up unobserved effects. Second, it is well-known that (G)ME is able to handle “ill-posed” (underdetermined) inverse problems. As a result, the FE estimation can dispense with the need to “within-transform” the data—which, in any case, would preclude the estimation of the effects of at least some important determinants of migration. Third, and in contrast with all other gravity models of migration found in the literature, the FE-GME procedure allows for the provision of individual *slopes* on the bilateral linkage factors (distance, contiguity, and migrant stock), rather than just individual intercepts. What is particularly interesting about this approach is that it can be instrumental in “unchaining” the gravity model from its overly restrictive assumptions of homogeneity, quasi-symmetry, and (log-) linearity in the bilateral dimension.

To illustrate its potentials, the three-way FE gravity model will be estimated for an exploratory analysis by using cross-sectional data on aggregate inter-state migration in Mexico during the period 1995-2000.<sup>3</sup> While a more comprehensive and substantive analysis of migration in Mexico might be of interest to many researchers, such an endeavor is beyond the scope of the this paper (and relegated to future research). Yet, despite the partial nature of the present analysis, it is found that introducing fixed effects in both the unilateral and the bilateral dimensions has important empirical implications and should, therefore, deserve attention in the estimation of cross-section (as well as panel-data) gravity models of aggregate migration.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the standard gravity model and further develops the three-way FE gravity model of aggregate migration, along with a brief discussion of some of its accommodating features. Section 3 presents the GME estimator used to overcome the “ill-posed” nature of the model. In Section 4, the FE model is applied for the exploratory analysis of inter-state migration in Mexico. This section begins with an overview of the variables included in the empirical model, presents some basic descriptive statistics and preliminary empirical observations, and then moves on to discuss various results from the econometric estimation. Finally, Section 5 formulates some concluding comments.

## 2. THE GRAVITY MODEL OF MIGRATION

### 2.1 The standard gravity model

The use of modified gravity models became common in the migration literature, beginning in the 1960s, by including several other variables besides those of the basic gravity model such as population sizes in origins and destinations and the distances between origin and destinations (e.g., Greenwood and Hunt, 2003, p. 28). Much empirical research on aggregate place-to-place migration has involved estimating standard log-linearized (double-log) equations of the following form:

$$\ln M_{ij} = a_0 + \sum_k a_{1k} \ln X_{k,i} + \sum_k a_{2k} \ln X_{k,j} + b \ln D_{ij} + \sum_h c_h Z_{h,ij} + u_{ij} \quad (1)$$

where  $M_{ij}$  is the bilateral (gross) migration flow from origin  $i$  to destination  $j$ , and  $u_{ij}$  is an idiosyncratic error term. The variables  $X_{k,i}$  and  $X_{k,j}$  refer to population, income, unemployment rate, and other unilateral characteristics of origins and destinations, which are usually designated as “push” and “pull” factors. The variable  $D_{ij}$  refers to geographical distance, which is typically taken as the most important bilateral (gravity) factor separating origins and destinations.

Bilateral (gross) migration flows between  $i$  and  $j$  are assumed to be a positive function of net *expulsion* forces operating in the origin  $i$  (i.e., expulsion minus retention) and net *attraction* forces operating in the destination  $j$  (i.e., attraction minus repulsion), while it is inversely related to the friction caused by the distance between  $i$  and  $j$ . Distance is typically used as a proxy for both economic (e.g., transportation and information) costs and non-economic (e.g., psychic) costs of moving, where higher costs should impede migration between  $i$  and  $j$  (i.e., it is expected that  $b \leq 0$ , ceteris paribus). Some researchers would further include a set of “preference” variables,  $Z_{h,ij}$ , such as a contiguity/common-border indicator and migrant stock, where the latter ideally should reflect positive “network externalities”. These bilateral linkage factors should facilitate migration between  $i$  and  $j$  (i.e., it is expected that  $c_h \geq 0$ , ceteris paribus).

The standard cross-section gravity model has typically been estimated by using the convenient OLS estimator. However, since the explanatory variables included in Equation (1) can impossibly capture all the heterogeneity present in the data, the OLS estimator (and other conventional estimators, see below) are likely to produce biased and inconsistent results due to the likelihood of model misspecification.

## 2.2 The three-way FE gravity model

Given that the standard gravity model of aggregate migration in Equation (1) is likely to be plagued with several deficiencies, as already pointed out in the introductory section, a FE model eliminates the assumption that all determinants of migration flows are observed by the “omnipotent econometrician”.

Specifically, the following “three-way” FE model is proposed, written (in vector form) as

$$m_{ij} = a_0 + v_i + \omega_j + \mathbf{x}_i \mathbf{a}_1 + \mathbf{x}_j \mathbf{a}_2 + \mathbf{z}_{ij} \mathbf{b}_{ij} + \epsilon_{ij} \quad (2)$$

where lowercase letters denote log-transformed (non-dummy) variables. Importantly, the cross-section gravity model takes the form of a triple-indexed ( $i, j$ , and  $ij$ ) model. The unknown parameters  $v_i$  and  $\omega_j$  denote the unilateral two-way fixed effects associated with origin  $i$  and destination  $j$ , respectively. In other words,  $v_i$  accounts for all origin-specific (i.e., destination-invariant) and  $\omega_j$  for all destination-specific (i.e., origin-invariant) unobserved influences, which are assumed to have an intercept effect on current migration flows only. The vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$  contain origin- $i$  and destination- $j$  characteristics, such that

$$\mathbf{x}_i \mathbf{a}_1 = (x_{i1} \ x_{i2} \ \dots \ x_{iK}) \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1K} \end{pmatrix}; \quad \mathbf{x}_j \mathbf{a}_2 = (x_{j1} \ x_{j2} \ \dots \ x_{jK}) \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2K} \end{pmatrix}$$

and the vector  $\mathbf{z}_{ij}$  stands for the set of bilateral impediment (distance) and preference factors (contiguity, migrant stock), such that

$$\mathbf{z}_{ij} \mathbf{b}_{ij} = (C_{ij} \ s_{ij} \ d_{ij}) \begin{pmatrix} \lambda_{ij} \\ \eta_{ij} \\ \delta_{ij} \end{pmatrix}$$

where  $\lambda_{ij}$ ,  $\eta_{ij}$ , and  $\delta_{ij}$  are the individual *flow-specific* effects (individual slopes) of contiguity ( $C_{ij}$ ), migrant stock ( $s_{ij}$ ), and distance ( $d_{ij}$ ), respectively. Assuming further that  $\mathbf{b}_{ij} = \boldsymbol{\beta} + \boldsymbol{\vartheta}_{ij}$ , where  $\mathbb{E}(\boldsymbol{\vartheta}_{ij}) = \mathbf{0}$  by definition, it follows that

$$\mathbf{b}_{ij} = \boldsymbol{\beta} + \boldsymbol{\vartheta}_{ij} = \begin{pmatrix} \lambda + \kappa_{ij} \\ \eta + \mu_{ij} \\ \delta + \tau_{ij} \end{pmatrix}; \quad \mathbb{E}(\boldsymbol{\vartheta}_{ij}) = \mathbb{E} \begin{pmatrix} \kappa_{ij} \\ \mu_{ij} \\ \tau_{ij} \end{pmatrix} = \mathbf{0}$$

and substitution into Equation (2) yields

$$m_{ij} = a_0 + \mathbf{x}_i \mathbf{a}_1 + \mathbf{x}_j \mathbf{a}_2 + \mathbf{z}_{ij} \boldsymbol{\beta} + v_i + \omega_j + \mathbf{z}_{ij} \boldsymbol{\vartheta}_{ij} + \epsilon_{ij} \quad (3)$$

which includes the following “composite” error term:

$$\begin{aligned} \xi_{ij} &= v_i + \omega_j + \mathbf{z}_{ij} \boldsymbol{\vartheta}_{ij} + \epsilon_{ij} \\ &= v_i + \omega_j + \kappa_{ij} c_{ij} + \mu_{ij} s_{ij} + \tau_{ij} d_{ij} + \epsilon_{ij} \end{aligned} \quad (4)$$

where it is assumed that  $\epsilon_{ij}$  is uncorrelated with the fixed effects and observed variables in the model.

The FE allow the unobserved *flow-specific* heterogeneity, contained in the vector  $\boldsymbol{\vartheta}_{ij}$ , to *interact* with the observed bilateral linkage factors.<sup>4</sup> For example, the *individual* effect of distance ( $d_{ij}$ ) along the  $i$ - $j$  path depends on the unobserved effect  $\tau_{ij}$  (which may or may not be tied to  $v_i$  and/or  $\omega_j$ ), while the *population-averaged* effect of distance is  $\mathbb{E}(\delta_{ij}) = \mathbb{E}(\delta + \tau_{ij}) = \delta + \mathbb{E}(\tau_{ij}) = \delta$ . Furthermore, asymmetry in spatial relationships requires that  $\tau_{ij} \neq \tau_{ji}$ ,  $\kappa_{ij} \neq \kappa_{ji}$ , and  $\mu_{ij} \neq \mu_{ji}$ .<sup>5</sup>

### 2.3 Econometric implications

The model presented in Equation (3) has four important and related econometric implications. First, after controlling for origin- and destination-fixed effects ( $v_i, \omega_j$ ) and flow-fixed effects ( $\boldsymbol{\vartheta}_{ij}$ ), it is more realistic to maintain the hypothesis that  $\mathbb{E}(\epsilon_{ij} | \mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij}, v_i, \omega_j, \boldsymbol{\vartheta}_{ij}) = 0$  and, hence,  $\mathbb{E}(m_{ij} | \mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij}, v_i, \omega_j, \boldsymbol{\vartheta}_{ij}) = \alpha_0 + \mathbf{x}_i \boldsymbol{\alpha}_1 + \mathbf{x}_j \boldsymbol{\alpha}_2 + \mathbf{z}_{ij} \boldsymbol{\beta}$  (see Wooldridge, 2002, 2005, in a panel-data framework).

Second, by explicitly conditioning on  $\boldsymbol{\vartheta}_{ij}$ , apart from  $v_i$  and  $\omega_j$ , it is underscored that the dummy-interaction terms represent unobserved heterogeneity that is *non-randomly* distributed as well as *distinctively* related to each one of the bilateral linkage variables. When, say,  $\tau_{ij}$  is negatively correlated with  $d_{ij}$ , OLS is likely to underestimate the (negative) effect of distance.<sup>6</sup>

Third, the FE model deviates in a fundamental way from the conventional log-linearized version of the gravity model (implying common and strictly monotonic relationships). By the very nature of the varying slopes ( $\mathbf{b}_{ij}$ ), the fixed-effects model provides a good approximation to (unspecified) “non-linear” relationships; that is, (i) there are as many slopes as there are individuals (migration flows), and (ii) each of the individual slopes is the sum of two parts, a “direct” effect (contained in  $\boldsymbol{\beta}$ ) and an “indirect” effect (contained in  $\boldsymbol{\vartheta}_{ij}$ ) due to omitted variables (Swamy and Tavlak, 1995). Allowing for nonlinearities makes the model less subject to specification errors and should, therefore, be helpful in obtaining more consistent estimates of the population-averaged *direct* effects of the bilateral linkage factors.

Finally, the GME estimator (to be discussed shortly) used in this paper to implement the FE gravity model does *not* require any “within”-transformation of data—which would wipe out all state-specific (e.g., Egger, 2005) or flow-specific influences (e.g., Mayda, 2009). So, in contrast with most conventional approaches, FE-GME allows the recovery of the coefficients of *all* explanatory variables included in the gravity equation.

Apart from these implications for estimation, it should be mentioned that the GME estimator applied in this paper is similar in kind to a dummy-variables type of estimator. As a result, the unobserved bilateral effects,  $\boldsymbol{\vartheta}_{ij} = (\kappa_{ij}, \mu_{ij}, \tau_{ij})$ , can be treated as parameters to *estimate*, rather than just “nuisance parameters”. The advantage of this approach is that it provides useful information on the heterogeneity of the effects of the bilateral linkage variables (i.e., more information can be extracted than any standard model would permit). In this respect, I exploit the notion that, even though the individual effects may not be estimable in a precise way, their distribution or *variation* with other variables may be informative (Cameron and Trivedi, 2005, p. 704).<sup>7</sup>

### 2.4 What can be gained from introducing bilateral fixed effects?

It is worth pointing to a number of ways in which the three-way FE gravity model can be helpful in resolving, at least partly, the problems related to spatial heterogeneity, the potential endogeneity of the migrant-stock variable, and the heteroskedasticity in the migration data.

#### *Spatial heterogeneity*

Introducing FE as related to distance can be helpful in accounting for unobserved spatial heterogeneity and, hence, in counteracting common criticisms concerning the inadequacy of a single physical distance variable to properly represent spatial relationships, particularly if the analysis works with sizeable geographical areas (e.g., states or countries) as migration-defining boundaries. For example, migratory



moves over relatively short distances between neighboring areas are said to be poorly represented by the distances between the areas' centroids, which in many instances overstate the actual distances over which a large proportion of the moves take place. The inclusion of a contiguity (common-border) dummy would only partly resolve this problem, since "contiguity" signifies different things in different locations, given that the geographical units vary widely in size, that populations and "economic functions" are unevenly distributed over space within each geographical unit, etc. For example, in the case of the Mexican states, contiguity of Distrito Federal and the state of México does not mean the same thing as contiguity of Chihuahua and Durango.

An important advantage of the proposed FE model is that it dispenses with the overly restrictive assumptions of constancy and symmetry of the distance and contiguity parameters. Accordingly, the flow-specific fixed effects returned by the model may be reflective of significant unobserved (or unmeasured) heterogeneity of both geographical units and migrant populations.

#### *Endogeneity of the migrant stock*

Another problem, particularly in a cross section, is the difficulty in measuring the role of pre-existing migration networks, which can be important in guiding later migrants' destination choices. Networks, whether family or friends, carry information about potential destinations (job opportunities, housing conditions, etc.), and can aid adjustment and assimilation of newcomers at the destination. Hence, the "cost" of moving to a particular destination is less when there is already an established network of previous migrants (e.g., Carrington et al., 1996).

A common approach is to include the *migrant stock* as a proxy for network effects (e.g., Nelson, 1959; Greenwood, 1969, 1970; Levy and Wadycki, 1973; Kau and Sirmans, 1979). However, the migrant-stock coefficient may also reflect the effect of *unobserved* factors which affected previous migration and continue to affect current migration. This, obviously, creates an *identification* problem. In other words, current migration and migrant stock (which contains information about previous migration) may be *jointly* determined by unobserved linkage factors that have a persistent bearing on migration, being largely responsible for the *path dependence* (and self-reinforcing nature) of migratory flows. Thus, having the migrant-stock variable on the right-hand side of the gravity equation may imply that the convenient OLS estimator is subject to *endogeneity bias* (omitted variables). As a result, the OLS estimate of the migrant-stock coefficient is likely to overstate the "true" network effect.<sup>8</sup>

To overcome this problem, attempts have been made in the literature to use *both* the migrant stock *and* lagged migration flows in the cross-section gravity equation, where the former is interpreted as capturing network externalities and the latter as controlling for "habit persistence" (e.g., Dunlevy, 1993), "cumulative causation" (e.g., Massey, 1990), or "herd behavior" (e.g., Bauer et al., 2007). Including lagged migration flows, however, reduces the gravity model to a simple *partial-adjustment* model, which is marred by several factors.<sup>9</sup> Besides, it is surprising to see that not much progress has been made ever since the 1970s; in new jargon, the partial-adjustment mechanism introduced by Dunlevy and Gemery (1977) is now simply labeled herd behavior. So, the recent stock-flow approach appears to be just "old wine in new bottles".

Introducing fixed effects as related to migrant stock can offer an appealing alternative to including lagged flows in controlling for unobserved factors responsible for potential herd migration and, thus, in obtaining a more consistent estimate of the population-averaged network effect—to the extent that the effect of the migrant stock on current migration can be reasonably decomposed into two parts, a *direct* network effect, and an *indirect* (unobserved) herd effect captured by the flow-fixed effect. Moreover, introducing flow-specific effects relaxes the (unrealistic) assumption that persistence of migration is equally strong along *all* migration paths. However, there might be strong persistence along *some* "geographical channels" (e.g., Carrington et al., 1996), while persistence might be weak, or absent, along *others*—a point which has been generally overlooked in previous literature.

#### *Heteroskedasticity*

The heteroskedasticity problem associated with the log-linear form of the gravity model was already acknowledged in an early paper by Flowerdew and Aitkin (1982). More recently, Santos Silva and Tenreiro (2006) further pointed to the *bias*es that may sneak in due to heteroskedasticity. Therefore, they recommend the use of the Poisson level specification of the gravity model. Nonetheless, I stick with the log-linear form, because the individual slope coefficients actually imply that the composite error term ( $\xi_{ij}$ ) term in Equation (3) almost certainly accounts for heteroskedasticity in the migration data.

To get a sense of how this works, it should be noted that, *ceteris paribus*, observed migration flows are

likely to be larger between contiguous locations ( $C_{ij} = 1$ ) than between non-contiguous locations ( $C_{ij} = 0$ ), for higher levels of the migrant stock ( $s_{ij}$ ), and/or for shorter distances ( $d_{ij}$ ). Therefore, given that  $\xi_{ij} = v_i + \omega_j + \kappa_{ij}C_{ij} + \mu_{ij}s_{ij} + \tau_{ij}d_{ij} + \epsilon_{ij}$ , and assuming  $\kappa_{ij} > 0$ ,  $\mu_{ij} > 0$ , and  $\tau_{ij} < 0$  (indicating “above-average” effects), larger migration flows are likely to be associated with larger values of the composite error, and vice versa. Thus, the FE gravity model with individual slope coefficients on the bilateral linkage factors allows the error variance to be non-constant and directly related to the size of the migration flows. Moreover, the error variance adjusts itself “independently” of the conditional mean (i.e., without any prior restrictions) and, hence, makes allowance for potential *over-dispersion* (the occurrence of the latter is essentially an empirical matter, as theory generally does not provide any guidance concerning the pattern of the error variance and potential departures from the Poisson assumption).<sup>10</sup>

### 3. ESTIMATION OF THE FIXED-EFFECTS GRAVITY MODEL

The introduction of three-way FE into the gravity model creates a challenging estimation problem. Specifically, Equation (3) allows three full sets of flow-specific dummies to interact with contiguity, migrant stocks, and distance, besides the origin- and destination-fixed effects, which are assumed to absorb the unobserved heterogeneity in the cross section. This, obviously, implies that the model is *over-parameterized* (i.e., the number of unknown parameters is far much greater than the number of observations). Therefore, “classical” estimation methods cannot be generally used to identify the parameters of the model. In order to “resolve” this identification problem, I use a novel FE application of the *Generalized Maximum Entropy* (GME) estimator.

#### 3.1 The Generalized Maximum Entropy estimator

The literature on entropy econometrics is too vast to review here (see, for example, Golan et al., 1996; Golan, 2006; Golan and Maasoumi, 2008; for some interesting performance properties of GME, see Shen and Perloff, 2001). However, it is noteworthy to list a number of advantages that the GME method has over classical estimation methods, which will be fully exploited in this paper. First, GME is a shrinkage estimator, which brings to bear a combination of non-sample (prior) and sample information in order to regularize the “ill-posed” (over-parameterized) nature of the inverse problem, as illustrated by the famous Jaynes’ die problem (Jaynes, 1963; see also Golan et al., 1996; Paris and Howitt, 1998). Second, GME uses minimal distributional assumptions, whereas classical methods rely on specific asymptotic distributions in order to make statistical inferences. Third, earlier simulations have shown that GME outperforms many standard estimators (including the OLS and uninformative-prior Bayesian) in terms of mean squared error and average bias. Fourth, GME allows one to impose (in)equality restrictions, which do not compromise the estimation (and which is easy to implement). Finally, GME is generally found to be superior to classical estimation methods in extracting information from “noisy” data.

For expository purposes, I present a simplified version of the three-way FE gravity model, including distance as the sole bilateral linkage factor:

$$\begin{aligned} m_{ij} &= a_0 + v_i + \omega_j + a_1x_i + a_2x_j + \delta_{ij}d_{ij} + \epsilon_{ij} \\ &= a_0 + v_i + \omega_j + a_1x_i + a_2x_j + (\delta + \tau_{ij})d_{ij} + \epsilon_{ij} \\ &= a_0 + a_1x_i + a_2x_j + \delta d_{ij} + \xi_{ij} \end{aligned} \quad (5)$$

where  $\xi_{ij} \equiv v_i + \omega_j + \tau_{ij}d_{ij} + \epsilon_{ij}$  is the composite” error term,  $\delta + \tau_{ij} \leq 0$  (i.e., migration is deterred by distance, ceteris paribus), and  $\tau_{ij} \neq \tau_{ji}$  (asymmetry or directional heterogeneity).

To implement GME, the model in Equation (5) is re-parameterized and converted into a constrained (primal) optimization problem, where the objective function consists of the joint entropy in Equation (6) below (i.e., the sum of the entropies for coefficient distributions and error distributions). This objective function is to be maximized, subject to the relevant data-consistency, (in)equality, and normalization (adding-up) constraints.

#### Re-parameterization

The constant parameters in  $\theta = (a_0, a_1, a_2)$ , the unobserved state-fixed effects  $(v_i, \omega_j)$ , and the individual flow-specific effects  $(\delta_{ij})$  are defined as linear combinations of different sets of unknown probability vectors (of dimension  $Q \geq 2$ ):  $\mathbf{p}_{a_0} = (p_{a_0}^1, \dots, p_{a_0}^Q)$ ,  $\mathbf{p}_{a_1} = (p_{a_1}^1, \dots, p_{a_1}^Q)'$ ,  $\mathbf{p}_{a_2} = (p_{a_2}^1, \dots, p_{a_2}^Q)'$ ,  $\mathbf{p}_{v_i} = (p_{v_i}^1, \dots, p_{v_i}^Q)'$ ,  $\mathbf{p}_{\omega_j} = (p_{\omega_j}^1, \dots, p_{\omega_j}^Q)'$ ,  $\mathbf{p}_{\delta_{ij}} = (p_{\delta_{ij}}^1, \dots, p_{\delta_{ij}}^Q)'$ , and a common parameter support vector



$\mathbf{z} = (z^1, \dots, z^Q)'$ . The idiosyncratic error component,  $\epsilon_{ij}$ , is defined as a linear combination of a set (of dimension  $G \geq 2$ ) of unknown probabilities  $\mathbf{p}_{\epsilon_{ij}} = (p_{\epsilon_{ij}}^1, \dots, p_{\epsilon_{ij}}^G)'$  and a common error support vector  $\mathbf{z}_\epsilon = (z_\epsilon^1, \dots, z_\epsilon^G)'$ . Then, the coefficients on the observables, the unobserved heterogeneity effects, and the error terms are re-parameterized as  $a_0 = \mathbf{p}'_{a_0} \mathbf{z}$ , etc.,  $v_i = \mathbf{p}'_{v_i} \mathbf{z}$ , etc., and  $\epsilon_{ij} = \mathbf{p}'_{\epsilon_{ij}} \mathbf{z}_\epsilon$ , respectively.

### Optimization problem formulation

The GME formulation of the (simplified) FE gravity model is as follows:

$$\begin{aligned} \text{Max}_{\mathbf{p}} H(\mathbf{p}) = & -\mathbf{p}'_{a_0} \log(\mathbf{p}_{a_0}) - \mathbf{p}'_{a_1} \log(\mathbf{p}_{a_1}) - \mathbf{p}'_{a_2} \log(\mathbf{p}_{a_2}) - \sum_{i=1}^N \mathbf{p}'_{v_i} \log(\mathbf{p}_{v_i}) \\ & - \sum_{j=1}^N \mathbf{p}'_{\omega_j} \log(\mathbf{p}_{\omega_j}) - \sum_{i \neq j} \sum_{j \neq i} \mathbf{p}'_{\delta_{ij}} \log(\mathbf{p}_{\delta_{ij}}) - \sum_{i \neq j} \sum_{j \neq i} \mathbf{p}'_{\epsilon_{ij}} \log(\mathbf{p}_{\epsilon_{ij}}) \end{aligned} \quad (6)$$

such that

$$m_{ij} = (\mathbf{p}'_{a_0} \mathbf{z}) + (\mathbf{p}'_{a_1} \mathbf{z}) x_i + (\mathbf{p}'_{a_2} \mathbf{z}) x_j + (\mathbf{p}'_{v_i} \mathbf{z}) + (\mathbf{p}'_{\omega_j} \mathbf{z}) + (\mathbf{p}'_{\delta_{ij}} \mathbf{z}) d_{ij} + (\mathbf{p}'_{\epsilon_{ij}} \mathbf{z}_\epsilon), \quad \forall i, j (i \neq j) \quad (7)$$

$$\sum_{i=1}^N (\mathbf{p}'_{v_i} \mathbf{z}) = 0; \quad \sum_{j=1}^N (\mathbf{p}'_{\omega_j} \mathbf{z}) = 0 \quad (8)$$

$$\sum_{i \neq j} \sum_{j \neq i} (\mathbf{p}'_{\epsilon_{ij}} \mathbf{z}_\epsilon) = 0 \quad (9)$$

$$\mathbf{p}'_{\delta_{ij}} \mathbf{z} \leq 0, \quad \forall i, j (i \neq j) \quad (10)$$

$$\sum_{q=1}^Q p_{a_0}^q = 1; \quad \sum_{q=1}^Q p_{a_1}^q = 1; \quad \sum_{q=1}^Q p_{a_2}^q = 1 \quad (11)$$

$$\sum_{q=1}^Q p_{v_i}^q = 1; \quad \sum_{q=1}^Q p_{\omega_j}^q = 1, \quad \forall i, j$$

$$\sum_{q=1}^Q p_{\delta_{ij}}^q = 1; \quad \sum_{g=1}^G p_{\epsilon_{ij}}^g = 1, \quad \forall i, j (i \neq j)$$

Equation (6) represents the entropy objective, which is subject to the data-consistency constraints in Equation (7). The constraints in (8) and (9) preserve zero-mean origin-/destination-fixed effects and zero-mean realized errors, respectively. The constraint in (10) imposes a weak inequality restriction, in order to guarantee that all distance effects are returned as negative (or zero) values. Finally, the normalization constraints in (11) ensure that all probabilities add up to one. After solving the entropy optimization problem in (6) through (11), the parameter estimates and the error terms can be recovered as  $\hat{a}_0 = \hat{\mathbf{p}}'_{a_0} \mathbf{z}$ , etc.,  $\hat{v}_i = \hat{\mathbf{p}}'_{v_i} \mathbf{z}$ , etc., and  $\hat{\epsilon}_{ij} = \hat{\mathbf{p}}'_{\epsilon_{ij}} \mathbf{z}_\epsilon$ , respectively, where  $\hat{\mathbf{p}}_{a_0}$ , etc.,  $\hat{\mathbf{p}}_{v_i}$ , etc., and  $\hat{\mathbf{p}}_{\epsilon_{ij}}$  are the corresponding estimated probabilities.

For the “full” model, in line with Equation (3), two non-negativity constraints on the contiguity and migrant-stock coefficients have been added, along with two sets of “stochastic” constraints, where both contiguity and migrant stock are “forced” to exert *opposite*—compared to distance—effects, *ceteris paribus*, on migration along each path. Specifically, I experiment with the following “non-linear” constraints on the individual migrant-stock and contiguity slopes, respectively:

$$\eta_{ij} = \pi_{ij}^\eta \delta_{ij} + v_{ij}^\eta, \quad \pi_{ij}^\eta \leq 0 \quad (12)$$

$$\lambda_{ij} = \pi_{ij}^\lambda \delta_{ij} + v_{ij}^\lambda, \quad \pi_{ij}^\lambda \leq 0 \quad (13)$$

Although the choice of these constraints is to a certain extent arbitrary, they are flexible enough to allow for a wide range of relationships.<sup>11</sup>

Also, for simplicity, the individual flow-specific effects  $\mathbf{b}_{ij}$  are estimated *directly* (rather than estimating  $\boldsymbol{\beta}$  and  $\boldsymbol{\vartheta}_{ij}$  separately), where the estimated idiosyncratic deviations ( $\hat{\boldsymbol{\vartheta}}_{ij}$ ) from the population-averaged effects ( $\hat{\boldsymbol{\beta}}$ ) are simply derived as  $\hat{\boldsymbol{\vartheta}}_{ij} = \hat{\mathbf{b}}_{ij} - \mathbb{E}(\hat{\mathbf{b}}_{ij})$ , assuming that  $\hat{\boldsymbol{\beta}} = \mathbb{E}(\hat{\mathbf{b}}_{ij})$ . Yet, it is felt that this procedure should not compromise the final results.

### Support ranges (prior/non-sample information)

For the purpose of estimation, a common support vector for the coefficients is used. Since I have little prior knowledge about the “true” value of the coefficients, the support vector is set as  $\mathbf{z} = (-100, 100)'$  for both the coefficients on the observed variables and the unobserved heterogeneity effects. These “conservative” bounds represent a range wide enough (roughly 100 times the ranges of the magnitudes of the OLS estimates) to include “all possible outcomes”. Furthermore, the support vector for the error term

$\epsilon_{ij}$  is defined as  $\mathbf{z}_\epsilon = (-3\hat{\sigma}_m, 0, 3\hat{\sigma}_m)'$ , in keeping with the widely-adopted “three-sigma” rule, where  $\hat{\sigma}_m$  is the standard deviation of the empirical distribution of the dependent variable (Pukelsheim, 1994). Finally, the (in)equality restrictions can also be considered as regularizing prior information.

#### *Approximate asymptotic standard errors*

If it can be established that the GME estimator of the (population-average) parameters contained in the vector  $\boldsymbol{\theta}$  is consistent and asymptotically normally distributed, then  $\sqrt{N(N-1)}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega})$  (e.g., Golan et al., 2001). Let  $L$  be the number of parameters, including the intercept, then the approximate asymptotic covariance matrix is  $\hat{\boldsymbol{\Omega}}_L = [\mathbf{A}'_{N(N-1) \times L} \hat{\mathbf{V}}_{N(N-1)}^{-1} \mathbf{A}_{N(N-1) \times L}]^{-1}$ , where  $\hat{\mathbf{V}}_{N(N-1)}$  is a diagonal matrix with elements  $\hat{\sigma}_{\xi_{ij}}^2 = \hat{\sigma}_v^2 + \hat{\sigma}_\omega^2 + \hat{\sigma}_\lambda^2 C_{ij} + \hat{\sigma}_\eta^2 s_{ij}^2 + \hat{\sigma}_\delta^2 d_{ij}^2 + \hat{\sigma}_\epsilon^2 (\forall i, j; i \neq j)$ , and  $\mathbf{A}_{N(N-1) \times L} = \partial f / \partial \boldsymbol{\theta}$  is the gradient matrix. The approximate asymptotic standard errors are calculated as the square roots of the diagonal elements of matrix  $\hat{\boldsymbol{\Omega}}_L$ . It should be noted that the standard errors are calculated for the *population-average* effects of the observed unilateral push and pull factors and bilateral linkage factors only, whilst the individual unobserved effects are relegated to the composite error term ( $\xi_{ij}$ ).

### 3.2 Relationship with earlier entropy-based studies of migration

To conclude this section, it is worth emphasizing that the GME approach applied in this paper is *econometrically* based (i.e., applied in a regression framework) and, therefore, intrinsically different from earlier applications of entropy/information-theoretic techniques in the migration literature. Essentially, earlier maximum-entropy (ME, not GME) applications were primarily concerned with recovering, predicting, or updating migration flows from incomplete data (e.g., “matrix balancing”), whilst mostly relying on the exclusive use of distance as an “explanatory” variable (e.g., Wilson, 1970; Willekens, 1977; Plane, 1982; Pooler, 1985; Roy, 1987; Roy and Flood, 1992).

To exemplify my point, I refer specifically to Roy and Flood (1992). In the first part of their paper, bilateral migration flows are predicted for different “homogeneous” (age or skill) groups by applying entropy/information-theoretic techniques. In the second part of their study, a gravity model (similar to the standard gravity model presented in this paper) is estimated, on the basis of their predicted flows, by applying the convenient OLS estimator. In contrast to their study, the present paper estimates the gravity model *directly* by applying a FE-GME estimator on the basis of observed migration flows.

## 4. EXPLORATORY EMPIRICAL ANALYSIS

In this section, the three-way FE gravity model is applied for an exploratory analysis using cross-sectional data on inter-state migration flows (1995-2000) and migrant stocks (1995) in Mexico. The data set used is fairly representative of the literature on aggregate migration. Data sources are provided in Appendix A.

In my presentation of the empirical analysis, the discussion will be grouped under the following headings: (1) the variables included in the empirical model; (2) basic descriptive statistics; (3) some preliminary empirical observations; (4) the results from the econometric estimation; (5) an analysis of the estimated individual bilateral effects; and (6) a look at the pattern (variable dispersion) of the composite residuals.

### 4.1 Variables

The dependent variable in the empirical model is defined as the natural logarithm of gross migration flows (MIGR), from each of the Mexican origin states  $i$  to all possible destination states  $j$  during the period 1995-2000. This yields a total of 992 bilateral migration flows (note that the diagonal elements in the matrix of inter-state migration flows are set to zero).

The gravity model includes a conventional set of (unilateral) “push” and “pull” factors at origins and destinations, comprising population (POP), unemployment rate (UR), Human Development Index (HDI), share of manufacturing in total employment (MAN), and population density (DENS). In keeping with the most common specification found in the literature, the model includes origin and destination populations. The rationale for including these variables is that they act primarily as a *scale* (control) variables. The HDI, which is a composite indicator of income (GDP per capita), health (life-expectancy rate), and education (literacy and school-enrollment rates), measures the general “development status” of Mexico’s states. The unemployment rate is intended as a measure of job opportunities, while the share of manufacturing in total employment is used as an indicator of job opportunities not yet sufficiently captured by the

unemployment rate. Population density is used as a measure of urbanization—though this variable may also be considered as a “catch-all” for job opportunities in the formal and informal service industry, the supply of a wide range of public and private amenities, etc.

Furthermore, I include three bilateral linkage factors: distance, contiguity, and migrant stock. Distance (DIST) is measured as the km highway distance between the capital cities of the Mexican states, which acts as an impediment to inter-state migration. A contiguity or common-order indicator (CONTIG) is included to “correct” for potential short-distance moves between neighboring states. The migrant stock (STOCK) is intended to indicate the presence of pre-existing networks of “family and friends” in the destination, which are generally thought to facilitate inter-state migration.

Finally, and in agreement with Greenwood and Ladman (1978) and Greenwood et al. (1981), the model includes U.S. border-state dummies (USB), to account for the potential *maquiladora* effect (e.g., Kohout, 2009).<sup>12</sup> Besides, these dummy variables may also capture the possible “gateway function” of the U.S. border states; that is, prospective migrants to the U.S. (especially young males) may consider moving to the Mexican border region in preparation for migration to the U.S. (e.g., Zabin and Hughes, 1995; OECD, 2004). Both factors make the northern border states especially attractive and are widely held responsible for the apparent (net) “south-north drift” of migration in Mexico. So, it is expected that these states experience disproportionately much in-migration and disproportionately little (internal) out-migration.

## 4.2 Basic descriptive statistics

Table 1 presents basic descriptive statistics (and data sources) for all the variables included in the empirical gravity model, Table 2 presents the correlations between the bilateral linkage variables, and Table 3 presents the cross-tabulation according to the contiguity indicator.

What stands out from these tables is that: first, none of the potential migration flows is equal to zero, so that logarithmic transformation is not a problem here; second, the correlation between current migration and migrant stocks is only -0.207 (see further below); third, both current migration flows and migrant stocks are significantly stronger between contiguous states (see further below).

Other interesting statistics (not reported in the tables) reveal the importance of counter-streams in Mexico; that is, for every three people moving in one direction there are roughly two people moving in the opposite direction, on average (i.e., net migration is about 36% of gross migration). Or, more precisely:  $\sum_i \sum_j (MIGR_{ij} / MIGR_{ji}) / 992 = 1.46$ , or  $\sum_i \sum_j (MIGR_{ij} - MIGR_{ji}) / \sum_i \sum_j (MIGR_{ij} + MIGR_{ji}) = 0.364$ . Also, it was found that for every one person migrating from the U.S. border states into the south, there are 3.5 persons moving in the opposite direction. Thus, the U.S. border region has experienced substantial net in-migration during the sample period 1995-2000. Finally, more than 50% of total inter-state migration in Mexico takes place between contiguous states (of which a substantial part occurs between Distrito Federal and the state of México).

< Insert TABLE 1 about here >

< Insert TABLE 2 about here >

< Insert TABLE 3 about here >

## 4.3 Some preliminary empirical observations

Before turning to the results of the econometric estimation, it is worth taking a preliminary look at the cross-sectional relationships between current migration flows and migrant stocks, to highlight a number of ill-founded, though widespread, arguments about the interaction between distance (and contiguity) and migrant stock.

Previous studies have found distance to be a dominant factor in shaping the spatial distribution of migrants. Accordingly, many researchers (quite obstinately) adhere to the idea that both current and previous migration flows are primarily “distance-driven”; that is, if previous migrants tended to move to nearby places, then current migrants also tend to move to nearby places (Greenwood, 1997, p. 666). Put differently, distance is strongly correlated with past migration (and, hence, the migrant stock) because it has been exerting a reasonably stable direct influence on migration patterns through time (Renshaw, 1974). It has often been argued that, as a corollary, distance could simply serve as a proxy for migrant-stock linkages, or vice versa—where some researchers would even stretch this idea to the point that it is useless to include both distance and migrant stock as separate explanatory variables in the gravity equation, contending that they are just measuring the same thing and, hence, are “substitutable” for each other (erroneously assuming some kind of “one-to-one” relation between distance and migrant stock).

However, in the case of inter-state migration in Mexico (and very likely also for migration in many other instances of internal and international migration), the correlation between distance and the migrant-stock variable is rather low (-0.207), so distances and migrant stocks are far from perfectly correlated. Moreover, a preliminary analysis based on the weighted average distances of current migration (1995-2000) and migrant stocks (1995) reveals that there is a *systematic* tendency for current migrants to move to *more distant* places than the spatial distribution of migrant stocks (past migration) would suggest, as shown in panel A of Figure 1. The finding that current migrants (flows) tend to be more geographically dispersed than migrants in the past (stocks) also suggests that Greenwood's widely-adopted contention, that "If a 'migrant stock' (...) is not included in the model, distance reflects the importance of relatives and friends" (Greenwood, 1997, p. 666), might be seriously ill-founded. An examination of panel A in Figure 1 reveals that appreciable migratory shifts can already occur within a short (five-year) period of time.

< Insert FIGURE 1 about here >

Alternatively, one could argue that contiguity, rather than distance, exerts a stable direct influence on migration patterns through time. The cross-tabulation in Table 3 shows that both migration flows in 1995-2000 and migrant stocks in 1995 are statistically significantly greater for contiguous states (the correlation is 0.411). However, again, a preliminary simple analysis based on the distribution of current migration and migrant stocks over contiguous vs. non-contiguous states reveals a *systematic* tendency of "flow diversion" towards *non-contiguous* destinations, as is shown in panel B of Figure 1.

#### 4.4 Econometric results for the gravity equation

Table 4 presents the estimation results for various specifications of the gravity model. OLS estimates for the standard gravity model are reported in columns 1-3. GME estimates for the standard gravity model are given in column 4. GME estimates for three versions of the FE gravity model are provided in columns 5-7. The discussion of the results starts with a brief account of the OLS results. To assess the performance of the FE-GME estimator and to give an indicative judgment on the biases in the OLS estimates, the preferred FE-GME estimates in column 7 will be compared with their OLS counterparts in column 3.<sup>13</sup>

< Insert TABLE 4 about here >

##### *Analysis of OLS coefficients*

An examination of the OLS results in columns 1-3 of Table 4 reveals that the estimates are strongly tied to the model specification. The inclusion of the migrant-stock variable, in particular, leads to appreciable changes in the magnitudes of all coefficients and even to a remarkable sign reversal of the coefficient on unemployment in the origin.

Interestingly, and somewhat surprisingly, the coefficients on unemployment in the destination and HDI in both the origin and the destination become *larger* (in absolute value) after including migrant stock. This means that the established "Nelson-Greenwood hypothesis" (Nelson, 1959; Greenwood, 1970)—positing that the coefficients with the migrant-stock variable excluded should be interpreted as indicative of the cumulative effects of the other variables—is not generally supported by the OLS results. Besides, these coefficients suggest, quite unrealistically, "more-than-proportional" effects on current migration.

After controlling for the migrant stock, the distance effect (not surprisingly) becomes appreciably smaller in magnitude, with a dramatic downfall from -1.100 (statistically significant at the 1% level) to -0.066 (statistically insignificant). On the other hand, the coefficient on migrant stock is 0.632 (which is broadly in line with values typically found in earlier empirical studies; e.g., Greenwood, 1969, 1970; Levy and Wadycki, 1973; Kau and Sirmans, 1979; Cushing, 1986; Bruder, 2003; Fan, 2005). Overall, the results seem to confirm the usual suspicion that the distance effect tends to be "eaten" by the inclusion of the migrant-stock variable. Given that the standard gravity model suffers from specification errors, OLS is likely to underestimate the distance-decay parameter and to overestimate the network effect.

##### *Analysis of FE-GME coefficients*

The GME estimates reported in column 4 of Table 4 echo the OLS estimates in column 3, and serve primarily as a general "test" of the GME estimator. The GME estimates reported in columns 5 and 6 are for a *two-way* FE model with fixed effects in the unilateral (origin and destination) dimensions and for a *three-way* FE model with *symmetric* fixed effects in the bilateral dimension, respectively. Finally, the GME estimates reported in column 7 are for the three-way FE model with *asymmetric* flow-fixed effects (for

distance, contiguity, and migrant stock).<sup>14</sup>

Comparison of the preferred FE-GME (population-averaged) estimates in column 7 with those from their OLS counterparts in column 3 overwhelmingly demonstrates that it is worth applying the three-way FE estimator, leading to appreciable changes in *all* coefficients; that is, the estimates are noticeably sensitive to the type of FE, both quantitatively and qualitatively. If unobserved heterogeneity and asymmetry were unimportant, introducing fixed effects would have been unnecessary, and FE-GME and OLS would have yielded broadly similar results. This, obviously, is not the case. Also, all coefficients are indicating, more realistically, “less-than-proportional” influences on current migration. I now turn to a more detailed comparison of the FE-GME (column 7) and OLS (column 5) estimates, which shows numerous significant differences.

Beginning with the *bilateral linkage factors*, the coefficient on the migrant-stock variable has been considerably reduced to 0.103 (though still significant at the 1% level), which is about six times smaller than the OLS estimate (0.632). To the extent that the unobserved bilateral flow-specific effects—besides the unilateral origin-/destination-fixed effects—are captured by the FE-GME estimator, this estimate may represent a better indicator of the existence of network effects. This finding suggests that network effects (corrected, at least partly, for unobserved persistency effects) are considerably less important than has been usually reported in the literature. Furthermore, the estimate of the contiguity effect is 0.093 (and statistically insignificant), which is also appreciably smaller than its implausibly large OLS counterpart (0.452). The coefficient implies that, on average, migration between two contiguous states is only about 10% ( $=100 \times (e^{0.093} - 1)$ ) higher than migration between two non-contiguous states (in OLS this is 57% ( $=100 \times (e^{0.452} - 1)$ )). The finding of a weak contiguity effect should not be surprising, though, since the estimate is an indicator of the *average* effect, which conceals variations across pairs of contiguous states (as will be followed up in the next sub-section). Finally, the estimate of the distance effects is -0.204, which is about three times larger than the OLS estimate (-0.066). Moreover, this stronger distance-deterrence indicator becomes statistically significant (at the 5% level) and is thus not “eaten” by the inclusion of the migrant-stock variable.

Looking at the *pull factors* at the destinations, the most remarkable changes can be observed for the unemployment rate and the HDI. The repulsion effect of unemployment is estimated at -0.259, which is more than five times smaller than the OLS counterpart (-1.254). The attraction effect exerted by HDI is 0.821, which is—even though still appreciable—only 63% of the more-than-proportional effect on migrant arrivals suggested by the OLS estimate (1.312). The signs of the coefficients are in line with prior expectations and the estimates retain their statistical significance (at the 5% and 1% level, respectively). The changes for the other destination variables (population, share of employment in manufacturing, and population density) are less dramatic, but the numerical size of all coefficients generally tend to decline (to about 50% of their OLS counterparts).

Striking changes can also be noticed for the *push factors* at the origins, which alter substantive inferences in a number of instances. Most notable are, again, the changes in the estimates associated with unemployment and HDI. Unemployment and HDI are found to have little if any influence on migrant departures, where the coefficients become statistically insignificant. The impact of unemployment on migrant departures is 0.218, which is about half the size of the OLS estimate (0.402), while the expulsion effect of HDI is estimated at 0.199, which is even seven times smaller (in absolute value) than the OLS estimate (-1.386). How can such insignificant effects be explained/interpreted? In the case of unemployment, the insignificant result may be indicative of the (concealed) real-world heterogeneity of populations. While high unemployment may “drive out” *some* people from their home base, *others* may not be able to afford moving to other places (i.e., potential out-migrants in high-unemployment areas may be stuck in a “poverty trap”). Moreover, the unemployed constitute only a relatively small proportion of the total population, whilst a high unemployment rate is likely to be of less importance to those individuals who already have a job. Thus, high unemployment rates tend to encourage out-migration of those who are unemployed but exert little influence on the majority of the population (e.g., Cadwallader, 1992, p. 53). In the case of HDI, the insignificant result may be attributed to the fact that some people living in relatively developed/high-income/high-literacy states may generally have low incentives to out-migrate, whilst others may be less reluctant to move to other places in search for more attractive job opportunities or more adequate living/housing conditions.

The coefficients on the indicator variables for the *U.S. border states* also illustrate the gains from applying the FE-GME estimator. Migrant arrivals in the northern border states are estimated to be approximately 62% ( $=100 \times (e^{0.481} - 1)$ ) higher than in other states, everything else equal. The OLS estimator, in turn, clearly exaggerates the attraction of the northern border states, as the coefficient implies that the number of arrivals would be a dazzling 141% ( $=100 \times (e^{0.880} - 1)$ ) higher than in other states. On the other hand, retention forces in the northern border states are found to be 20%



( $=100 \times (e^{-0.227} - 1)$ ) stronger than in other states (but only marginally significant), *ceteris paribus*, whereas OLS returns an implausible low (insignificant) estimate of 6% ( $=100 \times (e^{-0.058} - 1)$ ) less migrant departures.<sup>15</sup>

As a final means of assessing the estimated coefficients, it is worth pointing at two observations regarding the FE-GME estimates in columns 5 and 6 of Table 4. First, controlling for two-way FE in the unilateral dimensions (column 5) appears to be just an “intermediate” step in accounting for correlated heterogeneity. Although most coefficients are altered in the expected directions, the high coefficient on the migrant-stock variable (0.769) remains particularly worrisome (which is even higher than the OLS value of 0.632). Second, the introduction of three-way FE in the bilateral dimension with *symmetry* placed on the individual slopes (column 6) tends to return severely crunched and mostly insignificant coefficients. Thus, the preferred FE-GME estimates reported in column 7 of Table 4 clearly illustrate the benefits of relaxing the symmetry assumption in the bilateral dimension.

### Variance decomposition

To assess the relative importance of the various components on the RHS of the gravity equation in determining the (log) size of migration—i.e., the dependent variable, I present a variance analysis as in [Abowd et al. \(1999\)](#) and [Combes et al. \(2008\)](#). Table 5 shows the explanatory power of each RHS variable by reporting, for all them, the standard deviations of their *effects* and their correlations with the dependent variable. Specifically, I compute the effect of each RHS variable by multiplying its coefficient by its value for each observation. For example, the effect of the unemployment rate on the dependent variable is equal to the estimated coefficient on this variable times the unemployment rate. Then, the variability of the effect of each RHS variable can be calculated. When the effect of a RHS variable has a large standard deviation and it is highly correlated with the dependent variable, this RHS variable has a strong explanatory power. When, on the contrary, the effect of a RHS variable has a small standard deviation and a small correlation with the migration flows, this RHS variable explains only a small fraction of the variation in the dependent variable.

< Insert TABLE 5 about here >

The results reported in Table 5 suggest that the heterogeneous parts of the distance and migrant-stock effects ( $\hat{\tau}_{ij}$  and  $\hat{\mu}_{ij}$ , respectively) are most successful in explaining the variation in observed migration flows. The heterogeneous component of the *distance* effect ( $\hat{\tau}_{ij}$ ) has the largest explanatory power; its standard deviation (0.622) is closest to that of the dependent variable (1.554) and its correlation with migration flows is very high at 0.685 (i.e., the third highest correlation). The components of the *migrant-stock* effect come second in importance; they show the highest correlations (0.715 for the “fixed” component, and 0.739 for the heterogeneous component), though their standard deviations are much lower (0.155 and 0.342, respectively).<sup>16</sup> Not quite surprisingly, the explanatory power of the *population* variables, taken together, is also substantial, with a standard deviation of 0.695 and a correlation of 0.668 (note that they act mainly as scale factors). Furthermore, *contiguity* has a relatively weak explanatory power, with a low standard deviation of 0.160 and a moderately high correlation of 0.481. Finally, socio-economic variables (unemployment rate, development status, share of manufacturing, and population density), the U.S. border-state indicators, and the state-fixed effects all explain only a small fraction of the variation in observed migration flows.

To give an idea of the importance of the varying effects of distance and migrant stock on the variation in migration flows, one can also examine the latter *net* of their heterogeneous components. The “net” (log) migration flows,  $\ln \text{MIGR}_{ij}^*$ , are calculated as follows:

$$\ln \text{MIGR}_{ij}^* = \ln \text{MIGR}_{ij} - \hat{\mu}_{ij} \ln \text{STOCK}_{ij} - \hat{\tau}_{ij} \ln \text{DIST}_{ij} \quad (14)$$

Table 6 compares variations in real/observed and “net” (log) migration flows. Depending on the measure of dispersion taken, the results suggest that the variations in unobserved individual (flow-specific) distance and migrant-stock effects account for 30 to 45% of the variation in (log) migration flows. A similar exercise for the observed effects of socio-economic “push and pull” factors in origins and destinations returns an explanatory power of only 5 to 10%.

< Insert TABLE 6 about here >



#### 4.5 Digging deeper: determinants of the individual bilateral effects

Figure 2 shows the kernel densities of the recovered individual slopes for distance ( $|\hat{\delta}_{ij}|$ ), contiguity ( $\hat{\lambda}_{ij}$ ), and migrant stock ( $\hat{\eta}_{ij}$ ). An examination of the distributions reveals wide variations in the flow-specific effects, so the effects of the bilateral linkage factors are not homogeneous. A remarkable (though expected) finding is the heavily (positively) skewed distribution of the contiguity effects, with a sharp concentration of the mass of the distribution at a value close to zero (in fact, about 63% of the estimated individual contiguity effects are equal to zero).

< Insert FIGURE 2 about here >

Since the individual effects and their distributions are not particularly informative as such, it is more instructive to ask: Where do the differences in bilateral effects come from? Therefore, armed with the individual slopes returned by the estimated FE gravity model, I examine whether some important “driving forces” behind their variations can be discovered. This is accomplished by regressing the flow-specific effects of distance and migrant stock on a set of relational variables reflecting differences between Mexican states and some spatial-configuration variables (see Mueser, 1989a, for a similar exercise).

Specifically, I use cross-state *disparities* in unemployment rate ( $\Delta \ln UR_{ij} = \ln UR_i - \ln UR_j$ ), development status ( $\Delta \ln HDI_{ji} = \ln HDI_j - \ln HDI_i$ ), and population density ( $\Delta \ln DENS_{ji} = \ln DENS_j - \ln DENS_i$ ). I also include distance (third-order polynomial), contiguity, and a “directionality” indicator as explanatory variables. The directionality indicator is defined as (e.g., Gabriel et al., 1993)

$$DIRECT_{ji} \equiv USB_j - USB_i = \begin{cases} 1, & \text{if } j \text{ is a border state and } i \text{ is a non-border state} \\ 0, & \text{if both } i \text{ and } j \text{ are either border or non-border states} \\ -1, & \text{if } j \text{ is a non-border state and } i \text{ is a border state} \end{cases} \quad (15)$$

where  $DIRECT_{ji} = 1$  indicates migration in the south-north direction,  $DIRECT_{ji} = -1$  indicates migration in the opposite direction, and  $DIRECT_{ji} = 0$  indicates “intra-regional” migration. For the individual slopes of contiguity, the average effect on migration will be compared with the effects for those Mexican states that have a large metropolitan zone close to and/or straddling the common border ( $ZM_{ij} = 1$ ).

##### Analysis of coefficients

Table 7 reports both OLS (columns 1.a and 2.a) and Tobit (columns 1.b and 2.b) results for the individual distance and migrant-stock effects, which are left-censored given the inequality constraints (about 6% of the individual effects hit the zero bound). Given the partial nature of the present analysis, I start the discussion with some general observations and then focus only on a few striking results.<sup>17</sup>

< Insert TABLE 7 about here >

The overall impression is that the results are quite satisfactory, where all coefficients are statistically significant (at the 5% or 1% level) and tend to have the expected sign. Moreover, the results provide supportive evidence of asymmetry (directional heterogeneity) of inter-state interactions. Finally, a large proportion of the variation in the individual effects (57% for distance, 72% for migrant stock) could be explained by only a limited set of variables (this, however, means, that 30-40% still remains unexplained). Such results are encouraging and, in a way, demonstrate that the individual effects obtained from FE-GME are not just “cheap talk” (or chance numbers rolling out from a gambling machine) but can, instead, be linked in a meaningful and informative way to the limited set of variables.

For the individual *distance* effects, it is found that disparities between states in terms of unemployment and development status tend to have a “compensatory” effect, making migrants less sensitive to distance if conditions in a destination are more favorable compared to those in the origin. Furthermore, migrants seem to be less hindered by distance when moving from relatively low-density (rural/amenity-poor) to relatively high-density (urban/amenity rich) areas, in conformity with the conjecture formulated in the introductory section. Also, the signs of the coefficients on directionality, contiguity, and migrant stock are generally in line with prior expectations.

Finally, the effect of distance varies perceptibly according to the distance travelled by migrants, as clearly shown in Figure 3. The signs and magnitudes of the coefficients associated with the third-order polynomial in distance indicate that, *ceteris paribus*, the impact of distance falls sharply for moves within a roughly 50-km radius of the home base, then increases for movements over medium-long distances up to about 1040 km, and further decreases again for longer-distance moves. The distance profile seems to

suggest (though, admittedly, only weakly) the presence of some sort of “reservation distance”, where people may be more reluctant to take on a costly resettlement in nearby locations. On the other hand, the significant decline in the deterring effect for longer-distance moves could perhaps be reflective of a *composition* effect; that is, long-distance migration flows may possibly have an age balance tended towards the young and/or carry higher human capital (e.g., Schwartz, 1973; Morgan and Robb, 1981; Schultz, 1982; Millington, 2000), containing migrants who may be primarily concerned with *functional* distance (opportunities) rather than with physical distance. Unfortunately, data limitations do not permit the identification of the distinctive spatial patterns of movement for different types of migrants.

< Insert FIGURE 3 about here >

Turning to the individual *migrant-stock* effects, it can be seen that the results exhibit a roughly similar picture. Socio-economic disparities in terms of unemployment and development status tend to lessen the importance of the migrant stock (migration history). Also, the distance profile of the migrant-stock effect turns out to be much less pronounced than for distance, as shown in Figure 3.

Perhaps the most important finding is that the coefficient on the migrant-stock variable is *negative* (and statistically significant at the 5% level). Such an outcome is in line with Davis et al. (2002), and suggests that on average the *marginal* effect of the migrant stock (past migrants) in a given destination on the current inflow of “compatriots” *decreases* with the *size* of the stock, holding other things equal. This decreasing marginal effect of migrant stock on the size of migration flows leaves room for different (not mutually exclusive) interpretations. First, the size of the migrant stock in a given destination is likely to be reflective of the *age structure* of the stock of previous migrants. The intuition here is that the larger is the stock of previous migrants in a destination, the longer is the “migration history” and, hence, the older is the population of past migrants living in that destination. Then, with increasing age of the migrant stock, the beneficial network externalities may be subject to a “depreciation function” of some sort (e.g., Dunlevy, 1993). Second, negative network externalities may arise due to increased *competition* for resources that might be getting in short supply (congestion effects), leading to depressed wages in a destination. Then, sending back information to those at home about, say, “a lack of job opportunities and poor living conditions” in a destination is expected to discourage further migration to that destination (e.g., Renshaw, 1974; Bauer et al., 2002, 2007; Warin and Blakely, 2009). Third, a larger migration stock in a destination means that people in the sending area are becoming increasingly integrated in migration networks. Then, those who migrated in the past may send *remittances* to their family and friends at home (e.g., Boucher et al., 2005; Taylor et al., 2008). To the extent that these remittances improve the living conditions of those “left behind”, the latter may have less incentives to follow the paths of those who recently emigrated. Fourth, current migration flows might have been diverted from the trajectories of migrants in the past simply because the *composition* of current flows might be different from that of previous flows, where “new” (e.g., highly-skilled) migrants might be less driven by historical experiences in their destination choice and, hence, bend away from “old” migration paths.

Furthermore, the migrant-stock effect becomes smaller with more pronounced discrepancies in terms of population density. This may be indicative of the fact that the attraction exerted by those destinations hosting a large number of previous migrants is likely to be weaker in more densely-populated areas. This finding is also consistent with Fussell and Massey (2004), who found that “friends-and-family” linkages are less pervasive in more urbanized destinations where social interactions are weaker.

Finally, Table 8 tabulates the *contiguity* effects for various “types” of common borders. While the overall effect of contiguity is found to be weak (0.093), and not statistically significant (migration between contiguous states is estimated to be at best 10% higher, on average, compared to migration between non-contiguous states), the effect for contiguous states with a *metropolitan zone* close to the common border (either in one state or in both states) is almost 17% ( $=100 \times (e^{0.155} - 1)$ ) higher than that for non-contiguous states. Moreover, for those contiguous states hosting a metropolitan zone *straddling* the common border, it is found that migration is about 32% ( $=100 \times (e^{0.276} - 1)$ ) higher than migration between non-contiguous states, *ceteris paribus*. These higher contiguity effects may thus largely capture the “local mobility” within large metropolitan zones and/or between these metropolitan zones and their rural-urban surroundings.

<Insert TABLE 8 about here>

#### *Variance decomposition.*

I perform a variance decomposition, as I did before for the gravity equation. The results reported in Table 9 suggest that migrant stock is the most important variable in the determination of the size of the bilateral

effects of distance and migrant stock; its standard deviation (0.052 and 0.054, respectively) is closest to that of the estimated bilateral effects (0.092 and 0.055, respectively) and its correlation is very high at 0.624 and 0.774, respectively. Furthermore, spatial variables play a moderate role, particularly in the case of the estimated distance effects, where contiguity stands out (relatively speaking). Finally, socio-economic differentials explain only a small fraction of the variation of the estimated individual effects. Interestingly, such findings raise some doubts about the importance of the “compensatory role” of socio-economic factors in reducing migrants’ sensitivity to distance, so that internal migration in Mexico is unlikely to be strongly positively affected (indirectly) by, say, increasing per-capita income gaps between origin and destination (as in [Ortega and Peri, 2009](#), for international migration).

<Insert TABLE 9 about here>

#### 4.6 Taking a glance at the pattern of the composite residuals

As a last step in the empirical analysis, it is worth eyeballing the pattern of the *composite* residuals resulting from the three-way FE model. In Figure 4, the squared values of the residuals, defined here as  $\hat{\tau}_{ij} = \text{MIGR}_{ij} - \widehat{\text{MIGR}}_{ij}^c$ , where  $\widehat{\text{MIGR}}_{ij}^c = \widehat{\text{MIGR}}_{ij}^c(e^{\xi_{ij}^2/2})$ , where  $\xi_{ij} = \hat{\nu}_i + \hat{\omega}_j + \hat{\kappa}_{ij}C_{ij} + \hat{\mu}_{ij}S_{ij} + \hat{\tau}_{ij}d_{ij} + \hat{\epsilon}_{ij}$ , are plotted against the predicted flows,  $\widehat{\text{MIGR}}_{ij}^c$ .<sup>18</sup>

< Insert FIGURE 4 about here >

Inspection of the pattern displayed in Figure 4 provides strong (qualitative) support for the conjecture that the dispersion of the composite residuals tends to increase with the level of the predicted flows. This illustrates that the proposed three-way FE gravity model is capable of accounting for the heteroskedastic nature (arbitrary variance) of the migration data in a quite satisfactory way. In other words, the FE gravity model in *logs* constitutes a viable alternative for the Poisson specification (in levels) in alleviating the bias that may arise from heteroskedasticity.

### 5. CONCLUDING COMMENTS

The purpose of this paper was to establish a novel fixed-effects (FE) estimator and to show its potentials for the estimation of generalized cross-section gravity models of aggregate place-to-place migration. The proposed three-way FE approach was applied for an exploratory analysis of inter-state migration in Mexico. To regularize the “ill-posed” (over-parameterized) nature of cross-section model, estimation was accomplished by using the non-classical Generalized Maximum Entropy estimator. The importance of introducing fixed effects was demonstrated by comparing the FE-GME estimates with those obtained from the convenient OLS estimator.

The most important findings can be summarized as follows. First, it was found that accounting for unobserved heterogeneity in all three dimensions ( $i$ ,  $j$ , and  $ij$ )—that is, relaxing the standard homogeneity and quasi-symmetry assumptions of the gravity model of aggregate migration—has an appreciable impact on the magnitudes of all estimated coefficients, whilst most of them retain their statistical significance and, in the end, appear to be more realistic. Second, after correcting for omitted-variables bias in the bilateral dimension, the estimate of the (population-averaged) elasticity of migration with respect to migrant stock, at around 0.10, is much lower than in previous empirical literature. Third, it is found that variations in the unobserved individual (flow-specific) distance and migrant-stock effects account for 30 to 45% of the variation in (log) migration flows, whereas observed socio-economic “push and pull” factors at origins and destinations (unemployment rates, incomes per capita, etc.) have only a small explanatory power of 5 to 10%. Finally, the individual bilateral effects could also be successfully related in a meaningful way to existing socio-economic disparities between Mexico’s states, some relevant spatial-configuration variables (including distance), and the size of the migrant stock, where the latter stands out.

Within the limited frame of the empirical analysis presented in this paper, the results, however, do represent something of a challenge to many of the claims that have been advanced in the previous literature about the preeminence of distance and/or migration networks in shaping migratory patterns at an aggregate scale. Furthermore, the results cast a shadow of doubts over the widespread contention that distance and migrant stock can be used interchangeably in the gravity equation as a proxy for “family-and-friends” linkages. Finally, the paper has shown that the three-way FE gravity model in logs is a viable alternative for the Poisson level specification in alleviating the problem of bias that may arise due to the heteroskedasticity in the migration data. The approach proposed in this paper should, therefore, deserve attention in the estimation of cross-section (as well as panel-data) gravity models.

## REFERENCES

- Abowd, John M., Francis Kramarz, David N. Margolis, 1999. 'High Wage Workers and High Wage Firms,' *Econometrica*, 67, 251-333.
- Allen, Jeremiah M. and B. Curtis Eaton. 2005. 'Incomplete Information and Migration: The Grass is Greener Across the Higher Fence,' *Journal of Regional Science*, 45, 1-19.
- Alvarez-Plata, Patricia, Herbert Brücker, and Boriss Siliverstovs. 2003. Potential Migration from Central and Eastern Europe into the EU-15—An Update. Report for the European Commission, DG Employment and Social Affairs, DIW, Berlin.
- Aroca, Patricia and William F. Maloney. 2005. 'Migration, Trade, and Foreign Direct Investment in Mexico,' *The World Bank Economic Review*, 19, 449-472.
- Baier, Scott L. and Jeffrey H. Bergstrand. 2007. 'Do Free Trade Agreements Actually Increase Members' International Trade?' *Journal of International Economics*, 71, 72-95.
- Balderas, J. Ulyses and Michael J. Greenwood. 2009. 'From Europe to the Americas: A Comparative Panel-Data Analysis of Migration to Argentina, Brazil, and the United States, 1870-1910,' *Journal of Population Economics*, DOI 10.1007/s00148-009-0243-x.
- Bauer, Thomas, Gil S. Epstein, and Ira N. Gang. 2002. Herd Effects or Migration Networks? The Location Choice of Mexican Immigrants in the U.S. IZA Discussion Paper No. 551, Bonn, Germany.
- Bauer, Thomas, Gil S. Epstein, and Ira N. Gang. 2007. 'The Influence of Stocks and Flows on Migrants' Location Choices,' *Research in Labor Economics*, 26, 199-229.
- Bavaud, François. 2002. 'The Quasi-Symmetric Side of Gravity Modelling,' *Environment and Planning A*, 34, 61-79.
- Boucher, Stephen R., Oded Stark, and J. Edward Taylor. 2005. A Gain with a Drain? Evidence from Rural Mexico on the New Economics of the Brain Drain. ARE Working Papers: University of California, Davis.
- Bruder, Jana. 2003. East-West Migration in Europe, 2004-2015. Thünen-Series of Applied Economic Theory, Working Paper No. 40, Rostock University, Germany.
- Cadwallader, Martin T. 1992. *Migration and Residential Mobility: Macro and Micro Approaches*. The University of Wisconsin Press.
- Calvet, Laurent and Etienne Comon. 2003. 'Behavioral Heterogeneity and the Income Effect,' *The Review of Economics and Statistics*, 85, 653-669.
- Cameron, A. Colin. and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Carrington, William J., Enrica Destragiache, and Tara Vishwanath. 1996. 'Migration with Endogenous Moving Costs,' *American Economic Review*, 86, 909-930.
- Cheng, I.-Hui and Howard J. Wall. 2005. 'Controlling Heterogeneity in Gravity Models of Trade and Integration,' *Federal Reserve Bank of St. Louis Review*, 87, 49-63.
- Cole, William E. and Richard D. Sanders. 1983. 'Interstate Migration in Mexico: Variations on the Todaro Theme,' *Journal of Development Economics*, 12, 341-354.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, 2008. 'Spatial Wage Disparities: Sorting Matters!' *Journal of Urban Economics*, 63, 723-742.
- Cushing, Brian J. 1986. 'Accounting for Spatial Relationships in Models of Interstate Population Migration,' *Annals of Regional Science*, 20, 66-73.
- Davis, Ben, Guy Stecklov, and Paul Winters. 2002. 'Domestic and International Migration from Rural Mexico: Disaggregating the Effects of Network Structure and Composition,' *Population Studies*, 56, 291-309.
- Dunlevy, James A. 1993. 'Migrant Stock vs. Lagged Migrant Flow as a Determinant of Migrant Settlement,' *Journal of Population Economics*, 6, 181-188.
- Dunlevy, James A. and Henry A. Gemery. 1977. 'The Role of Migrant Stock and Lagged Migration in the Settlement Patterns of Nineteenth Century Immigrants,' *The Review of Economics and Statistics*, 59, 137-144.
- Egger, Peter. 2005. 'Alternative Techniques for Estimation of Cross-Section Gravity Models,' *Review of International Economics*, 13, 881-891.
- Etzo, Ivan. 2008. Determinants of Interregional Migration in Italy: A Panel Data Analysis. MPRA Paper No. 8637, Munich Personal RePEc Archive.
- Fan, C. Cindy. 2005. 'Modeling Interprovincial Migration in China, 1985-2000,' *Eurasian Geography and Economics*, 46, 165-184.
- Flowerdew, Robin and Murray Aitkin. 1982. 'A Method of Fitting the Gravity Model Based on the Poisson Distribution,' *Journal of Regional Science*, 22, 191-202.
- Fukurai, Hiroshi, James B. Pick, and Edward W. Butler. 1987. 'Interstate Migration in Mexico, 1979-80: A Spatial Analysis,' *Sociology and Social Research*, 71, 312-322.
- Fussell, Elizabeth and Douglas S. Massey. 2004. 'The Limits to Cumulative Causation: International Migration from Mexican Urban Areas,' *Demography*, 41, 151-171.
- Gabriel, Stuart A., Janice Shack-Marquez, and William L. Wascher. 1993. 'Does Migration Arbitrage Regional Labor Market Differentials?' *Regional Science and Urban Economics*, 23, 211-233.
- Golan, Amos. 2006. 'Information and Entropy Econometrics—A Review and Synthesis,' *Foundations and Trends® in Econometrics*, 2, 1-145.
- Golan, Amos, George Judge, and Douglas Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley & Sons.



- Golan, Amos and Esfandiari Maasoumi. 2008. 'Information Theoretic and Entropy Methods: An Overview,' *Econometric Reviews*, 27, 317-328.
- Golan, Amos, Jeffrey M. Perloff, and Edward Z. Shen. 2001. 'Estimating a Demand System with Non-Negativity Constraints: Mexican Meat Demand,' *The Review of Economics and Statistics*, 8, 541-550.
- Gordon, Ian. 1985. 'Economic Explanations of Spatial Variation in Distance Deterrence,' *Environment and Planning A*, 17, 59-72.
- Gordon, Ian and Ian Molho. 1998. 'A Multi-Stream Analysis of the Changing Pattern of Interregional Migration in Great Britain, 1960-1991,' *Regional Studies*, 32, 309-323.
- Gordon, Ian and Roger Vickerman. 1982. 'Opportunity, Preference and Constraint: An Approach to the Analysis of Metropolitan Migration,' *Urban Studies*, 19, 247-261.
- Gordon, Peter and Peter Theobald. 1981. 'Migration and Spatial Development in the Republic of Mexico,' *Journal of Developing Areas*, 15, 239-250.
- Grant, Jason H. and Dayton M. Lambert. 2008. 'Do Regional Trade Agreements Increase Members' Agricultural Trade?' *American Journal of Agricultural Economics*, 90, 765-782.
- Greene, William H. 2003. *Econometric Analysis*. Prentice Hall, Pearson Education International (Fifth Edition).
- Greenwood, Michael J. 1969. 'An Analysis of the Determinants of Geographical Labor Mobility in the United States,' *The Review of Economics and Statistics*, 51, 189-194.
- Greenwood, Michael J. 1970. 'Lagged Response in the Decision to Migrate,' *Journal of Regional Science*, 10, 375-384.
- Greenwood, Michael J. 1978. 'An Econometric Model of Internal Migration and Regional Economic Growth in Mexico,' *Journal of Regional Science*, 18, 17-30.
- Greenwood, Michael J. 1997. 'Internal Migration in Developed Countries,' in M. Rosenzweig and O. Stark (eds.), *Handbook of Population and Family Economics*, Vol. 1B, North-Holland: Elsevier, pp. 647-719.
- Greenwood, Michael J. and Gary L. Hunt. 2003. 'The Early History of Migration Research,' *International Regional Science Review*, 26, 3-37.
- Greenwood, Michael J. and Jerry R. Ladman. 1978. 'An Economic Analysis of Migration in Mexico,' *Annals of Regional Science*, 12, 16-31.
- Greenwood, Michael J., Jerry R. Ladman, and Barry S. Siegel. 1981. 'Long-Term Trends in Migratory Behaviour in a Developing Country: The Case of Mexico,' *Demography*, 18, 369-388.
- Henderson, Daniel J. and Daniel L. Millimet. 2008. 'Is Gravity Linear?' *Journal of Applied Econometrics*, 23, 137-172.
- Jaynes, Edwin T. 1963. 'Information Theory and Statistical Mechanics,' in: K.W. Ford (ed.), *Statistical Physics*, New York: Benjamin, pp. 181-218.
- Jones, Richard C. 2001. 'Maquiladoras and U.S.-Bound Migration in Central Mexico,' *Growth and Change*, 32, 193-216.
- Kau, James B. and C.F. Sirmans. 1979. 'A Recursive Model of the Spatial Allocation of Migrants,' *Journal of Regional Science*, 19, 47-56.
- King, Jonathan. 1978. 'Interstate Migration in Mexico,' *Economic Development and Cultural Change*, 27, 83-101.
- Kohout, Michal. 2009. 'The Maquiladora Industry and Migration in Mexico: A Survey of Literature,' *Geography Compass*, 3, 135-153.
- Leblang, David, Jennifer Fitzgerald, and Jessica Teets. 2009. Defying the Law of Gravity: The Political Economy of International Migration. Unpublished Manuscript.
- Levy, Mildred B. and Walter J. Wadycki. 1973. 'The Influence of Family and Friends on Geographic Labor Mobility: An International Comparison,' *The Review of Economics and Statistics*, 55, 198-203.
- Massey, Douglas S. 1990. 'Social Structure, Household Strategies, and the Cumulative Causation of Migration,' *Population Index*, 56, 3-26.
- Mayda, Anna Maria. 2009. 'International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows,' *Journal of Population Economics*, DOI 10.1007/s00148-009-0251-x.
- Millington, Jim. 2000. 'Migration and Age: The Effect of Age on Sensitivity to Migration Stimuli,' *Regional Studies*, 34, 521-533.
- Morgan, James N. and Edward H. Robb. 1981. 'The Impact of Age Upon Interregional Migration,' *The Annals of Regional Science*, 15, 31-45.
- Mueser, Peter R. 1989a. 'The Spatial Structure of Migration: An Analysis of Flows Between the States in the USA Over Three Decades,' *Regional Studies*, 23, 185-200.
- Mueser, Peter R. 1989b. 'Measuring the Impact of Locational Characteristics on Migration: Interpreting Cross-Sectional Analyses,' *Demography*, 26, 499-513.
- Nelson, Peter B. 1959. 'Migration, Real Income and Information,' *Journal of Regional Science*, 12, 307-310.
- OECD (2004). *Economic Surveys 2002-2003: Mexico*, Vol. 2003-Supplement No. 1-Jan. 2004. Organization for Economic Co-operation and Development, Paris.
- Ortega, Francese and Giovanni Peri. 2009. The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005. NBER Working Paper Series (available at [http://www.nber.org/papers.w14833](http://www.nber.org/papers/w14833)).
- Paris, Quirino and Richard E. Howitt. 1998. 'An Analysis of Ill-Posed Production Problems Using Maximum Entropy,' *American Journal of Agricultural Economics*, 80, 124-138.
- Pedersen, Peder J., Mariola Pytlikova, and Nina Smith, N. 2008. 'Selection and Network Effects—Migration Flows Into OECD Countries 1990-2000,' *European Economic Review*, 52, 1160-1186.
- Plane, David A. 1982. 'An Information Theoretic Approach to the Estimation of Migration Flows,' *Journal of Regional Science*, 22, 441-456.
- Pooler, Jim. 1985. 'Derivation of a Generalized Family of Spatial Interaction Models: An Information Theoretic Approach,' *Modeling and Simulation*, 16, 199-207.

- Pukelsheim, Friedrich. 1994. 'The Three-Sigma Rule,' *American Statistician*, 48, 88-91.
- Raymer, James and Andrei Rogers. 2007. 'Using Age and Spatial Flow Structures in the Indirect Estimation of Migration Streams,' *Demography*, 44, 199-223.
- Renshaw, Vernon. 1974. 'A Note on Lagged Response in the Decision to Migrate,' *Journal of Regional Science*, 14, 273-280.
- Rogers, Andrei. 2008. 'Demographic Modeling of the Geography of Migration and Population: A Multiregional Perspective,' *Geographical Analysis*, 40, 276-296.
- Rogers, Andrei, Junwey Liu, Virgilio Partida, and Meg Tilton. 2006. Inferring Directional Flows from Net Migration Data: Mexico. Paper presented at the *Meeting of the Western Regional Science Association*, 22-25 February, Santa Fe, New Mexico.
- Roy, John R. 1987. 'An Alternative Information Approach for Modelling Spatial Interactions,' *Environment and Planning A*, 19, 385-394.
- Roy, John R. and Joe Flood. 1992. 'Interregional Migration Modeling via Entropy and Information Theory,' *Geographical Analysis*, 14, 16-34.
- Santos Silva, J.M.C. and Silvana Tenreyro. 2006. 'The Log of Gravity,' *The Review of Economics and Statistics*, 88, 641-658.
- SAS Institute, 2004. *SAS/ETS 9.1 User's Guide, Chap. 15: The Entropy Procedure*. Cary, NC: SAS Publishing.
- Schultz, T. Paul. 1982. 'Lifetime Migration Within Educational Strata in Venezuela: Estimates of a Logistic Model,' *Economic Development and Cultural Change*, 30, 559-593.
- Schwartz, Aba. 1973. 'Interpreting the Effect of Distance on Migration,' *Journal of Political Economy*, 81, 1153-1169.
- SEDESOL-CONAPO-INEGI (2007). *Delimitación de las Zonas Metropolitanas de México 2005*. Secretaría de Desarrollo Social; Consejo Nacional de Población; Instituto Nacional de Estadística, Geografía e Informática, México.
- Shen, Edward Z. and Jeffrey M. Perloff. 2001. 'Maximum Entropy and Bayesian Approaches to the Ratio Problem,' *Journal of Econometrics*, 104, 289-313.
- Shioji, Etsuro. 2001. 'Composition Effect of Migration and Regional Growth in Japan,' *Journal of the Japanese and International Economies*, 15, 29-49.
- Swamy, P.A.V.B. and George S. Tavlas. 1995. 'Random Coefficient Models: Theory and Applications,' *Journal of Economic Surveys*, 9, 165-196.
- Taylor, J. Edward, Richard Adams, Jorge Mora, and Alejandro Lopez-Feldman. 2008. 'Remittances, Inequality and Poverty: Evidence from Rural Mexico,' in J. DeWind and J. Holdaway (eds.), *Migration and Development Within and Across Borders*, International Organization for Migration (IOM) & Social Science Research Council (SSRC), Geneva/New York, pp. 103-130.
- Warin, Thierry and Andrew Blakely. 2009. Choice or Mimeticism in the Decision to Migrate? A European Illustration. CIRANO® Scientific Series Paper No. 2009s-38, Montréal, Canada.
- Willekens, Frans. 1977. *The Recovery of Detailed Migration Patterns from Aggregate Data: An Entropy Maximizing Approach*. IIASA Research Memorandum 77-58, International Institute for Applied Systems Analysis, Laxenburg, Austria.
- Wilson, Alan. G. 1970. *Entropy in Urban and Regional Modelling*. London: Pion.
- Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
- Wooldridge, Jeffrey M. 2005. 'Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models,' *The Review of Economics and Statistics*, 87, 385-390.
- Zabin, Carol and Sallie Hughes. 1995. 'Economic Integration and Labor Flows: Stage Migration in Farm-Labor Markets in Mexico and the United States,' *International Migration Review*, 29, 395-422.
- Zavodny, Madeline. 1997. 'Welfare and the Locational Choices of New Immigrants,' *Economic Review – Federal Reserve Bank of Dallas*, Q-II, 2-10.



TABLE 1: Variables and basic descriptive statistics

Variable	Definition	Mean	S.D.	Min	Median	Max
<i>Dependent variable</i>						
MIGR <sub>ij</sub> (1000)	Total number of out-migrants from origin state <i>i</i> to destination state <i>j</i> , 1995-2000	3.614	16.290	0.015	0.820	448.546
<i>Unilateral state characteristics</i> (“push and pull” factors)						
POP <sub><i>i</i></sub> (POP <sub><i>j</i></sub> )	Total population, 1995 (millions)	2.847	2.439	0.375	2.142	11.750
UR <sub><i>i</i></sub> (UR <sub><i>j</i></sub> )	Unemployment rate, average of 1990 and 2000	0.019	0.004	0.011	0.020	0.027
HDI <sub><i>i</i></sub> (HDI <sub><i>j</i></sub> )	Human Development Index = composite indicator including GDP per capita, life-expectancy rate, literacy rate, and school-participation rate, average of 1990 and 2000	0.685	0.114	0.421	0.708	0.864
MAN <sub><i>i</i></sub> (MAN <sub><i>j</i></sub> )	Share of manufacturing in total employment, 1995	0.229	0.084	0.053	0.243	0.375
DENS <sub><i>i</i></sub> (DENS <sub><i>j</i></sub> )	Population density, 1995 (thousands per square km)	0.259	0.986	0.005	0.047	5.717
USB <sub><i>i</i></sub> (USB <sub><i>j</i></sub> )	Dummy variable = 1 for U.S. border states, and 0 otherwise	0.219	—	—	—	—
<i>Bilateral linkage factors</i>						
CONTIG <sub>ij</sub>	Dummy variable = 1 for migration flows between contiguous states (sharing common border), and 0 otherwise	0.143	—	—	—	—
STOCK <sub>ij</sub> (%)	Migrant stock = number of previous migrants from the origin state <i>i</i> living in the destination state <i>j</i> , divided by population in the origin state, 1995	0.53	1.59	0.01	0.11	32.55
DIST <sub>ij</sub> (1000 km)	Geographical distance between states = km highway distance between the states’ capital cities	1.370	1.077	0.033	1.056	5.961

*Notes:* All continuous variables have been log-transformed in the empirical model. All independent variables are measured in 1995, hence, ante-dating “current” migration. However, data for some of the state characteristics were only available for the census years 1990 and 2000. The U.S. border states are defined as in [Rogers et al. \(2006\)](#) and [Raymer and Rogers \(2007\)](#), including Baja California Norte, Baja California Sur, Sonora, Chihuahua, Coahuila, Nuevo León, and Tamaulipas.

TABLE 2: Correlations between bilateral linkage variables

	MIGR	CONTIG	STOCK	DIST
MIGR	1			
CONTIG	0.242*** (0.000)	1		
STOCK	0.366*** (0.000)	0.411*** (0.000)	1	
DIST	-0.130*** (0.000)	-0.358*** (0.000)	-0.207*** (0.000)	1

*Notes:* Two-sided  $p$ -values are given in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1%, respectively.

TABLE 3: Cross-tabulation for contiguity

	CONTIG = 1	CONTIG = 0	Difference
MIGR (1000)	13.271	2.000	11.271*** (0.001)
STOCK (%)	2.124	0.264	1.861*** (0.000)
DIST (1000 km)	0.425	1.527	-1.101*** (0.000)
No. of observations	142	850	

*Notes:* Two-sided  $p$ -values are given in parentheses. The statistical significance of the difference in means is tested on the basis of a simple  $F$ -statistic from a parametric test (assuming unequal variances). \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1%, respectively.

TABLE 4: Estimation results from OLS and GME for standard and fixed-effects cross-section gravity model

	Standard gravity model				Fixed-effects gravity model		
	OLS		GME	FE-GME			
	(1)	(2)		(3)	(4)	(5)	(6)
Constant	-8.564*** (0.944)	-8.709*** (0.892)	<b>-1.505**</b> (0.753)	-1.554** (0.723)	0.095 (0.796)	0.488 (5.046)	<b>-0.091</b> (1.018)
ln POP <sub>i</sub>	1.034*** (0.041)	0.969*** (0.038)	<b>0.462**</b> (0.035)	0.465*** (0.035)	0.457*** (0.039)	0.175 (0.245)	<b>0.738***</b> (0.050)
ln POP <sub>j</sub>	0.808*** (0.040)	0.743*** (0.037)	<b>0.729***</b> (0.030)	0.733*** (0.030)	0.690*** (0.033)	0.204 (0.208)	<b>0.472***</b> (0.042)
ln UR <sub>i</sub>	-0.534*** (0.156)	-0.364** (0.149)	<b>0.402***</b> (0.118)	0.409*** (0.102)	0.138 (0.113)	-0.782 (0.725)	<b>0.218</b> (0.145)
ln UR <sub>j</sub>	-1.219*** (0.131)	-1.039*** (0.134)	<b>-1.254***</b> (0.090)	-1.261*** (0.099)	-0.972*** (0.109)	-1.517** (0.681)	<b>-0.259*</b> (0.139)
ln HDI <sub>i</sub>	-0.218 (0.209)	0.234 (0.192)	<b>-1.386***</b> (0.159)	-1.392*** (0.152)	-0.697*** (0.168)	-0.843 (1.071)	<b>0.199</b> (0.215)
ln HDI <sub>j</sub>	1.085*** (0.203)	1.069*** (0.186)	<b>1.312***</b> (0.131)	1.330*** (0.147)	0.691*** (0.161)	0.162 (1.001)	<b>0.821***</b> (0.204)
ln MAN <sub>i</sub>	-0.849*** (0.078)	-0.740*** (0.074)	<b>-0.365***</b> (0.060)	-0.369*** (0.056)	-0.216*** (0.061)	-0.421 (0.384)	<b>-0.310***</b> (0.078)
ln MAN <sub>j</sub>	-0.521*** (0.079)	-0.412*** (0.075)	<b>-0.106*</b> (0.060)	-0.108** (0.055)	-0.136** (0.061)	-0.366 (0.377)	<b>-0.064</b> (0.077)
ln DENS <sub>i</sub>	0.089*** (0.032)	0.166*** (0.030)	<b>0.062**</b> (0.026)	0.062*** (0.024)	-0.038 (0.026)	-0.063 (0.159)	<b>0.132***</b> (0.033)
ln DENS <sub>j</sub>	0.033 (0.031)	0.109*** (0.028)	<b>0.120***</b> (0.025)	0.118*** (0.024)	0.086*** (0.026)	0.008 (0.157)	<b>0.058*</b> (0.032)
USB <sub>i</sub>	0.923*** (0.120)	0.784*** (0.110)	<b>-0.058</b> (0.090)	-0.060 (0.085)	-0.386*** (0.094)	-0.325 (0.577)	<b>-0.227*</b> (0.119)
USB <sub>j</sub>	1.258*** (0.122)	1.119*** (0.112)	<b>0.880***</b> (0.077)	0.876*** (0.080)	0.882*** (0.088)	0.594 (0.549)	<b>0.481***</b> (0.112)
CONTIG <sub>ij</sub>	—	1.306*** (0.088)	<b>0.452***</b> (0.073)	0.450*** (0.076)	0.084 (0.084)	0.273 (0.478)	<b>0.093</b> (0.100)
ln STOCK <sub>ij</sub>	—	—	<b>0.632***</b> (0.025)	0.632*** (0.023)	0.769*** (0.025)	0.459*** (0.154)	<b>0.103***</b> (0.032)
ln DIST <sub>ij</sub>	-1.100*** (0.048)	-0.636*** (0.054)	<b>-0.066</b> (0.046)	-0.065 (0.045)	-0.115** (0.050)	-0.631** (0.290)	<b>-0.204***</b> (0.061)
Origin-/destination-specific fixed effects	No	No	<b>No</b>	No	Yes	Yes	<b>Yes</b>
Flow-specific fixed effects	No	No	<b>No</b>	No	No	Yes	<b>Yes</b>
Quasi-symmetry	Yes	Yes	<b>Yes</b>	Yes	Yes	Yes	<b>No</b>
R <sup>2</sup>	0.697	0.748	0.857	0.857	0.831	0.567	0.666
RMSE	0.861	0.786	0.592	0.592	0.642	1.520	0.961

Notes: Huber/White-robust standard errors are given in parentheses for the OLS estimator; approximate asymptotic standard errors are given in parentheses for the FE-GME estimator. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1%, respectively. The R<sup>2</sup> for the FE-GME estimator is based on the population-averaged effects only (i.e., the “restricted” R<sup>2</sup>, with all individual unobserved effects subtracted out) and calculated as the squared correlation between predicted and actual values. The FE-GME results are for error support  $\mathbf{z}_e = (-3\hat{\sigma}_m, 0, 3\hat{\sigma}_m)$ , in accordance with the three-sigma rule.

TABLE 5: Summary statistics for the variance decomposition—Gravity equation

	Std. dev.	Simple correlation with:												
		ln MIGR <sub>ij</sub>	Contiguity	$\hat{\lambda}$ CONTIG <sub>ij</sub>	$\hat{\kappa}_{ij}$ CONTIG <sub>ij</sub>	Migrant stock	$\hat{\eta}$ ln STOCK <sub>ij</sub>	$\hat{\mu}_{ij}$ ln STOCK <sub>ij</sub>	Distance	$\hat{\delta}$ ln DIST <sub>ij</sub>	$\hat{\tau}_{ij}$ ln DIST <sub>ij</sub>	State-fixed effects	$\hat{\omega}_i$	$\hat{\omega}_j$
ln MIGR <sub>ij</sub>	1.554	1.000	0.481	0.501	0.481	0.785	0.715	0.739	0.724	0.462	0.685	0.261	0.154	0.208
<b>Population</b>	<b>0.695</b>	<b>0.668</b>	-0.186	0.151	-0.186	0.318	0.420	0.240	0.167	0.320	0.100	0.182	0.197	0.077
$\hat{\alpha}_{POP,1}$ ln POP <sub>i</sub>	0.595	0.542	-0.113	0.108	-0.113	0.275	0.433	0.176	0.063	0.228	0.008	0.141	0.235	-0.005
$\hat{\alpha}_{POP,2}$ ln POP <sub>j</sub>	0.380	0.373	-0.170	0.108	-0.170	0.152	0.092	0.164	0.207	0.228	0.170	0.113	-0.008	0.148
<b>Socio-economic variables</b>	<b>0.249</b>	<b>0.242</b>	0.174	-0.066	0.174	-0.035	0.025	-0.059	-0.084	0.046	-0.107	0.301	0.062	0.328
$\hat{\alpha}_{UR,1}$ ln UR <sub>i</sub>	0.048	0.191	-0.129	0.080	-0.129	0.036	0.066	0.019	0.097	0.272	0.034	-0.046	-0.086	0.009
$\hat{\alpha}_{UR,2}$ ln UR <sub>j</sub>	0.057	-0.059	0.146	-0.080	0.146	-0.096	-0.158	-0.058	-0.112	-0.272	-0.051	0.226	-0.003	0.286
$\hat{\alpha}_{HDI,1}$ ln HDI <sub>i</sub>	0.035	-0.099	0.131	-0.067	0.131	0.102	0.234	0.031	-0.098	-0.179	-0.061	0.063	0.121	-0.015
$\hat{\alpha}_{HDI,2}$ ln HDI <sub>j</sub>	0.146	0.141	0.161	-0.067	0.161	-0.066	-0.110	-0.040	-0.098	-0.179	-0.061	0.357	-0.004	0.450
$\hat{\alpha}_{MAN,1}$ ln MAN <sub>i</sub>	0.151	-0.183	0.082	-0.064	0.082	-0.117	-0.195	-0.070	-0.092	-0.276	-0.027	0.075	0.123	-0.001
$\hat{\alpha}_{MAN,2}$ ln MAN <sub>j</sub>	0.031	-0.174	0.098	-0.064	0.098	-0.051	-0.040	-0.051	-0.134	-0.276	-0.074	0.032	-0.004	0.043
$\hat{\alpha}_{DENS,1}$ ln DENS <sub>i</sub>	0.177	0.322	0.008	0.032	0.008	0.080	0.246	-0.003	0.016	0.379	-0.086	-0.008	-0.013	0.001
$\hat{\alpha}_{DENS,2}$ ln DENS <sub>j</sub>	0.078	0.176	-0.061	0.032	-0.061	0.080	0.094	0.065	0.177	0.379	0.094	-0.012	0.000	-0.016
<b>US border-state indicators</b>	<b>0.223</b>	<b>0.159</b>	0.086	-0.033	0.086	-0.088	-0.187	-0.034	-0.060	-0.149	-0.027	0.107	0.030	0.112
$\hat{\alpha}_{USB,1}$ USB <sub>i</sub>	0.094	0.075	-0.051	0.070	-0.051	-0.212	-0.171	-0.209	-0.034	0.317	-0.126	0.044	0.066	0.004
$\hat{\alpha}_{USB,2}$ USB <sub>j</sub>	0.199	0.143	0.099	-0.070	0.099	0.001	-0.129	0.061	-0.051	-0.317	0.029	0.100	0.002	0.123
<b>Contiguity</b>	<b>0.160</b>	<b>0.481</b>	1.000	0.000	1.000	0.464	0.392	0.308	0.533	0.169	0.480	0.273	0.167	0.203
$\hat{\lambda}$ CONTIG <sub>ij</sub>	0.033	0.501	0.000	1.000	0.000	0.517	0.558	0.447	0.610	0.537	0.537	-0.003	-0.004	-0.001
$\hat{\kappa}_{ij}$ CONTIG <sub>ij</sub>	0.160	0.481	1.000	0.000	1.000	0.464	0.392	0.308	0.533	0.169	0.480	0.273	0.167	0.203
<b>Migrant stock</b>	<b>0.464</b>	<b>0.785</b>	0.464	0.517	0.464	1.000	0.847	0.970	0.855	0.366	0.859	-0.145	-0.070	-0.128
$\hat{\eta}$ ln STOCK <sub>ij</sub>	0.155	0.715	0.392	0.558	0.392	0.847	1.000	0.694	0.657	0.510	0.597	-0.059	0.097	-0.148
$\hat{\mu}_{ij}$ ln STOCK <sub>ij</sub>	0.342	0.739	0.308	0.447	0.308	0.970	0.694	1.000	0.861	0.264	0.893	-0.170	-0.139	-0.106
<b>Distance</b>	<b>0.698</b>	<b>0.724</b>	0.533	0.610	0.533	0.855	0.657	0.861	1.000	0.543	0.973	-0.207	-0.178	-0.122
$\hat{\delta}$ ln DIST <sub>ij</sub>	0.171	0.462	0.169	0.537	0.169	0.366	0.510	0.264	0.543	1.000	0.334	-0.136	-0.066	-0.119
$\hat{\tau}_{ij}$ ln DIST <sub>ij</sub>	0.622	0.685	0.480	0.537	0.480	0.859	0.597	0.893	0.973	0.334	1.000	-0.195	-0.181	-0.104
<b>State-fixed effects</b>	<b>0.386</b>	<b>0.261</b>	0.273	-0.003	0.273	-0.145	-0.059	-0.170	-0.207	-0.136	-0.195	1.000	0.602	0.788
$\hat{\omega}_i$	0.238	0.154	0.167	-0.004	0.167	-0.070	0.097	-0.139	-0.178	-0.066	-0.181	0.602	1.000	-0.017
$\hat{\omega}_j$	0.308	0.208	0.203	-0.001	0.203	-0.128	-0.148	-0.106	-0.122	-0.119	-0.104	0.788	-0.017	<b>1.000</b>

TABLE 6: Variations in real and "net" (log) migration flows

	Real $\ln \text{MIGR}_{ij}$	"Net" $\ln \text{MIGR}_{ij}$	Difference	Percentage
(Max-Min)/Min	3.806	2.038	1.768	46.5%
(P90-P10)/P10	0.825	0.474	0.351	42.5%
(P75-P25)/P25	0.381	0.235	0.146	38.3%
Coefficient of variation	0.227	0.157	0.070	30.8%

*Notes:* The "net" (log) migration flows are calculated as in Equation (14). Max, Min, P10, P90, P25, and P75 are the max, the min, the first decile, the last decile, the first quartile, and the last quartile, respectively.



TABLE 7: Determinants of flow-specific effects of distance and migrant stock—OLS and Tobit results

	Distance effects $ \hat{\delta}_{ij} $		Migrant-stock effects $\hat{\eta}_{ij}$	
	OLS (1.a)	Tobit (1.b)	OLS (2.a)	Tobit (2.b)
Constant	1.034*** (0.349)	0.578 (0.433)	0.504** (0.209)	0.250 (0.220)
$\Delta \ln UR_{ij}$	-0.027*** (0.007)	<b>-0.029***</b> (0.008)	-0.012*** (0.004)	<b>-0.013***</b> (0.004)
$\Delta \ln HDI_{ji}$	-0.023** (0.010)	<b>-0.024**</b> (0.010)	-0.013** (0.005)	<b>-0.013**</b> (0.005)
$\Delta \ln DENS_{ji}$	-0.013*** (0.002)	<b>-0.014***</b> (0.002)	-0.006*** (0.001)	<b>-0.006***</b> (0.001)
$DIRECT_{ji}$	-0.022*** (0.005)	<b>-0.024***</b> (0.005)	-0.010*** (0.002)	<b>-0.011***</b> (0.002)
$CONTIG_{ij}$	-0.070*** (0.008)	<b>-0.081***</b> (0.009)	-0.024*** (0.003)	<b>-0.030***</b> (0.004)
$\ln STOCK_{ij}$	-0.032*** (0.002)	<b>-0.034***</b> (0.002)	-0.028*** (0.001)	<b>-0.030***</b> (0.001)
$\ln DIST_{ij}$	-0.605*** (0.164)	<b>-0.407**</b> (0.164)	-0.319*** (0.140)	<b>-0.207**</b> (0.101)
$(\ln DIST_{ij})^2$	0.110*** (0.026)	<b>0.081***</b> (0.031)	0.057*** (0.015)	<b>0.040***</b> (0.015)
$(\ln DIST_{ij})^3$	-0.006*** (0.001)	<b>-0.005***</b> (0.002)	-0.003*** (0.001)	<b>-0.003***</b> (0.001)
$\hat{\sigma}$	—	0.063	—	0.030
Log-likelihood value	—	1207.2	—	1924.0
$R^2$	0.573	0.574	0.710	0.718
No. of observations	992	992	992	992

Notes: The difference variables are defined as follows:  $\Delta \ln X_{ji} = \ln X_j - \ln X_i = \ln(X_j/X_i)$ . The indexes  $i$  and  $j$  of  $\Delta \ln UR_{ij}$  have been reversed (in accordance with anticipated movements from high- to low-unemployment areas).  $DIRECT_{ji} = 1$  indicates south-north directionality of migration,  $DIRECT_{ji} = -1$  indicates north-south directionality of migration, and  $DIRECT_{ji} = 0$  indicates intra-regional migration. The predicted values from Tobit (with left censoring) are calculated as  $\hat{y}_{ij} = \Phi(\hat{z}_{ij})\hat{y}_{ij}^* + \hat{\sigma}\phi(\hat{z}_{ij}) \geq 0$ , where  $\hat{y}_{ij}^* = \mathbf{x}_{ij}\hat{\beta}$ ,  $\hat{z}_{ij} = \mathbf{x}_{ij}\hat{\beta}/\hat{\sigma}$ , and  $\hat{\sigma}$  is the estimated standard error of the residuals (Greene, 2003, p. 764). The partial effects can be calculated as  $\partial \mathbb{E}(y|\mathbf{x}_{ij})/\partial x_{ij,k} = \Phi(\hat{z}_{ij})\hat{\beta}_k$ . Instead of reporting the McFadden Pseudo- $R^2$ , which may not be the best measure of fit, I simply calculate the  $R^2$  as the square of the correlation between the predicted ( $\hat{y}_{ij}$ ) and observed ( $y_{ij}$ ) values. The Tobit estimates can be multiplied by appropriate adjustment factors to obtain the average partial effects (APE), which are roughly comparable to the OLS estimates. The APE scale factor is  $N^{-1}\Sigma_i \Phi(\hat{z}_{ij})$ , so the APE is  $[N^{-1}\Sigma_i \Phi(\hat{z}_{ij})]\hat{\beta}_k$ . The APE factors are 0.963 and 0.945 for distance and migrant stock, respectively. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1%, respectively.

TABLE 8: Differences in average contiguity effects – Cross-tabulation

	No. of observations	Average contiguity effect	Difference
All contiguous states	142	0.093	
Contiguous states without/with metropolitan zone close to <i>or</i> straddling the common border			
Without ( $ZM_{ij} = 0$ )	104	0.070	
With ( $ZM_{ij} = 1$ )	38	0.155	0.085*
Contiguous states without/with metropolitan zone straddling the common border			
Without ( $ZM_{ij} = 0$ )	128	0.073	
With ( $ZM_{ij} = 1$ )	14	0.276	0.203***

*Notes:* \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1%, respectively. The difference in means is tested on the basis of a simple  $F$ -statistic from a parametric test (assuming unequal variances). Identification of metropolitan zones in Mexico is based on SEDESOL-CONAPO-INEGI (2007). Six metropolitan zones are straddling seven Mexican state borders: La Laguna (Coahuila-Durango), Valle de México (D.F.-México/México-Hidalgo), Puerto Vallarta (Jalisco-Nayarit), La Piedad-Pénjamo (Michoacán-Guanajuato), Puebla-Tlaxcala (Puebla-Tlaxcala), and Tampico (Tamaulipas-Veracruz).

TABLE 9: Summary statistics for variance decomposition—Determinants of individual bilateral effects of distance and migrant stock

## (A) Individual distance-deterrence effects

	Std. dev.	Simple correlation with:									
		$ \hat{\delta}_{ij} $	<i>Socio-econ. differentials</i>	$\hat{\alpha}_{UR}$ $\Delta \ln UR_{ij}$	$\hat{\alpha}_{HDI}$ $\Delta \ln HDI_{ji}$	$\hat{\alpha}_{DENS}$ $\Delta \ln DENS_{ji}$	$\hat{\alpha}_{STOCK}$ $\ln STOCK_{ij}$	<i>Spatial variables</i>	$\hat{\alpha}_{DIRECT}$ $DIRECT_{ji}$	$\hat{\alpha}_{CONTIG}$ $CONTIG_{ij}$	$\hat{\alpha}_{DIST}$ $\ln DIST_{ij}^{(a)}$
$ \hat{\delta}_{ij} $	0.092	1.000	0.117	-0.009	0.004	0.118	0.624	0.557	-0.059	0.603	0.321
<i>Socio-econ. differentials</i>	<b>0.026</b>	<b>0.117</b>	1.000	0.110	0.211	0.898	-0.186	-0.124	-0.345	-0.000	-0.000
$\hat{\alpha}_{UR} \Delta \ln UR_{ij}$	0.009	-0.009	0.110	1.000	0.153	-0.277	-0.064	-0.007	-0.018	0.000	-0.000
$\hat{\alpha}_{HDI} \Delta \ln HDI_{ji}$	0.006	0.004	0.211	0.153	1.000	-0.078	-0.239	0.190	0.525	0.000	0.000
$\hat{\alpha}_{DENS} \Delta \ln DENS_{ji}$	0.026	0.118	0.898	-0.277	-0.078	1.000	-0.106	-0.164	-0.456	-0.000	0.000
<i>Migrant stock</i>											
$\hat{\alpha}_{STOCK} \ln STOCK_{ij}$	<b>0.052</b>	<b>0.624</b>	-0.186	-0.064	-0.239	-0.106	1.000	0.389	-0.209	0.558	0.141
<i>Spatial variables</i>	<b>0.039</b>	<b>0.557</b>	-0.124	-0.007	0.190	-0.164	0.389	1.000	0.361	0.821	0.619
$\hat{\alpha}_{DIRECT} DIRECT_{ji}$	0.014	-0.059	-0.345	-0.018	0.525	-0.456	-0.209	0.361	1.000	-0.000	-0.000
$\hat{\alpha}_{CONTIG} CONTIG_{ij}$	0.028	0.603	-0.000	0.000	0.000	-0.000	0.558	0.821	-0.000	1.000	0.240
$\hat{\alpha}_{DIST} \ln DIST_{ij}^{(a)}$	0.018	0.321	-0.000	-0.000	0.000	0.000	0.141	0.619	-0.000	0.240	1.000

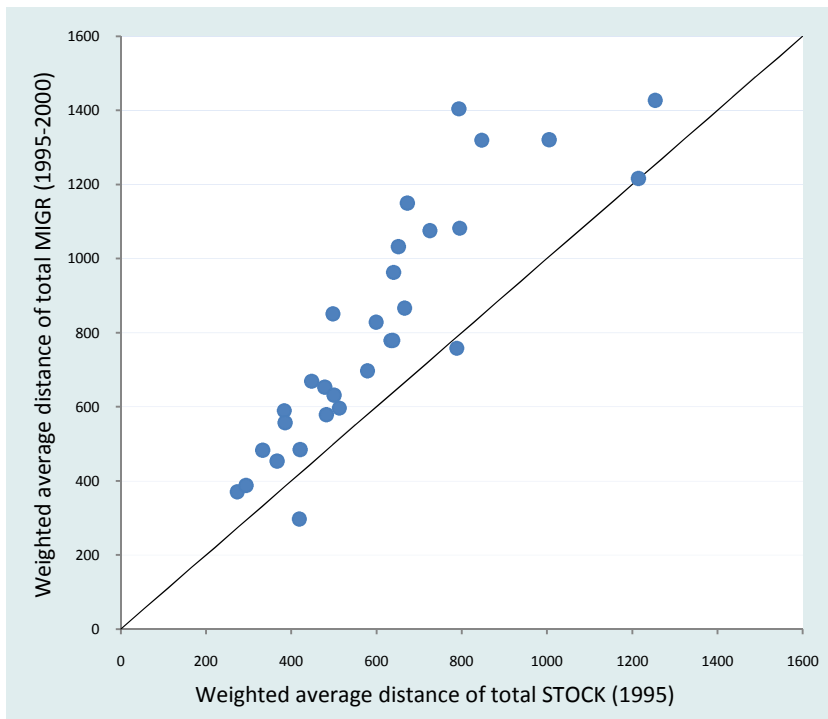
## (B) Individual migrant-stock effects

	Std. dev.	Simple correlation with:									
		$\hat{\eta}_{ij}$	<i>Socio-econ. differentials</i>	$\hat{\alpha}_{UR}$ $\Delta \ln UR_{ij}$	$\hat{\alpha}_{HDI}$ $\Delta \ln HDI_{ji}$	$\hat{\alpha}_{DENS}$ $\Delta \ln DENS_{ji}$	$\hat{\alpha}_{STOCK}$ $\ln STOCK_{ij}$	<i>Spatial variables</i>	$\hat{\alpha}_{DIRECT}$ $DIRECT_{ji}$	$\hat{\alpha}_{CONTIG}$ $CONTIG_{ij}$	$\hat{\alpha}_{DIST}$ $\ln DIST_{ij}^{(a)}$
$\hat{\eta}_{ij}$	0.055	1.000	0.008	-0.029	-0.076	0.041	0.774	0.334	-0.114	0.565	0.029
<i>Socio-econ. differentials</i>	<b>0.012</b>	<b>0.008</b>	1.000	0.130	0.271	0.875	-0.199	-0.123	-0.307	-0.000	-0.000
$\hat{\alpha}_{UR} \Delta \ln UR_{ij}$	0.004	-0.029	0.130	1.000	0.153	-0.277	-0.064	-0.007	-0.018	0.000	-0.000
$\hat{\alpha}_{HDI} \Delta \ln HDI_{ji}$	0.003	-0.076	0.271	0.153	1.000	-0.078	-0.239	0.211	0.525	0.000	0.000
$\hat{\alpha}_{DENS} \Delta \ln DENS_{ji}$	0.012	0.041	0.875	-0.277	-0.078	1.000	-0.106	-0.183	-0.456	-0.000	0.000
<i>Migrant stock</i>											
$\hat{\alpha}_{STOCK} \ln STOCK_{ij}$	<b>0.054</b>	<b>0.774</b>	-0.199	-0.064	-0.239	-0.106	1.000	0.091	-0.209	0.558	-0.228
<i>Spatial variables</i>	<b>0.017</b>	<b>0.334</b>	-0.123	-0.007	0.211	-0.183	0.091	1.000	0.401	0.491	0.672
$\hat{\alpha}_{DIRECT} DIRECT_{ji}$	0.007	-0.114	-0.307	-0.018	0.525	-0.456	-0.209	0.401	1.000	-0.000	-0.000
$\hat{\alpha}_{CONTIG} CONTIG_{ij}$	0.011	0.565	-0.000	0.000	0.000	-0.000	0.558	0.491	-0.000	1.000	-0.180
$\hat{\alpha}_{DIST} \ln DIST_{ij}^{(a)}$	0.013	0.029	-0.000	-0.000	0.000	0.000	0.141	0.672	-0.000	-0.180	1.000

<sup>(a)</sup>Third-order polynomial in  $\ln DIST_{ij}$

FIGURE 1: Spatial dispersion of current out-migrants and migrant stocks for the individual Mexican states

(A) Distance



(B) Contiguity

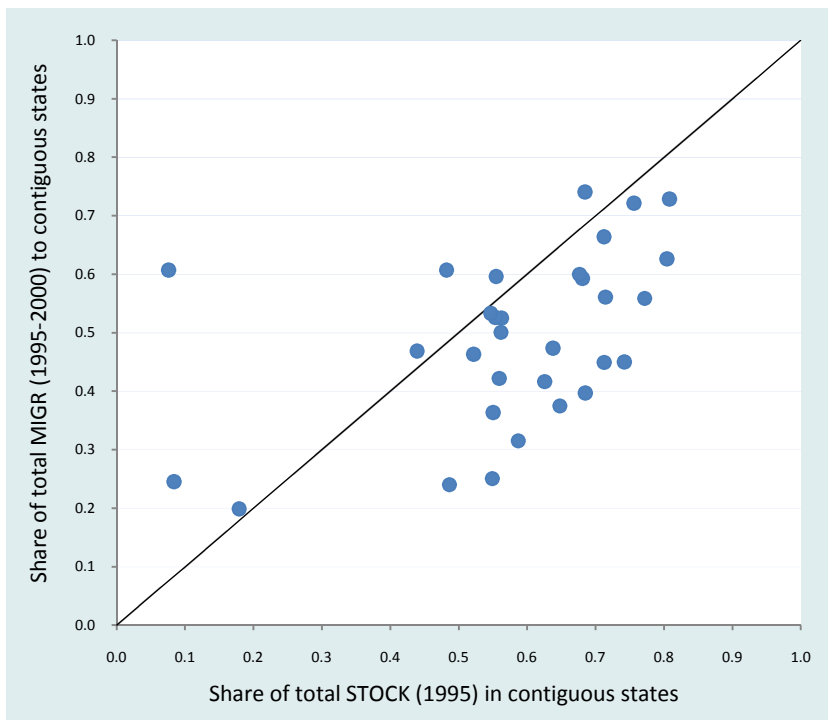


FIGURE 2: Kernel densities of estimated individual flow-specific effects

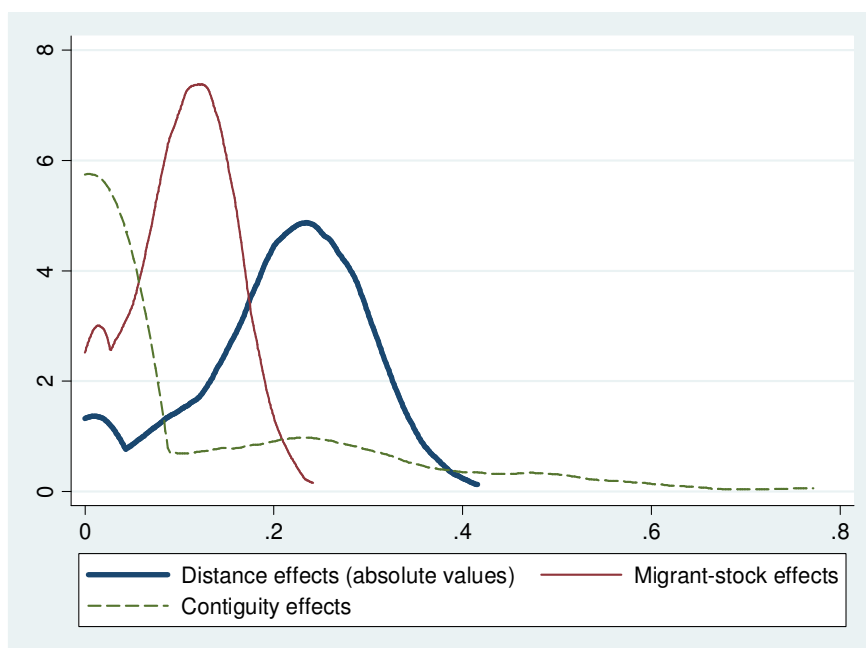
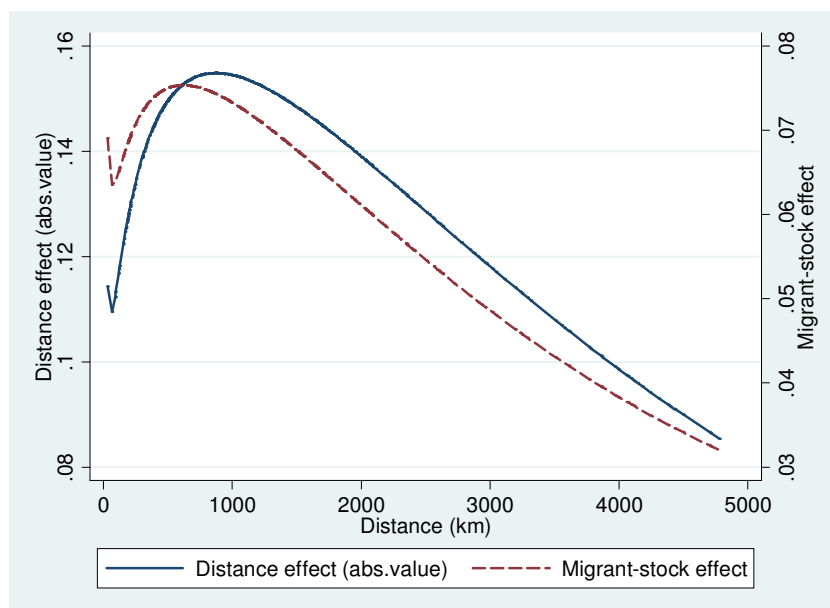
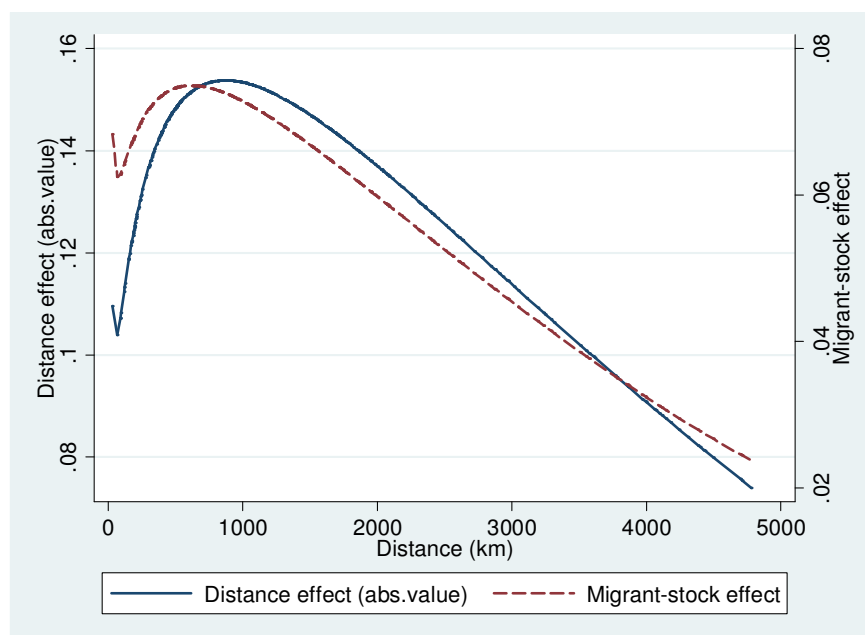


FIGURE 3: Distance profiles for distance-deterrence and migrant-stock effects

(A) Predicted conditional expectations



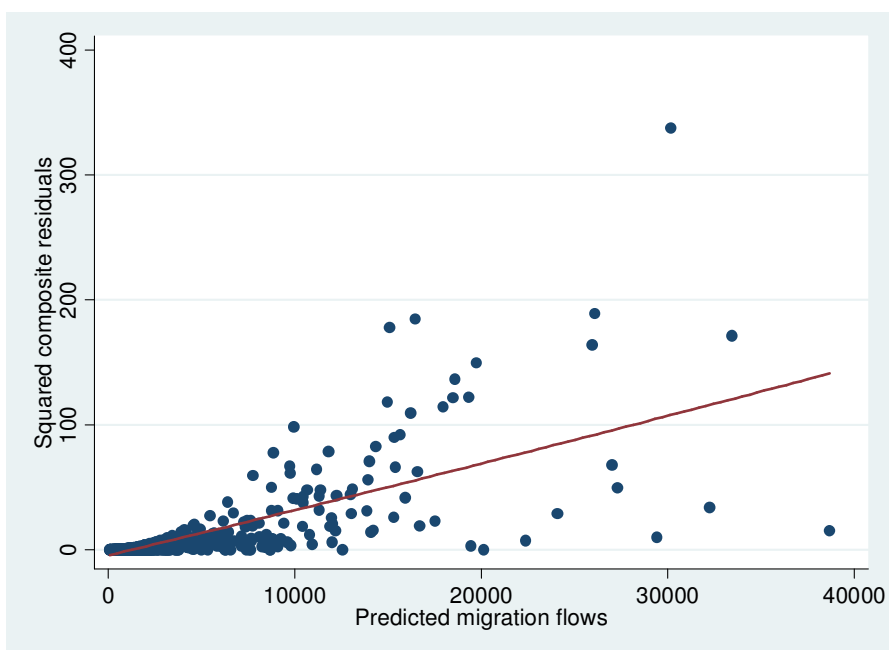
(B) Predicted unconditional expectations



*Notes:* The curves represent the predicted individual effects, based on the Tobit estimates in columns 1.b and 2.b of Table 7, evaluated at the sample mean of migrant stocks and all other variables (except distance, of course) set on zero. Inflection points for the varying distance effect are 50 km and 1040 km; inflection points for the varying migrant-stock effect are 70 km and 620 km. The *conditional* expectations of  $y$  (i.e., conditional on  $y > 0$ ) are  $\mathbb{E}(y|y > 0, \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)$ , where  $\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)$  is the inverse Mills ratio. These expectations tell us, for given values of  $\mathbf{x}$ , the expected values of  $y$  for the subpopulation where  $y$  is positive. The *unconditional* expectation of  $y$  (i.e., not conditional on  $y > 0$ ) are  $\mathbb{E}(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma) \cdot \mathbb{E}(y|y > 0, \mathbf{x})$ .



FIGURE 4: Arbitrary variance of composite residuals



Notes: On the vertical axis are the squared values of the composite residuals, which are calculated as  $\hat{\varepsilon}_{ij} = \text{MIGR}_{ij} - \widehat{\text{MIGR}}_{ij}^c$ , where  $\widehat{\text{MIGR}}_{ij}^c = \widehat{\text{MIGR}}_{ij}(e^{\hat{\xi}_{ij}^2/2})$ , where  $\hat{\xi}_{ij} = \hat{\nu}_i + \hat{\omega}_j + \hat{\kappa}_{ij}C_{ij} + \hat{\mu}_{ij}S_{ij} + \hat{\tau}_{ij}d_{ij} + \hat{\varepsilon}_{ij}$ . On the horizontal axis are the predicted migration flows,  $\widehat{\text{MIGR}}_{ij}^c$ . Values in the upper 1% tails of the distributions have been removed from the plot.

## APPENDIX A: DATA SOURCES

Data have been extracted from various sources:

- Data on inter-state migration and migrant stocks in Mexico are from the *Censos de Población* (INEGISIMBAD, CONAPO).
- Data on state characteristics are taken from the *Anuario Estadístico de los Estados Unidos Mexicanos* (INEGI).
- Data on the Human Development Index are taken from the *Population Reference Bureau*.
- Highway distances between the capital cities of the Mexican states are taken from the *Secretaría de Comunicaciones y Transportes* (SCT).

## APPENDIX B: MONTE CARLO SAMPLING EXPERIMENT

To give an idea of the sampling performances of OLS and FE-GME when dealing with an endogenous regressor, this Appendix presents results from a small-scale Monte Carlo experiment using simulated data.

### *Experimental design*

The data-generation process is assumed to be  $y_i = 1.0 + 0.2x_{1i} - 0.3x_{2i} + e_i$ . The two covariates are drawn randomly from a standard normal and uniform distribution, respectively; that is,  $x_{1i} \sim \mathcal{N}(0,1)$  and  $x_{2i} \sim \mathcal{U}(0,1)$ . The error term is drawn randomly so that the population correlation between the covariate  $x_{1i}$  and the error term  $e_i$  is equal to a predetermined level. To accomplish this, I make use of the following result:  $e_i = a \cdot x_{1i} + b \cdot \mathcal{N}(0,1)$ , where  $\rho_{x_1e} = a/\sqrt{a^2 + b^2}$ . For example, with  $a = 1$  and  $\rho_{x_1e} = 0.8$ , it follows that  $b = 0.75$ , so that  $e_i = x_{1i} + 0.75 \cdot \mathcal{N}(0,1)$ . Next, the standardized values of the generated errors ( $\varepsilon_i$ ) are calculated and added to the systematic part of the data-generating process. Thus,  $x_1$  is endogenous by construction; that is,  $\text{Cov}(x_{1i}, \varepsilon_i) \neq 0$  in  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ . To mimic the FE-GME estimator used in the main text, individual slopes on  $x_1$  are introduced; that is,  $y_i = \beta_0 + \beta_{1i} x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ . Note that this model is “ill-posed” (over-parameterized). The estimated mean (population-averaged) coefficient on  $x_1$  resulting from FE-GME is derived as  $\mathbb{E}(\hat{\beta}_{1i}) = N^{-1} \sum_{i=1}^N \beta_{1i}$ , where  $N$  is the sample size.

### *Estimation*

In implementing FE-GME, the parameter support is maintained at  $\mathbf{z}_\beta = (-10, -5, 0, 5, 10)'$ . The error support is consecutively set at  $\mathbf{z}_e = (-3\sigma_y, 0, 3\sigma_y)'$ , in keeping with the “three-sigma” rule (FE-GME3 $\hat{\sigma}$ ), and  $\mathbf{z}_e = (-10\sigma_y, 0, 10\sigma_y)'$ , which allows for a potentially reduced signal-to-noise ratio (FE-GME10 $\hat{\sigma}$ ). A total of 100 repeated samples are created, for sample sizes equal to 100, 500, and 1000, where  $\sigma_y$  (and, hence, the corresponding error support) is “updated” in every sampling replication. The 100 estimates of the parameters are treated as random drawings from their respective sampling distributions.

### *Results*

Table A.1 reports the expected values of the OLS and FE-GME parameter estimates. The latter are presented for different correlation levels and the two error support ranges. The results show the degree of bias for both estimators (recall that in a multiple-regression context it is generally impossible to sign the biases a priori). An examination of the limited results from the sampling experiments reveals that the convenient OLS estimator clearly fails when  $x_{1i}$  and  $e_i$  are correlated, whereas the competing FE-GME estimator yields considerably improved estimates, particularly when the sample size increases (as could be expected). Also, superior performance emerges for the widened error-support range defined as  $\pm 10\sigma_y$ . Thus, “loosening” the model appears to be a better choice for this sampling experiment. For correlations within the 0.4-0.8 interval and  $N = 1000$ , the OLS estimate of  $\beta_1$  is two to four times larger than the true parameter, whereas the finite-sample bias (either negative or positive) from FE-GME within the same interval is no more than about 25%.

### *Conclusion*

The results of this small-scale Monte Carlo experiment suggest that channeling the unobserved heterogeneity through the individual slope of the endogenous variable ( $x_1$ ) in a cross-section model is not a bad approximation, since the omitted-variables biases can be reduced to an “acceptable” level. Thus, the FE-GME estimator is clearly superior to the OLS estimator, particularly in the case of a relatively large sample (e.g.,  $N=1000$ ). Though one might raise the objection that the results are tied to the particular setup of the experiment, it is reassuring to find that the estimates reported for the “real-world” application in the main text are insensitive to the error-support range (and/or the parameter support space), which is suggestive of a robust interior solution.

TABLE B.1: Expected values of OLS and FE-GME parameter estimates (100 samples)

	Correlation ( $\rho_{x_1 \varepsilon}$ )						
	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<b>OLS</b>							
$N = 100$							
$\mathbb{E}(\hat{\beta}_0)$	1.047	1.068	1.089	1.110	1.131	1.153	1.176
$\mathbb{E}(\hat{\beta}_1)$	0.515	0.616	0.716	0.817	0.919	1.021	1.124
$\mathbb{E}(\hat{\beta}_2)$	-0.270	-0.271	-0.273	-0.275	-0.278	-0.282	-0.287
$N = 500$							
$\mathbb{E}(\hat{\beta}_0)$	0.998	1.002	1.006	1.010	1.015	1.019	1.025
$\mathbb{E}(\hat{\beta}_1)$	0.499	0.600	0.700	0.800	0.900	1.001	1.101
$\mathbb{E}(\hat{\beta}_2)$	-0.278	-0.278	-0.280	-0.281	-0.283	-0.286	-0.290
$N = 1000$							
$\mathbb{E}(\hat{\beta}_0)$	0.994	0.995	0.997	0.998	1.000	1.002	1.004
$\mathbb{E}(\hat{\beta}_1)$	0.505	0.606	0.707	0.810	0.914	1.019	1.126
$\mathbb{E}(\hat{\beta}_2)$	-0.282	-0.283	-0.284	-0.285	-0.287	-0.289	-0.292
<b>FE-GME3<math>\hat{\sigma}</math></b>							
$N = 100$							
$\mathbb{E}(\hat{\beta}_0)$	0.925	0.924	0.923	0.921	0.919	0.917	0.916
$\mathbb{E}(\hat{\beta}_1)$	0.319	0.381	0.443	0.504	0.564	0.625	0.686
$\mathbb{E}(\hat{\beta}_2)$	-0.148	-0.135	-0.122	-0.109	-0.097	-0.085	-0.075
$N = 500$							
$\mathbb{E}(\hat{\beta}_0)$	0.979	0.977	0.976	0.974	0.972	0.970	0.968
$\mathbb{E}(\hat{\beta}_1)$	0.322	0.384	0.445	0.506	0.565	0.624	0.682
$\mathbb{E}(\hat{\beta}_2)$	-0.263	-0.267	-0.260	-0.257	-0.255	-0.253	-0.251
$N = 1000$							
$\mathbb{E}(\hat{\beta}_0)$	0.986	0.987	0.987	0.988	0.989	0.991	0.992
$\mathbb{E}(\hat{\beta}_1)$	0.316	0.377	0.437	0.498	0.558	0.618	0.678
$\mathbb{E}(\hat{\beta}_2)$	-0.285	-0.288	-0.290	-0.293	-0.296	-0.300	-0.304
<b>FE-GME10<math>\hat{\sigma}</math></b>							
$N = 100$							
$\mathbb{E}(\hat{\beta}_0)$	0.838	0.830	0.882	0.815	0.808	0.801	0.795
$\mathbb{E}(\hat{\beta}_1)$	0.138	0.163	0.187	0.210	0.233	0.255	0.277
$\mathbb{E}(\hat{\beta}_2)$	-0.056	-0.041	-0.027	-0.013	-0.000	0.012	0.022
$N = 500$							
$\mathbb{E}(\hat{\beta}_0)$	0.934	0.927	0.920	0.912	0.905	0.898	0.892
$\mathbb{E}(\hat{\beta}_1)$	0.140	0.166	0.190	0.214	0.237	0.260	0.281
$\mathbb{E}(\hat{\beta}_2)$	-0.192	-0.179	-0.166	-0.154	-0.146	-0.129	-0.118
$N = 1000$							
$\mathbb{E}(\hat{\beta}_0)$	0.979	0.979	0.979	0.980	0.981	0.982	0.983
$\mathbb{E}(\hat{\beta}_1)$	0.136	0.160	0.184	0.207	0.230	0.252	0.274
$\mathbb{E}(\hat{\beta}_2)$	-0.272	-0.274	-0.276	-0.278	-0.281	-0.284	-0.287

Notes: The parameters of the data-generating process are set at  $\beta=(1.0, 0.2, -0.3)'$ . The parameter support is set at  $\mathbf{z}_\beta=(-10, -5, 0, 5, 10)'$ . The GME3 $\hat{\sigma}$  results are for error support  $\mathbf{z}_e=(-3\hat{\sigma}_y, 0, 3\hat{\sigma}_y)'$ ; the GME10 $\hat{\sigma}$  results are for  $\mathbf{z}_e=(-10\hat{\sigma}_y, 0, 10\hat{\sigma}_y)'$ .

## FOOTNOTES

---

<sup>1</sup> Some researchers have attempted to “resolve” the problem of heterogeneous people by taking into account differences in the profiles of the populations in the sending regions (e.g., in terms of the average educational attainment, or the demographic structure of a region’s population). However, such an approach, at best, only partly resolves the problem, given that the composition of the *migrant* population is likely to be noticeably different from the composition of the *region’s* population, which includes both (a minority of) “movers” and (a majority of) “non-movers” (e.g., Shioji, 2001).

<sup>2</sup> Researchers may have a continued interest in the use of cross sections for several reasons. First, numerous practical problems could arise in attempting to conduct (dynamic) panel-data studies of migration (e.g., due to lack of adequate panel data on an annual basis over an extended period of time). Second, the “identifying assumption” (i.e., the assumption that fixed effects ( $\alpha_i$ ) are time-invariant) may not be satisfied in reality, particularly if migration flows are observed over extended time intervals covering, say, five- or ten-year periods. In this respect, Mueser (1989b, p. 505) already pointed out that the unobserved factors or the prospective migrants’ *evaluations* of those factors may change within the time window employed. Third, additional problems with estimating (dynamic) panel-data models may arise due to, for example, the assumption of constant slopes ( $\beta_i = \beta$ ) (e.g., Alvarez-Plata et al., 2003). Obviously, such assumptions are likely to impose implausibly strong restrictions on the data.

<sup>3</sup> Other early studies dealing with inter-state migration in Mexico include Greenwood (1978), Greenwood and Ladman (1978), King (1978), Greenwood et al. (1981), Gordon and Theobald (1981); Cole and Sanders (1983), and Fukurai et al. (1987); a more recent example is Aroca and Maloney (2005). All these studies employ the convenient OLS estimator.

<sup>4</sup> The FE gravity model is *not* specified as  $m_{ij} = \alpha_0 + v_i + \omega_j + \theta_{ij} + \mathbf{x}_i\alpha_1 + \mathbf{x}_j\alpha_2 + \mathbf{z}_{ij}\beta + \epsilon_{ij}$ , where each migration flow has its own, unique intercept to absorb all omitted flow-specific effects. With a single cross section, the individual flow-specific effects ( $\theta_{ij}$ ) cannot be disentangled from the idiosyncratic error component ( $\epsilon_{ij}$ ). Conversely, the individual flow-specific effects ( $\theta_{ij}$ ) in Equation (3) *can* be isolated from the idiosyncratic error component ( $\epsilon_{ij}$ ).

<sup>5</sup> In a conceptual sense, the slopes on the bilateral linkage factors are treated as “unknown” (non-parameterized) functions of *both* observables *and* non-observables. For example,  $\delta_{ij} = \delta + \tau_{ij}$  can be written as  $\delta_{ij} = \phi(\mathbf{x}_{ij}, \zeta_{ij})$ , where  $\mathbf{x}_{ij}$  includes, say, distance and  $\zeta_{ij}$  represents the unobserved flow-specific effect. Obviously, this means that the FE model differs fundamentally from a constant-parameter model with squares and interactions, where the slopes would differ only by *observable* characteristics. In the empirical application below, potential nonlinearities will be examined in a second stage of the analysis.

<sup>6</sup> A roughly similar idea can be found in Calvet and Comon (2003), where  $\epsilon_{ij}$  creates (uncorrelated) heterogeneity in *all* directions and, say,  $\tau_{ij}$  introduces (correlated) heterogeneity *only* along  $d_{ij}$ .

<sup>7</sup> The three-way FE gravity model could be criticized for controlling too much, in the sense that all variation in the data is conditioned out by the fixed effects, which could lead to excessively reduced “signals”. However, it should be noted that the flow-fixed effects are not treated as “standalones” (intercepts) in the gravity equation but in *interaction* with the bilateral linkage factors.

<sup>8</sup> Further difficulties arise because network and herd effects are likely to be *complementary* and *reinforcing* each other (Bauer et al., 2002). That is, if there are positive network externalities, herd effects will be more pronounced, leading to relatively strong persistence of migration flows (path dependency). Conversely, if there are negative network externalities (e.g., due to increased competition for resources that are getting in short supply in a destination), herd effects will be less pronounced and a tendency emerges for new migrants to move to other locations (path diversion). As a result, it is difficult to disentangle their individual contributions to the persistence of migration. If we could hold the network effects constant, then any remaining difference would have to be due to the herd behavior, and vice versa. However, such an exercise is not possible in practice. Important in the context of the present analysis is that the unobserved bilateral effects are thus likely to be positively correlated with the (observed) levels of migrant stock.

<sup>9</sup> First, a partial-adjustment model may be plagued with the problem of “dominant variables”, where the lagged-flow variable accounts for so much of the variation in the dependent variable that the influence of the other variables in the model can no longer be properly estimated (e.g., Kau and Sirmans, 1979). Second, it is likely to be inappropriate to force an *identical* lag structure on all explanatory variables (e.g., Dunlevy and Gemery, 1977). By implication, the adjustment coefficient just represents a *compounded* effect which cannot be meaningfully traced to any underlying factor and, thus, leaves room for many different interpretations. Third, the herd effect, which is supposed to be captured by the adjustment coefficient, is *proportional* to the size of the previous migration flow, whereas it appears more reasonable to assume that path dependency is likely to be subject to a “depreciation function” of some sort (Dunlevy, 1993). Finally, entering lagged flows in the gravity equation often appears to be treated more as a “technical” control for data problems rather than being motivated by some theoretical considerations.

<sup>10</sup> The use of the log-linear form of the gravity model, rather than the Poisson (level) form, is thus by no means intended as a critique or rejection of the latter. In fact, the point is that it is simply not necessary to use a Poisson form to “correct” for heteroskedasticity (see below in the empirical section).

<sup>11</sup> It was found that the estimated coefficients of the push and pull factors in origin and destination are insensitive to variations in the specification of the stochastic constraints. Even though the mean (population-average) coefficients on the bilateral variables showed some numerical instability, the conclusions were not altered in a qualitative sense.

<sup>12</sup> Although Mexico has witnessed a growing decentralization of maquiladora jobs in the 1980s and 1990s, scattering from the U.S. border states into central Mexico, these jobs are still heavily concentrated along the U.S. border. Using data from INEGI, Jones (2001) reported that, in 1998, the U.S. border states were still hosting about 84% of the maquiladora jobs in Mexico.

<sup>13</sup> The GME estimation is implemented by using the non-linear optimization CONOPT3 solver of the GAMS software package. A limited version of the GME procedure is also available in the SAS software package (see, for example, SAS Institute, 2004), but was not considered suitable for the FE application.

<sup>14</sup> The FE-GME results are for the error support defined as  $\mathbf{z}_\epsilon = (-3\hat{\sigma}_m, 0, 3\hat{\sigma}_m)$ , in accordance with the three-sigma rule. To test the robustness of the estimates, the model was also run by using different parameter supports, ranging from  $(-10, 10)$  to  $(-1000, 1000)$ , and a widened error support vector defined as  $\mathbf{z}_\epsilon = (-10\hat{\sigma}_m, 0, 10\hat{\sigma}_m)$ , to allow for a potentially reduced signal-to-noise ratio. It was found that the results are not sensitive to changes in the setup of the estimation process, so the FE-GME estimator appears to return a robust “interior solution”. Such a finding should not be surprising, though, given the large number of observations ( $N = 992$ ). In Appendix B to this paper, I present a small-scale Monte Carlo experiment to illustrate the finite-sample properties of the FE-GME estimator in the case of an endogenous regressor.

<sup>15</sup> In passing, it can be noted that the relative impacts of population in origin and destination have been reversed, where the first is 0.738 (in OLS this is 0.462) and the second is 0.472 (in OLS this is 0.729). Accordingly, migration appears to be more “driven” by the origin population than by the destination population. Although theory generally does not provide any guidance as to what should be the relative magnitudes of these coefficients, it is felt that these changes are in the expected direction. For example, the larger is the origin population, the larger is the number of people “at risk” of migration, whereas the role of the size of the destination population is more difficult to understand (Schultz, 1982, p. 560)—unless being a proxy for unmeasured factors attracting migrants (Mueser, 1989b, p. 504).

<sup>16</sup> Note that the heterogeneity components of the distance and migrant-stock effects ( $\hat{\tau}_{ij}$  and  $\hat{\mu}_{ij}$ , respectively) display a very high correlation at 0.893.

<sup>17</sup> Given that the exact value of the individual bilateral effects are unknown, these “second-stage” regressions cannot be directly estimated by OLS or Tobit, but generally requires the application of some type of feasible generalized least squares (FGLS) estimator (see, for example, Combes et al., 2008). This issue, however, is beyond the scope of the present paper.

<sup>18</sup> Thus, the error variance is a quadratic function of the conditional mean, but the proportionality factor is arbitrarily related to the composite error term  $\xi_{ij}$  (arbitrary variance function); that is,  $\mathbb{V}[M_{ij}|\mathbf{x}] = \phi(\xi_{ij})\mu(\mathbf{x}_{ij}\boldsymbol{\beta})^2$ . This is a generalization of Case 3 in Santos Silva and Tenreiro (2006, p. 647), in which the log-linear model is consistent for the slope parameters.