Mostly Pointless Spatial Econometrics?

Stephen Gibbons (SERC, Department of Geography & Environment, LSE and CEP)
Henry G. Overman (SERC, Department of Geography & Environment, LSE and CEP)

October 2010

Mostly Pointless Spatial Econometrics?

Stephen Gibbons* and Henry G. Overman**

October 2010

* SERC, Department of Geography & Environment, London School of Economics
** SERC, Department of Geography & Environment, London School of Economics

Abstract

We argue that identification problems bedevil most applied spatial research. Spatial econometrics solves these problems by deriving estimators assuming that functional forms are known and by using model comparison techniques to let the data choose between competing specifications. We argue that in most situations of interest this, at best, achieves only very weak identification. Worse, in most cases, such an approach will simply be uninformative about the economic processes at work rendering much applied spatial econometric research 'pointless', unless the main aim is simply description of the data. We advocate an alternative approach based on the 'experimental paradigm' which puts issues of identification and causality at centre stage.

1. INTRODUCTION

The last two decades have seen economists become increasingly interested in geographical issues. This has been variously attributed to theoretical developments (the New Economic Geography), a growing interest in cities (the expansion of urban economics) or simply the greater availability of geo-referenced data. Regardless of the reason, the upshot has been greater interaction between economic geographers, regional scientists and economists in an attempt to understand spatial aspects of the economy. More recently, a similar process has seen mainstream econometric theorists becoming increasingly interested in spatial processes, traditionally the preserve of a specialised group of spatial econometricians. One might think that the next step would be convergence between the tools developed by spatial econometricians and the methods used by applied researchers to assess the extent to which models of spatial economics fit real world data. We argue that this is unlikely to happen because, while there may have been convergence between mainstream and spatial econometric theory, most applied research is taking a different path.

Most modern applied economic investigation is concerned with answering questions about causal relationships. If we increase an individual's years of education, what happens to their wages? If we decrease class sizes, what happens to student grades? These questions are fundamentally of the type "if we change x, what do we expect to happen to y". Just as with economics more generally, answering such questions is fundamental to increasing our understanding of spatial economics. When more skilled people live in an area, what do we expect to happen to individual wages? If a jurisdiction increases taxes, what do we expect to happen to taxes in neighbouring jurisdictions?

In an experimental setting, agents (individuals, firms, governments) would be randomly assigned different amounts of x and the outcomes y observed. Considering whether different levels of x are associated with different outcomes would then give us the causal effect of x on y. The fundamental challenge to answering these types of questions for (most) real world data is that we do not randomly assign x and observe outcomes y. Instead, we jointly observe x and y so we lack the counterfactual as to what would have happened if we were to change x. Fortunately, applied economics has come a long way in its effort to find credible and creative ways to answer such questions by constructing counterfactuals from observational data.

A good starting point for thinking about whether a particular question about causality can be answered and how to answer it, is to consider what an ideal experiment might look like. The experiment may not be feasible, but with the design in mind it is easier to think of ways to find sources of variation in the data that mimic or approximate the conditions of the ideal experiment. The 'experimental paradigm' (Angrist and Pishke, 2009) does this by using simple linear estimation methods, taking care to pinpoint and isolate sources of variation in x that can plausibly be considered exogenous to y. The aim of these methods is to mimic as far as possible the conditions of an experiment in which agents are randomly assigned different amounts of x and outcomes y observed. The central idea is to find otherwise comparable agents (e.g. twins) who for some reason have been exposed to different amounts of x. Whilst drawing on economic theory, a fundamental attraction of this strategy is that the assumptions required for identification of causal effects are usually clearly specified and understandable without reference to specific (and untested) economic theories. Put another way, the aim is to obtain plausible estimates of causal effects without relying

on a specific and restrictive set of functional form assumptions. This approach is particularly attractive in areas, like much of spatial economics, where we are far from having a structural model that closely captures the complexities that drive the processes for which we observe data.[4] Unfortunately, most applied research using spatial econometrics continues to ignore these insights into framing research questions and arriving at credible research designs.

Why is this the case? An important part of the answer, we suspect, is because the underlying spatial econometric theory has developed from time-series foundations, in such a way that questions about causality have never been centre stage. The standard approach to spatial econometrics has been to write down one of a number of spatial model specifications (e.g. the spatial autoregressive model), to assume that the equation accurately describes the data generating processes, and then to estimate the parameters by non-linear methods such as (quasi) maximum likelihood (ML). Because estimation is not always simple, much effort has gone in to developing techniques that allow estimation of parameters from a range of models for large data sets. Questions of identification (i.e. does x *cause* y) have been addressed by asking which of these assumed spatial processes best fit the data. While in principle this sounds straightforward, as we discuss further below problems of identification mean it is hard to distinguish between specifications that have very different implications for which causal relationships are at work.

In our opinion, the standard spatial econometric toolbox does not offer a solution to the problem of the identification of causal effects in a spatial setting but too many applied researchers proceed as if it does. Of course, much standard (i.e. non-spatial) analysis falls someway short of these lofty ideals. Finding sources of truly exogenous variation in x is difficult and often we have to settle for examining the effect of variation that is 'approximately' exogenous. Sometimes, it is difficult to get that far and we are left with correlations which may still, we hope, provide some insight in to the problem at hand. Despite these caveats however, for much applied research, identification of causal effects in line with the experimental paradigm would remain the gold standard to which most researchers claim to aspire. We will argue that this should also be the case in applied spatial research.

The rest of this paper is structured as follows. Section 2 provides a basic overview of standard spatial econometric models, while section 3 discusses problems of identification. Section 4 returns to the relationship between the spatial econometrics and experimental paradigms. Section 5 concludes.

2. SPATIAL ECONOMETRIC MODELS AND THEIR MOTIVATION

This section provides an introduction to spatial econometric models. It is not intended to be comprehensive. Rather, we provide enough background so that someone who is unfamiliar with spatial econometrics will be able to follow the arguments made later. We generally use the model terminology of LeSage and Pace (2009) and the reader is referred there for more details.

---

[4] The assumption of linearity and the reliance on simple linear methods may appear like quite a strong functional form assumption. However, as Angrist and Pishke 2009 (p.69) argue the assumption of a linear structural relationship - the Conditional Expectation Function (CEF) - is "not really necessary for a causal interpretation of regression". If the CEF is causal then, because the linear regression provides the best linear approximation to the CEF, the regression coefficients are informative about causality.

To develop ideas, start with a basic linear regression:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i \qquad\qquad (1)$$

where *i* indexes units of observation, y_i is the outcome of interest, $\mathbf{x}_i$ is a vector of explanatory variables, u_i is an error term and $\boldsymbol{\beta}$ is a vector of parameters. [5] In the most basic specification it is common to assume that outcomes for different units of observation are *independent* of each other. This is a strong assumption and there may be many reasons why outcomes are not independent, particularly when observations are for geographically referenced events, agents or places. In a spatial setting, this model is not very interesting. There are many contexts in which estimating and interpreting the parameters that characterise this dependence is of academic and policy interest. This might be, for example, because the dependence arises from causal spatial linkages that can be exploited for policy advantage, or because knowing these parameters increases understanding of the causes of spatial disparities.

Unfortunately, allowing for arbitrary correlations between observations is impractical because with *n* observations there are $n^2 - n$ possible relationships. However, if the data are spatial, i.e. they can be mapped to locations, we may use information on relative positions (and possibly direction) to restrict the nature of the connection between observations. For example, we might assume that outcomes at a given location depend on outcomes in other locations that are 'nearby' but not those further away. A simple way to capture these assumptions about the spatial dependence between observations is to write down a vector $\mathbf{w}_i$ where the jth element of $\mathbf{w}_i$ takes a value that is bigger, the more closely connected *j* is with *i* (e.g. $1/\text{distance}_{ij}$). If we have *n* observations, then multiplying $\mathbf{w}_i'$ by the *n*x1 vector of outcomes $\mathbf{y}$ gives us a value $\mathbf{w}_i'\mathbf{y}$ that spatial econometricians refer to as a spatial lag. For each observation, $\mathbf{w}_i'\mathbf{y}$ is a linear combination of all y_j with which the ith observation is connected. If, as is usually the case, we normalise $\mathbf{w}_i$ so that the elements sum to 1, then $\mathbf{w}_i'\mathbf{y}$ is a weighted average of the 'neighbours' of *i*.

What now if we want to understand whether changes in the outcome y_i are caused by changes in the places to which i is connected? A simple solution, proposed by Ord (1975) is to assume that the dependence between y_i and its spatial lag is linear and common across observations. This gives the spatial autoregressive model (SAR):

$$y_i = \rho \mathbf{w}_i'\mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + u_i \qquad\qquad (\text{SAR})$$

LeSage and Pace (2009) suggest a "time dependence motivation" for the SAR model. Imagine that fixed across time exogenous variables $\mathbf{x}_i$ determine outcome y_i. Now assume that when determining their own outcome, agents take in to account not only their own characteristics, but

---

[5] For ease of exposition we subsume the constant in $\mathbf{x}_i'\boldsymbol{\beta}$.

also recent outcomes for other agents located close to them. We might think of y_i as the price of a house, $\mathbf{x}_i$ as the (fixed) characteristics such as number of rooms and assume that when agreeing a sale price, people consider not only the characteristics of the house, but also the current selling price of nearby houses. In this case **β** captures the causal effect of house characteristics and ρ represents the causal effect of neighbouring prices (conditional on observed housing characteristics).

We could drop the assumption of dependence between y_i and the spatial lag of y_i and instead assume that y_i depends on spatial lags of the explanatory variables. If **X** denotes the matrix of explanatory variables and **γ** a vector of parameters, we have the spatial (lag of) X model (SLX):

$$y_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{w}_i^{'}\mathbf{X}\boldsymbol{\gamma} + u_i \qquad \text{(SLX)}$$

LeSage and Pace (2009) provide an 'externality motivation' for this model. In this case, we assume that the exogenous characteristics of nearby observations directly affect outcome y_i. Continuing with the housing example, we assume the characteristics of nearby houses, e.g. their size and state of repair, directly determine prices (rather than working through observed sales prices). Of course, an externality motivation could also be used to justify the SAR model if the externality works through the spatial lag of y_i

Next, drop the assumption of dependence on spatial lags of the explanatory variables and instead assume spatial dependence in the error process. We can use the SAR process to place structure on that dependency.[6] If **u** denotes the vector of residuals, this gives the spatial error model (SE):

$$y_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + u_i \; ; u_i = \rho\mathbf{w}_i^{'}\mathbf{u} + v_i \qquad \text{(SE)}$$

Finally, combining the SAR and SLX models gives us the Spatial Durbin Model (SD):

$$y_i = \rho\mathbf{w}_i^{'}\mathbf{y} + \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{w}_i^{'}\mathbf{X}\boldsymbol{\gamma} + u_i \qquad \text{(SD)}$$

which assumes dependence between y_i and the spatial lags of both the dependent and explanatory variables, but drops the assumption of spatial dependence in the error process.

The SD model can be motivated by a combination of the arguments above (causal effects from neighbouring outcomes and neighbouring characteristics). Alternatively, it can be motivated by simply re-arranging the SE model in a spatial Cochrane-Orcutt style transformation:

$$u_i = y_i - \mathbf{x}_i^{'}\boldsymbol{\beta} \qquad (2)$$

$$y_i - \mathbf{x}_i^{'}\boldsymbol{\beta} = \rho\mathbf{w}_i^{'}\mathbf{y} - \rho\mathbf{w}_i^{'}\mathbf{X}\boldsymbol{\beta} + v_i \qquad (3)$$

$$y_i = \rho\mathbf{w}_i^{'}\mathbf{y} + \mathbf{x}_i^{'}\boldsymbol{\beta} - \rho\mathbf{w}_i^{'}\mathbf{X}\boldsymbol{\beta} + v_i \qquad (4)$$

---

[6] More general specifications are available, but the SE model defined in the text is sufficient for our purposes.

An extension to this idea is used as another motivation for including spatial lags of **y** and **X** in a model, as a solution to the problem of omitted variables. Consider a situation where y_i depends on an exogenous variable (x_i) and unobservable factors (z_i). That is, the true model is:

$$y_i = x_i\beta + z_i \qquad\qquad (5)$$

Further, assume that the error term z_i is both spatially correlated and correlated with the observed x_i through an omitted variable, which is partly determined by x_i, such that:

$$z_i = \rho w_i'\mathbf{z} + \gamma x_i + v_i \qquad\qquad (6)$$

substituting $z_i = y_i - x_i\beta$ into (6) and rearranging gives

$$y_i = \rho \mathbf{w}_i'\mathbf{y} + (\gamma + \beta)x_i - \rho \mathbf{w}_i'\mathbf{x}\boldsymbol{\beta} + v_i \quad (7)$$

From (4) and (7) we see that the presence of a spatially correlated error term, whether or not this is correlated with $\mathbf{x}_i$, leads to the SD model involving a spatial lag in **y** and **X**. It is important to emphasise, however, that the in this motivation for the SD model, ρ cannot be interpreted as revealing anything about the causal effect of spatial lags of **y** or **X** on outcomes. The spatial lags of **y** and **X** are simply being used to control for spatial correlation in the error term.

These five processes are not exhaustive of all possible spatial econometric models, and we consider a particularly important generalisation further below, but for the moment they are sufficient for our purposes. In the text, we use the acronyms (SAR, etc) to refer to the specifications above.

Estimation of spatial econometric models using OLS leads to inconsistent parameter estimates if the models include a spatially lagged dependent variable and ρ is non-zero (e.g. the SAR and SD models). This inconsistency arises because of a mechanical link between u_i and $\mathbf{w}_i'\mathbf{y}$ for most standard specifications of $\mathbf{w}_i$. Standard errors are also inconsistently estimated for these models, as well as for models including a spatial lag in the error term (e.g. the SE model). OLS provides consistent estimates of the parameters if the spatial correlation occurs only through the error term (SE model) or exogenous characteristics (SLX model). In both cases standard errors are inconsistent, and OLS estimation of the SE model is inefficient. In contrast, Lee (2004) shows that (quasi) ML estimation provides consistent estimators for all these models *conditional on the assumption that the spatial econometric model being estimated is the true data generating process*. Alongside theoretical developments, advances in computational power and methods have made ML estimation feasible for large datasets.[7] As a result, ML estimation is preferred in the spatial econometrics literature. Inspection of the SAR and SLX models reveals they are nested within the SD model. We have also shown that the SE model can be rearranged to give the SD model. The fact that the SD model nests

---

[7] According to LeSage and Pace (2009, p.45) "these improvements allow models involving samples containing more than 60,000 US Census tract observations to be estimated in only a few seconds on desktop and laptop computers"

many of the other models provides an argument for estimating the SD model and then testing this against the nested models through the use of likelihood tests. This is the approach advocated by LeSage and Pace (2009). Model comparison techniques can be used to compare models based on different weight matrices and explanatory variables.

We have shown that contrasting motivations lead to different spatial econometric specifications. Taken at face value, this suggests we might learn about the underlying processes at work if we use ML estimation and model comparison to identify the correct specification for given data. Unfortunately, as we show in the next section, these different specifications will generally be impossible to distinguish without assuming prior knowledge about the true data generating process that we simply do not possess in practice. In short, contrasting motivations lead to spatial econometrics specifications that cannot usually be easily distinguished in applied research.

It is useful to see how these models are related in terms of their structure. Consider first the reduced form (which expresses y_i in terms of exogenous factors) of the SAR model. The way the model is written down, the only exogenous factors affecting y_i are $\mathbf{x}_i$ and u_i, so the only factors affecting $\mathbf{w}_i'\mathbf{y}$ (the spatial lag of y_i) are $\mathbf{w}_i'\mathbf{X}$ and $\mathbf{w}_i'\mathbf{u}$ (the spatial lags of **X** and $\mathbf{u}$, respectively). The spatial lag of y_i ($\mathbf{w}_i'\mathbf{y}$) also depends on the *second order* spatial lag of y_i ($\mathbf{w}_i'\mathbf{W}\mathbf{y}$), that is, on outcomes for the "neighbours of my neighbours". By repeated substitution (first for $\mathbf{w}_i'\mathbf{y}$, then for $\mathbf{w}_i'\mathbf{W}\mathbf{y}$ etc), we derive the reduced form as:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \rho\mathbf{w}_i'\mathbf{X}\boldsymbol{\beta} + \rho^2\mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \rho^3\mathbf{w}_i'\mathbf{W}^2\mathbf{X}\boldsymbol{\beta} + [\ldots] + v_i \qquad (8)$$

where $v_i = \rho\mathbf{w}_i'\mathbf{v} + u_i$, **W** is the matrix of stacked weight vectors ($\mathbf{w}_i'$) and $\mathbf{W}^2 = \mathbf{W}\mathbf{W}$.

Notice that, in the reduced form, the only thing that distinguishes this from the SLX model is the absence of terms in $\rho^n\mathbf{w}_i'\mathbf{W}^{n-1}\mathbf{X}\boldsymbol{\gamma}$ for *n*>1. As we explain in the next section, in practice these two models will often be hard to tell apart even if, in theory, both are identified. This is, of course, problematic because they have different implications for the economic processes at work.

It is also informative to derive the reduced form for the general SD model. Substituting for **y** we get:

$$
\begin{aligned}
y_i &= \rho\mathbf{w}_i'(\rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \mathbf{u}) + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}\boldsymbol{\gamma} + u_i \\
&= \rho^2\mathbf{w}_i'\mathbf{W}\mathbf{y} + \rho\mathbf{w}_i'\mathbf{X}\boldsymbol{\beta} + \rho\mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}\boldsymbol{\gamma} + v_i \qquad (9) \\
&= \rho^2\mathbf{w}_i'\mathbf{W}\mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}(\rho\boldsymbol{\beta} + \boldsymbol{\gamma}) + \rho\mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\gamma} + v_i \\
&= [\ldots] \\
&= \rho^n\mathbf{w}_i'\mathbf{W}^{n-1}\mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + \rho\mathbf{w}_i'\mathbf{W}\mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + \rho^2\mathbf{w}_i'\mathbf{W}^2\mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + [\ldots] + v_i
\end{aligned}
$$

where, for simplicity, v_i denotes the spatial lag terms in u_i. Under standard regularity conditions on ρ and $\mathbf{w}_i$, $\lim_{n\to\infty} \rho^n\mathbf{w}_i'\mathbf{W}^{n-1}\mathbf{y} = 0$ so we can ignore the spatially lagged term in **y**. Notice that, in the reduced form, the only thing that distinguishes this from the SLX model is the cross coefficient restrictions on the terms in $\mathbf{w}_i'\mathbf{W}^{n-1}\mathbf{X}$ for *n*>1.

7

In short, spatial interaction in y_i, spatial externalities in $\mathbf{x}_i$, or spatially omitted variables lead to different spatial econometric specifications. However, the reduced form for all these models is:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}\boldsymbol{\pi}_1 + \mathbf{w}_i'\mathbf{WX}\boldsymbol{\pi}_2 + \mathbf{w}_i'\mathbf{W}^2\mathbf{X}\boldsymbol{\pi}_3 + [\ldots] + v_i \qquad (10)$$

and the only differences arise from how many spatial lags of **X** are included, constraints on the way the structural parameters enter into the composite parameters $\boldsymbol{\Pi}$, and whether or not the error term is spatially correlated. It should be clear that distinguishing which of these models generates the data that the researcher has at hand is going to be difficult. This is the case, because the specification of W is (in most situations) arbitrarily chosen, and because all the spatial lags of **X** are just neighbour averages that are almost always very highly mutually correlated. We now consider these difficulties in detail.

3. THE REFLECTION PROBLEM AND IDENTIFICATION IN SPATIAL ECONOMETRIC MODELS

Readers familiar with the 'neighbourhood effects' literature, will see immediate parallels between the spatial econometrics models (SAR, etc) and 'linear-in-means' neighbourhood and peer effects models that appear in many branches of applied economics. The SAR model is identical in structure to what Manski (1993, 2000) termed an 'endogenous' neighbourhoods effects model:

$$y_i = \rho E[y_j \mid a] + \mathbf{x}_i'\boldsymbol{\beta} + u_i \qquad (11)$$

where y_i is the outcome of interest, $\mathbf{x}_i$ a vector of exogenous variables, u_i an error and *a* indexes areas (non-nested 'neighbourhoods'). This specification captures the direct effect of neighbours' expected outcome on own outcomes. The endogenous effects model uses *expected* outcomes to capture the idea that decisions are simultaneous. As agents do not observe outcomes before making their own decision they instead have to form an expectation of other outcomes in their neighbourhood. The endogenous effects model is essentially a SAR model where, for a neighbourhood of size *n*, the *j*[th] element of $\mathbf{w}_i$ takes value 1/n if *i* and *j* are in the same neighbourhood, zero otherwise. Using $\mathbf{w}_i$ so defined, we can write:

$$y_i = \rho\mathbf{w}_i'\mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + u_i \quad (11a)$$

(11a) is the empirical counterpart to (11) in the sense that $\mathbf{w}_i$ is a neighbourhood weighting vector that provides an estimate of the mean, conditional on the location of *i*.

There are three endogeneity problems if (11a) is estimated using data on y_i, $\mathbf{x}_i$ and neighbourhood averages $\mathbf{w}_i'\mathbf{y}$ (the sample analogue to $E[y_j \mid a]$). As usual, the explanatory variables $\mathbf{x}_i$ may be correlated with u_i, which is a general issue that we will come back to later in this paper. The first specifically spatial endogeneity problem is that, by construction, $\mathbf{w}_i'\mathbf{y}$ depends on $\mathbf{u}$ when $\rho \neq 0$. The second is that the error term may be spatially autocorrelated which provides another reason why $\mathbf{w}_i'\mathbf{y}$ depends on $\mathbf{u}$. For the moment, we focus on the first of these spatial endogeneity issues.

One possible solution is to use the reduced form and see if this provides estimates of both ρ and $\boldsymbol{\beta}$. As with the standard SAR model, repeated substitution (for $E[y_j \mid a]$), gives:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \rho\mathbf{w}_i'\mathbf{X}\boldsymbol{\beta} + \rho^2\mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \rho^3\mathbf{w}_i'\mathbf{W}^2\mathbf{X}\boldsymbol{\beta} + [\ldots] + u_i \qquad (12)$$

where, once again, we are using $\mathbf{w}_i'\mathbf{X}$ as the empirical counterpart to $E[\mathbf{X} \mid a]$. In contrast to the standard SAR model, repeated substitution leaves the error (u_i) unchanged, assuming $E[\mathbf{u}] = \mathbf{0}$, because the expectations operator passes through the linear spatial weights so that $E[\mathbf{w}_i'u_i] = \mathbf{w}_i'E[u_i] = \mathbf{0}$. As we see later, this will not be the case if the errors are correlated within neighbourhoods. Providing ρ is less than one in absolute value and using the fact that in the neighbourhoods literature (once we order observations by neighbourhood) **W** is typically block diagonal and idempotent allows us to simplify the reduced form as follows:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}\boldsymbol{\beta}\rho/(1-\rho) + u_i \qquad (13)$$

The reduced form models outcome y_i as a function of individual exogenous characteristics ($\mathbf{x}_i$) and the neighbourhood average of those characteristics ($\mathbf{w}_i'\mathbf{X}$). Armed with data on these, OLS will give a direct estimate of $\boldsymbol{\beta}$ (from the coefficient on $\mathbf{x}_i$). We can then back out ρ using the estimate of $\boldsymbol{\beta}$ and the fact that the coefficients on neighbourhood average characteristics $\mathbf{w}_i'\mathbf{X}$ equal $\boldsymbol{\beta}\rho/(1-\rho)$. That is, the reduced form allows for the separate identification of $\boldsymbol{\beta}$ and ρ.

Things are much more problematic if neighbourhood exogenous characteristics ($\mathbf{w}_i'\mathbf{X}$) *directly* influence individual outcomes. That is, if we allow for what Manski (1993) called 'exogenous' or 'contextual' neighbourhood effects. As is well known in the neighbourhoods literature and as demonstrated by Manski (1993), once we allow for direct effects from $\mathbf{w}_i'\mathbf{X}$ on y_i we can no longer separately identify $\boldsymbol{\beta}$ and ρ. To see why, note that the assumed relationship in the population becomes

$$y_i = \rho E[y_j \mid a] + \mathbf{x}_i'\boldsymbol{\beta} + E[\mathbf{X} \mid a]\boldsymbol{\gamma} + u_i \qquad (14)$$

where notation is identical to equation (11) and $E[\mathbf{X} \mid a]\boldsymbol{\gamma}$ captures the direct effect of neighbourhood characteristics on individual outcomes. As already noted, the endogenous effects model can be written as a SAR model with the spatial weight matrix block diagonal and idempotent. Similarly, allowing for exogenous neighbourhood effects gives the SD model with y_i determined by a spatial lag of both the endogenous and exogenous variables. As with the standard SD model (see equation (9)), repeated substitution gives the reduced form:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}(\boldsymbol{\gamma} + \boldsymbol{\beta}\rho)/(1-\rho) + u_i \qquad (15)$$

where the only difference with the standard SD model is that the error is unchanged.[8] OLS regression of y_i on individual characteristics ($\mathbf{x}_i$) and the neighbourhood average of those characteristics ($\mathbf{w}_i'\mathbf{X}$) gives two reduced form coefficients in three unknown structural coefficients. As before, OLS gives a direct estimate of **β**, but only the composite parameter vector $(\mathbf{\gamma}+\mathbf{\beta}\rho)/(1-\rho)$ is identified. That is to say, only the *overall* effect of neighbours' characteristics is identified, but not whether they work through exogenous or endogenous neighbourhood effects (i.e. through $\mathbf{w}_i'\mathbf{X}$ or $\mathbf{w}_i'\mathbf{y}$, respectively). This is the 'reflection' problem of Manski (1993).

Serious problems also arise if there is spatial correlation in the unobserved components u_i. This may happen because of sorting (unobservably similar agents tend to be co-located), common unobserved shocks or causal linkages between neighbours unobserved characteristics. For simplicity, assuming that neighbourhood exogenous characteristics ($\mathbf{w}_i'\mathbf{X}$) do *not* directly affect outcomes this gives:

$$y_i = \rho_1 E[y_j \mid a] + \mathbf{x}_i'\mathbf{\beta} + u_i \;\; ; u_i = \rho_2 E[u_j \mid a] + v_i \qquad (16)$$

Substituting $u_i = y_i - \rho_1 E[y_j \mid a] - \mathbf{x}_i'\mathbf{\beta}$ for u_i in $u_i = \rho_2 E[u_j \mid a] + v_i$ and rearranging gives:

$$y_i = (\rho_1 + \rho_2 - \rho_2\rho_1)E[y_j \mid a] + \mathbf{x}_i'\mathbf{\beta} - \rho_2 E[\mathbf{X} \mid a]\mathbf{\beta} + v_i \qquad (17)$$

Letting $\mathbf{\gamma} = -\rho_2\mathbf{\beta}$ and $\rho = (\rho_1 + \rho_2 - \rho_2\rho_1)$ and using $\mathbf{w}_i'\mathbf{X}$ and $\mathbf{w}_i'\mathbf{y}$ as empirical counterparts to the expectation terms we get:

$$y_i = \rho\mathbf{w}_i'\mathbf{y} + \mathbf{x}_i'\mathbf{\beta} + \mathbf{w}_i'\mathbf{X}\mathbf{\gamma} + u_i \qquad (18)$$

which, ignoring the parameter restrictions on ρ and **γ,** is identical to (14a) with the reduced form as in (15). Again, estimation provides two coefficients which identify **β** but do not allow separate identification of ρ and **γ** (which, in turn, means the underlying structural parameters in (16a) are not separately identified). Intuitively, this makes sense: with cross-sectional data how can you distinguish between something unobserved and spatially correlated driving spatial correlation in y from the situation where y is spatially correlated because of direct interaction between outcomes? Of course if, in addition to ruling out a direct impact for neighbourhood exogenous characteristics, we also rule out a direct impact from neighbourhood endogenous outcomes (i.e. set $\rho_1 = 0$ in (16)) then the coefficients are separately identifiable just as they were in the endogenous effects model. Notice, also, that (16) cannot be distinguished from the SD model involving spatial lags of the endogenous and exogenous variables without prior knowledge of the sign on coefficient **γ** in equation (14a) or other ad-hoc assumptions. Of course, all these conclusions are conditional on the maintained assumption that $\mathbf{x}_i$ and u_i are uncorrelated. We will return to this issue later.

Now we have outlined the reflection problem, it is useful to consider to what extent it applies to general applications of the SAR, SD and SAR plus SE models. This has been the subject of surprisingly

---

[8] See the discussion following equation (12).

little discussion (Lee, 2004 and Pinkse and Slade, 2010 are notable exceptions). In the most trivial case, if the spatial weights matrix is idempotent and interactions are contemporaneous then the reflection problem *must* apply if spatial lags of the exogenous variables also influence y_i because the correct model is exactly equation (14a). This makes it clear that spatial econometrics cannot help with the identification problem in the neighbourhood effects specification. Of course, one could take data on y_i, $\mathbf{w}_i^{'}\mathbf{y}$, and $\mathbf{w}_i^{'}\mathbf{X}$, estimate the SD model using ML and sampling variation would ensure you got estimators of **β**, **γ** and ρ. Clearly, however, it is pointless trying to reach *any* conclusions about the actual process at work on the basis of these estimates.

What if we are truly interested in spatial lags of observed, rather than expected y_i? Again, if the spatial weights matrix is idempotent the same problems arise as for the neighbourhood effects literature. Taking the SAR reduced form given in equation (8) and simplifying using the fact that W is idempotent gives the reduced form:

$$y_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{w}_i^{'}\mathbf{X}\boldsymbol{\beta}\rho/(1-\rho) + v_i \qquad (19)$$

where the fact that $v_i = \rho\mathbf{w}_i^{'}\mathbf{v} + u_i$ is the only difference with the neighbourhoods effect literature. As before OLS estimation identifies **β** and ρ. Clustering by the blocks in **W** provides robust standard errors, although we could increase efficiency by exploiting the fact that we are assuming the exact structure of dependence is known. We reach exactly the same conclusion for the SD and the SAR plus SE model. If we have cross section data, and are interested in situations where **W** is idempotent, switching from E[y_i] to y_i does nothing to solve the identification problem.

For researchers from the neighbourhood effects literature, the reflection problem raises such profound identification issues that they find it baffling that spatial econometrics generally ignores them. Most would assume that the ML estimators exploit some non-linearity to achieve identification. This is probably because the use of non-linearities to identify neighbourhood effects has been suggested by Brock and Durlauf (2001) in the neighbourhood effects literature – for example, when the outcome is a discrete variable, and the researcher is willing to impose a logit or probit functional form. More precisely, ML estimation uses a combination of cross coefficient restrictions and the structure of **W** to achieve identification. In the neighbourhoods effect literature W is idempotent. In traditional spatial econometric models , it is the assumption that most standard **W** matrices are not idempotent that allows identification.[9]

It is useful to think about the way in which this helps solve the identification problem. If all the assumptions are correct, then SAR models are identified even with an idempotent **W** matrix, because direct effects from spatial lags of X have been assumed away. There thus seems little point in considering them further. Instead, consider the SD model (and by extension the SE model as they have the same reduced form). Equation (10) showed how substitution for **y** gives the reduced form:

---

[9] Interestingly, the idea of putting more structure on neighbourhood effects (e.g. by assuming a hierarchical network) has recently been suggested as a way of solving the identification problem. See Lee (200x).

$$y_i = \rho^n \mathbf{w}_i^{'} \mathbf{W}^{n-1} \mathbf{y} + \mathbf{x}_i^{'} \boldsymbol{\beta} + \mathbf{w}_i^{'} \mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + \rho \mathbf{w}_i^{'} \mathbf{W} \mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + \rho^2 \mathbf{w}_i^{'} \mathbf{W}^2 \mathbf{X}(\boldsymbol{\beta}\rho + \boldsymbol{\gamma}) + [\ldots] + v_i \quad (20)$$

As discussed above, under standard regularity conditions on ρ and $\mathbf{w}_i$, $\lim_{n\to\infty} \rho^n \mathbf{w}_i^{'} \mathbf{W}^{n-1} \mathbf{y} = 0$ so we can ignore the spatially lagged term in y. The final part of this expression makes it clear how restrictions on the coefficients on $\mathbf{w}_i^{'} \mathbf{X}$, $\mathbf{w}_i^{'} \mathbf{W} \mathbf{X}$, $\mathbf{w}_i^{'} \mathbf{W} \mathbf{W} \mathbf{X}$ etc, could be exploited to help with identification (remembering that there are only three underlying structural parameters).

Clearly, whether these restrictions are useful depends crucially on two factors. The first is the extent to which the SD model and the spatial weights are correctly specified. If the higher order lags $\mathbf{w}_i^{'} \mathbf{W} \mathbf{X}$, $\mathbf{w}_i^{'} \mathbf{W} \mathbf{W} \mathbf{X}$ affect y directly, then they cannot provide additional information that is useful in identifying ρ. One good reason for worrying about this is that if the true **W** is unknown, the **W²** **W³** etc may be better specifications of the connectivity between i and its neighbours than **W** (e.g. $\mathbf{x}_i$ has an effect up to 5km, but $\mathbf{w}_i$ incorrectly restricts effects to within 2km). Secondly, the spatial lags of $\mathbf{x}_i$ will tend to be highly collinear. If collinearity means that the reduced form coefficients are not precisely estimated then this means the underlying structural parameters will not be precisely estimated.

In theory the degree of collinearity depends on sample size, sampling frame and how **W** changes as observations are added.[10] In practice, in reasonable sized samples and using standard specifications for **W**, $\mathbf{w}_i^{'} \mathbf{X}, \mathbf{w}_i^{'} \mathbf{W} \mathbf{X}$, etc are likely to be highly correlated for the simple reason that they are a weighted average (and consistent estimate of the mean) of $\mathbf{x}_i$ in some neighbourhood of *i*. As a result in many applications the parameters on $\mathbf{w}_i^{'} \mathbf{X}$, $\mathbf{w}_i^{'} \mathbf{W} \mathbf{X}$, etc are likely to be weakly identified. A nice way to illustrate this, which will resonate with researchers from the experimental paradigm is to consider using IV, rather than the reduced form, to estimate the SD model.

In fact, before theoretical and computational advances made ML estimation of the SD model possible for large samples, IV estimation was the only feasible approach. If we assume that the SD model is the correct spatial model, then IV estimation is attractive because the spatial structure of the data can be used to construct valid instruments for $\mathbf{w}_i^{'} \mathbf{y}$. Specifically, from equation (9) it is clear that, assuming the SD model is correct, $\mathbf{w}_i^{'} \mathbf{W} \mathbf{X}$ and $\mathbf{w}_i^{'} \mathbf{W} \mathbf{W} \mathbf{X}$ do not directly determine y_i, except via their effect on $\mathbf{w}_i^{'} \mathbf{y}$. Thus $\mathbf{w}_i^{'} \mathbf{W} \mathbf{X}, \mathbf{w}_i^{'} \mathbf{W} \mathbf{W} \mathbf{X}$, etc provide valid instruments for $\mathbf{w}_i^{'} \mathbf{y}$. This idea is the basis for the various IV estimators of spatial models that have been proposed.

For simplicity, if only the second order lag is used as an instrument, the first stage regression is:

---

[10] In theoretical analysis it is usual to distinguish between increasing domain asymptotics (adding observations expands the space over which we estimate) and infill asymptotics (increasing sample size means sampling more observations in a given area). The latter makes more sense in many micro-econometric settings, which is problematic because, general results are not available for MLE under infill asymptotics (see Lee (2004)). Our intuition is that consistency under infill asymptotics will be difficult to derive precisely because, for many common W matrices, infill asymptotics increases the correlation between x and its spatial lags.

$$\mathbf{w}_i'\mathbf{y} = \mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{w}_i'\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\beta}_3 + u_i \qquad (21)$$

Following an influential paper by Staiger and Stock (1997) applied researchers worry about the strength of the instruments in the first stage regression (and the implications that weak instruments have for the coefficient estimates in the second stage, which can be both biased and imprecisely estimated as a consequence). Given our discussion above it should be clear that as $\mathbf{w}_i'\mathbf{X}$ and $\mathbf{w}_i'\mathbf{W}\mathbf{X}$ will tend to be highly correlated, it is unlikely that $\mathbf{w}_i'\mathbf{W}\mathbf{X}$ will provide a very strong instrument conditional on $\mathbf{w}_i'\mathbf{X}$. There is simply very little meaningful information in the spatial average-of-spatial-averages, conditional on the spatial average of X in that location. This is in addition to the point we make above: that the assumptions on **W** and the *spatial* lag structure are crucial to ensuring $\mathbf{w}_i'\mathbf{W}\mathbf{X}$, $\mathbf{w}_i'\mathbf{W}\mathbf{W}\mathbf{X}$, etc satisfy the exclusion restrictions for a valid set of instruments.

We demonstrate the collinearity and weak instrument issues that occur in practice using a simple example based on census data for one year on the population of children in state schools in London at the end of primary (elementary) school. The data, described in detail in Gibbons and Telhaj (2008) is taken from an administrative database and records whether the child was registered for free school meals (FSM) at the time of the census, their scores in national science tests, and their residential postcode.[11] Each unique postcode generally identifies 10-20 contiguous houses. The data is aggregated to give mean FSM and mean science scores by postcode. We define (normalised) spatial weights using inverse distance weighting with a cutoff at 2000 metres.

Table 1 reports correlations between these variables and various 'spatial lags'. The top panel reports these for the full census sample. The first column shows that the variables (FSM, Science) are only moderately correlated with their spatial lags.[12] However, it is clear that, $\mathbf{w}_i'\mathbf{X}$, $\mathbf{w}_i'\mathbf{W}\mathbf{X}$ and $\mathbf{w}_i'\mathbf{W}\mathbf{W}\mathbf{X}$ are all highly correlated. It would be extremely unwise to include all these variables in a regression and, conditional on $\mathbf{w}_i'\mathbf{X}$ (which we assume directly influences y_i), the higher order spatial lags will clearly be weak instruments. A smaller sample buys more independent variation, but not much more, as evidenced by the correlations for a 10% sample reported in the lower panel. What is more, this independent variation arises simply from sampling variation (noise) which is of no value as a source of identification of causal parameters.

---

[11] Free school meals are offered to children of families who are in receipt of a range of welfare benefits.

[12] The Moran's I – which is the coefficient from a regression of $\mathbf{w}_i'\mathbf{X}$ on $\mathbf{x}_i$ - are 0.11 for FSM and 0.07 for Science Scores.

Table 1: Example correlations between spatial lags

100% census

	FSM	$\mathbf{w}_i'$FSM	$\mathbf{w}_i'$WFSM	$\mathbf{w}_i'$WWFSM
$\mathbf{w}_i'$FSM	0.3378	1.0000		
$\mathbf{w}_i'$WFSM	0.3508	0.9844	1.0000	
$\mathbf{w}_i'$WWFSM	0.3373	0.9763	0.9980	1.0000
$\mathbf{w}_i'$WWWFSM	0.3328	0.9692	0.9947	0.9991
	Sci	$\mathbf{w}_i'$Sci	$\mathbf{w}_i'$WSci	$\mathbf{w}_i'$WWSci
$\mathbf{w}_i'$Sci	0.2671			
$\mathbf{w}_i'$WSci	0.2859	0.9688		
$\mathbf{w}_i'$WWSci	0.2637	0.9501	0.9950	
$\mathbf{w}_i'$WWWSci	0.2554	0.9330	0.9867	0.9977

10% sample

	FSM	$\mathbf{w}_i'$FSM	$\mathbf{w}_i'$WFSM	$\mathbf{w}_i'$WWFSM
$\mathbf{w}_i'$FSM	0.2773	1.0000		
$\mathbf{w}_i'$WFSM	0.4303	0.9140	1.0000	
$\mathbf{w}_i'$WWFSM	0.3637	0.9261	0.9872	1.0000
$\mathbf{w}_i'$WWWFSM	0.3630	0.9002	0.9847	0.9962
	Sci	$\mathbf{w}_i'$Sci	$\mathbf{w}_i'$WSci	$\mathbf{w}_i'$WWSci
$\mathbf{w}_i'$Sci	0.2246	1.0000		
$\mathbf{w}_i'$WSci	0.4127	0.8669	1.0000	
$\mathbf{w}_i'$WWSci	0.3190	0.8938	0.9691	1.0000
$\mathbf{w}_i'$WWWSci	0.3193	0.8410	0.9717	0.9865

If further evidence is needed, consider the results from IV estimation of the SD model in which we instrument $\mathbf{w}_i'\mathbf{y}$ with $\mathbf{w}_i'\mathbf{WX}$. Note that $\mathbf{w}_i'\mathbf{WX}$ is unlikely to be a valid instrument because it will be correlated with unobservables (u_i) that determine science scores (an issue to which we return below). But here we are simply making a point about weak instruments (and as $\mathbf{w}_i'\mathbf{WX}$ proxies for unobservables this will make it appear like a stronger instrument).

Table 2: Example spatial IV

	OLS	IV
$\mathbf{w}_i'\mathrm{Sci}$	0.97	15.18
	(0.02)	(7.44)
FSM	-11.67	-7.32
	(0.35)	(2.54)
$\mathbf{w}_i'\mathrm{FSM}$	7.94	461.13
	(1.28)	(237.00)
Constant	2.24	-822.81
	(1.43)	(432.00)
First stage		
Partial R-squared	-	0.001
F test	-	4.11
Obs	37968	37968

The results are shown in Table 2. The first column reports OLS estimates, which indicate that children's scores tend to move one-for-one with neighbours' scores. This obviously isn't because of neighbourhood effects, but because neighbours share many characteristics like income, parental education, local school quality, etc. in common. Column 2 shows results when we implement the traditional spatial econometrics approach and use the higher order lags of x (in this case $\mathbf{w}_i'\mathbf{WFSM}$) as instruments. The coefficient on the spatial lag jumps up to an implausible magnitude, and becomes only marginally significant. But the first stage diagnostics are of more interest than the coefficients themselves, and are shown at the bottom of the table. The partial r-squared of the instruments (i.e. the share of the variance in $\mathbf{w}_i'\mathrm{Sci}$ that is not explained by FSM and $\mathbf{w}_i'\mathrm{FSM}$, but that is explained by $\mathbf{w}_i'\mathbf{WFSM}$ is a mere 0.1%). The F-statistic of the excluded instrument in the first stage is 4, which is well below the simple rule-of thumb minimum of 10 (Staiger and Stock 1997) and the tabulated critical values of Stock and Yogo 2005. In short, there is a fundamental weak identification problem with this model (coupled in this case with a problem of implausible exclusion restrictions).

To summarise, in theory if **W** is not idempotent identification is possible, but the parameters in the SD model are likely only to be weakly identified in many applications. In addition to the weak instruments problem there are also good reasons to doubt the validity of the exclusion restrictions that make higher order spatial lags of $\mathbf{x}_i$ valid instruments. First, and most fundamentally, we cannot be sure that the included variables $\mathbf{x}_i$ are exogenous. In many applied situations $\mathbf{x}_i$ is likely to be endogenous, a problem that has essentially been ignored by much of the spatial econometrics literature, to focus instead on the issue that by construction $\mathbf{w}_i'\mathbf{y}$ depends on $\mathbf{u}$ (providing $\rho \neq 0$)**.** If $\mathbf{x}_i$ is endogenous then, as usual, we cannot get consistent estimators without finding some suitable set of instruments for $\mathbf{x}_i$. In addition, if there is spatial autocorrelation in the errors, then the endogeneity of $\mathbf{x}_i$ rules out the possibility of using spatial lags of $\mathbf{x}_i$ as instruments for $\mathbf{w}_i'\mathbf{y}$.

Of course, the problem of endogenous $\mathbf{x}_i$ is not unique to spatial econometrics. A more specific problem is that, for the exclusion restrictions on the spatial lags of $\mathbf{x}_i$ to be valid $\mathbf{w}_i$ must be correctly specified. If it is not then $\mathbf{w}_i'\mathbf{WX}$, etc will simply pick up problems arising from the misspecification of $\mathbf{w}_i$. It is not clear how this assumption could ever satisfactorily be evaluated.

It should be clear that there are fundamental problems in using the spatial structure of the data to generate instruments that allow estimation of the SD model. The problems are even more profound if we allow for spatial autocorrelation in the error terms, and drop the assumption of the exogeneity of $\mathbf{x}_i$. These changes give rise to the general spatial model:

$$y_i = \rho \mathbf{w}_i'\mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\mathbf{X}\boldsymbol{\gamma} + u_i \qquad (22)$$

$$u_i = \rho \mathbf{w}_i'\mathbf{u} + v_i \qquad (23)$$

$$x_{ik} = \mathbf{z}_i'\boldsymbol{\delta}_1 + \delta_2 u_i + e_{ik} \qquad (24)$$

where x_{ik} is an element of $\mathbf{x}_i$ and $\mathbf{z}_i$ is a vector of factors determining x_{ik}, which may include elements of $\mathbf{x}_i$. This general spatial model presents all the challenges described above, plus the additional problem of one (or more) explicitly endogenous explanatory variables. Typical spatial ML methods simply assume away (24) and treat $\mathbf{x}_i$ as exogenous. True, the parameters of this model could all, in principle, be estimated by ML techniques or spatial IV techniques, imposing all the restrictions that are implied by the specification of W and the way the model is written down. Nevertheless, this mechanical approach to estimation will not appeal to many applied researchers who view minimal assumptions on functional forms and strong identification as necessary conditions to infer anything about causality.

A better approach is to more precisely delineate the research question, and focus on the key parameter of interest, whether this parameter is relevant for policy or because of its theoretical importance. The design-based, experimental paradigm insists that a satisfactory strategy must use theoretical arguments or informal reasoning to make a case for a source of exogenous variation that can plausibly used to identify this parameter of interest. In addition, this mode of research would expect rigorous empirical testing to show demonstrate the validity of these assumptions, as far as possible. We now consider these issues in more detail.

4. THE EXPERIMENTAL PARADIGM AND SPATIAL ECONOMETRICS

The discussion so far has been critical of the spatial econometrics approach, particularly regarding the crucial issue of identification of causal parameters. Others have made similar arguments although perhaps not as forcefully (see for example McMillen 2010). Of course, any alternative approach also has to solve the identification problems that bedevil spatial analysis. Our argument is that these problems are so fundamental that they must sit at centre stage of good applied work, not be shunted to the sidelines through the use of ML that assumes knowledge of the appropriate functional forms and spatial weights. In this section we argue that spatial research would be best

served by turning away from the application of generic spatial models and from the obsession with trying to distinguish between observationally equivalent models using contestable parameter restrictions that emerge only from the fairly arbitrary way the assumed model is specified. Instead, we advocate using strategies that have been carefully designed to answer well-defined research questions. In short, whilst not ignoring lessons from spatial econometrics, applied research should focus on the use of identification strategies that are at the core of the design-based experimental paradigm.

To simplify, start by assuming that we have some spatial data and are interested in estimating:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i \qquad \text{(21)}$$

When a spatial econometrician estimates (21) they would check whether the residuals are spatially correlated. We think this is good practice, which should be extended to non-spatial econometric treatment of spatial data. What if spatial correlation is detected? The default approach of most applied econometricians would be to assume the presence of a spatially correlated unobserved variable and to cluster the errors using pre-defined spatial units, content to err on the side of caution to avoid Type I errors (perhaps reporting standard errors under different clustering assumption). If one is willing to impose more structure, e.g. assume the errors follow the SE model, ML gives efficient estimates of $\boldsymbol{\beta}$ and correct standard errors. In structural modelling if theory places structure on the spatial correlation this may be useful. When working with reduced forms any efficiency gain depends on how well the assumed $\mathbf{w}_i$ approximate the true $\mathbf{w}_i$ governing the error process. Given the true $\mathbf{w}_i$ are unknown, the spurious precision offered by ML is a poor reason to implement it in practice.

If spatial correlation arises from an unobserved variable correlated with $\mathbf{x}_i$ we have an omitted variable problem and estimates of $\boldsymbol{\beta}$ are biased. Spatial autocorrelation in $\mathbf{x}_i$ and in the omitted variables is likely to exacerbate this bias relative to the case where the variables do not exhibit spatial autocorrelation. Increasingly, following the derivation in section 2, the presence of spatially correlated omitted variables is used in spatial econometrics to justify estimation of the SD model, in order to 'solve' the omitted variable problem. However, if this was a general solution, it would also work for non-spatial panel data. The equivalent of (5)-(7) with panel data is:

$$y_{it} = \beta x_{it} + z_{it} \qquad \text{(22)}$$

$$z_{it} = \rho z_{it-1} + \gamma x_{it} + v_{it} \qquad \text{(23)}$$

$$y_{it} = \rho y_{it-1} - \rho \beta x_{it-1} + (\gamma + \beta) x_{it}' + v_{it} \qquad \text{(24)}$$

This equation can be estimated consistently by ML or non-linear least squares, or estimates of the various parameters retrieved from the OLS coefficients.

Although endogeneity problems of this type might be mitigated by this strategy, it is certainly not a complete fix. To see this, modify the set up in equations (5)-(7) slightly to cope with more general endogeneity in that x_i is partly determined by the omitted variable (f_i). In this case we have:

$$y_i = \beta x_i + z_i \qquad\qquad (25)$$

$$z_i = \rho \mathbf{w}_i' \mathbf{z} + \gamma f_i + v_i$$

$$x_i = f_i + u_i$$

$$y_i = \rho w_i' \mathbf{y} + (\gamma + \beta) x_i - \rho \beta w_i' \mathbf{x} + (v_i - \gamma u_i)$$

The error term now has a component u_i that is negatively correlated with f_i, so the coefficients cannot be estimated consistently by OLS, NLS or ML. In this more general setting, the SD model does not provide a solution that gives consistent estimates for the parameter of interest (β). See, for example, Todd and Wolpin (2003) for a related discussion in the context of 'value-added' models in the educational literature. In short, the SD model should not be seen as a general solution to the omitted variable problem in spatial research, and it is a mistake to proceed as if it is one. A better solution is to treat this as a standard endogeneity problem that makes $\mathbf{x}_i$ correlated with the error, and to bring to bear tools for dealing with such problems. In the applied microeconometric literature this would usually mean adopting one of two approaches based on either instrumental variables or some kind of differencing (e.g. the use of fixed effects or discontinuities).

What about the situation where we are interested in estimating the SAR or SD model to test for the presence of direct spatial interaction between outcomes y_i ? It is hard to imagine situations in which this *is* the true data generating process because simultaneous decisions based on y must rely on expectations (as in the neighbourhoods effect literature), but let us suppose that estimation of ρ is the goal. As argued above, in most situations, we find ML solutions unconvincing, fundamentally because of the unverifiable restrictions on $\mathbf{w}_i$. In some settings, the spatial econometrics literature may offer interesting insights in to the potential for using specific, restrictions on $\mathbf{w}_i$ to achieve identification, where these restrictions arise from the institutional context (e.g. from the directed structure of friendship networks, or the spatial scope of area targeted policies). However, for most applied problems, however, uncertainty about functional forms and spatial weights mean alternative strategies are more appropriate.

One possibility is to consider IV. As we argued above, *if* the SAR model is correctly specified then $\mathbf{w}_i' \mathbf{X}$ provides an instrument for $\mathbf{w}_i' \mathbf{y}$. In practice, researchers working in most areas of modern micro-econometric practice would expect very careful arguments to justify the exclusion of $\mathbf{w}_i' \mathbf{X}$ from the estimating equation.[13] Institutional arrangements might provide exogenous variation in

---

[13] Of course, if $\mathbf{w}_i' \mathbf{X}$ can be excluded then this solves the identification problem for ML as well. Even in this case, we still think that the case for switching to ML is weak because it relies on precise knowledge of $\mathbf{w}_i$.

$\mathbf{w}_i'\mathbf{X}$ that has no direct influence on y_i. For instance, a researcher might argue that there are no direct impacts from a policy intervention in neighbouring districts, but the policy does have effects via neighbouring outcomes. Researchers also use similar ideas to argue that certain 'natural experiments' directly induce exogenous variation in $\mathbf{w}_i'\mathbf{y}$. In both cases IV is justified, although a researcher employing this strategy often faces difficulty convincing others that there are no direct effects from neighbours' policies (i.e. justifying the exclusion restrictions).

Given the difficulty in justifying the exclusion restrictions on $\mathbf{w}_i'\mathbf{X}$, coupled with the conceptual problems in thinking about what the SAR model implies about underlying causal relationships, we argue for abandoning the SAR model altogether. We advocate the path taken by most recent neighbourhood effects literature and argue for estimation of reduced form SLX models in $\mathbf{x}_i$ and spatial lags of $\mathbf{x}_i$, rather than any attempt at direct estimation of the SAR or SD model. Given the identification problems the reduced form approach is simply far more credible and honest. The composite reduced form parameter that describes the influence of neighbours characteristics *or* outcomes is itself a useful and policy relevant parameter. With this in hand judgements can be made based on theory and institutional context about the likely channels through which the effects operate, without imposing the untestable assumptions on model structure at the outset that are implicit in spatial econometric approaches to estimating specifications involving $\mathbf{w}_i'\mathbf{y}$. Setting to one side the challenge of estimating ρ directly also leaves the researcher free to focus on the remaining threats to consistent estimation of the composite parameters in the reduced form, which are still formidable.

The key challenge that remains is the one discussed above. That is, the fact that $\mathbf{x}_i$ and $\mathbf{w}_i'\mathbf{X}$ are unlikely to be exogenous, and will be correlated with the unobserved determinants of y_i via causal linkages or because of the sorting of agents across geographical space. This issue is generally ignored in ML-based spatial econometrics in which the main focus of attention is consistent estimation of models under very strong (but poorly justified) assumptions. Estimation of the reduced form SLX models would instead force researchers to focus on finding sources of exogenous variation in $\mathbf{x}_i$ and $\mathbf{w}_i'\mathbf{X}$ with which to identify their corresponding parameters: which is in itself a challenge.

How then should researchers working on spatial empirical analysis proceed? There are many potential examples of 'natural experiments' in the spatial context which offer channels for identification of interesting spatial parameters (e.g. Redding and Sturm, 2009). Some settings also offer explicit sources of randomisation due to institutional rules and processes (Sacerdote 2001, Giorgi et al 2008). Field experiments designed specifically for purpose are also clearly very useful. However, big ones like the Moving to Opportunity Programme are rare, costly and often suffer from unavoidable design flaws, and small ones tend to suffer from concerns about external validity. It would also be very difficult to design experiments to answer many spatial questions and we do not advocate this as a general way forward. However, the standard toolkit of IV and differencing based

strategies employed by researchers in many other fields of applied economics can be used effectively, if applied carefully and with attention to the identification of specific causal parameters rather than an arbitrarily specified system of equations.

If we want to use an IV strategy to get consistent estimates of the parameters of interest in these reduced form SLX models, we need instruments that satisfy the usual relevance and exclusion restrictions. In the highly unlikely situation that we know the structure of the spatial dependency spatial econometricians might argue for the same strategy discussed above for $\mathbf{w}_i'\mathbf{y}$. That is, to use higher order spatial lags of $\mathbf{x}_i$ as instruments. However, using lags as instruments isn't often a good idea in the time-series context when causality runs in only one direction. For example, using historical city populations as instruments for current city populations to studying agglomeration economies can only work if the researcher is sure that whatever unobservable factors made a city big in the past are not what makes it big in the present. In the spatial case, bi-directional causality makes this kind of strategy even less compelling. Further, as discussed above, we expect weak instrument problems even in situations where the exclusion restrictions are valid. Finally, in most applications the true spatial weights are unknown raising considerable uncertainty about the exclusion restrictions.[14] For all these reasons, we believe that standard IV strategies which require the researcher to pay careful attention to the omitted variables and to clearly justify the validity of instruments represent a more appropriate way to address the problem of spatially correlated omitted variables[15]. In contrast, we have little faith in generic spatial econometric solutions which rely on generic assumptions about the functional form and spatial weights, with little focus on careful research design.

There are many examples of these kinds of instrumenting strategies applied to spatial problems, by researchers working outside the traditional spatial econometric mould. For example, Michaels (2008) uses the grid-like planning of the US highway network to predict whether towns experienced exogenous improvements in market access as the network developed. Luechinger (2007) use the sites of installation of SO2 scrubbers and prevailing wind directions to predict pollution levels. Gibbons, Machin and Silva (2008) use distances to school admissions district boundaries to predict levels of choice and competition in school markets. These strategies may not be without their problems, but at least provide some hope of uncovering causal relationships in the spatial context, which off-the-shelf spatial econometrics techniques do not.

A good alternative to IV is differencing to remove relevant omitted variables, either through difference-in-difference, fixed effects or regression discontinuity designs. In this case, the fact that the unobserved component is spatially correlated helps because it suggests that spatial differencing

---

[14] Note, however, that while the spatial structure of the data doesn't help, neither does it especially hinder the search for a suitable instrument. If we have an instrument that is independent of u_i, then it is also independent of $\mathbf{w}_i'\mathbf{u}$, (unless the weights are endogenous) so the fact that x_i and u_i are both spatially correlated is irrelevant (aside from the implications for standard errors).

[15] Control function approaches may also be equally valid, but require instruments too, and generally require more assumptions than IV.

(of observations with their "neighbours") is likely to be effective. Holmes (1998) provides an early example. Gibbons, Machin and Silva (2009) provide more recent discussion. Other differencing strategies drawing on a "case-control" framework may also be appropriate, for example the 'million dollar plant' analysis of Greenstone et al (2008) which compares the effect of large plant relocations on the destination counties, using their second ranked preferences – revealed in a real estate journal feature – as a counterfactual. Both Busso, Gregory and Kline (2010) and Kolko and Neumark (2010) evaluate the effects of spatial policies by comparing policy-treated areas with control areas that were treated in later periods, as a means to generating plausible counterfactuals. Of course, differencing can also be combined with instrumenting as discussed in, for example, Duranton, Gobillon and Overman (2009).To summarise, different economic motivations lead to spatial econometric specifications that will be very hard to distinguish in practice. Add to the mix the fact that in (nearly) all applications we face uncertainty about the endogeneity of $\mathbf{x}_i$, the appropriate functional form and spatial weights and it becomes clear why many applied researchers find ML or IV estimation of some assumed spatial econometric specification uninformative. Instead, we support a focus on attempting to solve identification problems using empirical strategies that have been carefully designed for the specific application. Further, if empirical strategies cannot be devised that satisfactorily identify the causal impact of the spatial lag in the endogenous variable (i.e. most applied situations) then we advocate a reduced form approach paying particular attention to the problems raised by endogeneity of the $\mathbf{x}_i$.

So far we have said little about the role of theory. Many spatial econometricians are defensive about the role theory plays in the construction of their empirical models and see comments about the lack of theory as a misguided criticism of their work (e.g. see Fingleton and Corrado in this journal volume). But the role played by theory is not our main criticism, that is the failure to adopt a careful research design that solves the problems specific to the research question being addressed, and the lack of attention to finding credible sources of random or exogenous variation in the explanatory variables of interest. This is not to say we do not think that theory is very important. Theorising, of a formalistic or more heuristic type, is of course essential in organising thoughts about how to design a research strategy and theory and assumptions at some level are necessary for any empirics. Theory is also useful once you have these causal parameter estimates to hand, when it comes to making predictions about general equilibrium effects, as long as it is made clear that these predictions are valid only for that theoretical view of the world.

Consistent with our overall approach, we argue that testing theories means correctly estimating the coefficients on specific causal variables (as suggested by the theory). This provides another point of contrast to most applied spatial econometrics where the role of theory is to derive a generic functional form with ML applied to give the parameters that ensure the best "fit" to readily available data. For example, to test the predictions of NEG models, our approach insists on a research strategy to identify whether market potential has a causal impact on wages while recognising that no model is going to completely explain the spatial distribution of wages. This contrasts strongly with the applied spatial econometrics approach which uses the extent to which different spatial econometric models 'fit' the data as a way to test competing theories. This has the unfortunate side effect of encouraging the inclusion of endogenous variables in empirical specifications as, for obvious reasons, these tend to increase the fit of the spatial model with the data.

In many spatial economic problems, theory may thus play an important role in identifying variables for which we would like to know the causal effects. But empirical implementation requires careful research design if the results are to have any general scientific credibility or to be considered trustworthy for policy making. It is surely wrong to use specialised theory alone to impose specific restriction on the research design (e.g. by assuming away potentially confounding sources of variation) unless you have reasonable confidence that the theory is correct and that it is demonstrably so to a general audience. Unfortunately, this is the role played by theory in most applied spatial econometrician research. Theory is used to justify the inclusion of a spatial lag, assumptions are made about the form of the spatial weight matrix (possibly derived from theory), ML is used to achieve 'identification' and then model 'fit' is used as a basis for testing theory which justified the inclusion of the spatial lag. It should be clear by now that, for most spatial problems, we simply do not find this a convincing approach. Without wishing to weigh further into the vigorous debate on structural versus experimental approaches to empirical work (e.g. Journal of Economic Perspectives, Vol 24 (2) 2010) we simply make the point that whatever method is adopted (structural, experimental, qualitative or any other) any empirical research that aims to find out if x causes y needs to find a source of exogenous variation in x.

5. CONCLUSIONS

We have argued that identification problems bedevil most applied spatial econometric research. Most spatial econometric theorists are aware of these problems but downplay their importance by deriving estimators assuming that functional forms are known and by using model comparison techniques to choose between competing specifications. While this raises interesting theoretical and computation issues that have been the subject of a growing literature it does *not* provide a toolbox that gives satisfactory solutions to these problems for applied researchers interested in causality and the economic processes at work. It is in this sense that we think of much applied research using standard spatial econometric techniques as "pointless".

Paradoxically, we think that using the standard spatial econometric specifications (adapted, as we have done throughout the text, to reinforce the focus on the causal factors that drive outcome y_i) helps clarify identification problems for those researchers who are interested in causality. In particular, we think that closer attention to model specification will be helpful in understanding the exclusion and relevance assumptions that underlie IV approaches. Spatial econometrics also provides important insights on the correct interpretation of model parameters that we may identify from the IV or some other suitable estimation strategy. In short, there are lessons to be learnt from the spatial econometrics literature but for most applied researchers the appropriate strategy should be based on the experimental paradigm which puts issues of identification and causality at centre stage.

REFERENCES

Angrist, J. and S. Pishke. 2009. *Mostly Harmless Econometrics.* Princeton, NJ: Princeton University Press.

Brock, W. A. and S. N. Durlauf. 2001. "Interactions-based models," in: J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics*, edition 1, volume 5, chapter 54, pages 3297-3380 Elsevier.

Busso, M., J. Gregory and P. M. Kline (2010) "Assessing the Incidence and Efficiency of a Prominent Place," National Bureau of Economic Research Working Paper 16096

Dall'erba, S. and J. Le Gallo. 2008. "Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis," *Papers in Regional Science*, 87(2), 119-244.

Duranton, G., L. Gobillon and H. G. Overman. 2009. "Assessing the effects of local taxation using microgeographic data" Processed LSE.

Fingleton, B. 2008. "Testing the NEG Model: Further Evidence from Panel Data," *Spatial Economics Research Centre Discussion Paper* no. 0005.

Gibbons, S., S. Machin and O. Silva (2008) "Choice Competition and Pupil Achievement", *Journal of the European Economic Association*, 6(4) 912-947

Gibbons, S., S. Machin and O. Silva (2009) "Valuing School Quality Using Boundary Discontinuities", London School of Economics Spatial Economics Research Centre SERCDP0018Gibbons, S. and S. Telhaj (2008) "Peers and Achievement in England's Secondary Schools", London School of Economics Spatial Economics Research Centre SERCDP0001

De Giorgio, G., M Pellizzari and S. Redaelli (2010) "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics* 2 (2) 241–75

Greenstone, M, R. Hornbeck, and E. Moretti (2010) " Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants", *Journal of Political Economy*, 18(3) 536-598

Holmes, T. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, 106, 667-705.

Kolko, J. and D. Neumark (2010) "Do enterprise zones create jobs? Evidence from California's enterprise zone program," *Journal of Urban Economics* 68 (1) 1-19

Lee,L. 2004. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models," *Econometrica*, 72, 1899-1926.

LeSage, J. and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. New York: CRC Press.

Luechinger, S. (2009) "Valuing Air Quality Using the Life Satisfaction Approach," *The Economic Journal*, 119 (536) 482-515

Manski, C. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531-542.

Manski, C. 2000. "Economic Analysis of Social Interactions," *The Journal of Economic Perspectives*, 14 (3), 115-136

McMillen, D.P. (2010) "Issues in Spatial Data Analysis," *Journal of Regional Science*, 50(1), 119-141

Michaels, G. (2008) "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90 (4) 683-701

Ord, J.K. 1975. "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120-126.

Pinske, J. And M. E. Slade. 2010. "The Future of Spatial Econometrics," *Journal of Regional Science*, 50(1), 103-117.

Redding, S.J. and D. Sturm. (2008) " The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98 (5) 1766–97

Sacerdote, B. (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116 (2) 681-704

Staiger and Stock. 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557-586.

Todd, P. and K. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, 113, F3-F33.