Comment on ”Multinational Firms and Backward Linkages” by Ping Lin and Kamal Saggi

Michael Keane

University of Oxford

April 2005

Online at https://mpra.ub.uni-muenchen.de/63475/
MPRA Paper No. 63475, posted 10 April 2015 16:54 UTC
Comment on "Multinational Firms and Backward Linkages" by Ping Lin and Kamal Saggi

by

Michael P. Keane
Yale University

Any attempt to infer the effect of foreign direct investment (FDI) on the productivity of domestic firms must first confront a number of severe econometric challenges. I believe these challenges are so great that the literature has barely begun to tackle them. In particular, the econometric specifications estimated in the current literature do not seem to be closely tied to any underlying economic theory that specifies the mechanisms through which productivity spillovers might operate. Given the severe econometric problems this literature must confront, I expect that progress will require econometric modeling based rather tightly on maintained economic theory. Any inferences so obtained will then be dependent on strong a priori identifying assumptions regarding economic structure.

Of course, this situation is not unique to identifying FDI spillover effects. We always need a priori identifying assumptions in order to learn anything of interest from data, beyond perhaps the simplest of descriptive statistics. Data cannot simply "speak" and answer our questions about interesting aspects of economic behavior without our bringing some a priori information to bear. However, I think the need for more a priori structure if we are...

---

Michael P. Keane is professor of economics in the Department of Economics at Yale University.

1. By "data" I mean the joint distribution of observed variables. To use the language of the Cowles Commission,

Suppose ... [an econometrician] is faced with the problem of identifying ... the structural equations that alone reflect specified laws of economic behavior. ... Statistical observation will in favorable circumstances permit him to estimate ... the probability distribution of the variables. Under no circumstances whatever will passive statistical observation permit him to distinguish between different...
to make any progress in estimating FDI spillover effects is particularly striking.

For the most part, the existing literature on FDI spillovers takes the approach of estimating production functions in which the total factor productivity (TFP) of the domestic firm in a particular industry/country is allowed to be a function of some measure of the FDI directed by multinational corporations (MNCs) into that industry. For instance, the market share of foreign-owned MNC affiliates in the industry, or the average foreign equity participation across all firms in an industry, are commonly used FDI measures. To fix ideas, consider a Cobb-Douglas production function:

\[ y_t = A_t K_t^a L_t^b \]

where \( y_t \) is output (or value added) of firm \( i \) at time \( t \), \( A_t \) is TFP, and \( K_t \) and \( L_t \) are the capital and labor inputs, respectively. Letting TFP be determined by

\[ \ln A_t = \pi_t + \psi_t F_{it} + \eta_t + \phi_t + \epsilon_t \]

where \( F_{it} \) is the measure of foreign presence in the industry \( j \) to which firm \( i \) belongs, \( \eta_t \) is an industry-specific effect, \( \phi_t \) is a firm-specific effect, and \( \epsilon_t \) is a firm/time-specific idiosyncratic productivity shock, we obtain the equation:

\[ \ln y_{it} = \pi_t + \psi_t F_{it} + \alpha \ln K_t + \beta \ln L_t + \eta_t + \phi_t + \epsilon_t \]

The basic approach is to estimate an equation like 1.1 and to view a significant positive estimate of the coefficient \( \eta_t \) on the FDI variable as evidence of spillovers.

Now, some of the key difficulties in estimating the parameters of equation 1.1 are common to the literature on estimating production functions in general and are not specific to estimating FDI spillover effects. Good discussions of these problems can be found in Marschak and Andrews (1944), Griliches and Mairesse (1995), and Klette and Griliches (1996).

The most obvious problem is endogeneity of the capital and labor inputs, since these will generally be chosen in response to productivity as determined by \( \eta_t + \phi_t + \epsilon_t \). The standard solutions to this problem are some combination of (i) differencing the data to eliminate the time-invariant components of productivity and/or (ii) finding instruments that are assumed to be correlated with factor input choices but uncorrelated with productivity (such as input prices). We can then estimate equation 1.1 using instrumental variables (IV) to deal with the endogeneity of the factor inputs.

Instruments that might typically be used for the capital and labor inputs are measures of wage rates and the price of capital. A more recent alternative to (ii) is the Olley and Pakes (1996) procedure that uses investment as an indicator of the productivity shock.

A major problem with this general approach is the failure to deal seriously with heterogeneity in firms' production processes. As Lipsy and Sjöholm cogently point out in chapter 2 of this volume, it seems unlikely that we can make sense of FDI spillovers in a modeling framework where all they do is shift TFP. If knowledge spillovers occur, it seems likely that FDI will alter the production functions of domestic firms in much more subtle and extensive ways. For example, perhaps a knowledge spillover will lead to a more capital-intensive production process or enable a reorganization of the domestic firm so that it can take advantage of economies of scale. In such cases, spillovers will affect the share parameters in equation 1.1, not just TFP. This highlights the point that no specific theory of how knowledge spillovers affect production technology underlies specifications as in equation 1.1.

Furthermore, a striking aspect of equation 1.1 is the assumption of fixed technology parameters even across industries. Yet, if we look at manufacturing industries where FDI occurs, there is ample evidence of substantial heterogeneity in firms' technologies even when we look at firms within narrowly defined industries. Evidence is provided in a series of studies by Feinberg and Keane (2001, 2003a, 2003b), Feinberg, Keane, and Bognanno (1998), and the chapter by Feinberg and Keane in this volume. In these studies, we find that MNC affiliates operate in very different ways even within narrowly defined industries: the shares of capital, labor, materials, and intermediates imported from parents differ quite substantially across firms.

This suggests three problems: First, if technological heterogeneity is prevalent for affiliates, it is likely to be prevalent for domestic firms as well. If share parameters are heterogeneous across firms/industries (i.e., some firms use more capital-intensive production processes than others), then an instrumental variables technique applied to equation 1.1, using supply-side instruments like wage rates and the price of capital, will not generally resolve the problem of endogeneity of the capital and labor inputs. We might think that with heterogeneous coefficients we would estimate the mean coefficient values in the population of firms, which would still be of interest. However, as I discuss in the appendix, this is true only under very special assumptions that we would not expect to hold in the present case. In general, with
heterogeneity on the production function parameters, the IV approach will not consistently estimate behaviorally interpretable parameters.²

Second, if the impact of FDI on TFP differs across firms, industries, and/or FDI has more general impact on technology than just shifting TFP, then estimation of specifications as in equation 1.1 using instrumental variables procedures will not in general identify the parameter \( \pi_1 \) and, furthermore, it is no longer even clear how a parameter like \( \pi_1 \) could be interpreted.

Third, Feinberg and Keane (chapter 10 in this volume) document that MNC affiliates in developing countries are organized in very different ways. That is, if we look at affiliates within an industry/country, some are organized to use a substantial volume of imported intermediates from parents while others are not, some are organized to ship substantial volumes of intermediates back to parents while others are not, some are organized to trade heavily with the parent’s affiliates in other countries while others are not, and so on. The existing econometric literature takes no account of the possibility that the form of affiliate organization may play a key role in whether and how knowledge is transferred to local firms. This again stems from the general failure to base empirical work on specific underlying theories that specify the mechanisms through which knowledge transfers arise.

On this issue, the case study literature, such as Moran (2001), is quite informative, because it suggests that the nature of the interaction between parents and affiliates does heavily influence whether and how knowledge is transferred. As Moran (chapter 11 in this volume) notes “these investigations showed that there is a fundamental difference in performance between subsidiaries that are integrated into the global or regional sourcing networks of the parent multinationals, and subsidiaries that are oriented toward protected domestic markets.”

Lipsey and Sjöholm (chapter 2 in this volume) argue that the case study literature addresses the question of whether there are examples where technology was transferred from MNCs to domestically owned firms, while the “statistical studies ask whether on average domestically owned firms gain . . . from the presence of foreign-owned firms.” However, it is important to recognize that, when heterogeneous effects are present, an econometric analysis that assumes homogenous effects does not in general answer this “on average” question. In general, the biases induced by ignoring heterogeneity are much more fundamental (see appendix IA). I would argue instead that the case study literature might be used to help guide the development of theoretical models of the mechanisms through which knowledge transfers occur, and that these in turn might be used as the basis for structural economic analysis (see below).

Another general problem in the estimation of production functions that I would like to point out is what I will call the “price times quantity” (PQ) problem. “This problem was stressed by Griliches and Mairesse (1995) and Klette and Griliches (1996), and is also discussed by Lipsey and Sjöholm (in this volume). The main problem is that we do not actually have data on real output quantities (or real value added) \( y_q \) for purposes of estimating equation 1.1. Rather, we typically have data only on sales revenues, meaning we see only \( PQ \) and not quantity itself.

In order to deal with the PQ problem, it has been typical in the literature on production function estimation to simply use industry-level price indices to deflate nominal sales revenue data so as to construct real output. But Griliches and Mairesse (1995) as well as Klette and Griliches (1996) have pointed out that this procedure is valid only in perfectly competitive industries, so that price is exogenous to firms. If firms have market power, then any change in inputs will shift both price and quantity, with the price effect depending both on the elasticity of demand facing the firm and the elasticity of output with respect to the input, both of which are, in general, firm specific. In the present context, the point is that, if FDI seems to shift revenue, we do not know if this is because it shifted productivity or prices. These considerations are important here because we would expect FDI to occur in industries where firms do indeed have market power. The PQ problem has received a great deal of attention recently in the industrial-organization (IO) literature (see, for example, Katayama, Lu, and Tybout 2003 and Levinsohn and Melitz 2002).

The only general solution to the problem of endogenous output prices is to estimate the production function jointly with an assumed demand system. To my knowledge, the only study where such an approach has been fully implemented is Feinberg and Keane (2003a), which estimates the production technology of US parents and their manufacturing affiliates in Canada.⁴ In the FDI spillover context, one would need to estimate the production functions for domestic (i.e., host country) firms jointly with a demand system.

In this regard, Lipsey and Sjöholm (in this volume) make the interesting point that “if the foreign investors’ . . . superiority consists of knowledge about . . . marketing a product . . . it will not be visible in . . . production

---

3. Intuitively, the problem arises because, if production function coefficients differ across firms, then a model that assumes homogeneous parameters will sweep the firm-specific part of the parameters, along with capital and labor input terms they multiply, into the error term. Also, with heterogeneous production function parameters, “exogenous” factor price changes will have different effects on factor inputs for different firms. Thus, changes in factor inputs induced by “exogenous” changes in the wage rate or price of capital are still correlated with the firm-specific parameters that enter the error term. The only way to deal with the problem is to model the heterogeneity explicitly. Although in a different context, Feinberg and Keane (2003a) is the only study I know of that does this.

4. The solutions proposed by Katayama, Lu, and Tybout (2003) as well as Levinsohn and Melitz (2002) are somewhat more specialized and do not require complete specification of the demand system.
function comparisons.” Another way to put it is that FDI spillovers may operate by affecting the demand system rather than the production function. Given the PQ problem, this means they actually may show up in production function comparisons, but not in any interpretable way. It is disconcerting that a knowledge spillover that enables firms to enhance market power could show up spuriously as a productivity enhancement, since the two have very different welfare implications.

Next, I focus on an issue that is more specific to the literature on estimating FDI spillover effects: FDI is likely to be endogenous. One key source of endogeneity is that MNCs may be attracted to invest in countries that have high productivity. A country’s solid legal framework, low risk of appropriation (see Wang and Blomström 1992), and a relatively free/well-managed economy could make it more attractive for FDI. All these factors would presumably lead to more productive domestic firms as well. An econometric study that failed to account for these factors might falsely conclude that FDI enhanced productivity, when in fact FDI was just attracted to countries that were pursuing productivity-enhancing policies.

But there are other stories where the endogeneity bias can go either way. In the theoretical literature on FDI and spillovers (e.g., Ethier and Markusen 1996 and Petit and Sanna-Randaccio 2000), one point that is stressed is that FDI is more likely when a country/industry has better protection of intellectual property. This could lead us to see more FDI in industries/countries where spillovers are smaller, leading to a negative correlation between host country productivity and MNC penetration. On the other hand, if intellectual property protection is good for productivity-enhancing investment in general, or if it is associated with a country having a generally favorable economic climate, then MNCs’ tendency to invest in countries with strong intellectual property protection could lead to a spurious positive correlation between domestic firm productivity and FDI. Furthermore, a key point is that MNC entry into an industry will generally change the market structure, leading to either more or less competition depending on the entry and exit behavior of domestic firms. This in turn can lead to either an increase or decrease in research and development (R&D) activity by domestic firms (see Veurkers and Vanden Houte 1990). Also, if domestic firms have market power, they may engage in inefficient management practices that are reformed when foreign competition is introduced. Thus, it is perfectly possible that FDI could affect productivity of domestic firms for reasons that have nothing to do with knowledge spillovers.

Such endogeneity problems are the focus of Gordon Hanson’s insightful comment in this volume. He argues that, given the endogeneity of FDI, we cannot really infer anything causal from the positive correlation between FDI and domestic firm productivity found in the literature. As Hanson discusses, one way we might try to address these endogeneity problems is to find instruments or “natural experiments” that generate exogenous changes in FDI. Such approaches are currently quite fashionable in applied economics. However, I am not optimistic that we can make much headway on estimating FDI spillover effects if we pursue this course. One basic problem is that truly exogenous factors driving FDI are hard to come by.

The theoretical literature suggests that firms choose between FDI or exports as a means to serve a foreign market based on such factors as tariffs, transport costs, wage rates, and other costs in the host country. Similar factors influence vertical FDI. But would a change in tariffs be exogenous with respect to FDI? It seems likely that tariff changes are associated with other changes in the economic environment—for example, countries that lower tariffs often engage in broad range of economic liberalizations at the same time. In order to conclude that any change in productivity associated with a tariff-induced change in FDI was due solely to the change in FDI itself, we would have to control for all these other concomitant changes in the economic environment.

Factors like restrictions on capital investment, intellectual property protection, expropriation risk, political stability, and economic freedom are presumably important determinants of FDI as well. The data also contain a great deal of variation along these dimensions. Indeed, Lipsey and Sjöholm (in this volume) point out that more than 1,500 policy changes making regulations more favorable to FDI occurred during the 1991–2002 period, along with 100 changes making regulations less favorable. Do any of these regulator changes provide “natural experiments?” I expect not, precisely because they are likely to be accompanied by other economic policy changes and institutional changes.

Another problem with inferring causality from FDI/productivity correlations is the problem of reverse causality. FDI itself may lead to changes in regulations affecting FDI and also to changes in economic policy and/or in political/economic institutions. These policy changes might enhance productivity, quite apart from any direct effects of FDI. Thus, it seems highly implausible to me that we can ever find natural experiments where FDI is made more or less attractive while holding all other things equal.

Even if we could find a perfectly exogenous instrument that shifts FDI while not being associated with any other (unmeasured) changes in the economic environment, we would still face a number of problems. First, if production function heterogeneity exists, then the biases discussed earlier (and in the appendix) would prevent two-stage least squares (2SLS) from giving a consistent estimate of coefficient $\pi_i$. Further, even if we could use such an approach, the theoretical (or behavioral) assumptions that underlie exogeneity assumptions...
instrument to estimate the causal effect of FDI on productivity, we would not have isolated the mechanism through which the productivity enhancement occurred. Is it knowledge spillovers, or is it the response of domestic firms to enhanced competition, which, as we noted earlier, could lead to enhanced R&D or to adoption of more efficient management practices?

All of this suggests that we are unlikely to make progress in estimating FDI spillover effects using the simple strategy of estimating production functions with instrumental variables techniques. I believe that to make any progress we will need to model additional data beyond measures of FDI, output, and factor inputs. The preceding discussion suggests it is important to also look at industry structure (entry and exit), measures of R&D spending, and, perhaps, to attempt to form more direct measures of spillovers. Severe problems of endogeneity and reverse causality suggest that we should not attempt to model effects of FDI on productivity in a single equation framework but, rather, view both variables as endogenous outcomes that are jointly determined, along with such factors as market structure and R&D, as part of a structural system.

In this regard, I am struck by the fact that there are several parallel literatures to the FDI spillover literature that seem to simply reorganize what is on the left- and right-hand sides of the estimating equations. For instance, there is literature that examines the determinants of FDI (see, for example, Brainard 1997 and Neven and Siotis 1996), where the determinants are things like R&D intensity, labor productivity, tariff and nontariff barriers, and legal and economic institutions. There is literature in growth theory that examines the effects of such factors as institutions, the legal environment, and corruption on economic growth and/or productivity. Studies in this literature estimate equations like equation 1.1 except that now the FDI measure is replaced by measures of institutions (see, for example, Acemoglu, Johnson, and Robinson 2001). There is also both theoretical and empirical literature examining how FDI and market structure influence R&D spending (see, for example, Veglers and Vanden Houte 1990 and Wang and Blomström 1992). Furthermore, there is literature in development that examines the effect of FDI and growth on institutions.

Putting it all together, we have literature that looks at the effect of FDI on productivity, literature that looks at the effect of productivity and institutions on FDI, literature that looks at the effect of institutions on productivity, literature that looks at how FDI and market structure affect R&D, and literature that looks at the effects of productivity and FDI on institutions. Perhaps it would make more sense to model how productivity, FDI, R&D, market structure, and economic institutions evolve together in a dynamic process, rather than having separate literatures that examine the determinants of each of these variables individually while treating all the others as exogenous.

Development of a dynamic structural model of FDI and productivity would obviously be a major undertaking, but one can see where it might have advantages. For instance, thinking about a dynamic model would immediately lead one to ask: Why should the contemporaneous value of FDI enter equation 1.1? Knowledge transfer is presumably a gradual process, and there are costs to adopting new technologies (see Teece 1976 for a discussion). One would think that, if FDI does transfer knowledge, then it should raise productivity with a lag. These types of considerations may lead to timing restrictions that could aid identification.

In summary, I have tried to highlight some of the severe econometric problems facing the literature on FDI spillovers. Given these problems, it is very difficult to draw any reliable conclusions regarding causality from the existing econometric work in this area. I hope that by bringing additional theory and data to bear it will be possible to make further progress. In my view, the most urgently needed step is to stop treating the mapping from FDI to productivity as a black box. Researchers should attempt to formulate realistic theoretical models of the mechanisms through which FDI might generate knowledge spillovers and attempt to bring these models to the data.

are often quite strong, and they are often left implicit. The failure of wages and the price of capital to provide valid instrumental variables for the estimation of equation 1.1, even if they are completely exogenous from individual firms' point of view, is a good illustration of the potential pitfalls that come from failing to explicitly record a structural model in order to examine whether an instrument is valid. As the appendix shows, wages and the price of capital are not valid instruments for estimation of this model unless a very stringent assumption on production function homogeneity is made.
Appendix IA

If the share parameters $\alpha$ and $\beta$ in equation 1.1 differ across industries, then the capital and labor inputs enter the error term of the homogenous coefficients model. That is, we can write:

$$
\ln y_i = \pi_0 + \pi_F \ln F_i + \tilde{\alpha}_i \ln K_i + \tilde{\beta}_i \ln L_i + \sum \tilde{\eta}_i + \tilde{\phi}_i + \varepsilon_i
$$

where we have let $\alpha_i = \tilde{\alpha}_i + \bar{\alpha}$ and $\beta_i = \tilde{\beta}_i + \bar{\beta}$, where $\bar{\alpha}$ and $\bar{\beta}$ are the mean share parameters across all firms/industries, and $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ are the firm $i$ specific deviations from those means.

Note that the error term, which appears in square brackets in equation IA.1, contains the capital and labor inputs multiplied by the firm-specific share parameters. To give a simple illustration of the problems this creates, suppose that all firms face isoelastic demand functions of the form:

$$
P_i = P_0 y_i^{\gamma}
$$

where $0 < \gamma < 1$, and further suppose that production is constant returns to scale (CRS), so that $\alpha_i + \beta_i = 1$. In that case, the firm’s optimal choice of labor input is given by:

$$
\ln L_i = \frac{1}{\gamma} \left[ \frac{(1 - \alpha_i(1 - g))}{\ln (1 - \alpha_i + \alpha_i (1 - g)) \ln \alpha_i} + \ln (1 - g) P_0 + (1 - g) \ln A_s - (1 - \alpha_i (1 - g)) \ln w_i + \alpha_i (1 - g) \ln \pi_s \right]
$$

where $w_i$ is the wage rate facing firm $i$ at time $t$, and $P_0$ is the rental price of capital. Let us assume that the log wage and price of capital are “exogenous” in the sense that:

$$
E[\tilde{\alpha}_i|\ln w_i, \ln \pi_s] = E[\tilde{\eta}_i|\ln w_i, \ln \pi_s] = E[\tilde{\phi}_i|\ln w_i, \ln \pi_s] = 0
$$

so these variables would, at first sight, appear to be plausible instruments.

Note that in the CRS case we have $\beta_i = (1 - \alpha_i) - \bar{\alpha}$, we see that the error term in IA.1 can be written as:

$$
(\tilde{\alpha}_i) (1 - \alpha_i) \ln w_i,
$$

where the variance of $\tilde{\alpha}_i$ is zero (i.e., unless there is no heterogeneity in the share parameters across firms).

This argument implies that, in estimating equation IA.1, the supply price of labor is not a valid instrument for the labor input, since it will be correlated with the error term IA.3. The same type of argument shows that the supply price of capital is not a valid instrument for the capital input. In fact, given how the capital and labor inputs enter the error term, there are no valid instruments by construction. Any variable that alters firms’ choices of capital and labor inputs will be correlated with the error term. The only way to deal with this problem is to actually model the distribution of the firm-specific parameters and estimate this distribution.

Of course, there are conditions under which instrumental variables can be used to uncover the mean values of the coefficients in a random coefficients model. These conditions are discussed in Wooldridge (1997) and Card (1999, 1927-20). To understand why these conditions do not hold in the present case, it is useful to cast our model into their framework. Let $\xi$ denote the composite error term in equation IA.1. Wooldridge noted that, so long as

$$
E[\xi|z] = \text{constant}
$$

where $z$ is the vector of instruments, then 2SLS applied to equation IA.1, using $z$ as the vector of instruments, will only lead to inconsistency in the estimate of the intercept. The slope coefficients of the model are still consistently estimated. However, we have seen that $\xi$ contains the term

$$
(\tilde{\alpha}_i) (1 - \alpha_i) \ln w_i,
$$

and we therefore have that:

$$
E[(\tilde{\alpha}_i) (1 - \alpha_i) \ln w_i | \ln w_i] = (1 - \alpha_i) \xi_i \ln w_i.
$$

Thus, the expected value of $\xi_i$ varies with the potential instrument $\ln w_i$.

Intuitively, the real source of the problem is that the effect of the potential instrument (price of labor) on the endogenous variable (the labor input) is heterogeneous across firms, as we can see in equation IA.2. In equation

6. For an example, but in a different context, see Feinberg and Keane (2003a).
IA.2, the common part of the effect of the wage on the labor input just derives from the \((1 - \alpha(1 - g)) \ln w_t\) term. This produces a component of \(\xi_t\) of the form \((1 - \alpha(1 - g)) \ln w_t\), and this term is mean zero conditional on \(\ln w_t\). Thus, we see that it is the random part of the coefficient on \(\ln w_t\) in the labor demand function of equation IA.2 that creates the problem.

References


