Ann Reg Sci (2008) 42:375–389
DOI 10.1007/s00168-007-0166-y

ORIGINAL PAPER

On the role of distance for outward FDI

Peter Egger

Received: 1 January 2006 / Accepted: 1 June 2007 / Published online: 1 November 2007

Abstract This paper focuses on the estimation of three distance-related effects on outward FDI. (1) Distance harms vertical multinationals, since they engage in trade. (2) It makes non-trading multinationals better off than exporters. (3) This positive effect on horizontal FDI is expected to rise with bilateral parent and host market size. The use of panel data and related econometric methods is highly recommended to avoid parameter bias from endogenous, unobserved, time-invariant effects. A unified estimation approach to assess all three hypotheses then has to rely on instrumental variable techniques for generalized least-squares methods. In the empirical analysis of 1989-1999 bilateral US outward FDI stocks at the industry level, it is shown that testing and accounting for autocorrelation is extremely important for parameter inference. In sum, the paper lends strong support to the theory of horizontally organized multinationals as outlined in Markusen and Venables (J Int Econ 52(2):209–234).

JEL Classification C3 · F14 · F15

1 Introduction

It is this paper's purpose to shed further light on the role of distance for foreign direct investment (FDI). For this, I apply the simplifying assumption that trade costs are associated with distance.[1] Given this, previous research motivates two distance-related effects on FDI: a positive one in models of horizontal FDI, where the decision of going multinational is determined by the trade-off between proximity to the market

[1] For instance, Hummels (2001) and Limao and Venables (2001) build on such an approach and independently find that trade costs (c.i.f./f.o.b.) rise with distance at an elasticity of 0.3. See Anderson (2000) and Anderson and van Wincoop (2004) for an overview of the related literature.

P. Egger (✉)
Ifo Institute and University of Munich, Schackstrasse 4, 80539 Munich, Germany
e-mail: egger@ifo.de

and concentration of production facilities at the firm-level (Brainard 1997), and a negative one in models of vertical FDI, since multinational enterprise (MNE) activity and trade are complementary there. Bearing in mind the overwhelming support of previously found size and factor endowment parameters (not necessarily of distance) in favor of horizontal FDI, the paper additionally shows—in a stylized, three-factor variant of the model of horizontal MNEs—that distance not only positively affects horizontal FDI per se but that its marginal effect also rises with bilateral country size.

Geographical distance is known to be one of the most important obstacles to FDI.[2] Looking at the determinants of US foreign affiliate sales in levels, Carr et al. (2001) and Markusen and Maskus (2002) estimate a distance parameter of -1.5. Evaluated at the data mean, this suggests that a one percent increase in distance is associated with a decline in US foreign affiliate sales by -0.79% percent. Running alternative specifications on the same data, Blonigen et al. (2003) even identify parameter estimates of in between -2.1 and -3.1, which imply elasticities of -1.11 and -1.64, respectively. Cross-sectional inference on outward FDI (rather than foreign affiliate sales) of the member countries of the Organisation for Economic Cooperation and Development (OECD) in Blonigen et al. (2003) points to an elasticity of FDI with respect to distance of about -0.35 when evaluated at the data mean. Martín and Velázquez (1997) report one of -0.6.[3]

This evidence nourishes two different interpretations. First, it could be seen as an indication for (a dominance of) *vertical* multinational firm activity, because of the association of sheer distance with trade costs and the notion that vertical multinationals engage in trade (Helpman 1984). Second, it could indicate that foreign plant set-up costs are positively correlated with distance so that a negative distance parameter could also arise in the case of *horizontal*, nontrading multinationals (Markusen and Venables 2000). However, since empirical evidence generally tends to lend much more support on the horizontal view of FDI (see Carr et al. 2001; Markusen and Maskus 2002),[4] one is tempted to favor the latter interpretation over the former. Since the range of available parameter estimates is huge and their interpretation is crucial from a theoretical perspective, thorough inference on the role of distance and a better understanding from a theoretical viewpoint are required.

This paper investigates three empirically testable hypotheses of the distance effect on FDI: (1) FDI declines with distance if relative factor endowment differences are sufficiently large and if vertical rather than horizontal FDI takes place; (2) FDI rises with distance, if factor endowment differences are minimal and horizontal FDI

[2] The role of distance for trade is now well understood from research on the so-called gravity equation (see Anderson 1979; Bergstrand 1985, 1989; Anderson 2000; Anderson and van Wincoop 2004).

[3] Obviously, the range previously estimated elasticities is huge, which is partly due to biased parameter estimates. There are two major sources of this bias. First, the estimates tend to be based on cross-sectional data, and we know that these are often biased due to omitted variables. This bias can be avoided with panel data at hand. Second, previous research often employs data in levels rather than logarithms, while recent research indicates that specifications using FDI or foreign affiliate sales in levels are typically rejected against their alternatives in logarithms (see Mutti and Grubert 2004).

[4] Also, this shows up in much higher bilateral intra-OECD FDI figures than in FDI of the OECD economies with non-OECD members.

dominates; (3) FDI should rise even more strongly with distance, the larger the parent and host markets are together.

These hypotheses are tested in a panel data set of bilateral US outward FDI in seven manufacturing industries over the period 1989–1999. The panel econometric setting allows to control for all unobserved influences which are industry-by-host-country specific. In this way, time-invariant legal or market access variables can be conveniently controlled for. However, there are good reasons to believe that explanatory variables such as market size and, especially, bilateral distance are correlated with time-invariant unobserved factors such as legal, cultural, institutional and other determinants. If so, not only the between (i.e., the cross-sectional) estimator but also the random effects panel data estimator is biased.[5] Unfortunately, although the fixed effects estimator naturally overcomes this type of bias, it renders the second of the three hypotheses impossible to test and it does not allow the computation of a marginal distance effect. To circumvent this problem, I apply the two-stage generalized least squares (GLS) methods outlined in Hausman and Taylor (1981); Amemiya and MaCurdy (1986), and Breusch et al. (1989) to retrieve the main effect of distance and estimate the parameters consistently. Moreover, I relax the assumption that the stochastic shocks are uncorrelated over time to obtain efficient parameter estimates, following Baltagi and Wu (1999). In this regard, Baltagi (2001, p. 81) mentions that modeling shocks as uncorrelated over time is "*a restrictive assumption for economic relationships, like investment ..., where an unobserved shock this period will affect the behavioral relationship for at least the next few periods.*"

2 Theoretical background

Whereas the first hypothesis (of a negative impact of distance on vertical outward FDI) and the second hypothesis (of a positive impact of distance on horizontal outward FDI) are derived in previous research (see Markusen 2002, for an overview), as long as we directly associate distance with trade costs, the third hypothesis has not yet been addressed. To derive it, assume a highly stylized, one differentiated goods sector (Dixit and Stiglitz 1977), two-country, three-factor (labor L, human capital H, physical capital K) model of exporting national firms and locally producing horizontal multinationals only.[6] The distinction between H and K is essential to establish a theory of FDI, where headquarters serve their affiliates not only with (invisible) firm-specific assets but also with (visible) capital. The assumption of differentiated goods and the endogenous decision to go multinational renders an analytical treatment of general equilibrium models in many cases impossible. However, the third hypothesis, above, may be investigated analytically once we are willing to assume symmetric countries

[5] Previous research mostly relies on pooled OLS or weighted least-squares (Carr et al. 2001; Blonigen et al. 2003), where the same arguments apply. Also the fixed industry and time effects estimates reported in Hanson et al. (2002) are potentially affected by omitted cross-sectional effects, and their estimates should not be interpreted as classical 'Within' estimates.

[6] I do not consider vertical MNEs (Helpman 1984), complex MNEs (Yeaple 2003; Grossman et al. 2006), or export-platform MNEs (Ekholm et al. 2007).

in every respect (absolute factor endowments, trade costs t, and foreign plant set-up costs g).[7]

Given the symmetry, one can skip the country indices in the theoretical exposition. It is useful to adopt the assumption of access to the same production technology for each country's exporters (n) and horizontal MNEs (m) as in Markusen and Venables (2000); to restrict the matrix of input coefficients associated with clearing of all three factor markets for analytical tractability, so that only L is used in production; and to choose L's factor reward as the numéraire ($w_L = 1$). Then, the locally sold quantities of each horizontally differentiated variety in equilibrium are

$$x = s^{\varepsilon-1}E = \frac{E}{[n(1+t^{1-\varepsilon})+2m]} \tag{1}$$

where s is the constant elasticity of scale (CES) price index under complete symmetry, ε is the elasticity of substitution between varieties, and $E = L + w_H H + w_K K$ is national income (GNP) with w_H and w_K as the respective factor rewards of human and physical capital. In the denominator of the last term in (1), i.e., in $s^{\varepsilon-1}$, $2m$ indicates that there is two-way FDI, and t refers to iceberg type transport costs ($t-1$ units are lost during transportation across borders). Note that under our assumptions $x^* t^{1-\varepsilon}$ is the export volume of each n-type firm in equilibrium.

Further, let both H and K be required to generate firm-specific assets on the one hand and to set-up plants on the other (it is a common notion in the literature that firm set-up uses skilled labor intensively while goods production uses unskilled labor intensively, see Markusen 2002), but let the capital requirement of horizontal MNEs be higher than that one of national exporters for two reasons. First, they have to establish production facilities both at home and abroad. Second, foreign plant set-up requires even more capital than domestic plant set-up (the difference between the two being $g-1$). Clearing of all three factor markets under these conditions requires [8]

$$\begin{aligned} L &= x[n(1+t^{1-\varepsilon})+2m] \\ H &= n+m \\ K &= n+(1+g)m. \end{aligned} \tag{2}$$

Of course, the equilibrium number of firms is determined by the last two conditions, and we have $m^* = \frac{K-H}{g}$ and $n^* = \frac{(1+g)H-K}{g}$.[9] Inserting into the market clearing condition for L yields

$$x^* = \frac{gL}{gH(1+t^{1-\varepsilon})+(t^{1-\varepsilon}-1)(H-K)}. \tag{3}$$

7 This is sufficient, since an interaction term between bilateral overall country size and distance is to be derived.

8 For the ease of presentation, unitary input coefficients in production, firm set-up and domestic plant set-up are assumed.

9 Note the obvious restriction on the difference between endowments with H and K to ensure coexisting national exporters and MNEs in equilibrium.

If the number of active firms is sufficiently large, firms apply a fixed mark-up over marginal costs, so that each variety sells at $p^* = \frac{\varepsilon}{\varepsilon - 1}$ in equilibrium. Finally, free entry and exit of firms eliminates all profits (π) exceeding fixed costs for both types of firms:

$$\begin{aligned} \pi_n^* &= \frac{\varepsilon}{\varepsilon - 1}(1 + t^{1-\varepsilon})x^* - w_H - w_K = 0 \\ \pi_m^* &= \frac{\varepsilon}{\varepsilon - 1}2x^* - w_H - (1+g)w_K = 0. \end{aligned} \tag{4}$$

Accordingly, the equilibrium capital rental is determined as

$$w_K^* = \frac{\varepsilon}{\varepsilon - 1}\frac{1 - t^{1-\varepsilon}}{g}x^* = \frac{\varepsilon}{\varepsilon - 1}\frac{L(1 - t^{1-\varepsilon})}{gH(1 + t^{1-\varepsilon}) + (t^{1-\varepsilon} - 1)(H - K)}. \tag{5}$$

Note that a country's equilibrium outward FDI in the sense of its physical capital delivery to foreign affiliates in this model is simply $gm^* w_K^* = (K - H)w_K^*$. Hence, the comparative statics of FDI with respect to t (associated with distance) becomes

$$\frac{\partial \text{FDI}}{\partial t} = \frac{\varepsilon(K - H)Lt^{-\varepsilon}}{A}\left[1 + \frac{(1 - t^{1-\varepsilon})[(1+g)H - K]}{A}\right] > 0, \tag{6}$$

where $A = gH(1 + t^{1-\varepsilon}) + (t^{1-\varepsilon} - 1)(H - K)$ has been used. Further, let v denote a simple scaling factor to change "initial" world factor endowments (indexed by 0) and, hence, bilateral country size, and rewrite the marginal effect of FDI with respect to t:

$$\begin{aligned} \frac{\partial \text{FDI}}{\partial t} &= v\frac{\varepsilon(K_0 - H_0)L_0 t^{-\varepsilon}}{A_0}\left[1 + \frac{(1 - t^{1-\varepsilon})[(1+g)H_0 - K_0]}{A_0}\right] \\ &\Longrightarrow \frac{\partial^2 \text{FDI}}{\partial t \partial v} > 0. \end{aligned} \tag{7}$$

Accordingly, we may conclude that horizontal FDI does not only positively depend on the level of trade costs (distance), but that this marginal effect rises with bilateral country size. The latter effect is due to the home market effect induced by transport costs.

3 Econometric specification

Based on numerical simulations, it can be shown (see Egger and Pfaffermayr 2005) that the above model of trade and horizontal FDI motivates a specification which accounts for overall country size ($G_{ij}^+ = \text{GDP}_i + \text{GDP}_j$), relative country size ($r_{ij}^{+/-} = \text{GDP}_i / \text{GDP}_j$,[10] not necessarily similarity in country size), parent-to-host physical

[10] Since horizontal FDI is local market seeking, we would expect r_{ij}^+ in this case. See also Barrios et al. (2004).

capital endowments ($k_{ij}^{+} = K_i/K_j$), parent-to-host human capital endowments ($h_{ij}^{+} = H_i/H_j$), parent-to-host labor endowments ($l_{ij}^{+/-} = l_i/l_j$),[11] and two distance terms: the main effect (D_{ij}^{+}) and the interaction term with bilateral size ($(G_{ij} \times D_{ij})^{+}$). To capture the possible presence of vertical MNEs at sufficiently different relative factor endowments, a third distance-related term should arise accounting for the fact that distance harms vertical FDI ($[\Delta(K/L)_{ij} \times D_{ij}]^{-}$, with $\Delta K/L = |K_i/L_i - K_j/L_j|$). Of course, empirically we expect horizontal and vertical MNEs to coexist. In sum, the following empirical model for U.S. outward FDI in industry k of host country j and year t can be formulated (skipping the index i since the US is the only parent economy considered):

$$F_{kjt} = \beta_0 + \beta_1 r_{jt} + \beta_2 k_{jt} + \beta_3 h_{jt} + \beta_4 l_{jt} \\ + \beta_5 (G_{jt} \times D_j) + \beta_6 (\Delta(K/L)_{jt} \times D_j) + \beta_7 D_j + u_{kjt}, \quad (8)$$

where the main effect of G_{jt} has been excluded due to its irrelevance in the application below. (Of course, G_{jt} is allowed to exert an impact through the interaction term with D_j). All variables are real figures and expressed in logs, and the error term can be written as

$$u_{kjt} = \mu_{kj} + \nu_{kjt} \quad (9)$$

with μ_{kj} as the (fixed or random) unobserved industry–host country effects, which capture all time-invariant, industry-specific legal or market access issues. ν_{kjt} is the remainder error term.

However, there are good reasons to believe that explanatory variables such as market size and, especially, bilateral distance are correlated with time-invariant unobserved factors of influence like legal, cultural, institutional and other determinants.[12] In this case, not only the between estimator, but also the random effects estimator is biased. Unfortunately, the consistent fixed effects estimator renders the second of the three hypotheses impossible to test. To recover the main effect of distance and estimate the parameters consistently, the two-stage generalized least squares (GLS) methods outlined in Hausman and Taylor (1981); Amemiya and MaCurdy (1986), and Breusch et al. (1989) are applied. Moreover, following the aforementioned argument by Baltagi (2001), the assumption that the stochastic shocks (ν_{kjt}) are uncorrelated over time is relaxed to obtain efficient parameter estimates. Since the available panel is unbalanced and unequally spaced, the empirical analysis builds on Baltagi and Wu (1999). The estimation strategy proceeds in the following steps:

11 Note that given other factor endowments and bilateral total labor endowment, a reallocation of labor from the parent to the host induces higher production costs in the parent relative to the host. A negative impact of l_{ij} on outward FDI points to the importance of low-cost seeking vertical FDI, whereas a positive one implicitly supports the importance of horizontal FDI.

12 Examples of these variables would be the rule of law or the quality of the legal system in the host country with an expected positive impact, a common language between the parent and the host country with an expected positive effect, and geographical or climatic factors.

1. Prais–Winsten transform the data as suggested in Baltagi and Wu (1999).[13]
2. Obtain the Amemiya (1971) type residuals (u^*), let ρ denote the autocorrelation parameter, and define

$$g_i = (1-\rho^2)^{1/2}\left(1, \frac{1-\rho^{(t_{i,2}-t_{i,1})}}{(1-\rho^{2(t_{i,2}-t_{i,1})})^{1/2}}, \ldots, \frac{1-\rho^{(t_{i,n_i}-t_{i,n_i-1})}}{(1-\rho^{2(t_{i,n_i}-t_{i,n_i-1})})^{1/2}}\right)' \quad (10)$$

with $P_{g_i} = g_i(g_i'g_i)^{-1}g_i'$ and $Q_{g_i} = I_{n_i} - P_{g_i}$. Estimate the Within variance component by

$$\widehat{\sigma}_\varepsilon^2 = u^{*\prime}\mathrm{diag}(Q_{g_i})u^* \Big/ \left(\sum_{i=1}^{N}(n_i-1)\right), \quad (11)$$

where N refers to the number of cross sections and n_i is the number of observations in cross-section i. The Within transformed model according to Baltagi and Wu (1999) is

$$y_{i,t_{i,j}}^{W**} = y_{i,t_{i,j}}^* - g_{i,j}\left(\sum_{s=1}^{n_i} g_{i,s}y_{i,t_{i,s}}^*\right) \Big/ \left(\sum_{s=1}^{n_i} g_{i,s}^2\right). \quad (12)$$

3. In the presence of correlation between (some of) the explanatory variables (X_2^*) and the unobserved effects (μ_{kj}), we have to average the within residuals over time (i.e., to construct pseudo-averages) and to run 2SLS of these residuals on the time-invariant, Prais–Winsten transformed variables (Z_2^*) with the Prais–Winsten transformed, doubly-exogenous variables (X_1^*, Z_1^*) as instruments.[14] This regression obtains (1) a parameter estimate for the time-invariant variables and (2) produces residuals, which are used to derive the second required variance component. Let the residuals from this second regression be η^*. An estimate of the second required variance component is

$$\widehat{\sigma}_\omega^2 = \eta^{*\prime}\mathrm{diag}(P_{g_i})\eta^*. \quad (13)$$

Accordingly, an estimate for the cross-sectional variance component is

$$\widehat{\sigma}_\mu^2 = \left(\eta^{*\prime}\mathrm{diag}(P_{g_i})\eta^* - N\widehat{\sigma}_\varepsilon^2\right) \Big/ \sum_{i=1}^{N} g_i'g_i, \quad (14)$$

[13] The Prais–Winsten approach can be used as a remedy for for "pure autocorrelation" (i.e., autocorrelatd residuals). Autocorrelation in the residuals also arises in case of an omission of a relevant lagged dependent variable in the model. The latter would lead to a parameter bias. However, the static fixed effects estimator is still a good approximation of the short-run impact in this case (see Egger and Pfaffermayr 2004). With the data at hand, using a lagged dependent variable leads to a dramatic loss of observations due to missing data in the panel. Here, we use a first-order autoregressive model. While the approach may be used with higher order autocorrelations in principal, the unbalancedness of the panel prevents their use, here.

[14] According to Cornwell et al. (1992), we call the variables correlated with μ_{kj} *singly exogenous* and the uncorrelated ones *doubly exogenous*.

which gives

$$\widehat{\omega}_i^2 = g_i' g_i \widehat{\sigma}_\mu^2 + \widehat{\sigma}_\varepsilon^2 \quad \text{and} \quad \widehat{\theta}_i = 1 - \left(\frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\omega}_i^2} \right)^{1/2}. \tag{15}$$

4. Finally, pre-multiply the Prais–Winsten transformed data according to Fuller and Battese (1973, 1974) by $\sigma_\varepsilon \Omega^{*-1/2}$ to get $y^{**} = \sigma_\varepsilon \Omega^{*-1/2} y^*$ with typical elements

$$y_{i,t_{i,j}}^{**} = y_{i,t_{i,j}}^{*} - \theta_i g_{i,j} \left(\sum_{s=1}^{n_i} g_{i,s} y_{i,t_{i,s}}^{*} \right) \Big/ \left(\sum_{s=1}^{n_i} g_{i,s}^2 \right). \tag{16}$$

5. Running 2SLS on the transformed model with the proper set of instruments (A) yields the consistent and efficient AR(1) estimators in the spirit of Hausman and Taylor (1981, HT), Amemiya and MaCurdy (1986, AM), and Breusch et al. (1989, BMS). Note that the appropriate instruments for these estimators under AR(1) are $A_{HT} = [\widetilde{X}_1^*, \widetilde{X}_2^*, \overline{X}_1^*, Z_1^*]$, $A_{\text{AM}} = [\widetilde{X}_1^*, \widetilde{X}_2^*, X_1^{\tau *}, Z_1^*]$, and $A_{\text{BMS}} = [\widetilde{X}_1^*, \widetilde{X}_2^*, \overline{X}_1^*, \widetilde{X}_1^{\tau *}, \widetilde{X}_2^{\tau *}, Z_1^*]$, where "$*$" refers to Prais–Winsten transformed variables, "~" indicates Within transformed variables according to (12), and "–" denotes pseudo-averages over time ($\overline{y}_i^* = \text{diag}(P_{gi}) y_i^*$) of the *doubly exogenous* variables. Finally,

$$X_1^{\tau *} = \begin{bmatrix} X_{1,11}^* & \cdots & X_{1,T1}^* \\ \vdots & \ddots & \vdots \\ X_{1,1N}^* & \cdots & X_{1,TN}^* \end{bmatrix} \otimes \iota_T \tag{17}$$

and similarly for $\widetilde{X}_1^{\tau *}$ and $\widetilde{X}_2^{\tau *}$. In our application, the set of time-invariant, *singly exogenous* variables $[Z_2^*]$ comprises D_j^*, and there are neither time-variant, *singly exogenous* variables $[X_2^*]$ nor, time-invariant *doubly exogenous* ones $[Z_1^*]$. Since there are obviously more columns in X_1 than in Z_2, the HT–AR(1), AM–AR(1) and BMS–AR(1) are more efficient than (fixed effects) FE–AR(1).

4 Data and estimation results

The data base comprises a panel of US outward FDI stocks (Bureau of Economic Analysis) in seven industries[15] and 69 countries[16] over the period 1989–1999. To

[15] Food and kindred products, Chemicals and allied products, Primary and fabricated metals, Industrial machinery and equipment, Electronic and other electric equipment, Transportation equipment, Other manufacturing.

[16] The included host countries are: Argentina, Australia, Austria, Brazil, Canada, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Gabon, Germany, Ghana, Greece, Guatemala, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Kenya, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Russia, Senegal, Singapore, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Trinidad and Tobago, Tunisia, Turkey, United Kingdom, Uruguay, Venezuela, Zaire, Zambia, Zimbabwe.

construct the explanatory variables, I rely on the World Bank's World Development Indicators. Specifically, real GDP (in 1995 US dollars), gross fixed capital formation (to construct real capital stocks based on the perpetual inventory method in Leamer 1984), tertiary school enrolment shares (as a proxy of skilled labor endowments), labor force, and the bilateral greater circle distance between capitals of the host countries and the US (own calculations) are used.

Table 1 presents the results of pooled OLS, random effects (RE), fixed effects (FE), HT, AM, and BMS regressions with (bottom) and without (top) considering the underlying AR(1) process. The results can be summarized as follows.

The consistent FE results illustrate that the US invests mainly in large host countries, which is consistent with market-seeking horizontal outward FDI. Also the results for the relative factor endowment variables are consistent with the above model of horizontal FDI, particularly showing that outward FDI rises with the parent-to-host capital endowment ratio (Egger and Pfaffermayr 2005; Bergstrand and Egger 2007). Similarly, $\widehat{\beta}_4 > 0$ underpins that U.S. outward FDI is on average not low-cost seeking (vertical). Independent of whether the AR(1) process is ignored or not, the pooled OLS and RE models perform poorly (see the Hausman test for the latter). Especially, all parameters related to factor endowment variables ($\widehat{\beta}_2$, $\widehat{\beta}_3$, $\widehat{\beta}_4$, $\widehat{\beta}_6$) are affected. This confirms our view that the correlation between variables such as the sheer geographical distance and unobserved determinants such as legal standards, institutional factors or market access regulations as captured by the industry-host country specific error component may lead to biased parameter estimates. Noteworthily, $\widehat{\beta}_5 > 0$ lends support to the second distance-related hypothesis outlined in Sect. 2 relating to horizontal outward FDI. Additionally, $\widehat{\beta}_6 < 0$ indicates that, in line with the third distance-related hypothesis, distance exhibits a smaller positive and eventually a negative impact on US outward FDI at sufficiently large factor endowment differences, where vertical, trading MNEs are more likely to exist or even dominate. As mentioned before, the main effect of distance (hypothesis one), cannot be assessed with the fixed effects estimator.

This effect is successfully retrieved by the HT, AM, and BMS models in the upper bloc of results of Table 1. The negative sign of $\widehat{\beta}_7$ motivates the two explanations from above: distance is either important for foreign plant set-up costs, or vertical MNEs (or more complex types of trading MNEs as introduced in Ekholm et al. 2007; Yeaple 2003, or Grossman et al. 2006) dominate. However, three concerns have to be raised with respect to those estimates. First, the HT estimator is almost rejected at 10% (the corresponding p-value is lower than 0.11), also showing up in an obvious difference to the FE model in terms of $\widehat{\beta}_2$, $\widehat{\beta}_4$, and $\widehat{\beta}_6$.[17] Second, the bias resulting from the relative weak exogeneity of the instruments is even larger in the potentially more efficient AM and BMS models (the instruments are highly relevant as indicated by the partial R^2 figures which reflect the explanatory power of the instruments in the first stage model). This shows up in the significant Hausman test statistics reported in the bottom row of the upper bloc of results in the respective columns

[17] Note that there is no better result to achieve with the specification at hand.

Table 1 Estimation results on the role of distance for US outward FDI

		OLS	RE	FE	HT	AM	BMS
$\ln(r_{jt})$	β_1	−1.731***	−0.996***	−1.554***	−1.261***	−0.728***	−1.098***
		0.129	0.167	0.226	0.200	0.243	0.191
$\ln(k_{jt})$	β_2	−0.286	2.105*	15.648***	22.941***	7.757***	11.959***
		0.463	1.164	4.550	3.618	2.858	3.814
$\ln(h_{jt})$	β_3	−0.068	0.109	0.332***	0.211**	0.219**	0.250**
		0.077	0.098	0.112	0.107	0.111	0.102
$\ln(l_{jt})$	β_4	0.250***	0.500***	2.391***	0.751***	0.771***	0.761***
		0.045	0.087	0.404	0.154	0.199	0.146
$\ln(G_{jt}) \times \ln(D_j)$	β_5	0.076***	0.330***	0.367***	0.356***	0.345***	0.360***
		0.012	0.013	0.014	0.013	0.015	0.013
$\ln \Delta(K/L)_{jt} \times \ln(D_j)$	β_6	0.135***	−0.187	−1.793***	−2.606***	−0.846***	−1.348***
		0.052	0.133	0.518	0.415	0.326	0.438
$\ln(D_j)$	β_7	−3.212***	−9.186***	–	−5.182***	−9.144***	−8.972***
		0.326	0.465	–	1.901	0.733	0.926
Constant	β_0	15.046***	6.537**	–	−33.837**	3.190	−2.144
		0.987	2.807	–	15.736	5.209	7.158
Standard error		1.613	0.578	0.580	0.567	0.420	0.560
Instrument relevance (partial R^2)		–	–	–	0.93	0.94	0.96
Hausman and Hausman–Taylor tests		–	χ^2 (6)=177.01***	–	χ^2 (5)=8.98	χ^2 (8)=173.74***	χ^2 (8)=139.27***
		Variance component and average GLS weights for HT, AM, and BMS: $\hat{\sigma}_\varepsilon^2 = 0.333$, $\hat{\sigma}_\mu^2 = 9.245$, $\hat{\theta} = 0.935$					

Table 1 continued

		AR(1) models					
		OLS	RE	FE	HT	AM	BMS
$\ln(r_{jt})$	β_1	−1.091***	−1.118***	−1.624***	−1.584***	−1.610***	−1.473***
		0.173	0.176	0.244	0.199	0.204	0.392
$\ln(k_{jt})$	β_2	−0.035	−0.063	5.012***	4.755***	4.607***	3.929***
		0.126	0.119	0.931	0.714	0.809	1.247
$\ln(h_{jt})$	β_3	0.301***	0.348***	0.009	0.028	0.015	0.006
		0.086	0.088	0.171	0.138	0.141	0.269
$\ln(l_{jt})$	β_4	−0.341	0.478	0.219*	0.255***	0.255***	0.182
		1.115	1.131	0.115	0.091	0.093	0.162
$\ln(G_{jt}) \times \ln(D_j)$	β_5	0.233***	0.270***	0.021***	0.023***	0.022***	0.024***
		0.021	0.017	0.001	0.001	0.001	0.001
$\ln \Delta(K/L)_{\mathrm{jt}} \times \ln(D_j)$	β_6	0.108	0.021	−0.504***	−0.480***	−0.455***	−0.385***
		0.129	0.130	0.107	0.082	0.094	0.144
$\ln(D_j)$	β_7	−7.137***	−7.955***	–	0.021***	0.020***	0.021***
		0.616	0.527	–	0.001	0.001	0.001
Constant	β_0	11.571***	10.111***	–	−1.504***	−1.301***	−2.137***
		2.864	2.693	–	0.405	0.435	0.634
Standard error		0.431	0.471	0.479	0.461	0.461	0.461
Instrument relevance (partial R^2)		–	–	–	0.87	0.97	0.99
Hausman and Hausman–Taylor tests		–	χ^2 (6)=14.19***	–	χ^2 (5)=0.97	χ^2 (8)=0.87	χ^2 (8)=0.99
		Variance component and average GLS weights for AR(1) HT, AM, and BMS: $\hat{\sigma}_\varepsilon^2 = 0.328$, $\hat{\sigma}_\mu^2 = 9.114$, $\hat{\theta} = 0.962$					

No of observations is 2,767; No of host country × industry cross-sections is 341. All time-variant variables are assumed as doubly-exogenous, time-invariant distance is singly-exogenous. χ^2 test statistics in the RE columns are Hausman (1978) tests, and those in the HT columns are Hausman and Taylor (1981) over-identification tests. Test statistics in the AM columns are with respect to the HT models and those in the BMS columns are with respect to the corresponding AM models (see Baltagi and Khanti-Akom 1990). The estimated autocorrelation parameter amounts to 0.728 and is highly significant according to the Baltagi and Wu (1999) LBI test statistic of 2.73. Partial R^2 figures are with respect to distance in the first stage regression (see Shea 1997).

*** Significant at 1%; ** Significant at 5%; * Significant at 10%

(see Baltagi and Khanti-Akom 1990, for their use in a similar context).[18] Third, all models discussed so far assume zero autocorrelation of the error term although, in fact, the estimated autocorrelation parameter (assuming AR(1)) amounts to $\widehat{\rho} = 0.728$ with a Baltagi and Wu (1999) LBI test statistic of 2.73.

Taking this result into account and noting its potential consequence not only for efficiency but also for the point estimates in finite samples,[19] all models are estimated assuming an AR(1) data generating process for stochastic shocks. The data are first Prais-Winsten transformed as outlined in Baltagi and Wu (1999), and then pooled OLS, RE, FE, HT, AM, and BMS are run as described in Sect. 3. Concerning the consistent fixed effects model, it is obvious that all point estimates besides $\widehat{\beta_1}$ are considerably smaller than before. However, their signs do not significantly change. Regarding OLS and RE, they perform now somewhat better than before. This already indicates that part of the correlation between the regressors and the unobserved industry-host country effects is generated by the ignorance of the AR(1) process. This also shows up in a much better performance of the HT, AM and BMS models, because the weak exogeneity of the instruments in the upper bloc of the table is also mainly due to omitted autocorrelation. On the one hand, HT is now very close to the FE model, showing up in a particularly low Hausman and Taylor (1981) test statistic. But also the Hausman tests of HT versus AM and AM versus BMS in the bottom row of the lower table do not any more reject.[20]

However, in this example the consideration of the AR(1) models has also an important consequence for the economic interpretation. The main effect of distance is positive, so that actually all parameters now strongly support the relevance of horizontal outward FDI in general[21] and at low relative factor endowment differences between the US and a respective host in particular.[22] This also provides further support for the general findings by Carr et al. (2001) and Markusen and Maskus (2002), but from a different angle regarding specification (focusing on the role of distance), econometric methods (panel rather than cross-section methods), and data (US outward FDI rather than affiliate sales).[23]

[18] This danger has already been pointed out by Metcalf (1996).

[19] It is not surprising that the coefficient estimates differ across the estimated models (OLS, RE, FE, HT, AM, and BMS) within the top and the lower panel of Table 1. The reason is that OLS and RE rely on stronger assumptions than FE regarding the correlation of the included observable variables and the unobservable industry–host–country components. Moreover, HT, AM, and BMS use different sets of instruments that may be more or less appropriate in terms of their relevance and exogeneity. However, that the parameters within a particular model, say BMS, differ between the top and bottom parts of Table 1 is only possible in limited samples. This difference will decrease as the sample becomes larger. However, the latter does not mean that the estimates at the bottom of Table 1 will converge to the ones at the top.

[20] It should be mentioned once again that an additional inclusion of bilateral overall GDP does not improve the explanatory power of the FE, HT, AM, or BMS models. Moreover, fixed time effects would not contribute significantly in this data set.

[21] Note that the marginal effect of distance evaluated at the sample mean amounts to 0.255 and is significant at 1% in the HT-AR(1) model.

[22] Recall that $\widehat{\beta_5} > 0$ but $\widehat{\beta_6} < 0$.

[23] The investigation of affiliate sales rather than FDI is consistent with the two-factor framework in Carr et al. (2001) and Markusen and Maskus (2002) and their focus on (intangible) knowledge-capital rather than (tangible) physical capital.

5 Conclusions

This paper focuses on the role of distance for outward FDI. It shows that, when associating distance with trade costs, general equilibrium theory on trade and multinational firms motivates at least three hypotheses regarding its impact on FDI. First, vertical FDI declines with distance due to rising trade costs. Second, horizontal FDI for the same reason rises with distance. Third, the latter effect positively depends on bilateral country size, which is shown in a highly stylized, three-factor version of a model with horizontal multinationals under perfect symmetry. In a specification motivated from such a horizontal model and a panel data-set of bilateral US outward FDI at the industry level, the three distance-related hypotheses are investigated in particular.

In the application, the use of panel econometric methods turns out to be useful to overcome the endogeneity bias from omitted, time-invariant determinants, which are likely related to legal, institutional, and cultural factors that are difficult to observe. However, a compulsory investigation of the mentioned hypotheses and the computation of a marginal effect of distance is impossible by fixed effects estimation. The reason is that for this purpose the estimation of a time-invariant variable's parameter (distance) is essential. This motivates the application of models developed by Hausman and Taylor (1981); Amemiya and MaCurdy (1986) and Breusch et al. (1989). However, the original versions of these models assume zero autocorrelation of the error term. Since shocks in FDI models such as in other investment models are likely correlated over time and applications typically involve missing values, the paper recommends to adopt the procedure outlined in Baltagi and Wu (1999) for estimation. The results suggest that an omitted autocorrelation process may induce and enforce correlation between the regressors and the unobserved effects. This not only leads to a larger deviation of the random effects estimator from its fixed effects counterpart, but it also likely reduces the instrument quality in Hausman and Taylor (1981); Amemiya and MaCurdy (1986), or Breusch et al. (1989) models that do not account for autocorrelation.

The results lend very strong support to the three distance-related hypotheses. Apart from this, they implicitly point to a dominance of horizontal US outward FDI. They underpin not only the empirical relevance of the model by Markusen and Venables (2000), but they also are well in line with the recent findings of Carr et al. (2001) and Markusen and Maskus (2002), though being based on a different specification, econometric method and data set.

Acknowledgments I would like to thank the editor in charge, Börje Johansson, and two anonymous referees for numerous helpful comments on an earler version of the manuscript.

References

Amemiya T (1971) The estimation of the variances in a variance-components model. Int Econ Rev 12(1):1–13

Amemiya T, MaCurdy TE (1986) Instrumental-variable estimation of an error components model. Econometrica 54(4):869–881

Anderson JE (1979) Theoretical foundation for the gravity equation. Am Econ Rev 69(1):106–116

Anderson JE (2000) Why do nations trade (so little)? Paci Econ Rev 5(2):115–134
Anderson JE, van Wincoop E (2004) Trade costs. J Econ Lit 42(3):691–751
Baltagi BH (2001) Econometric analysis of panel data. Wiley, Chichester
Baltagi BH, Khanti-Akom S (1990) On efficient estimation with panel data: an empirical comparison of instrumental variable estimators. J Appl Econ 5(4):401–406
Baltagi BH, Wu PX (1999) Unequally spaced panel data regressions with AR(1) disturbances. Econ Theory 15(6):814–823
Barrios S, Görg H, Strobl E (2004) Multinational enterprises and new trade theory: evidence for the convergence hypothesis. Open Econ Rev 14(4):397–418
Bergstrand JH (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. Rev Econ Stat 67(3):474–481
Bergstrand JH (1989) The generalized gravity equation, monopolistic competition, and factor proportions theory in international trade. Rev Econ Stat 71(1):143–153
Bergstrand JH, Egger P (2007) A knowledge-and-physical-capital model of international trade, foreign direct investment, and foreign affiliate sales: developed countries. J Int Econ (forthcoming)
Blonigen BA, Davies RB, Head K (2003) Estimating the knowledge-capital model of the multinational enterprise: Comment. Am Econ Rev 93(3):980–994
Brainard LS (1997) An empirical assessment of the proximity-concentration trade-off between multinational sales and trade. Am Econ Rev 87(4):520–544
Breusch TS, Mizon GE, Schmidt P (1989) Efficient estimation using panel data. Econometrica 57(3):695–700
Carr D, Markusen JR, Maskus KE (2001) Estimating the knowledge-capital model of the multinational enterprise. Am Econ Rev 91(3):693–708
Cornwell C, Schmidt P, Wyhowski D (1992) Simultaneous equations and panel data. J Econ 51(1–2):151–181
Dixit AK, Stiglitz JE (1977) Monopolistic competition and optimum product diversity. Am Econ Rev 67(3):297–308
Egger P, Pfaffermayr M (2004) Estimating short run and long run effects in static panel models. Econ Rev 23(3):199–214
Egger P, Pfaffermayr M (2005) Trade, multinational sales, and FDI in a three-factors model. Rev Int Econ 13(4):659–675
Ekholm K, Forslid R, Markusen JR (2007) Export-platform foreign direct investment. J Europ Econ Assoc 5(4):776–795
Fuller WA, Battese GE (1973) Transformations for estimation of linear models with nested error structure. J Am Stat Assoc 68(343):626–632
Fuller WA, Battese GE (1974) Estimation of linear models with cross-error structure. J Econometrics 2(1):67–78
Grossman GM, Helpman E, Szeidl A (2006) Optimal integration strategies for the multinational firm. J Int Econ 70(1):216–238
Hanson GH, Mataloni R, Slaughter MJ (2002) Expansion strategies of US multinational firms. In: Rodrik D, Collins S (eds) Brookings trade forum 2001. The Brookings Institution, Washington DC pp 245–294
Hausman JA, Taylor WE (1981) Panel data and unobservable individual effects. Econometrica 49(6):1377–1398
Helpman E (1984) A simple theory of international trade with multinational corporations. J Polit Econ 92(3):451–471
Helpman E, Krugman PR (1985) Market structure and foreign trade. MIT Press, Cambridge
Hummels D (2001) Toward a geography of trade costs, working paper. Purdue University Available at: http://www.mgmt.purdue.edu/faculty/hummelsd/research.htm
Leamer EE (1984) Sources of international comparative advantage: Theory and evidence. MIT Press, Cambridge
Limao N, Venables AJ (2001) Infrastructure, geographical disadvantage, transport costs and trade. World Bank Econ Rev 15(3):451–479
Markusen JR (2002) Multinational firms and the theory of international trade. MIT Press, Cambridge
Markusen JR, Maskus KE (2002) Discriminating among alternative theories of the multinational enterprise. Rev Int Econ 10(4):694–707
Markusen JR, Venables JA (2000) The theory of endowment, intra-industry and multinational trade. J Int Econ 52(2):209–234

Martín C, Velázquez F (1997) The determining factors of foreign direct investment in Spain and the rest of the OECD: lessons for CEECs. CEPR discussion papers no. 1637, CEPR, London
Metcalf GE (1996) Specification testing in panel data with instrumental variables. J Econ 71(1–2):291–307
Shea J (1997) Instrument relevance in multivariate linear models: a simple measure. Rev Econ Stat 79(2):348–352
Yeaple SR (2003) The complex integration strategies of multinational firms and cross-country dependencies in the structure of foreign direct investment. J Int Econ 60(2):293–314