Simultaneous-equations Analysis in Regional Science and Economic Geography

Timo Mitze
Andreas Stephan

Simultaneous-equations Analysis in Regional Science and Economic Geography

Timo Mitze and Andreas Stephan[1]

January 2013

Abstract

This paper provides an overview over simultaneous equation models (SEM) in the context of analyses based on regional data. We describe various modelling approaches and highlight close link of SEMs to theory and also comment on the advantages and disadvantages of SEMs.We present selected empirical works using simultaneous-equations analysis in regional science and economic geography in order to show the wide scope for applications. We thereby classify the empirical contributions as either being structural model presentations or vector autoregressive (VAR) models. Finally, we provide the reader with some details on how the various models can be estimated with available software packages such as STATA, LIMDEP or Gauss.

Keywords: Structural Equation Models, Regional Science and Economics, Empirical Applications, Software

JEL code: C33, C87

[1] Address for correspondence: Timo Mitze, Department for Border Region Studies, University of Southern Denmark, Alsion 2, DK-6400 Sønderborg. Phone.: +45/6550/1770, Email: tmitze@sam.sdu.dk. Andreas Stephan, Jönköping International Business School, PO Box 1026, 55111 Jönköping, Sweden, Phone: +46/36/101760, Email: andreas.stephan@jibs.hj.se. We are indebted to two anonymous referees for valuable comments. Moreover, we would like to thank Cheryl Adams, William Greene, Inessa Love and Gustavo Sanchez for helpful advice on the software section in this paper.

1. Introduction

The specification and estimation of simultaneous-equations models (henceforth *SEMs*) has a long tradition in economics. Although being originally established in the field of macroeconomics, meanwhile various applications can also be found in applied regional science and economic geography. Among many others, one prominent example is the Carlino and Mills (1987) study on the simultaneous evolution of regional population and employment densities, which gave rise to the famous "chicken-or-egg" quest for causality within the framework of regional adjustment models. And indeed, specifying and estimating a SEM has pretty much to do with *getting causal relationships right*. This is why applied economists and econometricians generally valorize the SEM approach for its capacity to formulate an explicit structural model with more than just one endogenous variable and the statistical power to control for correlated residuals among the individual equations of the system. While the first argument is of crucial importance for the *consistency* of the estimated model parameters, the second point is mainly concerned with the notion of estimation *efficiency*.

Nonetheless, these advantages do not come without costs. Among the main disadvantages for applied work with SEMs is the fact that estimation involves a much higher degree of complexity compared to single equation approaches. Moreover, in finite samples the potential bias stemming from an erroneously estimated variance-covariance matrix of the SEM can be larger compared to its potential gain in estimation efficiency. Given that the merits and pitfalls of SEM estimation are not always straightforward to see, this paper aims at providing some guidance for applied researchers in terms of model selection and application. Since recent empirical contributions in the field of regional science and economic geography are increasingly dominated by the use of panel data, in the following we will focus on the specification and estimation of simultaneous-equations models for this type of data.

Of course, this overview cannot provide a full-fletch methodological depiction for the estimation of SEMs with panel data. Here, the technically interested reader is, for instance, referred to Krishnakumar (1996) as well as Baltagi (2008). Additionally, there is also a growing demand from applied researchers to explicitly incorporate spatial structures into the SEM framework. Although a technical presentation of spatial SEMs would clearly exceed the scope of this paper, we present some empirical applications that deal with this issue later on. [2]

The remainder of this paper is organized as follows: The next section gives a brief overview of key conceptual issues related to model specification and estimation. Particularly, we show how an equation system can be constructed based on a set of individual equations and we review different methods to estimate this system empirically. Section 3 then summarizes two strands of empirical applications of SEMs, namely the estimation of *structural models* in Section 3.2. as well as *vector autoregressive* approaches in Section 3.3. Both sections also aim at linking standard SEM estimation with the fast growing spatial econ-

---

[2] According to the knowledge of the authors, the only explicit treatment of specifying and estimating spatial simultaneous equation models is given by Gebremariam (2007).

ometrics approach. Section 4 then discusses some software tools that are currently available to applied researchers and Section 5 finally provides a brief summary and outlook of future challenges in this research field.

2. Model Setup

2.1 Specification

The starting point for the specification of a SEM is a situation that is characterized by the mutual interdependence of economic variables. In many cases, the empirical model can be directly derived from economic theory. Consider, for instance, the joint specification of a firm's factor demand system to choose optimal levels of labor and capital inputs, the interrelatedness between population and regional house price dynamics or, more generally, the simultaneous prediction of supply and demand factors in one or several markets or regions. In all of these settings, the estimation of a SEM is a promising strategy to gain insights with regard to the equilibrium relationships among these variables as well as their temporal dynamics. Focusing on the long-run or equilibrium relationship typically involves the specification of a *static* SEM, while the integration of a short-run perspective calls for an explicit *dynamic* simultaneous equation model (DSEM). Since the latter approach can be seen as a generalization of a static case, we will take a closer look at this universal DSEM specification in the following.

Let's consider the *structural form* of a system of M dynamic equations, where its m-th equation can be written as

$$y_{i,t,m} = \alpha + \sum_{j=0}^{l} \beta'_j Y_{i,t-j,m} + \sum_{j=0}^{k} \gamma'_j X_{i,t-j,m} + u_{i,t,m}, \quad \text{with} \quad u_{i,t,m} = \mu_{i,m} + v_{i,t,m}, \quad (1)$$

for $i = 1, \ldots, N$ (cross-sectional dimension), $t = 1, \ldots, T$ (time dimension) and $m = 1, \ldots, M$ (number of equations). In the context of regional science and economic geography, the cross-sectional dimension typically comprises a set of regions or an associated measure of interregional interactions (such as migration, commuting or knowledge flows), which are observed for a fixed period of time. $y_{i,t,m}$ is the endogenous variable and $Y_{i,t,m}, \ldots, Y_{i,t-j,m}$ denote current and lagged endogenous explanatory variables of the system including the lagged endogenous variable of the m-th equation itself. Analogously, X is a $1 \times K$ vector of all further (unmodelled) K explanatory regressors, $u_{i,t,m}$ is the combined error term, which is composed of the two error components, $\mu_{i,m}$ as the unobservable individual effects and $v_{i,m}$ is the remainder error term. Both $\mu_{i,m}$ and $v_{i,t,m}$ are assumed to be i.i.d. residuals with standard normality assumptions as

$$\begin{aligned}
E(v_{i,t,m}v_{j,s,m}) &= 0, && \text{(for either } i \neq j \text{ or } t \neq s, \text{ or both)}, \\
E(\mu_{i,m}\mu_{j,m}) &= 0, && \text{(for } i \neq j), \\
E(\mu_{i,m}v_{j,t,m}) &= 0, && (\forall \quad i, j, t),
\end{aligned} \qquad (2)$$

where j and s have the same dimension as i and t, respectively. The first two assumptions state that the homoscedastic error terms are mutually uncorrelated over time and across cross-sections. Furthermore the unobserved individual heterogeneity is random and uncorrelated between individuals. The third assumption rules out any correlation between the individual effects and the remainder of the disturbance term. One has to note, that these assumptions hold for the error components of the m-th equation of the system, while one may still allow for cross error correlations between different equations of the system. Stacking the observations for each endogenous $y_{i,t}$ and exogenous variable $x_{i,t}$ as well as the error term $u_{i,t}$ according to

$$y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{iT} \\ \vdots \\ y_{NT} \end{pmatrix}, \quad x = \begin{pmatrix} x_{11} \\ \vdots \\ x_{iT} \\ \vdots \\ x_{NT} \end{pmatrix}, \quad u = \begin{pmatrix} u_{11} \\ \vdots \\ u_{iT} \\ \vdots \\ u_{NT} \end{pmatrix}, \qquad (3)$$

allows us to simplify the notation of eq.(1) in the following way:

$$\begin{aligned}
y_m &= R_m \xi_m + u_m \\
u_m &= \mu_m + v_m,
\end{aligned} \qquad (4)$$

where $R_m = (Y_m, X_m)$ and $\xi = (\beta', \gamma')$. Further stacking the equations into the form usually considered in a system – or multiple equation – analysis yields[3]

$$y = R\xi + u, \qquad (5)$$

where $y' = (y_1', \ldots, y_M')$ and similar for ξ and u. R is defined as

$$R = \begin{bmatrix} R_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_M \end{bmatrix}. \qquad (6)$$

---

[3] If Y_m is excluded from R_m in each equation, then the system reduces to a seemingly unrelated regressions (SUR) framework as proposed by Zellner (1962).

As typically done in the single equation model with panel data (see Baltagi, 2008, for details), we assume that both μ and ν are standard normal errors with the zero mean and covariance matrices for the error components as $\Sigma_\mu = [\sigma^2_{\mu_{(m,l)}}]$ (with $m = 1, \ldots, M$ and $l = 1, \ldots, M$) for the unobserved individual effects, and $\Sigma_\nu = [\sigma^2_{\nu_{(m,l)}}]$ for the remainder error term, respectively. These models are then typically called one-way error component models and treating μ as a random variable results in a random effects model (REM). Alternatively, if the unobservable individual effects are treated as a vector of fixed effects, the model is labeled fixed effects model (FEM). The difference between the REM and FEM estimation can be best described in terms of the assumed exogeneity of X with respect to μ in each equation as

$$
\begin{aligned}
E(X_{i,t,m}\mu_{j,m}) &= 0, & \text{REM} \quad (\forall \quad i,j,t), \\
E(X_{i,t,m}\mu_{j,m}) &= \rho_{i,j,m}, & \text{FEM} \quad (\forall \quad i,j,t),
\end{aligned}
\tag{7}
$$

where $\rho_{i,j,m}$ is a measure of correlation with $\rho_{i,j,m} \neq 0$. Thus, while the REM assumes that the vector of X variables is uncorrelated with the individual effects, the FEM specification allows these variables to be potentially correlated. For the case that one of these assumptions is true with respect to the empirical data at hand, different estimation techniques will be needed in order to get consistent and efficient regression results. As will be shown in the next section, the estimation procedure becomes more complicated in a SEM given the inclusion of explanatory endogenous variables Y.

2.2   Estimation

As Krishnakumar (1996) points out, a straightforward estimation of the coefficients for each structural equation of a SEM by ordinary (OLS) or generalized least squares (GLS) will lead to inconsistent estimates since the explanatory endogenous variables of the equation are correlated with both error terms. The reason for this is that the vector of endogenous (contemporaneous and lagged) explanatory variables $\sum_{j=0}^{l} Y_{i,t-j,m}$ in the right hand side of eq.(1) is correlated with both the unobservable individual effects $\mu_{i,m}$ and the remainder error component $\nu_{i,t,m}$. To see why this is the case, one has to bear in mind that the structural model has statistical noise and an individual effects in every structural equation. Computing its *reduced form* for each equation will yield a representation of the individual effects which is a linear combination of the individual effects in the other structural equations. As Cornwell et al. (1992) note, the solution for every endogenous variable will therefore in general involve every structural error and also the individual effect from every equation. The nature of this solution implies that every endogenous variable should be correlated with every structural error and hence also with the individual effects.

In such cases, instrumental variables (IV) are an appropriate solution for estimation purposes. In the static case, basically all contemporaneous and lagged values of the exogenous explanatory variables (X) are used as instruments for the set of endogenous variables. The researcher then has to decide whether to estimate the system in an equation-by-equation (two-stage least squares, 2SLS) or simultaneous manner (three-stage least squares, 3SLS). Since the former method does not employ all information available to the researcher it is also called *limited information* estimation, while the latter approach is called *full information* estimation. Theoretical and empirical contributions to static SEM estimation with panel data and full information include, for instance, Baltagi (1980, 1981 and 2008), Baltagi & Chang (1994, 2000), Prucha (1984), Hsiao (1986), Balestra & Krishnakumar (1987), Krishnakumar (1988), Bjorn & Krishnakumar (2007), Cornwell et al. (1992) as well as Park (2005).

For the case of fixed individual effects, Cornwell et al. (1992) have shown that essentially the same results for the simultaneous-equations model compared to the single equation case, namely that the model can be estimated as outlined above (2SLS or 3SLS) after a *within* transformation. Again note, the use of instrumental variables (rather than OLS) is necessary for the case that we have to deal with endogeneity of right-hand side regressors. Among the standard estimators in a REM world are Baltagi's (1980 and 1981) EC-SUR and EC-3SLS for the system case.[4] More general approaches in this field – which comprise the EC-SUR and EC-3SLS as special cases – are system extensions of the Hausman-Taylor (HT, 1981) type single equation estimators: HT-SUR and HT-3SLS (a convenient overview is given in Cornwell et al., 1992, Ahn & Schmidt, 1999, as well as Park, 2005 among others).

In the case of dynamic panel data estimators, the instrumentation problem is even more complex, since appropriate instruments for the lagged regressors of the set of endogenous variables have to be found as well. The reason is that, since the endogenous variables $y_{i,t,m}$ in each equation is a function of $\mu_{i,m}$, also its own lag $y_{i,t-1,m}$ is a function of $\mu_{i,m}$ and thus $y_{i,t-1,m}$ as right-hand side regressor is correlated with the error term. Given this additional estimation problem, the application of DSEM greatly benefits from recent advances in the (single equation) estimation of dynamic panel data models (see, e.g., Arellano, 2003). Starting from the IV-based estimation procedure proposed by Anderson & Hsiao (1981), Holtz-Eakin et al. (1988), Arellano & Bond (1991) as well as Blundell & Bond (1998) have proposed estimation methods based on the generalized method of moments (GMM), which allow the researcher to employ a richer set of instruments compared to the standard IV approach. Dealing with the general example of GMM estimation, which comprises standard 3SLS as a special case, the main difference between the single- and multiple-equations GMM rests on the specification of the weighting matrix for two-step efficient GMM estimation. This can be seen from the definition of the system GMM (henceforth SGMM) estimator $\hat{\Phi}_{SGMM}$ for the *M*-equation system as (see e.g. Hayashi, 2000, for details):

$$\hat{\Phi}_{SGMM} = \left(S'_{ZX} (V^S)^{-1} S_{ZX} \right)^{-1} S'_{ZX} (V^S)^{-1} S_{Zy} \tag{8}$$

---

[4] Where EC denotes *Error Components* and refers to the two components of the error terms in a panel data model as

with

$$S_{ZX} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} Z'_{i1}x_{i1} & \\ & \ddots & \frac{1}{N}\sum_{i=1}^{N} Z'_{iM}x_{iM} \end{bmatrix}$$

and

$$S_{Zy} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} Z'_{i1}y_{i1} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} Z'_{iM}y_{iM} \end{bmatrix}.$$

The above equations are basically the SGMM operationalization of the stylized system presentation given in eq.(5) and eq.(6). In empirical terms, the two-step efficient weighting matrix V^S has the following form

$$\hat{V}^S = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{i1}^2 Z_{i1}Z'_{i1} & \cdots & \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{i1}\hat{u}_{iM} Z_{i1}Z'_{iM} \\ \vdots & & \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{iM}\hat{u}_{i1} Z_{iM}Z'_{i1} & \cdots & \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{iM}^2 Z_{iM}Z'_{iM} \end{bmatrix}, \tag{9}$$

where the individual equations' $\hat{u}_{i,m}$ are based on consistent IV-based first stage estimates.[5] Thus, while single equation or equation-by-equation estimation assumes a block diagonal weighting matrix as

$$\hat{V} = diag(\sum_{i=1}^{N} \hat{u}_{i1}^2 Z_{i1}Z'_{i1}, \ldots, \sum_{i=1}^{N} \hat{u}_{iM}^2 Z_{iM}Z'_{iM}), \tag{10}$$

the SGMM weighting matrix $(\hat{V}^S)$ in eq.(9) fully exploits cross error correlations in the residuals. Giving that certain assumptions (conditional homoscedasticity and identical instruments across equations) hold, the SGMM approach reduces to the more familiar 3SLS notation (for details see Arellano, 2003). As Hayashi (2000) shows, joint estimation is asymptotically more efficient as long as at least one equation of the system is overidentified and the error terms are related to each other. However, the asymptotic results only hold if the model is correctly specified, that is, all the model assumptions are satisfied. Moreover, the asymptotic results may not be true for small samples (see Hayashi, 2000).

---

outlined in eq.(1).

[5] In comparison to this, one-step estimation replace the first step residuals by an identity or related transformation matrix.

In a recent Monte Carlo simulation exercise, Mitze (2011) has analyzed the finite sample properties of different full and limited information estimators. The simulation study especially focusses on data settings relevant for empirical applications in regional science and economic geography which often face small N and T settings. The results that in such small sample settings simple estimators are also among the best: The standard fixed effects (FEM) estimator with valid exogenous instruments either estimated by 2SLS manner or 3SLS ranks best in terms of estimation bias. This result particularly holds for data settings, where the unobserved fixed effects make up a dominant part of the overall error term. For constellations with a high persistence in the autoregressive parameter of the endogenous variables as well as a small time dimension, e.g. $T = 5$, the simultaneous-equations estimator based on the SGMM orthogonality conditions for the level equation according to the Blundell & Bond (1998) estimator (LEV-SGMM) performs best. This estimator generally also ranks best in terms of efficiency (measured by means of the root mean square error). While the latter two estimators may thus be seen as a good choice for empirical applications, when right hand side endogeneity and simultaneity matter, SGMM techniques for orthogonality conditions based on the regression equations in first differences, which are still a common tool in dynamic panel data settings, perform generally rather weak.

Generally, the superiority of system estimator relative to equation-by-equation estimation is expected to be more prevailing, the longer is the time horizon *T* of the respective dataset. Same results are reported in Park (2005) indicating that for finite samples – although computationally more demanding – system estimator may not necessarily be more efficient than single-equation alternatives.

3.   Empirical Applications

3.1   A Rough Classification

In this section, we present selected empirical works using simultaneous-equations analysis in regional science and economic geography in order to show the wide scope for applications. We thereby classify the empirical contributions as either being structural model presentations or vector autoregressive (VAR) models. The main difference between these two model presentations lies in ex-ante classification of variables as being endogenous or exogenous in the system. A structural DSEM typically starts from a predefined distinction of endogenous and exogenous (or predetermined) variables and then uses the latter to set up an IV-estimator for the former. This could be either done in an equation-by-equation fashion (2SLS) or in a simultaneous setting (3SLS). Based on the estimation results, the long- and short-run dynamics of the model, e.g. with respect to policy changes, is then analysed by means of multiplier analysis (Bardsen et al., 2005).

However, as Rickman (2010) points out, this approach of structural modelling has recently been criticized for various reasons. One critical argument is the rather ad-hoc classification of endogenous and exogenous variables used in the IV-setup to instrument the contemporaneous endogenous explanatory

variables in the respective equations. An alternative path to the structural approach is thus to start from an unrestricted VAR perspective, where each variable is treated as endogenous. The VAR then specifies each variable of the *M*-equation system as a function of own lagged values as well as time lags from the other variables of the system. A further advantage of the VAR methodology is that the dynamic properties of the system can be analyzed with the help of impulse-response functions. The latter approach may be seen as advancement compared to the standard multiplier analysis in structural DSEM modeling (see, e.g., Stein & Song, 2002, for an overview). On the contrary, unrestricted (or reduced form) VAR models are typically limited in terms of their economic interpretation. This has led researchers to work with structural VARs, which impose ex-ante coefficient restrictions derived from theoretical assumptions. This, however, brings back in the necessity to impose theoretical restrictions ex-ante as in standard structural DSEM approach.

3.2   Structural Models

The roots of structural analysis in regional science and economic geography lie in complicated nature of causes and consequences of many regional economic phenomena, which ideally call for a theory-grounded fully specified model (see Holmes, 2010). Having a long history in dealing with these concepts, modern macroeconomic theory and macroeconometric practice has therefore served as a good source of inspiration for regional scientists. As Rickman (2010) points out in his contribution to the 50 year anniversary volume of the *Journal of Regional Science*, one way to go ahead in the field of regional science is to use structural macroeconomic approaches for the conduct of regional and policy analyzes as an alternative to the merely *descriptive* modeling tradition in the field.[6]

Although the initial example of the frequently cited Carlino & Mills (1987) study has already shown, regional scientists and economic geographers are indeed aware of this issue. However, striving through the recent empirical literature nevertheless shows that the number of empirical applications of structural DSEM is still rare and mostly bound to the continuous work of individual authors and their thematic adopters. Without attempting to give an all-encompassing account of applications in the field, let's give an example for this observed "clustering" phenomenon: In a series of papers, Driffield and coauthors (see Driffield & Girma, 2003, Driffield & Taylor, 2006, Driffield & De Propris, 2006) use a multiple-equation extension to the standard Anderson & Hsiao (1981) estimator to analyze the impact of regional FDI on wage spillovers. For instance, estimating a simultaneous-equations model for skilled and unskilled labor based on plant level data for the UK, Driffield & Girma (2003) argue that "(t)he estimates highlight the importance of employing a simultaneous equation estimator, as there are sizable differences in the cross-wage coefficients between the two estimators, these differences being highly significant in the case of unskilled workers. Once one allows for simultaneity, the impact of wages in other occupational groups becomes significantly greater than has previously been reported, where single equation studies often fail to

---

[6] With "descriptive" being defined in line with Holmes (2010) as a type of explorative empirical analysis, able to identify correlations of variables but not causal effects. The latter would need either a structuralist- or experimentalist-model approach.

find this impact, particularly in terms of the effect of skilled wages on unskilled workers".[7] Driffield & De Propris (2006) likewise estimate a SEM to quantify the importance of regional industrial clusters on inward FDI activity.

But not only the work of certain groups of researchers (where Driffield and coauthors are surely just one example out of many), also certain research topics seem to be prevailingly attracted by the simultaneous-equation approach. The chicken-or-egg debate on "Do jobs follow people?" or "Do people follow jobs?" is an excellent example with respect to this phenomenon. Not only that most of the classical work is done based on a SEM framework (see e.g. Okun, 1968, Muth, 1971, Salvatore, 1980, Bilger et al., 1991, Carlino & Mills, 1987), starting with Boarnet (1994), increasing attention has been given to the inclusion of spatial effects in a SEM framework. In a follow-up study to Boarnet, Henry et al. (2001) compare several spatial econometric approaches to the estimation of multiple-equation models of small regional development. Using data for French rural communities, the authors can establish that the inclusion of spatial autoregressive terms can clarify the interpretation of such regional adjustment models.

In a similar manner, Jeatney et al. (2010) estimate a spatial simultaneous-equations model of population migration and housing price dynamics. The authors show that both the structural relationship between population migration and housing prices as well as the spatial interdependence inherent in the data has to be properly addressed within a unifying model framework. After controlling for simultaneity and spatial autocorrelation, "the results show that neighborhoods are likely to experience an increase in their housing values if they gain population and they are more likely to lose population if they experience an increase in housing values".[8] Gebremariam et al. (2011) apply a spatially extended structural model for the joint estimation of employment, income and migration patterns in the Appalachia region in the U.S. Finally, using survey data for British students with the aid of a Geographical Information System (GIS), Faggian & McCann (2004) are able to analyze the spatial pattern of student migration within a simultaneous-equations model of human capital knowledge flows and regional knowledge assets.

Another application of structural SEM using German city level data is Kemmerling & Stephan (2002). They estimate, similarly as Cadot et al. (2006) do for French regions, a SEM which describes the production function of manufacturing in the city and also models the determinants of local infrastructure investments, where the simultaneity between equations exists as local infrastructure investments also depend on a cities' manufacturing productivity but local infrastructure investments on the other hand spur local manufacturing sector output.[9] The endogeneity of private sector production and public infrastructure investments at the regional and interregional level is also a matter of research in vector autoregressive time series models to which we turn next.

---

[7] See Driffield & Girma (2003), pp. 464-465.
[8] See Jeatny et al. (2010), p. 343.
[9] Focussing on fiscal policy interactions, Elhorst & Allers (2001) apply a linear expenditure system to Dutch municipality data in order to simultaneously estimate interactions in both taxation and different spending categories. They find that that a partial analysis of fiscal policy interactions risks to underestimate the degree of interaction, while the simultaneous equation approach yields the best empirical results.

3.3 Vector Autoregressive Models

Following the influential work of Sims (1980), the use of vector autoregressive (VAR) models has become a widespread empirical tool complementary to dynamic single equation specifications. The VAR approach starts from the general treatment of variables as endogenous for the specified system of interdependent equations and grounds specification issues such as (weak) exogeneity of variables and the direction of causality on empirical testing. Holtz-Eakin et al. (1988) adapted its estimation and analysis to the case of panel data. While the use of VAR models in time series analysis is a common standard, in a panel data context this is less so. This may explain while only recently, VAR models have reached the focus of regional modeling with first applications by Carlino & DeFina (1998, 2002). In these applications the authors examine whether monetary policy had symmetric effects across US states. Based on impulse-response functions the authors were then able to reveal that a set of core regions behaves quite similar to the US average, while certain other regions either reacted more or less sensitive to monetary policy shocks.

A very popular empirical application of the (panel) VAR model is the analysis of interregional migration pattern and local labor market trends (see, e.g., Blanchard & Katz, 1992, Decressin & Fatas, 1995, Möller, 1995, Lu, 2001, Mäki-Arvela, 2003, Partridge & Rickman, 2006, Alecke et al., 2010). In their seminal paper, Blanchard & Katz (1992) identify innovations in employment with shocks to labor demand. The authors find for US regional data that migratory responses are the dominant adaption mechanism to shocks in employment. The type of analysis proposed by Blanchard & Katz (1992) has been adapted to different data settings in the following. While Decressin & Fatas (1995) use European wide aggregate data, several authors also use individual data for European countries (e.g., Mäki-Arvela, 2003, for Finland, Möller, 1995, and Alecke et al., 2010, for Germany). Basically, these models also establish a link from interregional labor market disparities between regional migration responses.

As for the case of structural DSEM, estimation of Panel VAR models in a regional context is likely to call for an inclusion of spatial effects. Here, different solutions have been proposed in the literature. A straightforward way to do so is chosen by Jiwattanakulpaisarn et al. (2009), who apply spatial filtering techniques to Granger causality tests in a Panel VAR framework in order to analyze the direction of causality between highway infrastructure and aggregate employment for US states.[10] The advantage of the spatially filtered VAR approach is that – after decomposing each variable into a structural and purely spatial component – standard (non-spatial) estimators can be applied to the structural part of the model. However, there are also contributions to the literature that explicitly specify a Spatial VAR (SpVAR) in order to analyse the nature and direction of spatial dependence in the model.

The authors define the SpVAR as a vector autoregressive model which includes spatial as well as temporal lags among the stationary state variables and the model is estimated by means of dynamic spatial

---

[10] In a related application, Marquez et al. (2010) use a structural VAR approach to estimate interregional spillover effects from public capital.

panel data econometrics.[11] In an extension to this work, Beenstock & Felsenstein (2010) also discuss ways to estimate spatial vector error correction models (SpVECM), which allow extending the research focus to the analysis of dynamic processes in which spatially co-integrated variables are related in the long run. Finally, Di Giacinto (2010) has recently proposed the computation of space-time impulse response functions (STIRs) in order capture the complex spatial and temporal feedback effects allowed for by the SpVAR approach. STIRs are thus a convenient way to evaluate the extent of spatial spillover effects of structural shocks among variables. Di Giacinto (2010) also proposes a bootstrapping algorithm to compute confidence intervals.

Another application of the panel VAR approach is found in Di Giacinto et al. (2012). The purpose of their paper is to estimate coordinated and uncoordinated effects from regional infrastructure investments. In the first stage, the authors implement a common factor model recently put forward by Bai and Ng (2004), within their PANIC (Panel Analysis of Non-stationarity in Idiosyncratic and Common components) approach to unit roots and cointegration on panel data with a large cross-section dimension. By doing so they get a decomposition of individual regional time series of public capital in transport infrastructures into common and idiosyncratic components. In the second stage, the common components resulting from the PANIC procedure are used to estimate a set of VEC models for each Italian region, in order to separately identify the responses of main macro variables to unexpected shocks to common and idiosyncratic public capital. By comparing the different responses of GDP to the two types of shock, they draw evidence on the existence and the magnitude of network externalities.

4.   Software

As the brief review of recent applications in the literature has shown, the estimation of a SEM with panel data (either in a structural or time-series fashion) is a steadily growing subfield of applied economics and econometrics. Given the novelty of most of the approaches presented here, commonly used statistical software packages (e.g., SAS, Stata, LIMDEP, Gauss, MatLab, RATS, EViews) have not fully incorporated the plethora of new tools into their standard portfolio. In most cases, these packages cover a large variety of single-equation estimators for panel data models as well their limited information IV/2SLS regression, where each equation is estimated separately. However, 3SLS models are largely restricted to the *pooled* case (that is, ignoring unobserved individual effects).

To circumvent this problem and estimate a SEM by full information methods, the applied researcher may proceed as follows: Cornwell et al. (1992) propose to include individual effects or use *within*-type transformed variables before estimating the model in a simultaneous equation manner, e.g., by means of FEM-3SLS. The fixed effects transformation can be easily done by hand and most software packages offer three-stage least square estimators for systems of simultaneous equations. However, the reader has to note

---

[11] Alternatively, Mutl (2009) considers panel vector-autoregressive model with cross-sectional dependence in the disturbances.

that this estimation procedure will only yield correct standard errors, if the former least square dummy variable (LSDV) specification of the FEM model is implemented – that is, a set of binary dummy variables for each cross-sectional unit is included in the model. If the number of cross-sectional units is very large and a *within*-type transformation has to be done in advance to SEM estimation, a degrees of freedom correction needs to be done for obtaining the estimated standard errors.

If the research aims at using a limited information approach in a REM specification, LIMDEP and Stata have implemented Baltagi's EC2SLS estimator (e.g., in Stata use the *xtivreg* command and choose the *ec2sls* option). Nguyen (2010) has written the *xtsur* ado-file for Stata to implement a SUR model with unbalanced panel data as proposed in Bjorn (2004). Based on her paper jointly written with L. Zicchino, Inessa Love provides a user written Stata ado-file *pvar* for the estimation of VAR models with panel data, which also includes the computation of impulse-response functions (see Love & Zicchino, 2006). EViews also provides some tools to estimate (panel) VARs.

Finally, if the researcher wants to estimate a static or dynamic SEM by GMM manually, the first step needed is to stack the data for the individual equations by each variable to one vector and then apply two-step efficient GMM estimation as outlined above. Besides manually pre-processing the variables in the above described way, readily available software GMM-routines for estimating static and dynamic specifications can be applied for this purpose. GAUSS, for instance, offers the constrained Maximum Likelihood application, a general purpose program for producing estimates and statistical inferences for a variety of models, including multiple equations, panel series models. In Stata, the *gmm* command fits single equation models, but it can also be used to fit systems of equations as long as the user can derive the appropriate moment conditions.

5   Summary and Outlook

This paper highlights the importance and wide applicability of structural equation models in the context of regional science and economic geography. The different methods presented here have mostly been developed in the general context of panel data analysis. However, as the panel unit can be easily defined as a region or another geographic entity, many of these approaches can be easily adopted for the regional context. Most recent development in this field comprise dynamic structural SEMs as well as VAR models. Besides presenting recent empirical applications for both type of models and their spatial extensions, we also provide a short introduction to software solutions for the estimation of such models. Though there exist not so many ready-to-use routines in statistical software packages, many of the standard routines for SEM estimation can be adjusted to capture the panel nature of the data in the context of regional science and geographic analyses.

References

Ahn, S.; Schmidt, P. (1999): "Estimation of linear panel data models using GMM, in: Matyas, L. (Ed.): "Generalized Method of Moments", Cambridge, pp. 211-274.

Alecke, B.; Mitze, T.; Untiedt, G. (2010): "Internal migration, regional labour market dynamics and implications for German East-West disparities: results from a Panel VAR", *Review of Regional Research*, Vol. 30, No. 2, pp. 159-189.

Anderson, T. W.; Hsiao, C. (1981): "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, Vol. 76, pp. 598-606.

Arellano, M. (2003): "Panel Data Econometrics", Oxford University Press.

Arellano, M.; Bond, S. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, Vol. 58, pp. 277-297.

Balestra, P.; Krishnakumar, J. (1987): Full Information Estimations of a System of Simultaneous Equations with Error Component Structure", in: *Econometric Theory*, Vol. 3, pp. 223-246.

Baltagi, B. (1980): "On seemingly unrelated regressions with error components", in: *Econometrica*, Vol. 48, pp. 1547-1551.

Baltagi, B. (1981): "Simultaneous Equations with error components", in: *Journal of Econometrics*, Vol. 17, pp. 189-200.

Baltagi, B. (2008): "Econometric Analysis of Panel Data", 4th ed., Chichester.

Baltagi, B.; Chang, Y. (2000): "Simultaneous Equations with incomplete Panels", in: *Econometric Theory*, Vol. 16, pp. 269-279.

Bardsen, G.; Eitrheim, O.; Jansen, E.; Nymoen, R. (2005): "The Econometrics of Macroeconomic Modelling", Oxford.

Bilger, U.; Genosko, J.; Hirte, G. (1991): "Migration and Regional Labour Market Adjustment in West Germany", in: Stillwell, J. Congdon, P. (Eds.): "Migration Models: Macro and Micro Approaches", London, pp. 152-167.

Biorn, E. (2004): "Regression system for unbalanced panel data: a stepwise maximum likelihood procedure", in: *Journal of Econometrics*, Vol. 122, pp. 281-91.

Bjorn, E.; Krishnakumar, J. (2007): "Measurement errors and simultaneity", in: Matyas, L; Sevestre, P. (Eds.): "The Econometrics of Panel Data. Fundamentals and Recent Developments", Berlin.

Blundell, R.; Bond, S. (1998): "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, Vol. 87, pp. 115-143.

Boarnet, M. G. (1994): "The Monocentric Model and Employment Location", in: *Journal of Urban Economics*, Vol. 36(1), pp. 79-97.

Cadot, O.; Roller, L.-H.; Stephan, A. (2006): "Contribution to productivity or pork barrel? The two faces of infrastructure investment," *Journal of Public Economics*, Vol. 90(6-7), pp. 1133-1153.

Carlino, J.; Mills, E. (1987): "The determinants of county growth", *Journal of Regional Science*, Vol. 27, pp.39-54.

Cornwell, C.; Schmidt, P.; Wyhowski, D. (1992): "Simultaneous equations and panel data", *Journal of Econometrics*, Vol.51, pp. 151-181.

Decressin, J.; Fatas, A. (1995): "Regional labour market dynamics in Europe", *European Economic Review*, Vol. 39, pp.1627-1665.

Di Giacinto, V.; Micucci, G.; Montanaro, P. (2012): "Network effects of public transport infrastructure: Evidence on Italian regions", *Papers in Regional Science*, Vol. 91, pp. 515-541.

Driffield, N.; Girma, S. (2003): "Regional Foreign Direct Investment and Wage Spillovers: Plant Level Evidence from the UK Electronics Industry", *Oxford Bulletin of Economics and Statistics*, 65, pp. 453-474.

Driffield, N.; Taylor, K. (2006): "Wage spillovers, inter-regional effects and the impact of inward investment", *Spatial Economic Analysis*, Vol. 1, Issue 2, pp. 187-205.

Driffield, N.; De Propris, L. (2006): "The Importance of Cluster for Spillover from FDI and Technology Sourcing", *Cambridge Journal of Economics*, Vol.30, pp. 277-291.

Allers, M.; Elhorst, P. (2011): "A Simultaneous Equations Model of Fiscal Policy Interactions", in: *Journal of Regional Science*, Vol. 51, No. 2, pp. 271-291.

Gebremariam, G. (2007): "Modeling and Estimation Issues in Spatial Simultaneous Equations Models", paper presented at the 54th Annual North American Meetings of the Regional Science Association International (RSAI), Savannah, Georgia, November 7-10, 2010 for download at (last access date 17.01.2013) http://www.rri.wvu.edu/pdffiles/gebrewp2007-13.pdf.

Gebremariam, G.; Gebremedhin, T.; Schaeffer, P. (2011): "Employment, Income and Migration in Appalachia: A Spatial Simultaneous Equations Approach", in: *Journal of Regional Science*, Vol. 51, No. 1, pp. 102-120.

Hayashi, F. (2000): "Econometrics", Princeton.

Holtz-Eakin, D.; Newey, W.; Rosen, H.S. (1988): "Estimating vector autoregressions with panel data", *Econometrica*, Vol. 56, pp. 1371-1395.

Hsiao, C. (1986): "Analysis of Panel Data", Cambridge.

Kemmerling, A., Stephan, A. (2002): "The Contribution of Local Public Infrastructure to Private Productivity and Its Political Economy: Evidence from a Panel of Large German Cities", *Public Choice*, Vol. 113(3-4), pp. 403-424.

Krishnakumar, J. (1988): "Estimation of Simultaneous Equation Models with Error Components Structure", Lecture Notes in Economics and Mathematical Systems, Berlin et al.

Krishnakumar, J. (1996): "Simultaneous Equations", in: Matyas, L.; Sevestre, P. (Eds.): "The Econometrics of Panel Data. A Handbook of the Theory with Applications", 2nd Ed., Kluwer Academic Publishers.

Love, I.; Zicchino, L. (2006): "Financial Development and Dynamic Investment Behavior: Evidence from Panel Vector Autoregression", *The Quarterly Review of Economics and Finance*, Vol. 46, pp.190-210.

Lu, M. (2001): "Vector autoregression (VAR)—an approach to dynamic analysis of geographic processes", in: *Geografiska Annaler: Series B, Human Geography*, Vol. 83(2), 67-78.

Mäki-Arvela, P. (2005): "Regional Evolutions in Finland: Panel Data Results of a VAR Approach to Labour Market Dynamics", *Regional Studies*, Vol. 37, pp.423-443.

Marquez, M.; Ramajo, J.; Hewings, G. (2009): "Domestic and Cross-Border Effects of Public Capital: A SVAR Approach for the Spanish Regions", Discussion paper REAL 09-T-2.

Mitze, T. (2011): "Dynamic Simultaneous Equations with Panel Data: Small Sample Properties and Application to Regional Econometric Modelling", Chapter 7 in: Mitze, T.: "Empirical Modelling in Regional Science. Towards a Global Time-Space-Structural Analysis", Berlin et al., pp. 219–272.

Möller, J. (1995): "Empirische Analyse der Regionalentwicklung", in: Gahlen, B.; Hesse, H.; Ramser, H. (Eds.): "Standort und Region. Neue Ansätze zur Regionalökonomik", Tübingen.

Muth, R. (1971): "Migration: Chicken or Egg", *Southern Economic Journal*, Vol. 37, pp.295-306.

Nguyen, M. (2010): "XTSUR: Stata module to estimate seemingly unrelated regression model on unbalanced panel data", Statistical Software Component No. S456953, Boston College Department of Economics.

Okun, B. (1968): "Interstate population Migration and State Income Inequality: A Simultaneous Equation Approach", *Economic Development and Cultural Change*, Vol. 16, pp.297-313.

Park, S. (2005): "Estimating a Linear Simultaneous Equation Model with Panel Data", Working Paper Carleton University.

Partridge, M.D.; Rickman, D.S. (2006): "An SVAR Model of Fluctuations in U.S. Migration Flows and State Labor Market Dynamics", *Southern Economic Journal*, April 2006.

Prucha, I. (1984): "On the Asymptotic Efficiency of Feasible Aitken Estimator for Seemingly Unrelated Regression Models with Error Components", in: *Econometrica*, Vol. 52, pp. 203-207.

Rickman, D. (2010): "Modern Macroeconomics and Regional Economic Modeling", *Journal of Regional Science*, Vol. 50, No. 1, pp. 5-22.

Salvatore, D. (1980): "A Simultaneous Equations Model of Internal Migration with Dynamic Policy Simulations and Forecasting", *Journal of Development Economics*, Vol. 7, pp.231-246.

Vermeulen, W. ; van Ommeren, J. (2009): "Does land use planning shape regional economies? A simultaneous analysis of housing supply, internal migration and local employment growth in the Netherlands", in: *Journal of Housing Economics*, Vol. 18(4), pp. 294-310.

Zellner, A. (1962): "An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias", in: *Journal of the American Statistical Association*, Vol. 57, pp. 348-368.