# GRAVITY AND SPATIAL STRUCTURE: THE CASE OF INTERSTATE MIGRATION IN MEXICO*

**Ludo Peeters**
*Faculty of Business Economics, KIZOK Research Center for Entrepreneurship and Innovation, Hasselt University, Agoralaan (Building D), 3590 Diepenbeek, Belgium.*
*E-mail: ludo.peeters@uhasselt.be*

**ABSTRACT.** The estimation of gravity models of internal (aggregate) place-to-place migration is plagued with endogeneity (omitted-variable) biases if the unobserved effects of spatial structure are not accounted for. To address this econometric problem, this paper presents a more general specification of the gravity model, which allows for (bilateral) parameter heterogeneity across individual migration paths—along with (unilateral) origin- and destination-specific effects. The resultant "three-way fixed-effects" (3FE) model is applied for an analysis of interstate migration in Mexico based on cross-sectional data. To overcome parameter-dimensionality problems (due to limited or incomplete information), the 3FE model is estimated using the Generalized Maximum Entropy (GME) estimator. The empirical implications of this new modeling strategy are illustrated by contrasting the 3FE-GME estimates with those for the traditional and two-way fixed-effects (2FE) models. The former are far more plausible and intuitively interpretable than their traditional and 2FE counterparts, with parameter estimates changing in expected directions. The (average) effect of the migrant stock is markedly smaller than usually estimated, providing a more realistic measure of network-induced migration. Migration outflows from centrally located origins have significantly steeper distance decay. Path-specific distance effects exhibit directional asymmetries and spatial similarities.

## 1. INTRODUCTION

Gravity models are important for understanding the relationship between aggregate place-to-place migration and socioeconomic, demographic, and geographic factors. However, in empirically estimating a gravity model of migration, one of the trickiest, hardest to solve, and perhaps most important problems concerns the treatment of the role of spatial-structural factors in shaping aggregate migration patterns—an issue that keeps on puzzling many researchers working on internal (as distinct from international) migration. It has been recognized for many years by now that migration is not just a random process of relocation of people over space. That is, the observed migration pattern within a given geographical setting emerges from the migrants' choices which tend to be made within the context of an overarching *spatial structure* (Mueser, 1989). If migration flows are not independent over space, any standard (constant-parameter) gravity model of bilateral migration is likely to be misspecified due to its inability to capture the peculiarities of the hierarchical arrangement of the origins and destinations embedded in the particular migration setting under study—causing phenomena such as spatial

*I thank the co-editor, Steven Brakman, and three anonymous referees. I particularly thank one of the referees for providing exceptionally detailed and critical comments which allowed me to vastly improve the presentation. I also thank Amos Golan, Patricio Aroca, and Raúl Romo Viramontes for their encouragement, suggestions, and advice. Finally, I am grateful to Isidro Soloaga for providing me with most of the data used in the paper. The usual disclaimers apply.

Received: February 2010; revised: March 2012; accepted: April 2012.

 DOI: 10.1111/j.1467-9787.2012.00770.x

focus (Rogers and Raymer, 1998), spatial drift (Plane, 1999), and oriented channels (Carrington, Detragiache, and Vishwanath, 1996), involving *spatial heterogeneity* (spatially varying parameters and heteroskedasticity arising from unobserved, localized factors; see Anselin, 2010, p. 5). Thus, any failure to account for the spatial-structural effects on bilateral migrations may lead to biased parameter estimates and, hence, to misleading inferences (Pellegrini and Fotheringham, 2002).

Econometrically speaking, the underlying spatial structure of migration raises two *endogeneity* concerns. The first concern arises from the fact that, conditionally on migrating, people *choose* their destinations. As a result, destinations are endogenous and consequently so are all migration distances (Pinkse and Slade, 2010, p. 113). Or, put differently, the choice of a particular destination is *consubstantial* with (inextricably linked to) the choice of the distance to cover (Magrini and Lemistre, forthcoming, p. 2). The second endogeneity concern arises from the observation that spatial patterns of migration tend to be fairly persistent, path dependent, or *sluggish*, as Tobler (1995) termed it. Therefore, the use of "migrant stock" to assess the importance of "family-and-friends" linkages in inducing *network* migration is problematic, given that established stocks of in-migrants (which are a function of previous levels of migration) and current migration flows are likely to be *jointly* determined by the same underlying spatial structure—of which distance is just one component (for some recent studies on the effect of networks on international migration, see Bauer, Epstein, and Gang, 2007; Clark, Hatton, and Williamson, 2007; Pedersen, Pytlikova, and Smith, 2008; Hatton and Williamson, 2011). As people currently migrating from origin $i$ to destination $j$ may do so largely for the same reasons as those who migrated from $i$ to $j$ in the past, it is difficult to empirically disentangle network effects and spatially induced structural effects.

Several attempts have been made in earlier migration studies to account for the spatial structure of migration. The most prominent ones are the intervening-opportunities (Stouffer, 1960; Akwawua and Pooler, 2001) and competing-destinations models (Fotheringham, 1991; Hu and Pooler, 2002). These models extend the gravity equation by including some centrality or accessibility indicator to capture the "hierarchical structure of space" emanating from the spatially uneven distribution of opportunities (usually proxied by population counts) within the migration setting studied. Most of these early attempts have not been successful, though, typically producing counter-intuitive estimates of the distance parameter (Tiefelsdorf, 2003). Yet, such unexpected results are not surprising, as Denslow and Eaton (1984) already rightly noted that if the distance function is misspecified, the addition of spatial indicators (which are usually based on some distance measure themselves) may be picking up some of the effects of distance inappropriately, producing biases similar to those previously outlined in Renshaw (1974) with regard to the migrant stock. Therefore, followers of Casetti's (1972) expansion method have emphasized the "contextual" nature of the distance effect and attempted to account for spatially varying relationships (Eldridge and Jones, 1991; Fotheringham and Pitts, 1995). More recently, the focus has shifted to such methods as "geographically weighted regression" (GWR) (Fotheringham, Brunsdon, and Charlton, 2002), the use of Box–Cox transformations of distance (Tiefelsdorf, 2003), the specification of flexible distance-decay functions (de Vries, Nijkamp, and Rietveld, 2009), and the use of spatial econometrics/spatial filtering techniques (Tiefelsdorf and Griffith, 2007; Chun, 2008) to remove or isolate the spatial-structural effects in standard gravity models. Without refuting the specific virtues of these approaches, they do have some major shortcomings, though, in that estimation results are likely to be strongly tied to the assumed form of the distance-decay function or the spatial weights (spatial-dependence structure) pre-imposed on the data.

In this paper, I propose yet another approach to account for the spatial structure of migration and to address the challenge posed by the endogeneity problems mentioned above. This approach starts from the basic idea that any observed migration pattern within a given geographical setting is a realization of (relatively stable) spatial choices involving information that is largely *unobservable* to (or difficult to measure by) the analyst, making the explicit modeling of the entire spatial-dependence structure a nearly impossible task (Pinkse and Slade, 2010, p. 108).[1] As a result, omitted variables are likely to abound. To deal with this apparent difficulty, I present a more general specification of the gravity model, which allows for *heterogeneity* in the *bilateral slope parameters*. Specifically, my objective is to estimate a model in which (unobserved) spatially structured heterogeneity *interacts* with the (observed) linkage variables (such as distance, contiguity, and the migrant-stock proxy for networks) and may be *unconditionally correlated* with these linkage variables (see, for example, Wooldridge, 2004, for a similar model structure). To achieve this, the bilateral slope parameters will be estimated within a "fixed-effects" framework, by entering (bilateral) *path-specific dummy interactions* with each of the linkage variables—besides the more common (unilateral) origin- and destination-fixed effects, giving way to what I refer to as a "three-way fixed-effects" (3FE) model. By making explicit allowance for correlated spatial heterogeneity in the parameters of the linkage variables, the 3FE gravity model is expected to be superior in accounting for the heteroskedasticity (overdispersion) present in the data (Zietz, 2001) and to be helpful in removing a potentially important source of omitted-variable bias (Congdon, 1993; Berk and MacDonald, 2008).

Another distinct feature of the proposed modeling strategy is that all fixed effects are treated as parameters to be *estimated* rather than just an ancillary "nuisance" problem. This creates a complication, though, particularly when the researcher has only access to cross-sectional data. That is, the inclusion of dummy interactions with each of the linkage variables turns the estimation of the gravity model into an "ill-posed" (underdetermined) inverse problem (according to the traditional econometric point of view), so that classical estimation methods cannot generally be used. To overcome this parameter-dimensionality problem, the 3FE model is estimated using an (in)equality-constrained (restricted) variant of the Generalized Maximum Entropy (GME) estimator, which builds on the standard GME method introduced by Golan, Judge, and Miller (1996). An important advantage of using GME is that it opens up the possibility of "regularizing" inverse problems with incomplete information or limited (and noisy) data, while providing a unique and stable solution.

To illustrate the usefulness of adopting this new modeling strategy, the GME estimation of the proposed 3FE model is applied for an exploratory case-study analysis, employing cross-sectional data on interstate migration in Mexico during the five-year observation period 1995–2000. For the sake of comparison of the results across alternative model specifications, I also estimate the traditional (observables-only) and 2FE (including origin- and destination-fixed effects) specifications of the gravity model using the conventional OLS and Poisson Pseudo Maximum Likelihood (PPML) estimators (Santos Silva and Tenreyo, 2006; see also the early paper on using Poisson by Flowerdew and Aitkin, 1982). The analysis demonstrates that allowing for bilateral slope-parameter heterogeneity has important empirical implications, giving rise to appreciable changes, both in magnitude and in sign, of the estimated coefficients, where the estimates for the 3FE model are found to be far more plausible and intuitively interpretable than their

[1]Even though the notion of spatial structure is well accepted in the field of geography and regional sciences, much of the literature on internal migration is curiously ambiguous on the definition of spatial structure and how it should be measured. Mueser (1989) is one of the few empirical studies in which it has been explicitly recognized that the underlying forces governing the spatial patterns of migration are not well understood, which motivated him to introduce the notion of "generalized distance."

traditional and 2FE counterparts. Moreover, the 3FE-GME estimation results reveal substantial variation in the path-specific effects of the linkage variables.

The remainder of this paper is organized as follows. Section 2 outlines the 3FE specification of the gravity model. Section 3 sets out the (in)equality-constrained GME method used to estimate the unknown parameters of the 3FE model. Section 4 presents the empirical analysis of interstate migration in Mexico, beginning with an overview of the variables included in the empirical model. This overview is followed by a brief comparative analysis, showing the sensitivity of the parameter estimates across alternative model specifications (traditional, 2FE, and 3FE), along with some specification tests of the GME-estimated 3FE gravity model. The last piece of the empirical analysis expands on the heterogeneity of the estimated effects of the linkage covariates, focusing on the extent to which they are reflective of the impact of some special features of the spatial structure embedded in the migration setting. Finally, Section 5 provides a summary and some concluding remarks.

## 2. THE GRAVITY MODEL WITH THREE-WAY FIXED EFFECTS

In its most elementary form, the gravity model of migration stipulates that each (gross) migration flow from origin $i$ to destination $j$, denoted by $M_{ij}$ ($i \neq j; i, j = 1, \ldots, n$), is proportional to the product of the origin and destination population sizes ("masses"), denoted by $P_i$ and $P_j$, respectively, and inversely related to the distance separating origins and destinations, denoted by $D_{ij}$. The stochastic version of this elementary gravity equation, in *multiplicative* form, is

$$M_{ij} = \beta_0 P_i^{\beta_1} P_j^{\beta_2} D_{ij}^{\delta} u_{ij}, \tag{1}$$

where $\beta_0$, $\beta_1$, $\beta_2$, and $\delta$ are unknown parameters, and $u_{ij}$ is an idiosyncratic error with $E(u_{ij}|P_i, P_j, D_{ij}) = 1$, which is assumed to be statistically independent of the covariates $P_i$, $P_j$, and $D_{ij}$, from which it follows that $E(M_{ij}|P_i, P_j, D_{ij}) = \beta_0 P_i^{\beta_1} P_j^{\beta_2} D_{ij}^{\delta}$. This basic gravity model can be extended ("modified") by including additional (unilateral) origin and destination "push/pull" and (bilateral) origin-destination "linkage" indicators.

Rather than log-linearizing the model in Equation (1), which is common practice in the empirical migration literature, $M_{ij}$ is predicted "directly" by using a linear-exponential conditional mean function as a baseline specification, so that the gravity model can be written as

$$M_{ij} = e^{\mathbf{x}_{ij}\boldsymbol{\beta}} u_{ij}; \quad E(u_{ij}|\mathbf{x}_{ij}) = 1, \tag{2}$$

$$M_{ij} = e^{\mathbf{x}_{ij}\boldsymbol{\beta}} + r_{ij}; \quad E(r_{ij}|\mathbf{x}_{ij}) = 0, \tag{3}$$

where $\mathbf{x}_{ij}\boldsymbol{\beta}$ is used as a shorthand for the linear predictor, $\mathbf{x}_{ij} = [1\ \mathbf{x}_i\ \mathbf{x}_j\ \mathbf{z}_{ij}]'$, in which $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors containing the origin-push and destination-pull factors, respectively, $\mathbf{z}_{ij}$ is a vector of bilateral factors linking origin $i$ and destination $j$, and $e^{\mathbf{x}_{ij}\boldsymbol{\beta}}$ is the conditional expectation of $M_{ij}$ given $\mathbf{x}_{ij}$, denoted by $E(M_{ij}|\mathbf{x}_{ij})$.[2]

[2]Whether the error term enters the gravity model multiplicatively, as $u_{ij}$ or additively, as $r_{ij}$ is in fact irrelevant (Santos Silva and Tenreyo, 2006, p. 644). It can be shown that $u_{ij} = 1 + r_{ij}/e^{\mathbf{x}_{ij}\boldsymbol{\beta}}$, with $E(u_{ij}|\mathbf{x}_{ij}) = 1$ provided that $E(r_{ij}|\mathbf{x}_{ij}) = 0$ and $V(u_{ij}|\mathbf{x}_{ij}) = V(r_{ij}|\mathbf{x}_{ij})/e^{2\mathbf{x}_{ij}\boldsymbol{\beta}}$. Note that the multiplicative error term, $u_{ij}$ is homoskedastic (constant variance of the error distribution) if $V(r_{ij}|\mathbf{x}_{ij})$ or $V(M_{ij}|\mathbf{x}_{ij})$ is proportional to $[E(M_{ij}|\mathbf{x}_{ij})]^2$. That is, if $V(r_{ij}|\mathbf{x}_{ij}) = \lambda_0 e^{2\mathbf{x}_{ij}\boldsymbol{\beta}}$ it follows that $V(u_{ij}|\mathbf{x}_{ij}) = (\lambda_0 e^{2\mathbf{x}_{ij}\boldsymbol{\beta}})/e^{2\mathbf{x}_{ij}\boldsymbol{\beta}} = \lambda_0$.

*Heterogeneous Bilateral Slope Parameters*

To control for the correlated effects of the underlying spatial structure of migration and, hence, to properly identify the model parameters, some adjustment of the gravity model in Equation (2) or (3), is needed. This adjustment is achieved by allowing for slope-parameter heterogeneity, which yields the following three-way fixed-effects (3FE) specification of the gravity model:

$$M_{ij} = e^{\beta_0 + \boldsymbol{\beta}_1'\mathbf{x}_i + \boldsymbol{\beta}_2'\mathbf{x}_j + \boldsymbol{\beta}_{3,ij}'\mathbf{z}_{ij} + \upsilon_i + \omega_j} u_{ij}, \tag{4}$$

where $\beta_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_{3,ij}$, $\upsilon_i$, and $\omega_j$ are unknown parameters to be estimated.

The model in (4) differs from the traditional gravity model in two major respects. First, $\upsilon_i$ and $\omega_j$ are origin- and destination-fixed effects, respectively, which are assumed to absorb the heterogeneity along the unilateral dimensions that remains after conditioning on the observed push/pull factors in $\mathbf{x}_i$ and $\mathbf{x}_j$. Second, and more importantly, the model includes dummy interactions with each of the bilateral linkage factors in $\mathbf{z}_{ij}$, allowing for path-specific coefficients, $\boldsymbol{\beta}_{3,ij}$, so that each migration flow along a given $(i, j)$-path has its own set of slope parameters. In a certain way, accommodating such parameter heterogeneity can be viewed as an adequate means of approximating "nonlinear" gravity relationships (of an unknown form), which makes the model less subject to specification errors (as convincingly argued in, for example, Swamy and Tavlas, 1995).

For simplicity, the parameters in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ associated with $\mathbf{x}_i$ and $\mathbf{x}_j$, are assumed to be *constant* throughout this paper. Assuming further that $\boldsymbol{\beta}_{3,ij} \equiv \boldsymbol{\beta}_3 + \mathbf{b}_{ij}$, where $E(\mathbf{b}_{ij}) = 0$ or $E(\boldsymbol{\beta}_{3,ij}) = \boldsymbol{\beta}_3$, and $u_{ij} \equiv e^{\epsilon_{ij}}$, the 3FE model in (4) can be rewritten as

$$M_{ij} = \underbrace{e^{\beta_0 + \boldsymbol{\beta}_1'\mathbf{x}_i + \boldsymbol{\beta}_2'\mathbf{x}_j + \boldsymbol{\beta}_3'\mathbf{z}_{ij}}}_{[1]} \times \underbrace{e^{\mathbf{b}_{ij}'\mathbf{z}_{ij} + \upsilon_i + \omega_j}}_{[2]} \times \underbrace{e^{\epsilon_{ij}}}_{[3]}, \tag{5}$$

or, more compactly,

$$M_{ij} = e^{\mathbf{x}_{ij}\boldsymbol{\beta}} e^{\xi_{ij}} e^{\epsilon_{ij}} \equiv e^{\mathbf{x}_{ij}\boldsymbol{\beta} + \xi_{ij}} e^{\epsilon_{ij}} \equiv e^{\mathbf{x}_{ij}\boldsymbol{\beta} + \xi_{ij}} u_{ij}, \tag{6}$$

where three different components can be distinguished. Component [1] is the (global) *trend-surface* component (e.g., Tiefelsdorf, 2003; Chun, 2008), where all unobserved heterogeneity is "filtered out." The vector $\boldsymbol{\beta}$ in [1] contains, apart from the intercept, $\beta_0$, the constant parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ related to the origin and destination characteristics, $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and the "population-averaged" parameters $\boldsymbol{\beta}_3 = E(\boldsymbol{\beta}_{3,ij})$ associated with the bilateral covariates, $\mathbf{z}_{ij}$. Component [2] is the (local) *spatial-structural heterogeneity* component, which accounts for the additional variation in the expected number of migrant exchanges due to (unilateral) *level* heterogeneity, $\upsilon_i$ and $\omega_j$, and (bilateral) *slope* heterogeneity, $\mathbf{b}_{ij}$. Component [3] is a stochastic (random) i.d.d. *error term*, $e^{\epsilon_{ij}} \equiv u_{ij}$, capturing the *remaining* heterogeneity after both trend and spatial-structural effects have been controlled for. The basic intuition behind the 3FE model specification is that, conditionally on both $\mathbf{x}_{ij}$ and $(\mathbf{b}_{ij}, \upsilon_i, \omega_j)$, it is more reasonable to maintain that $E(u_{ij}|\mathbf{x}_{ij}, \mathbf{b}_{ij}, \upsilon_i, \omega_j) = 1$, from which it follows that the conditional mean of the migration count is $E(M_{ij}|\mathbf{x}_{ij}, \mathbf{b}_{ij}, \upsilon_i, \omega_j) = e^{\mathbf{x}_{ij}\boldsymbol{\beta} + \xi_{ij}}$ (see also Wooldridge, 2005, in a panel-data setting).

To get a better understanding of how the 3FE model works, it is instructive to focus for the moment on a model with (log) distance, $d_{ij}$, as the sole covariate (which, after all, is at the very core of the gravity equation). Assuming that $\delta_{ij} \equiv \delta + \tau_{ij}$, the 3FE model in Equation (5) reduces to

$$\begin{aligned} M_{ij} &= e^{\beta_0 + \delta_{ij} d_{ij} + \upsilon_i + \omega_j} e^{\epsilon_{ij}} \\ &= e^{\beta_0 + \delta d_{ij} + \upsilon_i + \omega_j} e^{\tau_{ij} d_{ij} + \epsilon_{ij}} \\ &= e^{\beta_0 + \delta d_{ij} + \upsilon_i + \omega_j} v_{ij}, \end{aligned} \tag{7}$$

Using this simplified version of the 3FE model, five of its key features can more easily be clarified. First, the model allows unobserved heterogeneity (arising from omitted spatial-structural information not contained in the distance measure), $\tau_{ij}$, to be *unconditionally correlated* with $d_{ij}$ and to *interact* with $d_{ij}$ (Wooldridge, 2004). This is important for two reasons: (a) failure to control for potentially correlated spatial effects implies that the (trend) parameter of distance, $\delta$, cannot be identified, even after controlling for origin- and destination-fixed effects, $\upsilon_i$ and $\omega_j$, since $\tau_{ij}$ is embedded in the error, $v_{ij}$, so that $E(v_{ij}|d_{ij}, \upsilon_i, \omega_j) \neq 1$; and (b) the inclusion of the interaction term $\tau_{ij}d_{ij}$ implies that the distance effect *depends* on unobserved heterogeneity, hence giving rise to *spatially uneven* (as opposed to spatially invariant) effects.[3] Second, the model does not require making *a priori* assumptions regarding the form of the unconditional distribution of the path-specific $\delta_{ij}$*'s* (such as the normal, log-normal, triangular, or any other form), while opening up the possibility of accounting for the bounded nature of their distribution (see further below). The only assumption implied here is that $E(\tau_{ij}) = E_d[E(\tau_{ij}|d_{ij})] = 0$, from which it follows that $E(\delta_{ij}) = E_d[E(\delta_{ij}|d_{ij})] = \delta$. However, this assumption (which is actually a definition) simply tells us that the parameter $\delta$ represents the *unconditional* (marginal) *mean* of the path-specific effects. Third, the model places no restrictions on the *conditional* distribution of $\delta_{ij}$ *given* $d_{ij}$, so that the assumption of mean independence (which is always implicitly present in a standard random-coefficients model) is relaxed, allowing the conditional mean effect, $E(\delta_{ij}|d_{ij})$, to be generally different from the unconditional mean effect, $E(\delta_{ij})$. This feature should be viewed as a key asset of the 3FE model, since mean-independence is not a credible assumption given the likely presence of correlated heterogeneity in the sample population. (As the conditional expectations of the path-specific effects are of special interest here, they will be estimated (in a second-stage estimation) on the basis of the estimated $\delta_{ij}$'s (see the empirical section later in the paper).) Fourth, by letting $\tau_{ij} \neq \tau_{ji}$ (or $\delta_{ij} \neq \delta_{ji}$), the model allows for *directional heterogeneity*, thus avoiding the heroic "quasi-symmetry" assumption typically underlying the standard gravity model of migration (Bavaud, 2002). Last, the path-specific $\delta_{ij}$'s (unlike the trend parameter $\delta$) have an obvious *elasticity* interpretation; that is, the elasticity of $E(M_{ij}|d_{ij}, \tau_{ij}, \upsilon_i, \omega_j)$ with respect to distance, $D_{ij}$, is $\partial \ln E(M_{ij}|\bullet)/\partial d_{ij} = \delta + \tau_{ij} \equiv \delta_{ij}$, where the unconditional mean distance elasticity, or "distance-decay rate," $\delta = E(\delta_{ij})$, can be consistently estimated as $N^{-1}\sum_i \sum_j \hat{\delta}_{ij} = N^{-1}\Sigma_i \Sigma_j(\hat{\delta} + \hat{\tau}_{ij}) = \hat{\delta} + N^{-1}\Sigma_i \Sigma_j \hat{\tau}_{ij} = \hat{\delta}$, with $N = n(n-1)$.

The basic model in Equation (7) can easily be extended to accommodate other (widely used) bilateral covariates, such as a *contiguity* (common-border) dummy, $C_{ij}$, and the (log) *migrant stock*, $s_{ij}$, so that

$$\text{(8)} \qquad \boldsymbol{\beta}'_{3,ij}\mathbf{z}_{ij} = \begin{bmatrix}\delta_{ij} & \lambda_{ij} & \eta_{ij}\end{bmatrix}\begin{bmatrix} d_{ij} \\ C_{ij} \\ s_{ij}\end{bmatrix}; \quad \boldsymbol{\beta}_{3,ij} = \boldsymbol{\beta}_3 + \mathbf{b}_{ij} = \begin{bmatrix}\delta \\ \lambda \\ \eta\end{bmatrix} + \begin{bmatrix}\tau_{ij} \\ \kappa_{ij} \\ \mu_{ij}\end{bmatrix},$$

where $\lambda_{ij} \equiv \lambda + \kappa_{ij}$ and $\eta_{ij} \equiv \eta + \mu_{ij}$ are the path-specific slope parameters associated with contiguity and migrant stock, respectively, $E(\lambda_{ij}) = \lambda$, $E(\eta_{ij}) = \eta$, or $E(\kappa_{ij}) = 0$,

[3]Attempts to make standard gravity models more flexible by allowing the coefficient on distance to vary with distance have not been pursued here for several reasons. First, there are an infinite number of ways to do so, as theory provides no guidance in choosing an appropriate form of the distance-decay function. Second, more flexible distance-decay functions essentially maintain the restrictive assumption of parameter homogeneity. Last, making allowance for path-specific slope parameters actually constitutes the most general way of introducing non-linear relationships of an unknown form.

$E(\mu_{ij}) = 0$, and $\lambda_{ij} \neq \lambda_{ji}$, $\eta_{ij} \neq \eta_{ji}$, so as to allow for potential directional asymmetries in these bilateral effects as well.

An important implication of this extension is that the proposed 3FE model does not only account for the endogeneity of distance (and contiguity), but can also be particularly helpful in addressing the recurrent concern about properly measuring the importance of networks of "family and friends," already settled in the destination, in facilitating migration due to lowered moving costs (brought about by the job-search assistance, help in finding housing, etc., provided by previous migrants to later migrants). Specifically, by controlling for the effects of the underlying spatial structure, the 3FE model is likely to return an improved (more realistic) measure of network-induced migration, hence reducing the usual ambiguity in interpreting the migrant-stock coefficient in standard models. To the extent that the network interpretation of the coefficient on the migrant-stock variable in the 3FE model is correct, its magnitude is expected to be much smaller than the overstated values reported in previous studies of internal migration (typically ranging between 0.6 and 0.9), which are likely to be confounding network effects and spatial-structure effects, and, by implication, providing sturdily overstated measures of network-induced migration.[4]

*Three-Way vs. Two-Way Fixed Effects*

It might be tempting to be content with just introducing origin- and destination-fixed effects (2FE) to account for the presence of unobserved spatial-structural effects. Some researchers would refer to the recent literature on gravity modeling of bilateral trade flows, following the influential paper of Anderson and van Wincoop (2003), in which the inclusion of exporter- and importer-fixed effects has become a standard approach to account for unobserved "multilateral resistance" (MR) to trade; that is, roughly speaking, for the fact that bilateral trade is determined by *relative* trade costs.

However, the adoption of a 2FE MR-type of approach to control for unobserved bilateral effects on migration flows is not considered appropriate for two major reasons. First, the 2FE approach places too much emphasis on the (relative) *unilateral* retentiveness and attractiveness of origins and destinations, whereas spatial structure is primarily about *bilateral* relations. Specifically, 2FE measure each origin's (destination's) *invariant* propensity (ability) to send (attract) migrants to (from) *any* destination (origin), exploiting only origin- and destination-specific information. The problem with 2FE is that path-specific effects are simply defined as the *sum* of origin- and destination-fixed effects, hence not allowing for interactions; that is, $\psi_{ij} = \upsilon_i + \omega_j$ and $\psi_{ik} = \upsilon_i + \omega_k$ $(k \neq j)$, such that $\upsilon_i$ is *common* to $\psi_{ij}$ and $\psi_{ik}$. This is an overly restrictive way of specifying spatial relations among locations (see also Pakko and Wall, 2001, in the context of bilateral trade), since it neglects the "idiosyncratic ties" (Mueser, 1989) between origins and destinations, not identified by location characteristics (whether observed or not), which may have an important bearing on the "resistance (or susceptibility) to migrate" between locations.

[4]Some researchers would include *lagged* migration *flows* as a proxy for network externalities (Anjomani and Hariri, 1992), or use both stocks and lagged flows to capture the effect of networks and "historical" migration choices, respectively (Bauer et al., 2007). Such an approach is not pursued here, as the resultant partial-adjustment model is generally known to be marred by several factors (Renshaw, 1974; Dunlevy and Gemery, 1977; Dunlevy, 1993). Besides, it is important to note that historical movements are largely reflective of the effects of spatial structure.

Second, since the 2FE enter the gravity model additively as "standalones" (varying intercepts), the model does not allow the unobserved effects to interact with the bilateral linkage variables, thus preventing the heterogeneous nature of the relations between origin-destinations pairs to manifest itself through varying slope parameters across migration paths. In effect, entering 2FE implies that the effects of distance, contiguity, and migrant stock, are assumed to be *independent* of any (not-controlled for) spatial-structural influences and, hence, to be *homogenous* over space.

To sum up, if unobserved path-specific factors are not accounted for, correlated specification errors (which are likely to be somehow related to observed linkage variables, such as distance and contiguity) sneak into the migration-cost function and end up in the error term, giving rise to potentially severe biases in the estimates of the model parameters.[5] On the other hand, the 3FE model controls for as much heterogeneity as possible (in the three dimensions of the gravity equation) and is, therefore, likely to return more reliable estimates (Baltagi, Egger, and Pfaffermayr, 2003). Though, whether the unobserved-effect interactions really matter remains an empirical question.

*Two Potential Criticisms of Fixed Effects*

Focusing on fixed-effects modeling may give rise to two potential criticisms. First, it could be argued that by relying on a huge number of bilateral fixed effects, one may run the risk of removing all the cross-sectional variation in the data, hence spuriously reducing the "signal-to-noise" ratio and making it hard to find any meaningful effects of the observables on the scale of migration even if they exist. However, in the present case, the path-specific fixed effects do not enter the 3FE gravity model as standalones (alongside the additive origin- and destination-specific effects), but are taken *in interaction with* each of the bilateral covariates and, hence, giving rise to varying slope parameters. Therefore, most of the cross-sectional variation in the sample population is exploited to identify the (unconditional mean) effects of the bilateral linkage variables.

Second, Congdon (2010, p. 757) has criticized the fixed-effects approach, contending that "it is not possible to model spatial autocorrelation ... if push and pull effects are treated as fixed effects." However, to the extent that correlated spatial effects *are* empirically relevant, it is anticipated that these will emerge (ex-post) as strong spatial similarities of the estimated path-specific parameters. The fact that the 3FE model does not require the specification of a particular distribution of the unobserved heterogeneity nor the imposition of a potentially too restrictive spatial structure on the data should rather be viewed as one of its attractive features. Or, as McMillen (2010, p. 135) put it, the results of models relying heavily on pre-imposed parametric structure should always be viewed with a great deal of skepticism, so that fixed effects are preferable to a spatial error model.

[5]Note that this problem was also recognized by Anderson and van Wincoop (2003, p. 180): "Errors can enter the model in many other ways..., about which the theory has little to say. In particular, it is possible that the trade cost function...is misspecified in that other factors than just distance and borders matter.... If this error term is correlated with **d** [distance] or **δ** [border dummy], our estimates will be biased. But this is a standard omitted variables problem that is not specific to the presence of multilateral resistance terms.... The simple fixed-effects estimator is not necessarily more robust to specification error". Although the problem of omitted-variables bias is not specific to the presence of multilateral resistance, the problem is there and, thus, cannot be simply brushed aside (even if the model is theoretically consistent). Recognition of this problem related to the unobserved effects in the bilateral dimension of the gravity equation invited Henderson and Millimet (2008) to use non-parametric methods, and Behrens, Ertur, and Koch (forthcoming) to apply spatial-econometric techniques.

## 3. THE GME ESTIMATOR

The introduction of parameter heterogeneity into the gravity model is informationally quite demanding and, therefore, creates a challenging estimation problem. Specifically, the 3FE gravity model allows three sets of dummies to interact with each of the bilateral linkage variables distance, contiguity, and migrant stock, besides the origin- and destination-fixed effects. As a result, the number of unknown fixed parameters to be estimated is far much greater than the number of observations, so the 3FE model is "ill-posed" (underdetermined). Whenever this occurs, classical estimators, such as OLS or PPML, cannot generally be used to estimate the parameters of interest. To overcome this apparent parameter-dimensionality problem, I use an (in)equality-constrained variant of the GME estimator, which builds on the standard GME estimator introduced by Golan et al. (1996).

### *Some Advantageous Properties of GME*

The GME estimator is a member of the class of information-theoretic estimators, which is based on the entropy-information measure introduced by Shannon (1948). The use of entropy-based methods has already a long history in the migration literature, ever since the seminal work of Wilson (1970) and subsequent papers by Plane (1982), Roy and Flood (1992), and others. However, in this paper the entropy approach is applied within a *regression* framework, which is, therefore, fundamentally different from conventional Maximum Entropy (ME) applications in earlier studies of migration—which are primarily concerned with "matrix balancing" (i.e., recovering, predicting, or updating matrices of bilateral migration flows) based on incomplete information.

Since the literature on entropy econometrics is too vast to review here, I just briefly summarize some of the desirable properties of GME that will be profitably exploited in estimating the 3FE gravity model.[6] First, GME is a "shrinkage-type" estimator, similar to Stein-like and empirical Bayesian estimators. In loose terms, this means that the estimation can be "improved," in terms of mean squared error (MSE), by incorporating external (nonsample) information (Shen and Perloff, 2001).[7] Second, shrinkage is typically aimed at regularizing "ill-posed" inference problems, where GME shrinking has proven to be suitable for handling underdetermined inverse problems and to perform well when collinearity is present. Third, unlike, say, Maximum Likelihood (ML), GME does not require strong distributional error assumptions, so GME is robust even if the errors are not normally distributed. Last, using GME makes it easy to impose linear, nonlinear, and inequality restrictions (e.g., sign restrictions) on the parameters consistent with economic-theoretic and/or other behavioral information.[8]

Of course, every estimation method has its *pros* and *cons*. Two potentially worrisome drawbacks of the GME estimator are related to the treatment of the multiplicative error

[6]Readers unfamiliar with GME are directed to Golan et al. (1996), Golan (2006), and Golan and Maasoumi (2008). Some good elementary expositions of GME can be found in Fraser (2000) and Campbell and Hill (2005).

[7]Of course, the possibility of beating traditional estimators on the MSE criterion does not mean that the GME estimator is unbiased. However, the biases are expected to be minimal (or nil) when the prior, nonsample information is theoretically or empirically consistent—i.e., if the "true" parameter values are contained in the pre-specified parameter space (see further below).

[8]Some researchers would criticize information-based (shrinkage) methods as relying on "subjective" prior information beyond the sample. It should be noted, though, that any modeling effort, also in a traditional econometric setting, involves subjective choices (e.g., functional form of the mean or variance function, error-term properties, etc.).

term in the presence of (incidental) zero-flow observations, and the estimation of the covariance matrix. These problems will be briefly discussed shortly.

*Formulation of GME Estimation Problem*

To implement GME, the estimation problem has to be converted into a constrained optimization problem, where the objective function consists of the *joint entropy* in Equation (9) below. The latter is to be maximized, subject to the data (and other) constraints. For simplicity, but without loss of generality, the path-specific slopes in $\boldsymbol{\beta}_{3,ij}$ (see Equation (8)) will be estimated *directly* (rather than $\boldsymbol{\beta}_3$ and $\mathbf{b}_{ij}$ separately), assuming that $E(\mathbf{b}_{ij}) = 0$, or $\boldsymbol{\beta}_3 = E(\boldsymbol{\beta}_{3,ij})$.

*Re-parameterization*

To arrive at GME estimation, it is required to re-parameterize the gravity model first, recasting the unknown parameters and errors in terms of unknown probabilities associated with some pre-specified sets of discrete support points. Let $\boldsymbol{\theta}$ denote the full set of unknown parameters, where the individual set for a given path (represented by a stacked column vector) is written as $\boldsymbol{\theta}_{ij} = [\beta_0\ \boldsymbol{\beta}_1\ \boldsymbol{\beta}_2\ \boldsymbol{\beta}_{3,ij}\ \boldsymbol{\psi}_{ij}\ \boldsymbol{\rho}_{ij}\ \boldsymbol{\zeta}_{ij}]'$, where $\boldsymbol{\beta}_1 = [\beta_{1,1}\ \beta_{1,2} \ldots \beta_{1,K}]'$ and $\boldsymbol{\beta}_2 = [\beta_{2,1}\ \beta_{2,2} \ldots \beta_{2,K}]'$, along with the intercept, are *common* to all migration paths, whereas $\boldsymbol{\beta}_{3,ij} = [\delta_{ij}\ \lambda_{ij}\ \eta_{ij}]'$ and $\boldsymbol{\psi}_{ij} = [\upsilon_i\ \ \omega_j]'$ are *varying* across migration paths (the elements $\boldsymbol{\rho}_{ij}$ and $\boldsymbol{\zeta}_{ij}$ contained in $\boldsymbol{\theta}_{ij}$ will be explained later).

Using a common set of three discrete support points for each unknown parameter $\theta_{ij}$ contained in $\boldsymbol{\theta}_{ij}$ (which may represent $\beta_{1,1}$, $\beta_{2,1}$, $\delta_{ij}$, $\upsilon_i$, and so on), given by the vector $\mathbf{z}_\theta = [z_{\theta,1}\ z_{\theta,2}\ z_{\theta,3}]'$, then, for each single $\theta_{ij}$, there exist unknown probabilities $p_{\theta_{ij}} \in [0, 1]$, with $p_{\theta_{ij},1} + p_{\theta_{ij},2} + p_{\theta_{ij},3} = 1$, such that

$$\theta_{ij} = z_{\theta,1}p_{\theta_{ij},1} + z_{\theta,2}p_{\theta_{ij},2} + z_{\theta,3}p_{\theta_{ij},3} = \begin{bmatrix} z_{\theta,1} & z_{\theta,2} & z_{\theta,3} \end{bmatrix} \begin{bmatrix} p_{\theta_{ij},1} \\ p_{\theta_{ij},2} \\ p_{\theta_{ij},3} \end{bmatrix} = \mathbf{z}'_\theta \mathbf{p}_{\theta_{ij}}.$$

Along similar lines, a common support vector for the unknown errors on the log scale, $\epsilon_{ij}$, is defined as $\mathbf{z}_\epsilon = [z_{\epsilon,1}\ z_{\epsilon,2}\ z_{\epsilon,3}]'$, so for each error term, there exist unknown probabilities $p_{\epsilon_{ij}} \in [0, 1]$, with $p_{\epsilon_{ij},1} + p_{\epsilon_{ij},2} + p_{\epsilon_{ij},3} = 1$, such that

$$\epsilon_{ij} = z_{\epsilon,1}p_{\epsilon_{ij},1} + z_{\epsilon,2}p_{\epsilon_{ij},2} + z_{\epsilon,3}p_{\epsilon_{ij},3} = \begin{bmatrix} z_{\epsilon,1} & z_{\epsilon,2} & z_{\epsilon,3} \end{bmatrix} \begin{bmatrix} p_{\epsilon_{ij},1} \\ p_{\epsilon_{ij},2} \\ p_{\epsilon_{ij},3} \end{bmatrix} = \mathbf{z}'_\epsilon \mathbf{p}_{\epsilon_{ij}}.$$

*Choice of Support Values*

The support vectors $\mathbf{z}_\theta$ and $\mathbf{z}_\epsilon$ are chosen to span the relevant parameter and error-term spaces. Because no *a priori* knowledge is generally available on the magnitudes and/or signs of (most of the) the unknown parameters in $\boldsymbol{\theta}_{ij}$, I specify a "conservative" (wide enough to be nonbinding) support range as $\mathbf{z}_\theta = [-100, 0, 100]'$, where the endpoints are symmetrically spaced around zero. Then, if the prior distribution is *uniform*, GME shrinks all parameter estimates towards the "prior mean" of zero—which, actually, corresponds to the null of "no effect." However, by using sufficiently wide supports along with a large enough sample population, the "posterior means" (the GME point estimates), $\hat{\boldsymbol{\theta}}_{ij}$, are likely to be dominated by the information contained in the data, so that the risk of distortions or biases (towards zero) is greatly reduced.

The support space for the unknown error terms, $\epsilon_{ij}$, is based on the so-called *three-sigma* rule (Pukelsheim, 1994), recommended by Golan et al. (1996). That is, the error support is specified as $\mathbf{z}_\epsilon = [-3\hat{\sigma}_m, 0, 3\hat{\sigma}_m]' = [-4.662, 0, 4.662]'$, where $\hat{\sigma}_m$ is the empirical standard deviation of the dependent variable in logs, $m_{ij} = \ln M_{ij}$ (see the summary statistics reported in Table 1 below). Given that $u_{ij} \equiv e^{\epsilon_{ij}}$, the implied support space for the multiplicative error, $u_{ij}$, is $\mathbf{z}_u = [0.0095, 1, 105.8]'$ with the lower bound of $u_{ij}$ being very close to zero, the "center" equal to 1, and the upper bound close to 100, hence allowing for *multiplicative heteroskedasticity* (of an unknown form) of the error term (see also Golan et al., 1996, p. 154).[9]

The selection of discrete support values for the errors, $\epsilon_{ij}$, goes not without a cost, however, in the sense that GME is not "zero-flow consistent" for *individual* observations. This problem arises from the simple fact that $\epsilon_{ij}$ is undefined at $u_{ij} \equiv e^{\epsilon_{ij}} = 0$, so that the inclusion of zero observations causes the GME optimization to be computationally infeasible (indefinite). (Note that this problem does not apply to the conditional mean, which is always strictly positive.) While being a potential drawback of GME, the presence of zero observations should not be too threatening *in practice*. First, the number of zero flows is typically very small (or zero flows may not be present at all) in studies of migration between sufficiently large regions (such as states, provinces, or counties) and wide time intervals (such as five- or ten-year observation periods). Second, adding an arbitrarily small value between zero and 1 (say, 0.1 or 0.01) to the zero observations is likely to have a negligible impact on the results, considering that (a) the 3FE model accounts for slope heterogeneity across migration paths, hence also covering the idiosyncratic nature of the zero-migration flows, and (b) the GME estimator (like PPML, but unlike gamma PML) places equal weights on observations.[10] With this in mind, it is hard to imagine that only a few minimally "over-predicted" *individual* zero counts would significantly alter the estimated *average* effects of the covariates included in the gravity model.[11]

*Constrained Optimization Problem*

All unknown parameters and errors of the 3FE gravity model can be estimated jointly by solving the following (primal) constrained optimization problem (for convenience, I use compact notation):

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \left\{ H(\mathbf{p}) = -\sum_{i \neq j}^{n} \sum_{j \neq i}^{n} \mathbf{p}'_{\theta_{ij}} \ln \mathbf{p}_{\theta_{ij}} - \sum_{i \neq j}^{n} \sum_{j \neq i}^{n} \mathbf{p}'_{\epsilon_{ij}} \ln \mathbf{p}_{\epsilon_{ij}} \right\}, \tag{9}$$

subject to

$$M_{ij} = e^{[(\mathbf{I} \otimes \mathbf{z}'_\theta)\mathbf{p}_{\varphi_{ij}}]' \mathbf{y}_{ij}} e^{\mathbf{z}'_\epsilon \mathbf{p}_{\epsilon_{ij}}}, \forall i, j (i \neq j), \tag{10}$$

[9]In unreported results, it was found that the estimates are insensitive to widened error support spaces. However, solving the GME optimization problem becomes infeasible for a lower error bound set at $1.0 \times 10^{-8}$.

[10]Running ahead of the empirical analysis presented below, where the issue of zero-migration flows is actually not of a concern (see the summary statistics in Table 1), it was found that adding a small value of, say, 0.01 to the observations for which $M_{ij} < 100$ (i.e., 62 cases out of a total of 992) had no discernable impact on the estimation results.

[11]One possibility for dealing with zero observations (assuming that there is something "special" about zero flows) is to estimate a model in the spirit of Heckman's sample-selection model (Greene, 1994), using GME in the second stage of the estimation process based on the truncated-at-zero sample. However, further investigation of this issue is outside the scope of the present paper.

$$\sum_{i=1}^{n} \mathbf{z}'_\theta \mathbf{p}_{\upsilon_i} = 0; \sum_{j=1}^{n} \mathbf{z}'_\theta \mathbf{p}_{\omega_j} = 0, \tag{11}$$

$$\mathbf{e}'\mathbf{p}_{\theta_{ij}} = 1;\ \mathbf{e}'\mathbf{p}_{\epsilon_{ij}} = 1, \forall i, j\, (i \neq j)\,, \tag{12}$$

$$\mathbf{z}'_\theta \mathbf{p}_{\delta_{ij}} \leq 0; \quad \mathbf{z}'_\theta \mathbf{p}_{\lambda_{ij}} \geq 0; \quad \mathbf{z}'_\theta \mathbf{p}_{\eta_{ij}} \geq 0, \forall i,\ j(i \neq j), \tag{13}$$

$$\mathbf{z}'_\theta \mathbf{p}_{\lambda_{ij}} = (\mathbf{z}'_\theta \mathbf{p}_{\rho_{\lambda,ij}})(\mathbf{z}'_\theta \mathbf{p}_{\delta_{ij}}) + \mathbf{z}'_\theta \mathbf{p}_{\zeta_{\lambda,ij}},\ \ \forall i, j(i \neq j), \tag{14}$$

$$\mathbf{z}'_\theta \mathbf{p}_{\eta_{ij}} = (\mathbf{z}'_\theta \mathbf{p}_{\rho_{\eta,ij}})(\mathbf{z}'_\theta \mathbf{p}_{\delta_{ij}}) + \mathbf{z}'_\theta \mathbf{p}_{\zeta_{\eta,ij}},\ \ \forall i, j(i \neq j), \tag{15}$$

where $\boldsymbol{\varphi}_{ij} = [\beta_0\ \boldsymbol{\beta}_1\ \boldsymbol{\beta}_2\ \boldsymbol{\beta}_{3,ij}\ \boldsymbol{\psi}_{ij}]'$ is a sub-vector of $\boldsymbol{\theta}_{ij} = [\boldsymbol{\varphi}_{ij}\ \boldsymbol{\rho}_{ij}\ \boldsymbol{\zeta}_{ij}]'$, $\mathbf{p}_{\varphi_{ij}}$ are the corresponding unknown probabilities, $\mathbf{y}_{ij} = [1\,\mathbf{x}_i\ \mathbf{x}_j\ \mathbf{z}_{ij}\ 1\ 1]'$, such that $M_{ij} = e^{\varphi'_{ij}\mathbf{y}_{ij}}e^{\epsilon_{ij}}$, and $\mathbf{e}$ is a 3-by-1 summation vector consisting of ones. The identity matrix $\mathbf{I}$ in Equation (10) has dimension $1 + 2K + 3 + 2$, and $\otimes$ is the Kronecker product.

GME maximizes the entropy objective $H(\mathbf{p})$ in Equation (9), subject to the *data-consistency* constraints in Equation (10), which basically represent the (re-parameterized) 3FE gravity model in Equation (5). Specifically, GME minimizes the joint entropy distance between the data and the state of complete uncertainty (uniform distribution) by assigning equal weights to the parameter (signal) and error (noise) entropies—i.e., by putting equal weights to prediction (bias) and precision (variance). The constraints in Equation (11) are aimed at preserving zero mean values of the origin- and destination-fixed effects, and the constraints in Equation (12) are *normalization* (adding-up) constraints, to ensure that GME returns proper probabilities (which, of course, should always sum up to one).

The constraints in Equations (13) and (14)–(15), which are important for improving the identification strength of GME estimation, require some additional explanation. These constraints are aimed at "regularizing" the ill-posedeness of the inverse problem through introducing *extraneous* (nonsample) information. Although ill-posed problems can be regularized in a variety of ways, a simple regularization is effected here by imposing a set of restrictions on the bilateral slope parameters, based on some "reasonable" assumptions about the solution space (as there is no received theory to apply). In general, solutions to regularized problems tend to be more stable (or efficient), whilst at the same time providing meaningful parameter estimates consistent with "prior beliefs" (Golan et al., 1996, p. 129; see also Campbell and Hill, 2005, 2006).

First, the *inequality restrictions* in Equation (13) are imposed on the path-specific effects of the bilateral linkage variables to bound the domain of their unconditional distributions (i.e., pulling the "posterior means" away from the common zero "prior mean" in only one direction), to ensure the "right signs," such that

$$\delta_{ij} \leq 0;\ \lambda_{ij} \geq 0;\ \eta_{ij} \geq 0,\ \forall i, j. \tag{16}$$

Here, the first inequality restriction implies that distance can only have a *negative* (or zero) ceteris-paribus effect on migration, by virtue of the expected distance decay. Conversely, the other two inequality restrictions ensure a *positive* (or zero) effect of both contiguity and the migrant stock. That is, migration counts between contiguous locations are likely to be greater than would be predicted on the basis of centroid-based distances, while the presence of a larger migrant stock is likely to indicate a greater propensity of migrants to move from $i$ to $j$, both in the past and in the present, than predicted on the basis of the covariates included in the gravity model.

Second, the *cross-parameter restrictions* in Equations (14)–(15) are mainly intended to ensure some degree of "internal consistency" among the path-specific bilateral slope parameters, such that

$$\lambda_{ij} = \rho_{\lambda,ij}\delta_{ij} + \zeta_{\lambda,ij}, \forall i, j, \tag{17}$$

$$\eta_{ij} = \rho_{\eta,ij}\delta_{ij} + \zeta_{\eta,ij}\,, \ \forall i, j, \tag{18}$$

where $\boldsymbol{\rho}_{ij} = [\rho_{\lambda,\,ij}\ \rho_{\eta,ij}]'$ are unknown parameters, re-parameterized as $\mathbf{z}'_\theta \mathbf{p}_{\rho_{\lambda,ij}}$ and $\mathbf{z}'_\theta \mathbf{p}_{\rho_{\eta,ij}}$, and $\boldsymbol{\zeta}_{ij} = [\zeta_{\lambda,ij}\ \zeta_{\eta,ij}]'$ represent *noise* terms, re-parameterized as $\mathbf{z}'_\theta \mathbf{p}_{\zeta_{\lambda,ij}}$ and $\mathbf{z}'_\theta \mathbf{p}_{\zeta_{\eta,ij}}$, allowing for *imperfections* to enter the mutual relations among the members *within* each parameter set $\{\delta_{ij}, \lambda_{ij}, \eta_{ij}\}, \forall i, j$. (Note that the parameters $\boldsymbol{\rho}_{ij}$ and $\boldsymbol{\zeta}_{ij}$ are included, together with $\boldsymbol{\varphi}_{ij}$, as additional elements of set $\boldsymbol{\theta}_{ij}$.) In other words, the path-specific effects are looked at as a set of *logically interrelated* effects, based on the intuition that they are likely to be *jointly* reflective of the influence exerted by the *same* underlying spatial structure. Accordingly, the introduction of these cross-parameter restrictions basically amounts to affirming that $\delta_{ij}$ contains (at least some) information about $\lambda_{ij}$ and/or $\eta_{ij}$, and vice versa. (Note that the restrictions in Equations (17)–(18) imply that $\eta_{ij} = (\rho_{\eta,ij}/\rho_{\lambda,ij})(\lambda_{ij} - \zeta_{\lambda,ij}) + \zeta_{\eta,ij}$, etc.) This "borrowing of strength" (pooling of information) from the expected interrelatedness of the path-specific bilateral effects, rather than treating them independently, is likely to provide improved parameter estimates. Nonetheless, it is imperative—in view of the bias-efficiency tradeoff involved—to specify the cross-parameter restrictions *in the most flexible way* (by allowing for varying slopes and adding noise terms) to minimize the risk of "over-regularization," which may cause potentially severe distortions or biases in the estimates.

After numerically solving the restricted GME optimization problem in Equations (9)–(15), the point estimates of the parameters of interest and errors can be recovered as $\hat{\theta}_{ij} = z_{\theta,1}\hat{p}_{\theta_{ij},1} + z_{\theta,2}\hat{p}_{\theta_{ij},2} + z_{\theta,3}\hat{p}_{\theta_{ij},3}$ and $\hat{\epsilon}_{ij} = z_{\epsilon,1}\hat{p}_{\epsilon_{ij},1} + z_{\epsilon,2}\hat{p}_{\epsilon_{ij},2} + z_{\epsilon,3}\hat{p}_{\epsilon_{ij},3}$, respectively. Even though the optimization problem is clearly ill-posed in the probability space, it is a well-known result that GME picks out a *unique* solution, since the number of unknowns is equal to the number of first-order conditions (Lagrange multipliers), hence making it possible to recover all point estimates.

*Approximate Asymptotic Standard Errors*

To give an idea of the statistical significance of the GME estimates, asymptotic standard errors (ASEs) are calculated for the subset of estimated trend-surface parameters $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_0\ \hat{\boldsymbol{\beta}}_1\ \hat{\boldsymbol{\beta}}_2\ \hat{\boldsymbol{\beta}}_3]'$, where the unobserved heterogeneity components [2] and [3] in Equation (5) are treated together as a "compound error term." The resulting ASEs are only approximate, for reasons discussed below.

Under some regularity conditions (summarized in Golan, Perloff, and Shen, 2001, Appendix, p. 550), the GME estimator is consistent and asymptotically normal, from which it follows that $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. Then, the ASEs can be derived as follows. Let $K$ be the number of elements in the subset $\hat{\boldsymbol{\beta}}$, then the $K$-by-$K$ variance matrix can be estimated as $\hat{\boldsymbol{\Omega}} = [\mathbf{A}'\hat{\mathbf{V}}^{-1}\mathbf{A}]^{-1}$, where $\hat{\mathbf{V}}$ is an $N$-by-$N$ diagonal matrix with elements $\hat{\sigma}^2_{ij} = \hat{\sigma}^2_{\xi_{ij}} + \hat{\sigma}^2_{\epsilon}$,

$$\hat{\mathbf{V}} = \begin{bmatrix} \hat{\sigma}^2_{12} & 0 & \cdots & 0 \\ 0 & \hat{\sigma}^2_{13} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}^2_{n,n-1} \end{bmatrix},$$

where

$$\hat{\sigma}^2_{\xi_{ij}} = \hat{\sigma}^2_{\upsilon} + \hat{\sigma}^2_{\omega} + \hat{\sigma}^2_{\tau} d^2_{ij} + \hat{\sigma}^2_{\kappa} C_{ij} + \hat{\sigma}^2_{\mu} s^2_{ij} + 2\hat{\sigma}_{\tau,\kappa} d_{ij} C_{ij} + 2\hat{\sigma}_{\tau,\mu} d_{ij} s_{ij} + 2\hat{\sigma}_{\kappa,\mu} C_{ij} s_{ij},$$

and $\mathbf{A} = \partial(\mathbf{x}_{ij}\hat{\boldsymbol{\beta}})/\partial\hat{\boldsymbol{\beta}}$ is the $N$-by-$K$ gradient matrix of the linear trend predictor. The variances $\hat{\sigma}^2_{ij}$ on the diagonal of matrix $\hat{\mathbf{V}}$ are different across observations by virtue of the variations in the bilateral covariates $d_{ij}$, $C_{ij}$, and $s_{ij}$. The covariances $\hat{\sigma}_{\tau,\kappa}$, $\hat{\sigma}_{\tau,\mu}$, and $\hat{\sigma}_{\kappa,\mu}$ reflect the "internal-consistency" relations among the bilateral effects along each migration path, which have been ensured by the cross-parameter restrictions in Equations (17)–(18). On the other hand, it is assumed that $\upsilon_i$ and $\omega_j$ vary only in the unilateral dimensions and are, therefore, uncorrelated with the bilateral effects. Finally, the ASEs are calculated by simply taking the square roots of the diagonal elements in matrix $\hat{\boldsymbol{\Omega}}$.

A legitimate concern can be raised about the validity of assessing the statistical significance of the GME-estimated parameters, as some of the regularity conditions under which GME is consistent and asymptotically normal may not be satisfied in reality. Specifically, since the compound error components $\xi_{ij}$ and $\xi_{rs}$ ($r \neq i$ or $s \neq j$) are unlikely to be spatially independent, failure to correct for the spatial correlation is likely to impart a downward bias on the covariance estimator (though, whether the practical implications will ultimately be big or small is difficult to predict). Therefore, inferences about the statistical significance of the GME estimates on the basis of the "heteroskedasticity-robust" variances in matrix $\hat{\mathbf{V}}$ are only *approximately* valid at best. While this is evidently an important issue for statistical inference, it is judged to be less severe than obtaining meaningless estimates with implausible magnitudes and/or counter-intuitive signs (having no practical significance). Moreover, the presence of spatial dependence in the sample population is an econometric issue that is not specific to GME; that is, the same problem naturally arises in the case of conventional OLS or PPML as well (where migration flows are typically assumed to be spatially independent).[12]

## 4. EMPIRICAL CASE-STUDY ANALYSIS

To illustrate its empirical potentials, the GME estimation of the 3FE gravity model is applied for a case-study analysis of internal migration in Mexico, using cross-sectional data on interstate migration for the five-year observation period 1995–2000. The data set is felt to be sufficiently representative of the literature on aggregate place-to-place migrations within a country. A short overview of the data sources is given in Appendix A of the paper. To help the reader, I also provide a base map of Mexico's states in Appendix B.

[12]Estimation of asymptotic (co)variances using procedures that account for spatial dependence is difficult (Wooldridge, 2002, p. 134), particularly when using a single cross section (Driscoll and Kraay, 1998, p. 559), and is beyond the scope of this paper. Note, however, that procedures to obtain SHAC standard errors (e.g., Kim and Sun, 2011) would also require making possibly strong assumptions about the (unknown) form of the spatial dependence, hence producing results that are also only approximately valid (see also Pinkse and Slade, 2010, p. 107, for some pertinent remarks on this issue). In addition, deriving the asymptotic properties of the (in)equality restricted GME estimator may actually not be possible at all. Finally, reliance on bootstrapping techniques is hardly an option in the case of a correlated sample population.

*Variables Included in Empirical Gravity Model*

The dependent variable in the empirical gravity model, $M_{ij}$, represents the (gross) interstate migration flows in Mexico over the observation period 1995–2000. A "migrant" from origin $i$ to destination $j$ is defined as any individual changing his/her *place of residence* from $i$ to $j$—i.e., any individual living in $i$ five years prior to the 2000 census and living in $j$ at the time of the 2000 census (no matter where this individual was born). Thus, those individuals that migrated from $i$ to $j$ at some point of time during the observation period but returned to $i$, re-migrated to $k \neq j$, or died before the 2000 census are not counted as migrants from $i$ to $j$. For the $n = 32$ states in Mexico (including the Distrito Federal), this yields a total of $N = n(n - 1) = 992$ potential migration flows.

The empirical model includes a conventional core set of social, economic, and demographic fundamentals, which previous studies have also found to be key forces driving bilateral migrations. The (unilateral) "push" and "pull" factors at origins and destinations included in the empirical model are:

| | |
|---|---|
| $P_i$ ($P_j$) | population size in origin state (destination state); |
| $HDI_i$ ($HDI_j$) | Human Development Index (HDI) in origin state (destination state); |
| $UR_i$ ($UR_j$) | unemployment rate in origin state (destination state); |
| $MAN_i$ ($MAN_j$) | employment share of manufacturing in origin state (destination state); |
| $PD_i$ ($PD_j$) | population density in origin state (destination state); |
| $B_i$ ($B_j$) | U.S. border-state dummy for origin state (destination state). |

The inclusion of population sizes, $P$, in origin and destination is common practice in gravity modeling of migration, in which these variables primarily act as "scaling factors." The variable *HDI* is a summary measure containing information on the income level (GDP per capita), the health situation (life-expectancy rate), and the level of educational attainment (literacy and school-enrollment rates) in Mexico's states. Thus, *HDI* is a much richer indicator of regional differences in "living standards" than, say, GDP per capita (which is also sometimes used as a proxy for wages). The variable *UR* is included as an indicator of labor-market conditions. The variable *MAN* is used as a measure of job opportunities not yet sufficiently captured by *UR*. The variable *PD* is intended as a proxy for the degree of urbanization or, alternatively, as a "catch-all" for job opportunities in the formal and informal service sectors, the supply of a wide range of public and private amenities, etc. Finally, in accordance with the specifications in Greenwood and Ladman (1978) and Greenwood, Ladman, and Siegel (1981), the empirical model also includes U.S. border-state origin and destination dummies, $B$. While these dummy variables may capture a multitude of state-specific unobserved effects, they are primarily intended to account for the *maquiladora* effect exerted by these northern border states, which may cause more migrants to be attracted and more residents to be retained.[13] Moreover, these indicator variables may also partly capture the potential "gateway function" of the northern border states, where prospective migrants to the U.S. may consider moving to the border region first in preparation for subsequent transmigration, either legally or illegally, to the U.S. (e.g., Zabin and Hughes, 1995; OECD, 2004).

[13]Even though Mexico has witnessed a growing decentralization of *maquiladora* jobs in the 1980s and 1990s, scattering from the U.S. border states into central Mexico, they are still heavily concentrated along the U.S. border in the sample period. Using data from INEGI, Jones (2001) reported that, in 1998, the U.S. border states were still hosting about 84 percent of the *maquiladora* jobs in Mexico.

The (bilateral) "linkage" variables included in the empirical model are:

$D_{ij}$ highway distance in kilometers between the states' capital cities;
$C_{ij}$ contiguity dummy indicating states sharing a common internal border;
$S_{ij}$ net accumulated stock of previous migrants to state $j$ coming from state $i$ (at the beginning of the five-year observation period), expressed as a proportion of the population in state $i$.

Distance, $D_{ij}$, is typically used as a proxy for "migration costs," including direct or pecuniary costs (expenses for transportation and lodging), opportunity costs (earnings forgone), psychic costs, and information costs (e.g., related to job- and/or housing-search efforts), which are assumed to increase with distance and to reduce the attraction of migration. The contiguity dummy, $C_{ij}$, is used to account for short-distance (perhaps just residential) moves across the shared border of neighboring states, which may not adequately be explained by the distance between the states' capital cities (which may be located far from each other). The migrant stock, $S_{ij}$, is defined as the (net) accumulated total number of *previous in-migrants* to $j$ born in $i$ and living in $j$ at the beginning of the observation period (1995), $I_{ij}$, expressed *relative* to the (1995) population of state $i$; that is, $S_{ij} = I_{ij}/P_i$. The migrant-stock variable has been widely used in the migration literature as a proxy for the linkages between prospective out-migrants in origin $i$ and their "family and friends" already settled in destination $j$, which are expected to facilitate further migration from $i$ to $j$ due to positive *network externalities* (through reducing the migration costs for the prospective "follower" migrants).[14]

Table 1 provides the definitions of the variables included in the empirical gravity model, along with basic descriptive statistics. From this table it is apparent that interstate migration flows in Mexico display a wide variation in size, ranging between a minimum number of 15 migrants (from Aguascalientes to Campeche) and a maximum number of about 450,000 migrants (from the Distrito Federal to the neighboring state of México). The distribution of $M_{ij}$ exhibits strong positive skewness, with a standard deviation of about 4.5 times the mean, suggesting the presence of strong (non-Poisson) "overdispersion." On the other hand, the distribution of the log-transformed migration flows within the sample population, $m_{ij}$, is fairly close to normal (with skewness = 0.344 and kurtosis = 2.773).

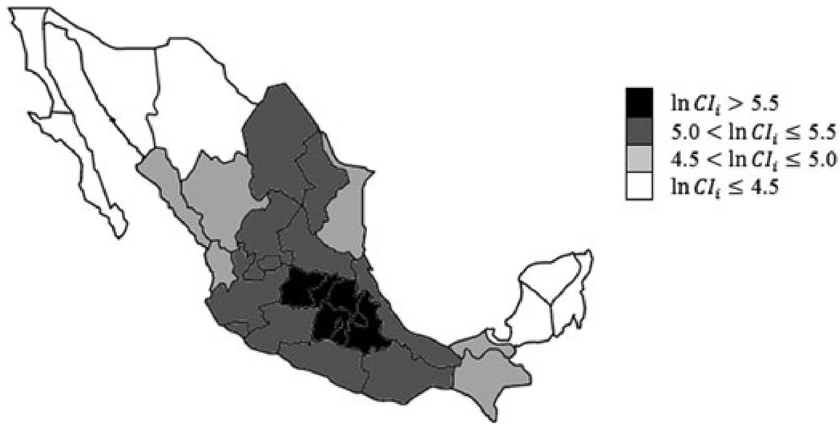
Before going into the discussion of the estimation results, it is worth taking a closer look at the hierarchy of the spatial structure in Mexico to get a better appreciation of some of the findings that will be reported below. For this purpose, I construct origin-specific *centrality indices*, defined here as $CI_i = \sum_{j=1}^{n}(P_j/D_{ij})$, where the summation is over all states, including the reference state $i$, in order to avoid ending up with "donut holes." The intra-state distance for the reference state $i$ is calculated as $D_{ii} = .67\sqrt{Area_i/\pi}$, where $Area_i$ is state $i$'s surface area in square kilometers (Head and Mayer, 2002, p. 11). The pattern depicted in Figure 1 suggests the presence of strong spatial structure, or "spatial dominance" (Akwawua and Pooler, 2001). (Since the distribution of the centrality indices is heavily (positively) skewed, their log values have been used in drawing the map in Figure 1.) Accordingly, Mexico City (or the Distrito Federal) and its surrounding states, hosting the main agglomerations of populations and economic activities, so dominate the

[14]The inclusion of distance along with migrant stock has sometimes been questioned, based on the ground that current and previous migration flows are both primarily distance-driven (Greenwood, 1997, p. 666). Some researchers would stretch this idea (quite obstinately) to the point that distances could also serve as a proxy for migrant-stock linkages, so that the two variables can be used interchangeably in the gravity model. However, given that distance is a poor indicator of the complex spatial structure of migration, it is hard to imagine that distances could possibly be a reasonable proxy for *network* linkages.

TABLE 1: Definition of Variables and Summary Statistics

| Variable | Definition | Mean | St. Dev. | Min. | 1st Qrt. | Median | 3rd Qrt. | Max. |
|---|---|---|---|---|---|---|---|---|
| Dependent variable | | | | | | | | |
| $M_{ij}$ | Number of migrants from $i$ to $j$ (1995–2000) | 3,614 | 16,290 | 15 | 294 | 821 | 2,562 | 448,546 |
| $m_{ij}$ | Natural log of $M_{ij}$ | 6.831 | 1.554 | 2.708 | 5.682 | 6.710 | 7.849 | 13.014 |
| Unilateral (origin and destination) push and pull factors | | | | | | | | |
| $P_i(P_j)$ | Population (1995) | 2,847,513 | 2,477,619 | 375,450 | 1,314,575 | 2,141,955 | 3,563,653 | 11,704,920 |
| $UR_i(UR_j)$ | Unemployment rate (avg. of 1990 and 2000) | 0.019 | 0.004 | 0.011 | 0.016 | 0.020 | 0.022 | 0.027 |
| $HDI_i(HDI_j)$ | Human Development Index (avg. of 1990 and 2000) | 0.685 | 0.114 | 0.421 | 0.610 | 0.708 | 0.769 | 0.864 |
| $MAN_i(MAN_j)$ | Employment share of manufacturing (1995) | 0.229 | 0.084 | 0.053 | 0.169 | 0.243 | 0.288 | 0.375 |
| $PD_i(PD_j)$ | Population density, per square kilometer (1995) | 259 | 986 | 5 | 24 | 47 | 104 | 5,717 |
| $B_i(B_j)$ | Dummy = 1 for U.S.-border state, and 0 otherwise | 0.219 | 0.414 | 0 | | | | 1 |
| Bilateral (origin-destination) linkage factors | | | | | | | | |
| $C_{ij}$ | Dummy = 1 for contiguous states, and 0 otherwise | 0.143 | 0.350 | 0 | | | | 1 |
| $I_{ij}$ | Absolute migrant stock (1995) | 16,014 | 93,057 | 49 | 735 | 2,204 | 9,614 | 2,761,510 |
| $S_{ij}$ | Relative migrant stock (1995) | 0.0053 | 0.0159 | 0.0001 | 0.0004 | 0.0011 | 0.0036 | 0.3255 |
| $D_{ij}$ | Interstate distance, in kilometers | 1,370 | 1,077 | 33 | 601 | 1,056 | 1,768 | 5,961 |

*Notes*: The number of observations (bilateral migration flows) is $N = n(n-1) = 32 \times (32-1) = 992$. All covariates (except dummies) enter log-transformed in the empirical model, and are measured at the beginning of the five-year observation period 1995–2000. Data for some of the state characteristics are only available for 1990 and 2000, in which case the averages of 1990 and 2000 have been used (as indicated in the table). The relative migrant stock (rate), $S_{ij}$, is defined as the absolute size of the stock of previous migrants (number), $I_{ij}$, divided by the 1995 population of state $i$ (number), $P_i$. The U.S.-border states have been identified as in Raymer and Rogers (2007), including Baja California Norte, Baja California Sur, Sonora, Chihuahua, Coahuila, Nuevo León, and Tamaulipas.




*Notes*: The distribution of the (log) centrality index shown in the map indicates strong spatial structure (or spatial dominance). The Moran's $I$ statistic, based on the first-order contiguity matrix (indicating global spatial autocorrelation), is estimated at 0.557 ($p < 0.001$).

FIGURE 1: Spatial Distribution of Centrality in Mexico.

spatial structure that they must have an effect on the conditional distribution of the distances over which the bilateral migration flows are being observed (Tiefelsdorf, 2003).

### *Estimation Results for the Gravity Equation*

The aim of the discussion of the estimation results is to address the following key questions. How sensitive are the parameter estimates to alternative specifications of the empirical gravity model? What are the empirical implications of imposing bounds and other restrictions on the path-specific bilateral slope parameters? To what extent can the heteroskedasticity in the migration flows be set within the broader context of the correlated effects of the underlying spatial structure (through bilateral slope-parameter heterogeneity), inducing extra variability of the conditionally expected number of migrants?

#### *Sensitivity of Parameter Estimates to Model Specification*

Table 2 analyzes the sensitivity of the parameter estimates to the choice of model specification. Here, three alternatives will be considered: (a) the traditional model (with observables only); (b) the 2FE model (with origin- and destination-specific effects); and (c) the 3FE model (with origin-, destination-, and path-specific effects). The first two models have been estimated using OLS, PPML, and GME; the third model has been estimated using only GME (for obvious reasons).

Columns 1–2 show the parameter estimates for the traditional model obtained using OLS (in logs) and PPML (in levels)—called here simply T-OLS and T-PPML, respectively. (The estimates of T-GME (in levels) have not been reported in Table 2, since they are identical to those of T-OLS (in logs) and, therefore, do not provide any additional useful information.) Columns 3–5 present the parameter estimates for the 2FE model obtained using OLS, PPML, and GME—called here simply 2FE-OLS, 2FE-PPML, and 2FE-GME, respectively. (The estimates of 2FE-GME (in levels) have been reported in Table 2 (Column 5), because they nicely illustrate the "image-recovery" feature of GME (Paris and Howitt, 1998, p. 132), making it possible to provide an informative picture of the

TABLE 2: Results for Gravity Equations—Sensitivity to Model Specifications

| | Traditional Model | | 2FE Model | | | 3FE Model | | |
|---|---|---|---|---|---|---|---|---|
| | OLS (1) | PPML (2) | OLS (3) | PPML (4) | GME (5) | UGME (6) | R1GME (7) | R2GME (8) |
| Origin | | | | | | | | |
| Log population [ln $P_i$] | 0.952*** | 1.230*** | | | 0.968*** | 0.701*** | 0.762*** | 0.765*** |
| | (0.023) | (0.066) | | | (0.025) | (0.059) | (0.039) | (0.026) |
| Log unemployment rate [ln $UR_i$] | –0.628*** | –0.229 | | | –0.650*** | 0.238 | 0.241* | 0.199** |
| | (0.084) | (0.190) | | | (0.080) | (0.193) | (0.128) | (0.083) |
| Log Human Development Index [ln $HDI_i$] | 0.068 | 0.523** | | | 0.115 | –0.003 | 0.196 | 0.221* |
| | (0.100) | (0.206) | | | (0.119) | (0.284) | (0.189) | (0.124) |
| Log employment share of manufacturing [ln $MAN_i$] | –0.361*** | –0.144 | | | –0.287*** | –0.125 | –0.321*** | –0.323*** |
| | (0.043) | (0.098) | | | (0.045) | (0.107) | (0.071) | (0.047) |
| Log population density [ln $PD_i$] | 0.178*** | –0.004 | | | 0.124*** | 0.198*** | 0.141*** | 0.147*** |
| | (0.019) | (0.033) | | | (0.019) | (0.044) | (0.030) | (0.020) |
| U.S. border-state dummy [$B_i$] | 0.486*** | 0.023 | | | 0.365*** | 0.022 | –0.181* | –0.199*** |
| | (0.052) | (0.114) | | | (0.065) | (0.155) | (0.103) | (0.070) |
| Destination | | | | | | | | |
| Log population [ln $P_j$] | 0.116*** | 0.137** | | | 0.134*** | 0.484*** | 0.438*** | 0.450*** |
| | (0.027) | (0.059) | | | (0.029) | (0.069) | (0.046) | (0.030) |
| Log unemployment rate [ln $UR_j$] | –0.091 | –0.464** | | | –0.228** | 0.024 | –0.250* | –0.225** |
| | (0.079) | (0.204) | | | (0.083) | (0.206) | (0.137) | (0.088) |
| Log Human Development Index [ln $HDI_j$] | –0.356*** | –0.429* | | | –0.254** | 0.306 | 0.746*** | 0.794*** |
| | (0.114) | (0.238) | | | (0.124) | (0.305) | (0.201) | (0.131) |
| Log employment share of manufacturing [ln $MAN_j$] | 0.052 | –0.307** | | | 0.064 | 0.011 | –0.100 | –0.073 |
| | (0.046) | (0.125) | | | (0.045) | (0.109) | (0.072) | (0.048) |
| Log population density [ln $PD_j$] | –0.018 | 0.059 | | | –0.083 | 0.142*** | 0.049 | 0.046** |
| | (0.019) | (0.053) | | | (0.019) | (0.045) | (0.030) | (0.021) |
| U.S. border-state dummy [$B_j$] | 0.078 | 0.439** | | | –0.102 | 0.382** | 0.471*** | 0.448*** |
| | (0.069) | (0.179) | | | (0.069) | (0.163) | (0.109) | (0.074) |

Continued

TABLE 2: Continued

| | Traditional Model | | 2FE Model | | | 3FE Model | | |
|---|---|---|---|---|---|---|---|---|
| | OLS (1) | PPML (2) | OLS (3) | PPML (4) | GME (5) | UGME (6) | R1GME (7) | R2GME (8) |
| Origin–Destination | | | | | | | | |
| Log migrant stock [ln $S_{ij}$] | 0.781$^{***}$ | 0.711$^{***}$ | 0.903$^{***}$ | 0.870$^{***}$ | 0.904$^{***}$ | 0.019 | 0.147$^{***}$ | 0.104$^{***}$ |
| | (0.021) | (0.051) | (0.019) | (0.027) | (0.019) | (0.043) | (0.029) | (0.020) |
| Log distance [ln $D_{ij}$] | 0.070$^{**}$ | –0.059 | 0.063$^{*}$ | –0.102$^{**}$ | 0.063 | –0.006 | –0.148$^{***}$ | –0.204$^{***}$ |
| | (0.035) | (0.072) | (0.035) | (0.042) | (0.037) | (0.081) | (0.056) | (0.039) |
| Contiguity dummy [$C_{ij}$] | 0.250$^{***}$ | 0.295$^{**}$ | –0.043 | –0.050 | –0.043 | 0.027 | 0.129 | 0.092 |
| | (0.058) | (0.119) | (0.039) | (0.056) | (0.042) | (0.134) | (0.092) | (0.064) |
| Origin- and destination-specific effects | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Path-specific effects (slope-parameter heterogeneity) | No | No | No | No | No | Yes | Yes | Yes |
| Sign restrictions | No | No | No | No | No | No | Yes | Yes |
| Cross-parameter (consistency) restrictions | No | No | No | No | No | No | No | Yes |
| Modified Park test p-value | 0.104 | | 0.176 | | 0.192 | 0.000 | 0.076 | 0.797 |
| GNR/WLS test p-value | | 0.313 | | 0.000 | | | | |
| Ramsey's (2df) RESET test p-value | 0.003 | | 0.936 | | | | | |
| Pregibon's (2df) LINK test p-value | | 0.000 | | 0.000 | 0.907 | 0.999 | 0.998 | 0.997 |

*Notes*: The number of observations is unvaryingly equal to 992. OLS and PPML estimation of the traditional and 2FE models was implemented using STATA; GME estimation of the models was implemented using GAMS. The GME estimates are for the parameter support $\mathbf{z}_\theta = [-100, 0, 100]'$ and error support (on the log scale) $\mathbf{z}_\epsilon = [-4.662, 0, 4.662]'$, or (on the raw or level scale) $\mathbf{z}_u \approx [0.01, 1, 100]'$. Heteroskedasticity-robust standard errors are given in parentheses. The GME estimates of the (bilateral) origin-destination effects are the unconditional (marginal) expectations of the path-specific effects. The modified Park test (Manning and Mullahy, 2001) and the GNR/WLS test (Santos Silva and Tenreyo, 2006) are for checking whether the standard deviation or the variance is proportional to the mean, respectively. Duan's (1983) smearing estimator was used to retransform the log-predictions in the case of OLS. Pregibon's LINK test (Pregibon, 1980) is an extension of the standard RESET test for linear models (Ramsey, 1969). The estimated intercepts have not been reported to save space. The T-GME estimates have not been reported either, since they provide no additional useful information (they are nearly identical to those of T-OLS).

$^{***}$Significant at 1 percent. $^{**}$Significant at 5 percent. $^{*}$Significant at 10 percent.

push and pull effects of the origin ($\hat{\boldsymbol{\beta}}_1$) and destination ($\hat{\boldsymbol{\beta}}_2$) characteristics. In contrast, 2FE-OLS and 2FE-PPML can only identify the parameters of the bilateral covariates ($\hat{\boldsymbol{\beta}}_3$), due to collinearity problems.) Finally, Columns 6–8 show the parameter estimates for the 3FE model obtained using three variants of GME: (a) unrestricted GME, with *no* restrictions placed on the path-specific slope parameters of the bilateral linkage variables (3FE-UGME); (b) restricted GME with the *sign* restrictions in Equation (16) included (3FE-R1GME); and (c) restricted GME with the *sign* restrictions in Equation (16) and *cross* restrictions in Equations (17)–(18) included (3FE-R2GME).

A comparison of the results in Table 2 reveals that the parameter estimates are strikingly different across the alternative model specifications, not only in magnitude but also in kind. Overall, the traditional and 2FE models perform poorly, returning several implausible or counter-intuitive coefficients with excessively large (or small) absolute sizes and/or unexpected signs. Another point to notice is that nothing can be gained from using GME in estimating the parameters of the traditional and 2FE models, as the (unreported) T-GME and (reported) 2FE-GME coefficients on the bilateral covariates are almost identical to those of T-OLS and 2FE-OLS, respectively. Focusing next on the results of the proposed 3FE model obtained using the three different variants of GME, it can be seen that the R2GME estimates (Column 8) are more stable, having appreciably smaller ASEs, not only relative to UGME (Column 6) but also relative to R1GME (Column 7). This finding indicates that using information beyond the data through introducing the regularizing sign and cross restrictions on the individual parameters of the linkage variables has been effective in reducing empirical risk.[15] On the other hand, using UGME implies that about one-third of the estimated distance-decay rates (ranging between $-0.161$ and $0.544$) have the "wrong" (positive) sign, making it difficult to attach any meaningful interpretation to the *average* of the distance effects. Using R1GME, in turn, the absolute sizes of the coefficients on distance and migrant stock look suspiciously similar, being just "mirror images" of one another (being simply pulled equally far away in the opposite direction from the prior zero mean). In contrast, the corresponding R2GME coefficients stray farther apart (covering a wider range), assigning somewhat greater importance (unconditionally) to the effect of distance and a weaker effect of the migrant stock (networks). (Note that the coefficients on the origin and destination characteristics change only little in moving from R1GME to R2GME.) In sum, even a cursory look at the estimation results across the columns in Table 2 instantly reveals that the R2GME estimates for the 3FE model are far more plausible and intuitively interpretable than their traditional and 2FE counterparts. Accordingly, the discussion that follows will be primarily focused on the 3FE-R2GME estimates (Column 8) and contrasting them with the traditional and 2FE estimates obtained using OLS and PPML (Columns 1–4) for some key explanatory variables.

To begin with the (constant) 3FE-R2GME coefficients on the origin-push and destination-pull characteristics ($\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$), it can be seen that most of them turn out to be statistically significant (at conventional levels) and to have the expected signs.[16]

[15]It is important to note, however, that it is difficult to explore the size of the potential biases that could have arisen by imposing the cross-parameter restrictions, since no useable benchmark is available from the empirical literature nor from any established theory. Nonetheless, at this point, the R2GME results are treated as the *preferred* results, with the *lowest empirical risk* (MSE)—in so far as the (loose) cross-parameter restrictions in Equations (17)–(18) are "true" or "nearly true."

[16]Unfortunately, attempts to replicate the T-PPML and 2FE-PPML results using GME have failed, due to computational problems when imposing a Poisson variance structure on the GME optimization. Note, however, that standard OLS and PPML cannot generally be used to estimate the 3FE gravity model, so any comparison of the results for the 3FE model with those for the traditional and 2FE models necessarily involves comparison across different estimators.

The coefficient on destination HDI (0.794) indicates that improved socio-economic conditions (higher income, higher education, etc.) cause a significant increase in the number of migrant arrivals. Evaluated at the sample mean of *HDI*, the coefficient implies that a one-standard-deviation increase in HDI (which amounts to an increase of 16.6 percent) raises in-migration by about 13 percent. Conversely, the marginally significant coefficient on origin HDI (0.221) suggests that changes in welfare conditions have only little effect on migrant departures, increasing only by about 3.5 percent. This minor effect at the origin could indicate that there may be two counteracting forces at work, where higher living standards tend to reduce the incentive for those well-off and to increase the ability for the relatively poor to out-migrate (Clark et al., 2007, p. 362). The result is also in line with other studies of internal migration in which an asymmetrical impact of per-capita GDP (which is a major component of the HDI) has been reported, having a stronger destination-pull than origin-push effect (e.g., Sasser, 2010). In contrast, the coefficients returned by T-OLS and T-PPML predict an unrealistic decline in the number of migrant arrivals, whereas T-PPML suggests a strong positive effect on migrant departures.[17]

The coefficients on the unemployment rate indicate that higher unemployment triggers out-migration at the origin (0.199) and dampens in-migration at the destination (−0.225), as expected. The estimated coefficients suggest that a 1 percent point increase in unemployment rate, evaluated at the sample mean of *UR* (implying a change from 1.9 percent to 2.9 percent, or a 52.6 percent growth of the number of unemployed) induces only a modest 10 percent increase in migrant departures at the origin and 12 percent decrease in migrant arrivals at the destination. The rather weak effect on departures is likely to be due to the (concealed) heterogeneity of the origin population; that is, those who are unemployed (and thus potentially pushed to out-migrate) represent only a small fraction of the total population, while a high unemployment rate is likely to be of little importance to those who have a job (Cadwallader, 1992, p. 53). The relatively small impact of the unemployment rate on migrant arrivals is somewhat more difficult to explain. One possible explanation is that officially reported unemployment rates tend to be focused on the urban and formal (modern industrial) sectors, whilst not revealing much about employment opportunities in the rural and informal economy in the destination. The latter may continue to attract a significant number of (unskilled or poverty-stricken) unemployed migrants, even with a higher level of "formal" unemployment in the destination. Conversely, T-OLS and T-PPML predict a strong retention effect at the origin (33 percent) and a substantial repulsion effect at the destination (24 percent) of unemployment, respectively. Such sizeable effects are likely to be overly large (Hunt, 2006).

Finally, the coefficients on the U.S. border-state dummies are also fully in line with prior expectations. The coefficient on the origin border dummy (−0.199) suggests that the U.S. border states are better able to retain their residents than other states, showing a differential retentivity of 18 percent $= (e^{-0.199} - 1) \times 100$. In contrast, T-OLS predicts an improbable expulsion effect at the origin of 63 percent $= (e^{0.486} - 1) \times 100$. The coefficient

[17]It was found (in unreported results) that the coefficients on destination HDI of both T-OLS and T-PPML change dramatically, both in sign and magnitude, when the migrant-stock variable is excluded from the regression. A similar finding with regard to the coefficient on destination GDP per capita was found in Pedersen et al. (2008). It is further interesting to note that by excluding migrant stock, the coefficients on distance and contiguity become excessively large, ranging between –0.563 (T-PPML) and –1.133 (2FE-OLS) for distance, and between 0.713 (2FE-PPML) and 1.306 (T-OLS) for contiguity.

TABLE 3: Estimated Unconditional Mean and Weighted Averaged Conditional Mean Effects (Preview)

| | Migrant Stock (1) | Distance (2) | Contiguity (3) |
|---|---|---|---|
| A: Unconditional mean effects | $\hat{\eta} = E(\hat{\eta}_{ij})$ 0.104 | $\hat{\delta} = E(\hat{\delta}_{ij})$ −0.204 | $\hat{\lambda} = E(\hat{\lambda}_{ij})$ 0.092 |
| B: Weighted averaged conditional mean effects | $\bar{\eta} = w.avg.[E(\hat{\eta}_{ij}\|s_{ij})]$ 0.050 | $\bar{\delta} = w.avg.[E(\hat{\delta}_{ij}\|d_{ij})]$ −0.143 | $\bar{\lambda} = w.avg.[E(\hat{\lambda}_{ij}\|ZM_{ij})]$ 0.086 |

*Notes*: The numerical values in Panel A are the same as those reported in Column 8 of Table 2. The weighted averaged conditional mean effects have been computed over all bilateral migration flows, using as weights $M_{ij}/\Sigma_i\Sigma_j M_{ij}$ (to be further explained below). The weighted averaged effects calculated directly on the basis of the R2GME estimates of the path-specific effects, $\hat{\eta}_{ij}$, $\hat{\delta}_{ij}$, and $\hat{\lambda}_{ij}$, are 0.038, −0.087, and 0.146, respectively.

on the destination border dummy (0.448) indicates that migrant arrivals are 56 percent $= (e^{0.448} - 1) \times 100$ higher (which is broadly similar to T-PPML). This disproportionally high attractivity of the northern border states is likely to be reflective of the gateway function of these states, providing a "jumping off" point for temporary or permanent migration to the U.S. (Greenwood et al., 1981, p. 378).

Turning to the R2GME coefficients on the bilateral linkage covariates ($\hat{\boldsymbol{\beta}}_3$) reported in Column 8 of Table 2, some caution is warranted. A simple comparison of the numerical values of these coefficients with those for the traditional and 2FE models reported in the other columns of Table 2 can be misleading, since the estimates may represent different quantities. Strictly speaking, the R2GME estimates can only be meaningfully interpreted as "average" effects across the sample population if there is *no* correlated heterogeneity; that is, if the underlying path-specific effects, $\hat{\boldsymbol{\beta}}_{3,ij}$, are *randomly* distributed around a common mean (mean independence). However, in the presence of *correlated* heterogeneity in the cross section, the R2GME estimates of the parameters $\boldsymbol{\beta}_3 = E(\boldsymbol{\beta}_{3,ij})$ merely represent the *unconditional mean* effects (as already explained before), which can therefore not generally be directly contrasted with the *average* effects across the sample population obtained for the traditional and 2FE models (assuming constant parameters, $\boldsymbol{\beta}_{3,ij} \equiv \boldsymbol{\beta}_3$). In order to provide a more balanced comparison across the model specifications in Table 2, one possible move is to report the *weighted average* effects of the bilateral linkage variables, calculated on the basis of the estimated path-specific effects, $\hat{\boldsymbol{\beta}}_{3,ij}$. Although a more detailed posterior (second-stage) analysis will be provided later in this paper, Table 3 gives a preview of the weighted average effects derived from the *expectations* of the individual effects, $\tilde{\boldsymbol{\beta}}_{3,ij} = E(\hat{\boldsymbol{\beta}}_{3,ij}|\bullet)$, which, in turn, are estimated *conditional* on some functions of observed covariates (see below). The weighted averages have a less ambiguous interpretation than the unconditional means and may therefore provide a more satisfactory comparison.

From the preview in Table 3, it is immediately apparent that the weighted averages of the conditional mean effects of the bilateral covariates differ appreciably from the unconditional mean effects (where the observed discrepancies can in a certain way be viewed as indirect evidence of the presence of correlated heterogeneity in the sample population). First, the weighted average effect of the migrant stock ($\bar{\eta} = 0.050$) is found to be roughly about 16 times smaller than the average effects obtained for the traditional and 2FE models (Columns 1–4 of Table 2). Although this estimated value may appear excessively small, at first sight, it is felt to be a greatly improved measure of the importance of networks of "family and friends" in inducing further migrations, given the persistent features of the spatial structure which are affecting current and previous migration flows in a fairly similar way.[18] Evaluated at the sample means of the migrant stocks and migration flows, the weighted average migrant-stock effect implies that for every 1,000 added to the stock of in-migrants a further 11.3 migrants would arrive in the following five-year observation period, or 2.3 migrants each next year. This annual number is roughly half the size of the 4.7 reported in Clark et al. (2007, p. 365) in their study of U.S. immigration—which these authors already called a "substantial" effect.[19] The smaller number found here can be explained by the fact that network effects may be more important in the context of international (as opposed to internal) migration.[20] Conversely, the estimates returned by the traditional and 2FE models greatly exaggerate the role of migrant stocks in inducing network migration. Evaluated at the sample means, the coefficients predict further migrations numbering 35 (traditional model) to 40 (2FE model) annually—or a dazzling 175 to 200 over a period of five years—per 1,000 added to the migrant stock. These incredible numbers also illustrate that the 2FE approach is inappropriate for dealing with the endogeneity problem associated with the migrant-stock variable. (In fact, the magnitudes of the migrant-stock coefficients returned by the 2FE model (Columns 3–5) are even larger than those emerging from the traditional model (Columns 1–2).

Second, the weighted average effect of distance ($\bar{\delta} = -0.143$) ascribes a greater role to distance as a deterrent to migration than 2FE-PPML ($\hat{\delta} = -0.102$), even though the magnitudes are not different from each other in a statistically significant way (the value of the weighted average is within the 95 percent confidence interval of the 2FE-PPML estimate). The smaller absolute size of the distance effect returned by 2FE-PPML is likely to be due to the fact that the endogeneity of the migrant stock is not properly accounted for.

Last, the weighted average effect of contiguity ($\bar{\lambda} = 0.086$) indicates that sharing a border has a relatively minor impact on migration flows between neighboring states, where the number of migrants is estimated to be only about 9 percent higher than between noncontiguous states. Conversely, the coefficients returned by T-OLS and T-PPML overstate the contiguity effect (estimated at 28 percent and 34 percent, respectively), particularly in view of the spatial coarseness of the analysis, while the 2FE model returns insignificant coefficients (having also the wrong signs).

*Specification Tests*

In concluding this empirical section, two tests are performed for checking the adequacy of the R2GME-estimated 3FE model. Since GME makes no *a priori* assumptions about the form of the variance-mean relationship underlying the data generating process, I start with performing the *modified Park test* (Manning and Mullahy, 2001,

[18]I have also made an attempt to address the endogeneity issue by instrumenting for the migrant stock (as suggested by some commentators on earlier versions of the paper), by using the *total* origin-specific migrant stock, $TS_i = \Sigma_j S_{ij}$, as a single instrument (based on the intuition that this instrument is likely to be reflective of the overall migration propensity at each origin-state $i$ in the past). However, IV estimation failed to produce sensible parameter estimates, heavily inflating several parameter estimates and producing an insignificant migrant-stock coefficient (though returning the expected positive sign for the coefficient on the destination HDI; see also footnote 17).

[19]Curiously, the number of 4.7 reported in Clark et al. (2007) is exactly the same as the number obtained here on the basis of the estimated unconditional mean effect ($\hat{\eta} = 0.104$) that is, $[0.104 \times (1000/16014) \times 3614]/5 = 4.7$

[20]The smaller effect of networks. on migration may also be due to the fact that migratory moves are defined as changes of place of *residence*, whereas migrant stocks are cross-tabulated by place of *birth* (see the definitions of migrant flows and migrant stocks in Appendix A). As a result, people migrating from $i$ to $j$ during the five-year observation period are not necessarily born in $i$ but in $k \neq i$ such that $s_{kj}$ might be more important than $s_{ij}$ in inducing (re-)migration from $i$ to $j$ during the observation period.

p. 471) to find out whether the variance structure of the estimated model belongs to the standard GLM class of "power-proportional" forms; that is, whether $V(M_{ij}|\mathbf{x}_{ij}, \mathbf{b}_{ij}, \upsilon_i, \omega_j) = \lambda_0[E(M_{ij}|\mathbf{x}_{ij}, \mathbf{b}_{ij}, \upsilon_i, \omega_j)]^{\lambda_1} = \lambda_0(e^{\mathbf{x}_{ij}\boldsymbol{\beta}+\xi_{ij}})^{\lambda_1}$. The modified Park test is implemented by estimating a regression of the form $\ln \hat{r}^2 = \ln \lambda_0 + \lambda_1 \ln \hat{M} + \epsilon$, where $\hat{M} = e^{\mathbf{x}\hat{\beta}+\hat{\xi}}$ and $\hat{r} = M - \hat{M}$, using OLS (based on a nonrobust covariance estimator). The parameter $\lambda_1$ is estimated at 2.009 (s.e. = 0.035), so the null $\lambda_1 = 2$ cannot be rejected, with a $p$-value of 0.797.[21] This test result strongly suggests that the variance-mean relationship implied by 3FE-R2GME matches with a gamma-class-variance, where the conditional variance of $M$ (and $r$) is proportional to the *square* of the conditional mean, $V(M|\bullet) \propto [E(M|\bullet)]^2 = e^{2(\mathbf{x}\beta+\xi)}$. This means that, after controlling for the spatial-structural heterogeneity component of the data generating process by conditioning $M_{ij}$. on both $(\mathbf{b}_{ij}, \upsilon_i, \omega_j)$ and $\mathbf{x}_{ij}$, the variability in the data can be adequately described by a gamma variance function. Such a pattern is indicative of the fact that the multiplicative (raw-scale) error, $u = 1 + r/e^{\mathbf{x}\beta+\xi}$, has constant variance (as shown in footnote 2; see also Manning and Mullahy, 2001, p. 465), hence satisfying one of the conditions put forward by Santos Silva and Tenreyo (2006, p. 644) for consistent estimation of the slope parameters of the gravity equation.[22]

Next, I perform *Pregibon's LINK test* (Pregibon, 1980), which is essentially a test for misspecification of the linear exponential mean of the R2GME-estimated 3FE model. The test is analogous to Ramsey's (1969) "two-degrees-of-freedom" (2df) RESET test for linear regression models (Wooldridge, 1997). In view of the results of the modified Park test above, the LINK test is performed by estimating a gamma GLM test regression (with scale parameter 1) of the form $M = e^{\alpha_0+\alpha_1(\mathbf{x}\hat{\beta}+\hat{\xi})+\alpha_2(\mathbf{x}\hat{\beta}+\hat{\xi})^2}\epsilon$, to "mimic" the GME estimation of the original 3FE model, and testing the statistical significance of the coefficient $\alpha_2$ on the squared term. The LINK test (based on GLM nonrobust standard errors) fails to reject the null $\alpha_2 = 0$, with a $p$-value of 0.997. This powerful test result indicates that "nonlinearities" are sufficiently accounted for (in a statistical sense) by the conditional mean of the 3FE model.[23] In contrast, both T-PPML and 2FE-PPML are unequivocally rejected at conventional levels. Furthermore, T-OLS is strongly rejected by the 2df RESET test, while 2FE-OLS, somewhat surprisingly, passes the test. This unexpected result for 2FE-OLS could be reflective of the fact that the standard RESET test may not always be successful in detecting neglected parameter heterogeneity (Zietz, 2001). Despite these mixed test results, it is fair to conclude that the generally poor performances of the traditional and 2FE models are, at least partly, due to neglected slope-parameter heterogeneity (and the erroneous assumption of the spatial independence of the migration flows).[24]

[21]The modified Park test for 3FE-UGME and 3FE-R1GME returned values for $\hat{\lambda}_1$ of 2.164 (*s.e.* = 0.042) and 2.061 (*s.e.* = 0.034), respectively. Although the value of $\hat{\lambda}_1$ for 3FE-UGME is significantly different from 2, it is practically very close to 2.

[22]I also performed the Gauss-Newton regression (GNR) test, proposed by Santos Silva and Tenreyo (2006, p. 646), to check the Poisson variance structure, $V(M|\bullet) \propto E(M|\bullet)$. The GNR test is implemented by estimating an auxiliary regression (*excluding* the constant term) of the form $\hat{r}^2/\sqrt{\hat{M}} = \lambda_0\sqrt{\hat{M}} + \lambda_0(\lambda_1 - 1) \ln \hat{M}\sqrt{\hat{M}} + \epsilon$, using Weighted Least Squares (based on a robust covariance estimator), which allows for testing the null $H_0 : \lambda_0(\lambda_1 - 1) = 0$, which is equivalent to $H_0 : \lambda_1 = 1$. The GNR/WLS test results indicate that T-PPML is not rejected ($p = .313$), while 2FE-PPML is strongly rejected ($p < .001$).

[23]In interpreting the test results, it should be noted that the realized errors (residuals or deviances) of 3FE-R2GME represent only a "small issue on the fringe", given that the unobserved heterogeneity in the data is almost entirely picked up by the spatial-structural heterogeneity component [2] in Equation (5).

[24]It should be recalled that the well-known property of PPML of being consistent for any pattern of heteroskedasticity (Gourieroux, Monfort, and Trognon, 1984) holds *only* if the conditional mean is correctly specified (see also Cameron and Trivedi, 2005, p. 669).

*Posterior Analysis of the Path-Specific Bilateral Effects*

Thus far the discussion of the results was mainly focused on the estimates of the trend-surface parameters in component [1] of the 3FE gravity model in Equation (5). In contrast, this section takes a closer look at the heterogeneity in the R2GME-estimated path-specific slope parameters and, more precisely, at the *conditional* expectations of the migration effects of distance, contiguity, and migrant stock.

Kernel densities for the R2GME-estimated path-specific effects are depicted in Figure 2 (together with those returned by R1GME and UGME, to illustrate their sensitivity to the imposed restrictions). The shapes of these unconditional (or marginal) distributions suggest substantial heterogeneity in the sample population. The frequent occurrence of bounded zero values for the individual distance and migrant-stock effects may be a possible concern. However, it is reassuring to see that most of those zeros pertain to migration flows between neighboring states. In addition, the piling up of individual contiguity effects at zero is not surprising, considering the spatial coarseness of the analysis.

Given the wide variation in these bilateral effects, it is worth examining whether they exhibit some (spatial) regularities, as simply knowing that the effects vary across migration paths is only of limited practical use. However, because the path-specific effects of the linkage variables are unlikely to be estimated in a precise way (given the limited/incomplete information used), it is more informative to look at their variation with some observed covariates (Cameron and Trivedi, 2005, p. 704). That is, by focusing on the *conditional expectations* across the sample population (which equates to inferring their most likely values for each migration path), the impact of the "intrinsic uncertainty" that exist in the underlying parameter estimates can be greatly reduced. In what follows, the conditional mean effects of the bilateral linkage variables will be estimated on the basis of some stylized conditional relationships.

*Conditional Mean Migrant-Stock (Network) Effects*

A widely recognized result in the migration literature is that network effects tend to become smaller with larger sizes of the migrant stock; that is, network externalities are expected to be subject to "diseconomies of size" of the established in-migrant population. To substantiate such a relationship, I estimate a standard gamma PML (GPML) regression (with log link), relating the estimated path-specific effects of migrant stock, $\hat{\eta}_{ij}$, to the (log) size of the stock, $s_{ij}$. From this GPML regression, informative estimates of the expected network effects *conditional* on the size of the stock, denoted by $\tilde{\eta}_{ij} = E(\hat{\eta}_{ij}|s_{ij})$, can be obtained.

The results reported in Panel A of Table 4 (Column 1) powerfully confirm the expected inverse relationship between the network effect and the size of the migrant stock. The plot in Panel A of Figure 3 shows that the network effect decreases rapidly with increasing stock size. Evaluated at the 10th, 50th, and 90th percentiles of the sample distribution of the migrant stock, the conditional mean effect is estimated at 0.178, 0.100, and 0.042, respectively (for the state of México, with the largest stock of in-migrants from the neighboring Distrito Federal, the effect is estimated at only 0.012). While such a decline in the effect of migrant networks has been usually explained as a manifestation of increasing job-search and settlement costs, or "diminishing returns to network externalities" (Clark et al., 2007, p. 363), it could also be explained by the fact that the presence of large migrant stocks may simply be due to strong spatially induced effects in the past—where the latter have occasionally been designated in the literature as "herd effects" (Bauer et al., 2007; Epstein, 2008), which have nothing to do with the presence of "family and friends" in the destination.

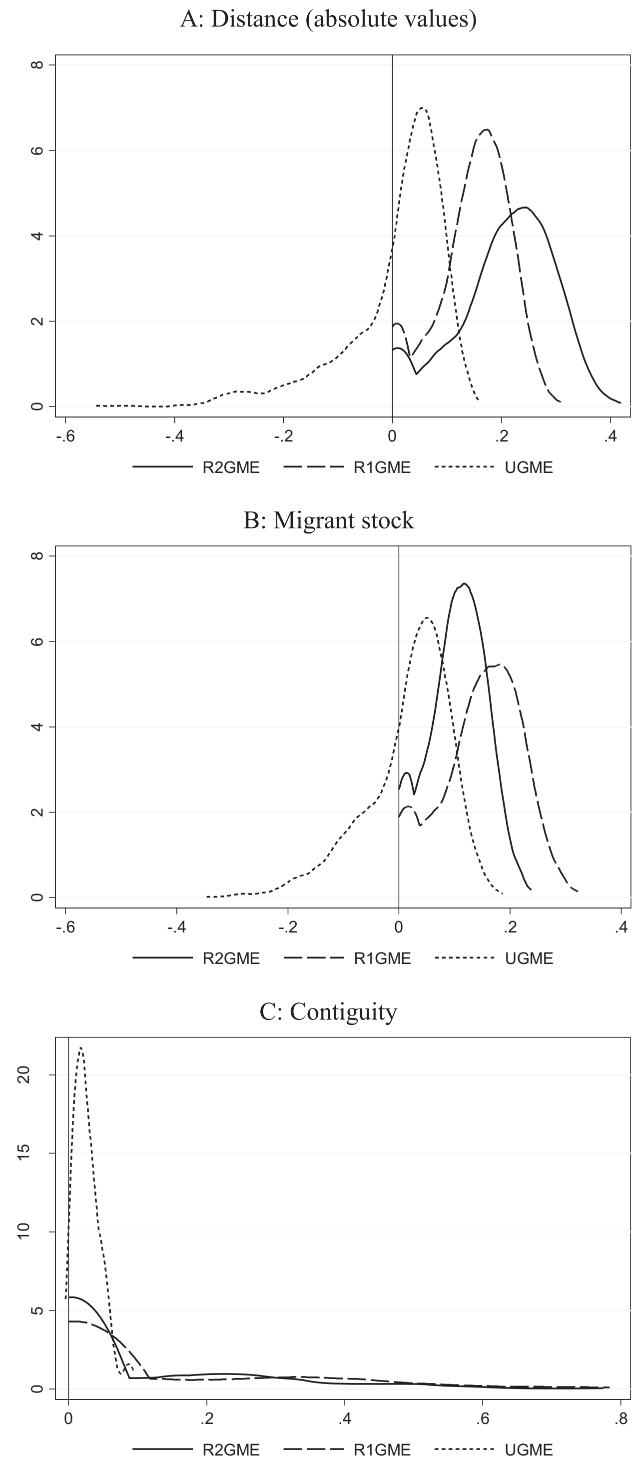



FIGURE 2: Kernel Densities of the Estimated Path-Specific Effects (Unconditional Distributions).

TABLE 4: Results for GPML Regressions—Conditional Mean Effects of Migrant Stock and Distance

| | (1) | (2) | (3) |
|---|---|---|---|
| A: Path-specific migrant-stock (network) effects [$\hat{\eta}_{ij}$] | | | |
| Log migrant stock [$\ln S_{ij}$] | $-0.375^{***}$ | | |
| | (0.018) | | |
| Intercept | $-4.868^{***}$ | | |
| | (0.130) | | |
| $R^2$ | 0.560 | | |
| B: Path-specific distance effects (in absolute value) [$\lvert\hat{\delta}_{ij}\rvert$] | | | |
| Log distance [$\ln D_{ij}$] | $3.711^{***}$ | $3.624^{***}$ | $3.731^{***}$ |
| | (0.426) | (0.421) | (0.426) |
| Log distance squared [$(\ln D_{ij})^2$] | $-0.252^{***}$ | $-0.237^{***}$ | $-0.253^{***}$ |
| | (0.030) | (0.030) | (0.030) |
| Log origin centrality [$\ln CI_i$] | | $0.240^{***}$ | |
| | | (0.028) | |
| Log relative centrality [$\ln RCI_{ij} = \ln(CI_i/CI_j)$] | | | $0.045^{***}$ |
| | | | (0.011) |
| Intercept | $-15.089^{***}$ | $-16.443^{***}$ | $-15.148^{***}$ |
| | (1.485) | (1.444) | (1.484) |
| $R^2$ | 0.284 | 0.332 | 0.298 |

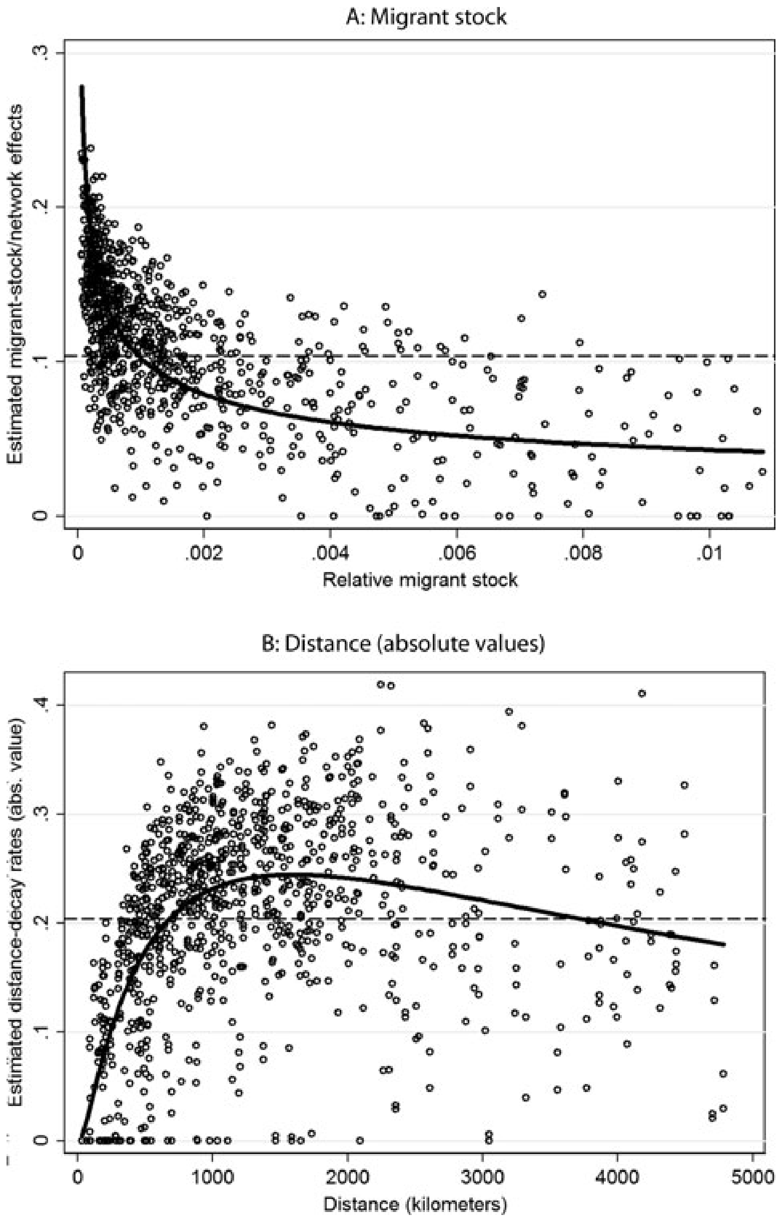
*Notes*: The results shown in the table are based on standard GPML regressions, where the dependent variables are the R2GME estimates of the path-specific migrant-stock (Panel A) and distance (Panel B) effects. Since the regressions are of the EDV type (having an estimated dependent variable), robust standard errors, given in parentheses, account for the sampling error in the dependent variable (e.g., Lewis and Linzer, 2005). The $R$-square is calculated as the squared correlation between the predicted and "actual" values.

$^{***}$Significant at 1 percent.

With the help of the estimated conditional means, which account for more than 50 percent of the heterogeneity in the marginal distribution ($R^2 = 0.560$), it is now possible to calculate the *weighted average* network effect. Given that $\widetilde{\Delta M}_{ij}/M_{ij} = \tilde{\eta}_{ij}\Delta S_{ij}/S_{ij} = \tilde{\eta}_{ij}\Delta I_{ij}/I_{ij}$ (holding $P_i$ constant), the weighted average is estimated at $\bar{\eta} = \sum_i \sum_j \widetilde{\Delta M}_{ij}/\sum_i \sum_j M_{ij} = \sum_i \sum_j \tilde{\eta}_{ij} M_{ij}/\sum_i \sum_j M_{ij} = 0.050$ (preview Table 3, Panel B, Column 1). This outcome tells us that, due to positive network externalities, the number of migrants increases, on average, by 0.5 percent following a uniform 10 percent increase in all migrant stocks across the sample population. This effect is only about half the size of the estimate of the unconditional mean, $\hat{\eta} = 0.104$ (Table 2, Column 8), where the latter is likely to be a heavily "inflated" measure of the network effect due to the strong positive (unconditional) correlation between migrant stocks and migration flows ($r = 0.742$). (Note that the segment of the solid curve below the dashed horizontal line in Panel A of Figure 4 represents about 93 percent of the total volume of interstate migrations in Mexico.)

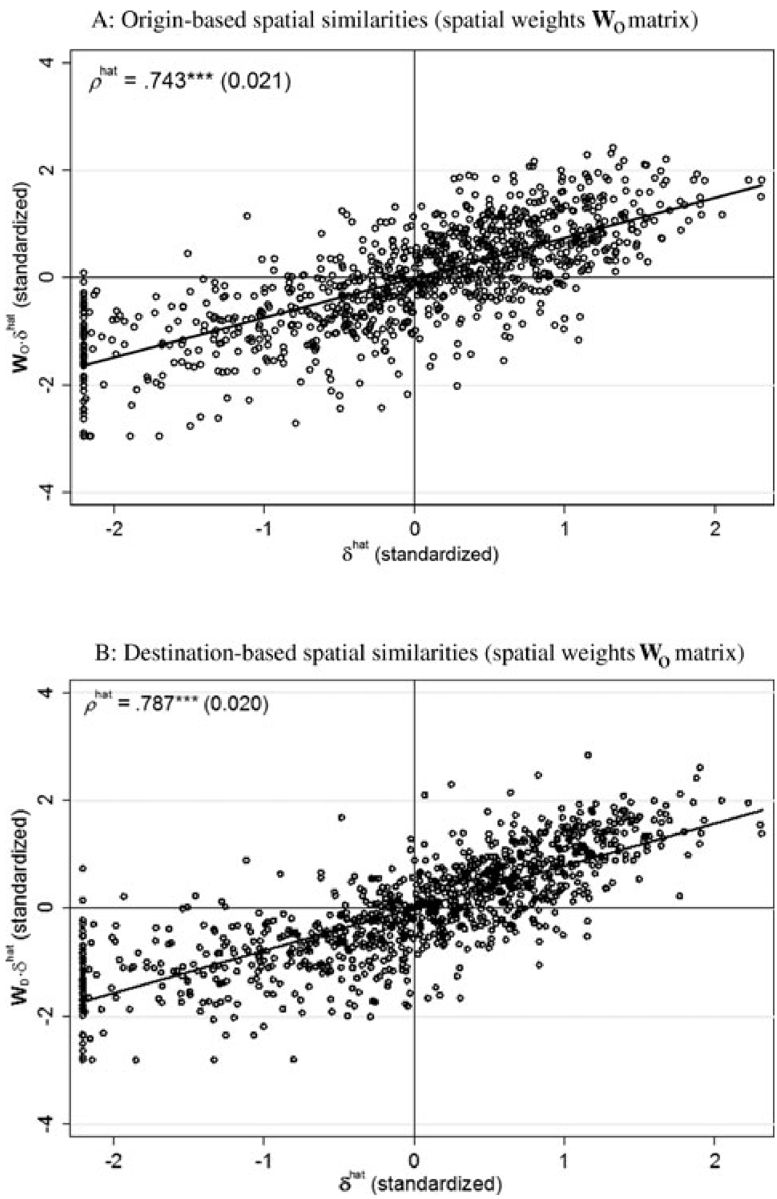
*Conditional Mean Distance Effects*

Another important recurring concern in studying internal migration is whether distance-decay rates exhibit plausible "map patterns" (Tiefelsdorf, 2003). Armed with the estimated path-specific distance effects, $\hat{\delta}_{ij}$, it is possible to provide reasonable estimates of the conditional mean decay rates, denoted by $\tilde{\delta}_{ij} = E(\hat{\delta}_{ij}|d_{ij}, A)$, where $A$ represents some observable (usually distance-based) measure of spatial structure. To accomplish this, three standard GPML regressions are estimated to examine whether certain spatial regularities come into view. In the first regression, I include as covariates




*Notes*: The dots represent the R2GME estimates of the individual, path-specific effects of migrant stock (Panel A), $\hat{\eta}_{ij}$, and distance (Panel B), $\hat{\delta}_{ij}$. The solid curves represent their conditional expectations, $\tilde{\eta}_{ij} = E(\hat{\eta}_{ij}|s_{ij})$ and $\tilde{\delta}_{ij} = E(\hat{\delta}_{ij}|d_{ij})$, respectively (based on the GPML results in Column 1 of Table 4). The dashed horizontal lines represent the corresponding unconditional expectations, $\hat{\eta} = E(\hat{\eta}_{ij})$ and $\hat{\delta} = E(\hat{\delta}_{ij})$, respectively (based on the R2GME results in Column 8 of Table 2). To improve the legibility of the plot in Panel A, the values in the 10 percent upper tail of the (heavily positively skewed) sample distribution of the bilateral migrant stocks have been excluded.

FIGURE 3: Estimated Path-Specific Effects and Unconditional vs. Conditional Mean Effects.




*Notes*: The spatial autoregressive models, estimated using OLS, are based on the standardized ($z$) values of the elements in the $N$-by-1 vector $\hat{\delta}$. Heteroskedasticity-robust standard errors are given in parentheses.

FIGURE 4: Spatial Similarities of Estimated Path-Specific Distance-Decay Rates.

only (log) *distance*, $d_{ij}$, and its *squared* value, $d_{ij}^2$, to allow for potentially nonmonotonic relationships. In the second regression, I add a measure of *origin centrality*, defined (like before) as $\ln CI_i = \ln \sum_{j=1}^{n}(P_j/D_{ij})$. In the third regression, I include an indicator of directionality. Although there is no obvious single way of accurately measuring the direction of migration, I use an index of *relative centrality* to proxy for flow directionality, simply defined as $\ln RCI_{ij} = \ln(CI_i/CI_j)$, where $\ln RCI_{ij} \gg 0$ for outward (center-to-periphery) movements, $\ln RCI_{ij} \ll 0$ for inward (periphery-to-center) movements, and $\ln RCI_{ij} = 0$ (or near zero) for other (center-to-center and periphery-to-periphery) movements.[25]

Looking at the regression results across the columns in Panel B of Table 4, some general findings stand out. The estimated coefficients on the two distance terms are roughly the same across the three regressions, such that $E_A[E(\hat{\delta}_{ij}|d_{ij}, A)] \approx E(\hat{\delta}_{ij}|d_{ij})$. Moreover, the coefficients on the constructed spatial-structure measures (Table 4, Columns 2 and 3) are found to be mutually consistent and to have plausible behavioral interpretations, providing forceful evidence of the "space warping engendered by contextual factors" (Eldridge and Jones, 1991). Thus, the R2GME-estimated individual distance-decay rates are found to be largely reflective of the hierarchical spatial arrangement of origins and destinations, after controlling for distances. The specific findings are as follows.

The estimated coefficients on the distance terms in the first regression (Table 4, Panel B, Column 1) suggest a *nonmonotonic* pattern of the conditional expectation of distance decay. In Panel B of Figure 3, the solid curve visualizes the estimated conditional means, $\tilde{\delta}_{ij} = E(\hat{\delta}_{ij}|d_{ij})$. The shape of this curve reveals that the absolute size of the decay rate is close to zero for short-distance moves, then sharply increases with distance up to a maximum of 0.245 at 1,580 km, and slowly decreases again thereafter.[26] That is, for moves over a short distance of, say, 100 kilometers, the number of migrants falls, ceteris paribus, by only 1.8 percent if the destination were located 50 percent farther away (at 150 kilometers). At larger distances of, say, 200, 500, and 1,000 km, the number of migrants declines, ceteris paribus, by 4.1 percent, 8.7 percent, and 11.6 percent, respectively, if distances are 50 percent larger.

Using the predicted conditional means, which account for almost 30 percent of the heterogeneity in the marginal distribution ($R^2 = 0.284$), the "system-wide" weighted average distance-decay rate (assuming a uniform 1 percent increase in all distances across the spatial system) is estimated at $\bar{\delta} = \Sigma_i \Sigma_j \tilde{\delta}_{ij} M_{ij} / \Sigma_i \Sigma_j M_{ij} = -0.143$ (preview Table 3, Panel B, Column 2), which means that migration is expected to decline on average by 7.2 percent, ceteris paribus, any time distance increases by 50 percent. This measure is smaller (in absolute value) than the estimated unconditional mean, $\hat{\delta} = -0.204$ (Table 2, Column 8), because of the negative (unconditional) correlation between distances and migration counts ($r = -0.130$). (Note that the segment of the solid curve (on the left side) below the dashed horizontal line in Panel B of Figure 4 represents about 64 percent of the total volume of interstate migrations in Mexico.)

The second regression (Table 4, Panel B, Column 2) returns a positive value for the coefficient on the origin-specific centrality index, which means that outflows from more centrally located origins exhibit a *steeper* distance decay, ceteris paribus. Evaluated at the 10th, 50th, and 90th percentiles of the sample distribution of the centrality index (and the sample mean of distance), the expected decay rates are estimated at 0.200, 0.246, and 0.301, respectively (for the most centrally located Distrito Federal, the decay rate is estimated at 0.379). These results suggest that migrants from central origins tend to be more reluctant to move to distant destinations, considering the greater number of opportunities (or alternatives) available in their direct vicinity, while migrants from remote origins tend to be willing to move over longer distances in their search for opportunities

[25]Although the relative-centrality measure is not a "perfect" indicator, a preliminary inspection revealed that its values do fairly represent the intended directionality of migratory moves (given the spatial dominance of the center).

[26]Using spline-regressions to estimate varying distance-decay rates for different tallies of distance, Mueser (1989) and, more recently, de Vries et al. (2009) reported quite similar decay patterns.

that may not be readily available in their direct vicinity. Importantly, these findings are in sharp contrast with the counter-intuitive results typically reported in earlier studies of internal migration (Fotheringham, 1991; Tiefelsdorf, 2003), including studies of interstate migration in Mexico (Greenwood et al., 1981; Denslow and Eaton, 1984).

The third regression (Table 4, Panel B, Column 3) returns also a positive value for the coefficient on the indicator of directionality. This indicates that outward (center-to-periphery) movements, having (large) positive values of $\ln RCI_{ij}$, exhibit a *steeper* distance decay than inward (periphery-to-center) movements, having (large) negative values of $\ln RCI_{ij}$, substantiating the prevalence of directional asymmetry in the effect of geographical distance—or *anisotropic trends*, as Fotheringham and Pitts (1995) called it. Evaluated at the 10th, 50th, and 90th percentiles of the sample distribution of the relative-centrality index (and the sample mean of distance), the decay rates are estimated at 0.229, 0.243, and 0.258, respectively. While these differences are not overwhelming in a practical (quantitative) sense, the directional asymmetry is statistically significant. Accordingly, a general pattern seems to emerge where out-migrants from the center of the spatial system tend to be more locally bounded (to move outward over shorter distances), other things being equal, while in-migrants into the center of the spatial system tend to originate from more distant places (to move inward over longer distances).

Finally, I examine the spatial autocorrelation or *spatial similarities* among the R2GME estimates of the path-specific effects of distance. One way of doing this is to estimate a basic spatial autoregressive model of the form $\hat{\boldsymbol{\delta}}^* = \rho \mathbf{W}\hat{\boldsymbol{\delta}}^* + \mathbf{v}$, where $\hat{\boldsymbol{\delta}}^*$ is an $N$-by-1 vector containing the standardized ($z$) values of the estimated path-specific distance-decay rates, $\hat{\delta}_{ij}$ $(i \neq j; i, j = 1, \ldots n)$, $\rho$ is the autoregressive coefficient on the spatial lag $\mathbf{W}\hat{\delta}^*$, and $\mathbf{v}$ is an $N$-by-1 vector of i.i.d. normally distributed random terms with zero mean and variance $\sigma^2$. The $N$-by-$N$ (row-standardized) spatial weights matrix $\mathbf{W}$ is constructed following the approach proposed by Fischer and Griffith (2008, p. 975) or Chun (2008, p. 321), where proximity is defined in terms of origin- or destination-related (first-order) contiguity relations. (By convention, the diagonal elements of the spatial weights matrix are set to zero.) Specifically, the spatial weights matrix $\mathbf{W}_{\mathrm{O}}$ links each individual $\hat{\delta}_{ij}$ to an *origin*-based neighborhood set; that is, an element $w_O(i, j; r, s)$ of $\mathbf{W}_{\mathrm{O}}$ defines an origin-destination path $(r, s)$ as being a "neighbor" of $(i, j)$ if origins $i$ and $r$ are contiguous and $j = s$. Likewise, the spatial weights matrix $\mathbf{W}_{\mathrm{D}}$ links each individual $\hat{\delta}_{ij}$ to a *destination*-based neighborhood set; that is, an element $w_D(i, j; r, s)$ of $\mathbf{W}_{\mathrm{D}}$ defines an origin-destination path $(r, s)$ as being a "neighbor" of $(i, j)$ if destinations $j$ and $s$ share a common internal border and $i = r$. The autoregressive models are estimated using OLS. The plots in Figure 4 provide compelling evidence of the presence of spatial similarities among the distance-decay rates for "nearby" flows sharing a "similar migration context," both at the origins (Figure 4, Panel A) and the destinations (Figure 4, Panel B), where the origin- and destination-based spatial autocorrelations are estimated at 0.743 and 0.787, respectively. These outcomes are clearly in line with Tobler's (1970, p. 236) first law of geography, in which "everything is related to everything else, but near things are more related than distant things." Evidently, if the effects of distance on migration differ across migration paths, and if these effects are very similar among neighboring paths, this must also lead to significant spatial dependencies among the levels of the corresponding migration counts, ceteris paribus.

### *Conditional Mean Contiguity Effects*

The coarse spatial scale at which migration flows have been analyzed may also raise some concern about the adequacy of using a crude contiguity indicator for measuring the impact on interstate migration of sharing a common internal border. From the estimated

TABLE 5: Results of $t$-Test (Unequal Variances)—Conditional Mean Effects of Contiguity

| Individual Contiguity Effects $[\hat{\lambda}_{ij}]$ | All | $ZM_{ij} = 0$ | $ZM_{ij} = 1$ | Difference |
|---|---|---|---|---|
| Average | 0.092 | 0.072 | 0.276 | 0.204*** |
| (St. dev.) | (0.159) | (0.142) | (0.196) | |
| No. of cases | 142 | 128 | 14 | |

*Notes*: The *Zonas Metropolitanas* are identified as in SEDESOL-CONAPO-INEGI (2007). Metropolitan zones straddling an internal state border ($ZM_{ij} = 1$) are: La Laguna (Coahuila-Durango), Valle de México (Distrito Federal-México and México-Hidalgo), Puerto Vallarta (Jalisco-Nayarit), La Piedad-Pénjamo (Michoacán-Guanajuato), Puebla-Tlaxcala (Puebla-Tlaxcala), and Tampico (Tamaulipas-Veracruz).

***Significant at 1 percent.

kernel density plot in Panel C of Figure 4, it could already be seen that the marginal distribution of the path-specific contiguity effects is heavily positively skewed, containing at least a (limited) number of very large values. Therefore, estimation of a single (constant) coefficient cannot possibly capture the variation across distinct pairs of neighboring states in a satisfactory way.

For example, "contiguity" is very likely to have a different meaning for people (prospective migrants) living in the large state of Chihuahua than for those living in the small Distrito Federal. Moreover, the border effect may also depend on the geographical distribution of populations and economic activities *within* the neighboring states. One such important "local" or "contextual" characteristic is whether or not contiguous states have a large metropolitan zone *straddling* the common internal border. To check whether this particular configuration leads to systematically stronger contiguity effects, the results of a simple $t$-test have been reported in Table 5. The results of this test suggest that the average contiguity effect for neighboring states sharing a metropolitan zone, which is estimated at 32 percent $= (e^{0.276} - 1) \times 100$, is four times the size of the average effect for those states that do not have a common metropolitan zone. Thus, migration counts between contiguous states appear to be significantly higher mainly in those instances of short-range (possibly just "residential") moves within the shared metropolitan zones.[27]

Finally, the weighted average effect of contiguity (averaged over the flows between contiguous states, $M^{c}_{ij}$), conditionally on having ($ZM_{ij} = 1$) or not having ($ZM_{ij} = 0$) a metropolitan zone on the shared internal border, $\tilde{\lambda}_{ij} = E(\hat{\lambda}_{ij}|ZM_{ij})$, is estimated at $\bar{\lambda} = (\tilde{\lambda}_1 \Sigma_i \Sigma_j M^{c}_{ij,1} + \tilde{\lambda}_0 \Sigma_i \Sigma_j M^{c}_{ij,0}) / \Sigma_i \Sigma_j M^{c}_{ij} = 0.086$ (preview Table 3, Panel B, Column 3), which means that migration counts between contiguous states are expected to be on average almost 9 percent higher than between noncontiguous states. The weighted average is very close to the value of the estimated unconditional mean, $\hat{\lambda} = 0.092$ (Table 2, Column 6), due to the relatively small number of flows (142 out of 992) involved.

## 5. SUMMARY AND CONCLUDING REMARKS

To address the important issue of spatial choices and the concomitant endogeneity of the bilateral covariates (including distance, contiguity, and migrant stock) on

[27]Apart from the apparent differences in magnitude of the contiguity effects, depending on the specific spatial context, important directional asymmetries are found to emerge as well. To give just one example, for the metropolitan zone of the Valle de México (covering the Distrito Federal and some parts of the neighboring state of México and the state of Hidalgo), it was found that the contiguity effect on outflows from the Distrito Federal towards the adjacent state of México is predicted to be 2.5 times greater than the effect in the opposite direction, $(e^{0.631} - 1)/(e^{0.314} - 1) = 2.4$. This finding suggests a tendency of sub-urbanization of metropolitan Mexico City (or net sprawl towards the suburban fringe).

internal place-to-place migration, this paper presented a generalization of the empirical gravity-model specification. The correlated unobserved effects of the spatial structure of the migration setting, which can be held responsible for the observed regularities in the way migrant populations are distributed over space, have been accounted for by entering fixed-effect (dummy) interactions with the bilateral linkage variables, allowing for heterogeneous (bilateral) slope parameters—besides the more usual (unilateral) origin- and destination-fixed effects. The resultant three-way fixed-effects (3FE) model was applied to cross-sectional data on interstate migration in Mexico during the five-year observation period 1995–2000. To overcome parameter-dimensionality problems, which are in fact always present in spatial data analyses (Pinkse and Slade, 2010), a "restricted" GME estimator was used as a suitable (practical) device to identify the model parameters. The empirical application illustrated several major benefits of accommodating the correlated effects of the underlying spatial structure of migration. The results further demonstrated that only introducing origin- and destination-fixed effects (2FE) is insufficient to mitigate the endogeneity problem related to the linkage variables included in the gravity equation.

The main findings of the empirical analysis can be summarized as follows. First, the parameter estimates are strikingly sensitive to the choice of model specification, where the coefficients for the 3FE model returned by R2GME (involving regularizing sign and cross restrictions on the individual-specific bilateral slope parameters) are far more plausible and intuitively interpretable than their traditional and 2FE counterparts, regardless of the estimator used (OLS, PPML, or GME), where parameter estimates generally change in the expected directions. Second, the (weighted) average effect of the migrant stock, calculated on the basis of the conditional means of the R2GME estimates of the path-specific effects, provides a more realistic—appreciably smaller—measure of network-induced migrations, suggesting that only 0.5 percent of migrants would arrive over a five-year period any time the migrant stock increases by 10 percent (where the actual percentage increase ranges between a minimum of 0.1 and a maximum of 2.8 depending on the size of the initial stock). Third, the spatial regularities (map patterns) emerging from the estimated path-specific distance effects are fully in line with prior expectations, where (a) migration outflows from more centrally located origins exhibit a significantly steeper distance decay, (b) "outward" and "inward" movements are influenced in an asymmetric way by the distance to cover, and (c) distance-decay rates for "neighboring" migration flows turned out to be markedly similar. These encouraging findings, which are clearly reminiscent of the notions of "migration space" (Plane, 1984) and "warped space" (Eldridge and Jones, 1991), are in sharp contrast with the counter-intuitive results usually reported in previous empirical studies of internal migration. Finally, the results convincingly support the view (Congdon, 1993) that the heteroskedasticity typically present in aggregate migration data can be set within the broader context of parameter heterogeneity arising from the correlated effects of the underlying spatial structure.

Although the analysis presented in this paper represents an important step forward to a better understanding of the determinants of interstate migration flows in Mexico, I make no claim whatsoever to have given a final word on the geography of migration, or the limitations of the estimation strategy adopted here. Certainly, many unanswered questions remain, leaving considerable scope for further work. For example, an important issue not explicitly addressed in this paper is the heterogeneity of the *migrant populations* (selection of migrants) in terms of socioeconomic and demographic characteristics, which is likely to be an important source of the variation in the path-specific bilateral effects (e.g., Millington, 2000). Unfortunately, though, the use of aggregate data puts obvious limits to the study of the spatial patterns of different "types" of migrants, as such data do not generally convey a great deal of information about "who moves to where." Also, much of the work presented here is still in an exploratory stage, and several points need more careful investigation. Nonetheless, the analysis has demonstrated the potential and flexibility of the GME estimator in dealing with spatially structured heterogeneity, tackling misspecification errors due to endogeneity and/or heteroskedasticity, and improving the identification strength of the estimation process by using additional (nonsample) information. This achievement should be seen as the most important methodological contribution of this paper (going beyond the particular case-study application), which, I hope, could appeal to other researchers for future work on gravity-type interaction models, both in cross-sectional and in panel-data settings.

## REFERENCES

Akwawua, Siaw and James A. Pooler. 2001. "The Development of an Intervening Opportunities Model with Spatial Dominance Effects," *Journal of Geographical Systems*, 3, 69–86.

Anderson, James E. and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170–192.

Anjomani, Ardeshir and Vida Hariri. 1992. "Migration Stock and the Issue of Competing and Complementary Flows in United States Interstate Migration," *Journal of Population Economics*, 5, 87–100.

Anselin, Luc. 2010. "Thirty Years of Spatial Econometrics," *Papers in Regional Science*, 89, 4–25.

Baltagi, Badi H., Peter Egger, and Michael Pfaffermayr. 2003. "A Generalized Design for Bilateral Trade Flow Models," *Economics Letters*, 80, 391–397.

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.

Behrens, Kristian, Cem Ertur, and Wilfried Koch. Forthcoming. "'Dual' Gravity: Using Spatial Econometrics to Control for Multilateral Resistance," *Journal of Applied Econometrics*, DOI: 10.1002/jae.1231.

Berk, Richard and John M. MacDonald. 2008. "Overdispersion and Poisson Regression," *Journal of Quantitative Criminology*, 24, 269–284.

Cadwallader, Martin T. 1992. *Migration and Residential Mobility: Macro and Micro Approaches*. University of Wisconsin Press.

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.

Campbell, Randall C. and R. Carter Hill. 2005. "A Monte Carlo Study of the Effect of Design Characteristics on the Inequality Restricted Maximum Entropy Estimator," *Review of Applied Economics*, 1, 53–84.

——. 2006. "Imposing Parameter Inequality Restrictions using the Principle of Maximum Entropy," *Journal of Statistical Computation and Simulation*, 76, 985–1000.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with Endogenous Moving Costs," *American Economic Review*, 86, 909–930.

Casetti, Emilio. 1972. "Generating Models by the Expansion Method: Applications to Geographical Research," *Geographical Analysis*, 4, 81–91.

Chun, Yongwan. 2008. "Modeling Network Autocorrelation Within Migration Flows by Eigenvector Spatial Filtering," *Journal of Geographical Systems*, 10, 317–344.

Clark, Ximena, Timothy J. Hatton, and Jeffrey G. Williamson. 2007. "Explaining U.S. Immigration, 1971–1998," *The Review of Economics and Statistics*, 89, 359–373.

Congdon, Peter. 1993. "Approaches to Modelling Overdispersion in the Analysis of Migration," *Environment and Planning A*, 25, 1481–1510.

——. 2010. "Random-Effects Models for Migration Attractivity and Retentivity: A Bayesian Methodology," *Journal of the Royal Statistical Society A*, 173, Part 4, 755–774.

Denslow, David A., Jr. and Peter J. Eaton. 1984. "Migration and Intervening Opportunities," *Southern Economic Journal*, 51, 369–387.

de Vries, Jacob J., Peter Nijkamp, and Piet Rietveld. 2009. "Exponential or Power Distance-Decay for Commuting? An Alternative Specification," *Environment and Planning A*, 41, 461–480.

Driscoll, John C. and Aart C. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *The Review of Economics and Statistics*, 80, 549–560.

Duan, Naihua. 1983. "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, 78, 605–610.

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.

Eldridge, J. Douglas and John Paul Jones, III. 1991. "Warped Space: A Geography of Distance Decay," *Professional Geographer*, 43, 500–511.

Epstein, Gil S. 2008. "Herd and Network Effects in Migration Decision-Making," *Journal of Ethnic and Migration Studies*, 34, 567–583.

Fischer, Manfred M. and Daniel A. Griffith. 2008. "Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union," *Journal of Regional Science*, 48, 969–989.

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.

Fotheringham, A. Stewart. 1991. "Migration and Spatial Structure: The Development of the Competing Destinations Model," in J. Stillwell and P. Congdon (eds.), *Migration Models: Macro and Micro Approaches*. London: Belhaven Press, pp. 57–72.

Fotheringham, A. Stewart, Chris F. Brunsdon, and Martin E. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Fotheringham, A. Stewart and Timothy C. Pitts. 1995. "Directional Variation in Distance Decay," *Environment and Planning A*, 27, 715–729.

Fraser, Iain. 2000. "An Application of Maximum Entropy Estimation: The Demand for Meat in the United Kingdom," *Applied Economics*, 32, 45–59.

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 Esfandiar 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.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–720.

Greene, William H. 1994. "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," Working Paper No. EC-94–10, Department of Economics, Stern School Business, New York University.

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 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 Behavior in a Developing Country: The Case of Mexico," *Demography*, 18, 369–388.

Hatton, Timothy J. and Jeffrey G. Williamson. 2011. "Are Third World Emigration Forces Abating?" *World Development*, 39, 20–32.

Head, Keith and Thierry Mayer. 2002. "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade," Working Paper No. 01, CEPII Research Center, Paris.

Henderson, Daniel J. and Daniel L. Millimet. 2008. "Is Gravity Linear?" *Journal of Applied Econometrics*, 23, 137–172.

Hu, Pingzhao and Jim Pooler. 2002. "An Empirical Test of the Competing Destinations Model," *Journal of Geographical Systems*, 4, 301–323.

Hunt, Jennifer. 2006. "Staunching Emigration from East Germany: Age and the Determinants of Migration," *Journal of the European Economic Association*, 4, 1014–1037.

Jones, Richard C. 2001. "Maquiladoras and U.S.-Bound Migration in Central Mexico," *Growth and Change*, 32, 193–216.

Kim, Min Seong and Yixiao Sun. 2011. "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix," *Journal of Econometrics*, 160, 349–371.

Lewis, Jeffrey B. and Drew A. Linzer. 2005. "Estimating Regression Models in Which the Dependent Variable is Based on Estimates," *Political Analysis*, 13, 345–364.

Magrini, Marie-Benoît and Philippe Lemistre. Forthcoming. "Distance-Income Migration Trade-Off of Young French Workers: An Analysis per Educational Level," *Regional Studies*, DOI: 10.1080/00343404.2011.579595.

Manning, Willard G. and John Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics*, 20, 461–494.

McMillen, Daniel P. 2010. "Issues in Spatial Data Analysis," *Journal of Regional Science*, 50, 119–141.

Millington, Jim. 2000. "Migration and Age: The Effect of Age on Sensitivity to Migration Stimuli," *Regional Studies*, 34, 521–533.

Mueser, Peter R. 1989. "The Spatial Structure of Migration: An Analysis of Flows Between the States in the USA Over Three Decades," *Regional Studies*, 23, 185–200.

OECD. 2004. *Economic Surveys 2002–2003: Mexico*, Vol. 2003–Supplement No. 1, Jan. 2004. Organization for Economic Co-operation and Development, Paris.

Pakko, Michael R. and Howard J. Wall. 2001. “Reconsidering the Trade-Creating Effects of a Currency Union,” *The Federal Reserve Bank of St. Louis Review*, 83, 37–45.

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. 2008. “Selection and Network Effects—Migration Flows into OECD Countries 1990–2000,” *European Economic Review*, 52, 1160–1186.

Pellegrini, Pasquale A. and A. Stewart Fotheringham. 2002. “Modelling Spatial Choice: A Review and Synthesis in a Migration Context,” *Progress in Human Geography*, 26, 487–510.

Pinkse, Joris and Margaret E. Slade. 2010. “The Future of Spatial Econometrics,” *Journal of Regional Science*, 50, 103–117.

Plane, David A. 1982. “An Information Theoretic Approach to the Estimation of Migration Flows,” *Journal of Regional Science*, 22, 441–456.

Plane, David A. 1984. “Migration Space: Doubly Constrained Gravity Model Mapping of Relative Interstate Separation,” *Annals of the Association of American Geographers*, 74, 244–256.

Plane, David A. 1999. “Migration Drift,” *The Professional Geographer*, 51, 1–11.

Pregibon, Daryl. 1980. “Goodness of Link Tests for Generalized Linear Models,” *Applied Statistics*, 29, 15–24.

Pukelsheim, Friedrich. 1994. “The Three-Sigma Rule,” *American Statistician*, 48, 88–91.

Ramsey, J.B. 1969. “Tests of Specification Errors in Classical Linear Least Square Regression Analysis,” *Journal of the Royal Statistical Society B*, 31, 350–371.

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 and James Raymer. 1998. “The Spatial Focus of U.S. Interstate Migration Flows,” *International Journal of Population Geography*, 4, 63–80.

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 Tenreyo. 2006. “The Log of Gravity,” *The Review of Economics and Statistics*, 88, 641–658.

Sasser, Alicia C. 2010. “Voting With Their Feet: Relative Economic Conditions and State Migration Patterns,” *Regional Science and Urban Economics*, 40, 122–135.

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.

Shannon, Claude E. 1948. “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 27, 379–423.

Shen, Edward Z. and Jeffrey M. Perloff. 2001. “Maximum Entropy and Bayesian Approaches to the Ratio Problem,” *Journal of Econometrics*, 104, 289–313.

Stouffer, Samuel A. 1960. “Intervening Opportunities and Competing Migrants,” *Journal of Regional Science*, 2, 1–26.

Swamy, P.A.V.B. and George S. Tavlas. 1995. “Random Coefficient Models: Theory and Applications,” *Journal of Economic Surveys*, 9, 165–196.

Tiefelsdorf, Michael. 2003. “Misspecification in Interaction Model Distance Decay Relations: A Spatial Structure Effect,” *Journal of Geographical Systems*, 5, 25–50.

Tiefelsdorf, Michael and Daniel A. Griffith. 2007. “Semi-Parametric Filtering of Spatial Autocorrelation: The Eigenvector Approach,” *Environment and Planning A*, 39, 1193–1221.

Tobler, Waldo R. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region,” *Economic Geography*, 46, 234–240.

——. 1995. “Migration: Ravenstein, Thornthwaite, and Beyond,” *Urban Geography*, 16, 327–343.

Wilson, Alan G. 1970. *Entropy in Urban and Regional Modelling*. London: Pion.

Wooldridge, Jeffrey M. 1997. “Quasi-Likelihood Methods for Count Data,” in M.H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics*, Vol. 2. Oxford: Blackwell, pp. 352–406.

——. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

——. 2004. “Estimating Average Partial Effects under Conditional Moment Independence Assumptions,” CEMMAP Working Paper CWP03/04, The Institute for Fiscal Studies, Department of Economics, UCL.

——. 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.

Zietz, Joachim. 2001. “Heteroskedasticity and Neglected Parameter Heterogeneity,” *Oxford Bulletin of Economics and Statistics*, 63, 263–273.

## APPENDIX A: DATA SOURCES

The data on interstate migration flows and bilateral migrant stocks in Mexico are estimations (extrapolations) compiled from the *Consejo Nacional de Población* (CONAPO: http://www.conapo.gob.mx), which, in turn, are based on the 1995 and 2000 censuses of the *Instituto Nacional de Estadística, Geografía e Informática* (INEGI, Censo General de Población y Vivienda, 1995, 2000: http://www.inegi.org.mx).

Migration flows (of persons over five years of age) are for the five-year observation period 1995–2000, while migrant stocks are for 1995 (at the beginning of the five-year observation period 1995–2000). Bilateral migration flows come in the form of simple counts, where movements between states are defined as changes of *place of residence* from state $i$ (the state where the migrant lived five years prior to the 2000 census) to state $j$ (the state where the migrant lives at the time of the 2000 census). (An obvious drawback of this definition is that observed migration flows mask whether, for example, a person counted as a migrant from $i$ to $j$ in the 2000 census moved from $i$ to $k$ in, say, 1996 and subsequently moved from $k$ to $j$ in, say, 1999.) Conversely, bilateral migrant stocks are cross-tabulated by *place of birth*; that is, the migrant stock in $j$ represents the (net) accumulated number of previous in-migrants to $j$ that are born in $i$ and are (still) living in $j$ in 1995. (This definition creates another problem, in the sense that persons moving from $i$ to $j$ during the observation period are not necessarily born in $i$, so that full consistency between migration flows and migrant stocks is not ensured.)

The data on (state-level) socioeconomic and demographic fundamentals (origin-push and destination-pull factors) come from two sources. Data on population sizes (for 1995), unemployment rates (for 1990 and 2000), and employment shares of manufacturing (for 1995) are taken from INEGI. Data on the Human Development Index (for 1990 and 2000) are extracted from the *Population Reference Bureau* (http://www.prb.org/Countries/Mexico.aspx).

Finally, highway distances (in kilometers) between the capital cities of Mexico’s states are calculated on the basis of information from the *Secretaría de Comunicaciones y Transportes* (SCT: http://www.sct.gob.mx).

## APPENDIX B

FIGURE B1: Base Map of Mexico’s States.


| *Code* | *State* | *Code* | *State* |
|---|---|---|---|
| 01 | Aguascalientes | 17 | Morelos |
| 02 | Baja California Norte | 18 | Nayarit |
| 03 | Baja California Sur | 19 | Nuevo León |
| 04 | Campeche | 20 | Oaxaca |
| 05 | Chiapas | 21 | Puebla |
| 06 | Chihuahua | 22 | Querétaro |
| 07 | Coahuila | 23 | Quintana Roo |
| 08 | Colima | 24 | San Luis Potosí |
| 09 | Distrito Federal | 25 | Sinaloa |
| 10 | Durango | 26 | Sonora |
| 11 | Guanajuato | 27 | Tabasco |
| 12 | Guerrero | 28 | Tamaulipas |
| 13 | Hidalgo | 29 | Tlaxcala |
| 14 | Jalisco | 30 | Veracruz |
| 15 | México | 31 | Yucatán |
| 16 | Michoacán | 32 | Zacatecas |