BLOCK RECURSIVE PATH MODELS

JAN DE LEEUW AND VIVIAN LEW

ABSTRACT. In this paper we discuss linear block recursive path models. They are introduced as techniques to linearly decompose variables organized in ordered blocks. It is shown how the algebra of path analysis leads to a calculus of direct and indirect effects, using only saturated models.

1. LINEAR BLOCK MODELS

Path analysis has been applied to health services research by, among others, Anderson and Aday [1]. In this paper we discuss the application of a particular class of path models to the ICS-II project on aging. The models we are interested in here are *linear block models.* These models have been discussed recently by Wermuth [4]. In the paper we review some of the main results on these models, and we suggest various extensions.

Suppose $\underline{x}_0, \cdots, \underline{x}_m$ are *blocks* of random variables, i.e. random vectors. We have special names for the variables in the first vector, and for those in the last vector. The variables in $\underline{x}_0$ are called *exogeneous* and the variables in $\underline{x}_m$ are the *outcome variables.* Moreover $\underline{\epsilon}_0, \cdots, \underline{\epsilon}_m$ is another sequence of (unobserved) vector variables, called the *disturbances* or the *innovations.*

All variables in this paper have expectations equal to zero, unless otherwise indicated. They also have finite variances. We assume the variables $\underline{x}_j$ and $\underline{\epsilon}_j$ are connected by a block-regression model of the form

$$\underline{x}_j = \sum_{\ell=0}^{j-1} B_{j\ell}\underline{x}_\ell + \underline{\epsilon}_j, \tag{1.1}$$

for $j = 1, \cdots, m$. For $j = 0$ the model simply says $\underline{x}_0 = \underline{\epsilon}_0$. We give a path diagram illustrating this model.

For interpretation purposes, we observe that the blocks $\underline{x}_j$ are ordered, for instance in time, or in some sort of "causal" order imposed by the researcher. Block regression models are often discussed in causal terms. This is somewhat misleading, because the causality involved in merely the order imposed by the researcher on the blocks. It has nothing to do, at least in principle, with the outcomes of manipulation, or the results of controlled experiments with some form of random assignment.

In general, $\underline{x}_j$ "depends" linearly on all "previous" blocks. The innovation $\underline{\epsilon}_j$ is the part of $\underline{x}_j$ not "determined" by the previous blocks.

To make the decomposition interesting (and *identified*, as we shall see below), we need some assumptions.

Assumption 1.1. The $\underline{\epsilon}_j$ are uncorrelated with all $\underline{x}_\ell$ for which $\ell < j$.

In fact, Assumption 1.1 implies a more interesting result, which is perhaps the most powerful property of these block models.

Theorem 1.2. *The $\underline{\epsilon}_j$ are uncorrelated with each other.*

Proof. Take $\underline{\epsilon}_j$ and $\underline{\epsilon}_\ell$. Suppose, without loss of generality, that $\ell < j$. According to Equation 1.1 $\underline{\epsilon}_\ell$ is a linear combination of $\underline{x}_0, \cdots, \underline{x}_\ell$, and we know that $\underline{\epsilon}_j$ is orthogonal to each of those. Thus $\underline{\epsilon}_j$ is orthogonal to $\underline{\epsilon}_\ell$. □

Assumption 1.1 also implies immediately, by the way, that the $B_{j\leftarrow\ell}$ can be found by regression of the variables in block j on the variables in all previous blocks. We formalize this in a theorem. We use the symbol Ω_j for the dispersion matrix of the $\underline{\epsilon}_j$. Observe that Ω_j is of the order of the number of variables in the block, and generally not diagonal (unless the block consists of a single variable). Thus

$$\mathbf{C}\left(\underline{\epsilon}_j, \underline{\epsilon}_\ell\right) = \delta^{j\ell}\Omega_j. \tag{1.2}$$

Theorem 1.3. *$B_{1\rightarrow j}, \cdots, B_{j-1\rightarrow j}$ can be found by projecting $\underline{x}_j$ on the space spanned by $\underline{x}_0, \cdots, \underline{x}_{j-1}$. Ω_j is the dispersion matrix of the anti-projections (the residuals).*

Proof. This is just Pythagoras in disguise, i.e. the (unique) orthogonal decomposition of a vector along a given subspace into a part in the subspace and a part orthogonal to the subspace. □

Remark 1.4. Theorem 1.3 gives us a constructive method to fit the unknown parameters. We also see that the Ω_j are *identified*, while the $B_{j\ell}$ are *identified* if the blocks are non-singular. Finally, we see that Model 1.1 is *saturated*, i.e. it can be fitted exactly to any set of variables which are partitioned into $m+1$ blocks.

2. Matrix Expression

The model can be rewritten in matrix notation by defining a number of matrices and vectors. This does not really add anything new, but it allows for compact formula, and for efficient computation in matrix-oriented environments. Here are the definitions.

$$\underline{x} \triangleq \begin{bmatrix} \underline{x}_0 \\ \underline{x}_1 \\ \cdots \\ \underline{x}_m \end{bmatrix}. \tag{2.1}$$

$$B \triangleq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ B_{10} & 0 & 0 & \cdots & 0 \\ B_{20} & B_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{m0} & B_{m1} & B_{m2} & \cdots & 0 \end{bmatrix}. \tag{2.2}$$

$$\underline{\epsilon} \triangleq \begin{bmatrix} \underline{\epsilon}_0 \\ \underline{\epsilon}_1 \\ \cdots \\ \underline{\epsilon}_m \end{bmatrix}. \tag{2.3}$$

$$\Omega \triangleq \begin{bmatrix} \Omega_0 & 0 & \cdots & 0 \\ 0 & \Omega_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_m \end{bmatrix}. \tag{2.4}$$

Thus Ω is the *direct sum* of the Ω_j. Ω is block-diagonal, but not necessarily diagonal. With these definitions, model 1.1 can be written simply as

$$\underline{x} = B\underline{x} + \underline{\epsilon}. \tag{2.5}$$

Definition 2.1. Define

$$A \triangleq (I - B)^{-1} \tag{2.6}$$

The *reduced form* of model 2.5 is

$$\underline{x} = (I - B)^{-1}\underline{\epsilon} = A\underline{\epsilon}. \tag{2.7}$$

Definition 2.1 means that we write the $\underline{x}_j$ as linear combinations of the successive innovations. Remember that the innovations are uncorrelated, i.e. we decompose $\underline{x}_j$ as linear combinations of uncorrelated (and in the normal case independent) blocks. Observed that A is still block triangular, with the identity matrix in the diagonal blocks.

The reduced form also implies a simple form for the dispersion matrix of the variables.

Theorem 2.2.

$$\mathbf{V}(\underline{x}) = A\Omega A'.$$

Proof. Trivial consequence of $\underline{x} = A\underline{\epsilon}$. □

As a further corollary, we have the following formula for the *concentration matrix*, i.e. the inverse of the dispersion matrix.

$$[\mathbf{V}(\underline{x})]^{-1} = (I - B')\Omega^{-1}(I - B). \tag{2.8}$$

3. Path Analysis

The key path analysis theorem for block recursive models is given first, in matrix notation.

Theorem 3.1. $A = I + B + B^2 + \cdots + B^m$.

Proof. By direct multiplication

$$(I - B)(I + B + B^2 + \cdots + B^m) = I - B^{m+1}.$$

Thus it suffices to show that $B^{m+1} = 0$, which is proved in the lemma below. □

Lemma 3.2. $B^{m+1} = 0$

Proof. Use induction over m. Obviously the lemma is true for $m = 0$. Suppose it is true for block-triangular matrices with m blocks. We have for a B with $m + 1$ blocks

$$B = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix},$$

where A has m blocks, and by the induction hypothesis satisfies $A^m = 0$. Direct multiplication shows

$$B^{m+1} = \begin{bmatrix} A^{m+1} & 0 \\ CA^m & 0 \end{bmatrix} = 0,$$

□

Because of this result, it is of some interest to look at the powers of B in more detail. We suppose $j > \ell$ and we look at submatrix (j, ℓ) of B^s. We find

$$[B^s]_{j\ell} = \sum\sum_{j>j_1>j_2<\cdots>j_s>\ell}\cdots\sum B_{jj_1}B_{j_1j_2}\cdots B_{j_s\ell} \tag{3.1}$$

Let us use a simple example with four blocks of variables. Thus

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ B_{10} & 0 & 0 & 0 \\ B_{20} & B_{21} & 0 & 0 \\ B_{30} & B_{31} & B_{32} & 0 \end{bmatrix}. \tag{3.2}$$

Moreover

$$B^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ B_{21}B_{10} & 0 & 0 & 0 \\ B_{31}B_{10} + B_{32}B_{20} & B_{32}B_{21} & 0 & 0 \end{bmatrix}, \tag{3.3}$$

and

$$B^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ B_{32}B_{21}B_{10} & 0 & 0 & 0 \end{bmatrix}. \tag{3.4}$$

Of course $B^4 = 0$. This means that

$$A_{10} = B_{10}, \tag{3.5}$$

$$A_{20} = B_{20} + B_{21}B_{10}, \tag{3.6}$$

$$A_{21} = B_{21}, \tag{3.7}$$

$$A_{30} = B_{30} + B_{31}B_{10} + B_{32}B_{20} + B_{32}B_{21}B_{10}, \tag{3.8}$$

$$A_{31} = B_{31} + B_{32}B_{21}, \tag{3.9}$$

$$A_{32} = B_{32}. \tag{3.10}$$

We can also look at the form of the dispersion matrix. Clearly

$$V_{j\ell} = \sum_{s=0}^{\min\{j,\ell\}} A_{js}\Omega_s A_{s\ell} \tag{3.11}$$

For our example this means

$$V_{00} = \Omega_0, \tag{3.12}$$

$$V_{10} = A_{10}\Omega_0, \tag{3.13}$$

$$V_{11} = A_{10}\Omega_0 A'_{10} + \Omega_1, \tag{3.14}$$

$$V_{20} = A_{20}\Omega_0, \tag{3.15}$$

$$V_{21} = A_{20}\Omega_0 A'_{10} + A_{21}\Omega_1, \tag{3.16}$$

$$V_{22} = A_{20}\Omega_0 A'_{20} + A_{21}\Omega_1 A'_{21} + \Omega_2, \tag{3.17}$$

$$V_{30} = A_{30}\Omega_0, \tag{3.18}$$

$$V_{31} = A_{30}\Omega_0 A'_{10} + A_{31}\Omega_1, \tag{3.19}$$

$$V_{32} = A_{30}\Omega_0 A'_{20} + A_{31}\Omega_1 A'_{21} + A_{32}\Omega_2, \tag{3.20}$$

$$V_{33} = A_{30}\Omega_0 A'_{30} + A_{31}\Omega_1 A'_{31} + A_{32}\Omega_2 A'_{32} + \Omega_3. \tag{3.21}$$

If we look at these expressions, we see a simple result arising.

Theorem 3.3. *For all* j

$$\Omega_j = V_{jj|j-1,\cdots,0}.$$

Moreover, if $j > \ell$ *then*

$$A_{j\ell} = V_{j\ell|\ell-1,\cdots,0} V^{-1}_{\ell\ell|\ell-1,\cdots,0}.$$

Proof. Again this is true for $j = 0$ by definition. □

Theorem 3.3 repeats what we already know, i.e that Ω_j is the residual variance after regressing $\underline{x}_j$ on $\underline{x}_0, \cdots, \underline{x}_{j-1}$. But it also gives a new interpretation of $A_{j\ell}$, because it turns out that these are the conditional regression coefficients of $\underline{x}_j$ on $\underline{x}_\ell$ if we control for $\underline{x}_0, \cdots, \underline{x}_{\ell-1}$.

4. Causal Analysis

In this section we discuss some definitions taken from, mainly, Sobel [3] and McDonald [2]. We relate our discussion to the discussion in Anderson and Aday [1]. There are a number of *effects* we can study in a particular path analysis. We first study these concepts in the contexts of blocks and variables, and we then reformulate them in terms of single variables.

In Anderson and Aday, the covariance between two variables is called the *total effect*. In terms of blocks, the total effect is $V_{j\ell}$, where again $j > \ell$. The *direct effects* are the $B_{j\ell}$ and the *indirect effects* are $A_{j\ell} - Bj\ell$. If we look at what is left after removing the direct and indirect effects from the total effect, we see that

$$V_{j\ell} - A_{j\ell}\Omega_\ell = V_{j\ell} - V_{j\ell|\ell-1,\cdots,0}, \tag{4.1}$$

which could be called the *joint* or *spurious effects.*

References

1. R. Anderson and L. A. Aday, *Access to Medical Care in the U.S.: Realized and Potential*, Medical Care **16** (1978), 533–546.
2. R. P. MacDonald, *Some Problems in the Application of Path Analysis*, Unpublished Research Report, University of Illinois, 1994.
3. M.E. Sobel, *Causal Inference in the Social and Behavioral Sciences*, A Handbook for Statistical Modelling in the Social and Behavioral Sciences (G. Arminger, C.C. Clogg, and M.E. Sobel, eds.), Plenum Press, New York, New York, 1993, pp. 1–38.
4. N. Wermuth, *On Block Recursive Linear Regression Equations*, Revista Brasileira de Probabilidade e Estatistica **6** (1992), 1–56, With Discussion.

UCLA Statistics Program, 8118 Mathematical Sciences Building, University of California at Los Angeles

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

E-mail address, Vivian Lew: `vlew@stat.ucla.edu`