Research Program on Forecasting

Identification of the Linear Factor Model

Benjamin Williams
The George Washington University

RPF Working Paper No. 2018-002
https://www2.gwu.edu/~forcpgm/2018-002.pdf

June 14, 2018

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
https://www2.gwu.edu/~forcpgm

Identification of the Linear Factor Model

Benjamin Williams*
George Washington University

June 5, 2018

Abstract

This paper provides several new results on identification of the linear factor model. The model allows for correlated latent factors and dependence among the idiosyncratic errors. I also illustrate identification under a dedicated measurement structure and other reduced rank restrictions. I use these results to study identification in a model with both observed covariates and latent factors. The analysis emphasizes the different roles played by restrictions on the error covariance matrix, restrictions on the factor loadings and the factor covariance matrix, and restrictions on the coefficients on covariates. The identification results are simple, intuitive, and directly applicable to many settings.

Keywords: Latent variables, factor analysis
JEL Codes: C38, C31, C36

*Address: 2115 G St NW, Washington DC 20052. Phone: (202) 994-6685. Email:bdwilliams@gwu.edu. This paper was motivated by work with James Heckman and has benefited greatly from his input. I am enormously thankful for his encouragement and advice. I also thank John Eric Humphries, Azeem Shaikh, and seminar participants at Temple University and the University of Chicago for helpful comments and discussions.

1 Introduction

In this paper, I develop a new approach to identification of linear factor models. Specifically, I study the model

$$Y_{ij} = \beta_j' X_i + \alpha_j' \theta_i + \varepsilon_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, J, \tag{1}$$

where X_i is a vector of observed regressors, θ_i is a vector of latent factors, and ε_{ij} is an idiosyncratic error. I consider identification based on the first and second moments of $M_i = (X_i', Y_{i1}, \ldots, Y_{iJ})'$ when the number of observations, n, is large and the number of dependent variables, J, is fixed.

Without the $\beta_j' X_i$ term, this is the standard linear factor model (Lawley and Maxwell, 1971; Spearman, 1904). Initially developed primarily as a joint model of multiple psychological evaluations, factor models have subsequently found many uses in economics. See Aigner et al. (1984) for a thorough review of the early literature on factor models in econometrics. In more recent years, factor models have been used as a way of flexibly but parsimoniously modeling unobserved heterogeneity.[1] The factor model has also been implemented as a generalization of the fixed effects model for panel data (see, e.g., Ahn et al., 2013; Bai, 2009; Moon and Weidner, 2010).

Standard conditions, due to Anderson and Rubin (1956) (henceforth AR), have been established for identification of the factor model (excluding observed covariates) when J is fixed, and these conditions are widely used. However, this well-known identification result relies on the assumption that the idiosyncratic errors are mutually uncorrelated. In many economic applications, this can be hard to justify. This is particularly problematic when j indexes time, as in a panel data model. However, this problem can also arise in cross-sectional data for various reasons. While large J results allow for idiosyncratic correlations (Bai, 2009; Chamberlain and Rothschild, 1983), few results are available for fixed J.[2]

Motivated by this problem, in Section 2 I study a factor model without observed covariates and develop an identification strategy that allows for correlation among the idiosyncratic errors. While correlation among the idiosyncratic errors could be modeled by including additional latent factors, in many cases this larger factor structure is not identified.[3] The main idea behind the identification strategy is to translate reduced rank restrictions implied by the factor structure into

[1] See, for example, Abbring and Heckman (2007); Cunha et al. (2005); Heckman et al. (2016); Khan et al. (2015).

[2] One important exception are the fixed J interactive fixed effects models for panel data (Ahn et al., 2001, 2013; Holtz-Eakin et al., 1988), though these models are not fully identified and rely on exclusion restrictions available only in the panel data context.

[3] Consider the simplest example, a model with 4 measurements and 1 factor where all pairs of idiosyncratic errors are uncorrelated except for one. The results in this paper can be used to show that the parameters of such a model are identified, including the nonzero correlation between idiosyncratic errors. However, if we introduce an additional latent factor to explain this idiosyncratic correlation, there are not enough measurements for the resulting two factor model to be identified using standard arguments.

determinantal equations that can be used to separate the variance of the factors from the variance of the idiosyncratic errors. I show through several examples how this straightforward approach can be used to derive identifying conditions for factor models with correlation among idiosyncratic errors.

This identification strategy also sheds light on other common aspects of factor models. First, I demonstrate how identification of the full model is possible in cases where the full distribution of M_i is not observed or identified. This is applicable, for example, in situations involving combination of data from multiple sources where all of the components of M_i are not observed in any single dataset. See Piatek and Pinger (2016) for an empirical application of this idea. Second, my results demonstrate how correlation among the latent factors affects identification requirements in a factor model.[4] Third, I show explicitly how overidentifying rank restrictions (e.g., a dedicated factor structure as in Conti et al., 2014) can alter the requirements for identification.

Because the factors are unobserved, an inherent problem with identification of factor models is the presence of an observationally equivalent model defined by $\theta_i^* = G\theta_i$ for an invertible matrix G. Goldberger (1972) cited this as an important factor in explaining why some economists are uncomfortable with the use of factor models. However, in many cases, as Goldberger (1972) noted and many since have demonstrated, normalizations which limit what matrices G produce an observationally equivalent model can be founded in theoretically motivated restrictions within an economic model. When such restrictions are not available, attention can instead be restricted to parameters, or combinations of parameters, that are identified without resolving this indeterminacy (Heckman and Scheinkman, 1987; Heckman et al., 2011; Pudney, 1981). I demonstrate this here by dividing identification into two steps where only the second step requires these normalizations.

I then show in Section 3 that the model of equation (1) can be viewed as a special case of a factor model without observed covariates and I show how this construction allows my results for identification of the factor model to be translated into new insights into identification of the more general model. One important finding relates to correlation between X_i and θ_i. In the interactive fixed effects model, there is typically no restriction on this correlation. This is typically the case in errors-in-variables models as well. In many other uses of the factor model, however, it is assumed that X_i and θ_i are independent. I show that these models can be viewed as different solutions to the problem of observational equivalence between a model with a latent factor $F_i = (X_i', \theta_i')$ and a model with a latent factor GF_i for an invertible matrix G. I then demonstrate that identification is possible under a range of alternative restrictions. I apply the results to an example to demonstrate the importance of the identifying conditions.

[4] For example, it does not generally increase the required number of measurements.

1.1 Related Literature

The results in this paper complement and build on an extensive literature on identification of models that take the form of equation (1) for fixed J. Many rules have been developed for specific cases that make it easier to apply the row deletion property of AR discussed below (Bollen, 1989; Dunn, 1973). Some have developed rules that are weaker than this property that apply when the factor loading matrix is restricted (O'Brien, 1994; Reilly, 1995; Reilly and O'Brien, 1996). In this literature, some consideration has been given to nonzero correlation among the idiosyncratic errors. See Bollen and Davis (2009) for a recent discussion of this literature. Similar rules have been developed for special cases in the applied econometrics literature (Carneiro et al., 2003; Cunha et al., 2005, 2010; Piatek and Pinger, 2016). Typically these rules do not allow for observed covariates explicitly or assume that observed covariates are uncorrelated with the latent factors.

Other results in the literature pertain to local identification or generic global identification. Bekker (1989) and Wegge (1996) provide conditions on local identification and statistical tests for determining whether these conditions are satisfied.[5] Shapiro (1985) and Bekker and ten Berge (1997) showed that if the Ledermann (1937) bound is satisfied then the factor model is generically identified in the sense that the set of parameter values for which the model is not identified has measure zero.[6] While the parameter values where identification fails are thus rare, standard inference also fails in neighborhoods around these values (Briggs and MacCallum, 2003; Cox, 2017; Ximénez, 2006). Thus full identification analysis is required to understand the finite sample distribution of estimators and test statistics.

Identification conditions relying on higher order moments have also been developed by Bonhomme and Robin (2009) and Ben-Moshe (2016), among others. These results extend the well-known identification argument for the errors-in-variables model of Reiersol (1950). The factor model has also been extended to semi- and nonparametric versions. Cunha et al. (2010), using results from the measurement error model of Hu and Schennach (2008), prove nonparametric identification of a nonseparable model. Freyberger (2017) studies nonparametric identification of a panel data model with interactive fixed effects and fixed J. These results require independent idiosyncratic errors.

1.2 Some matrix notation

The transpose of a matrix Q is denoted Q'. For two matrices, Q_1 and Q_2, $Q_1 \otimes Q_2$ denotes the Kronecker product. The $s \times s$ identity matrix is denoted I_s. In addition, I use the following

[5] Local identification of the parameters follows if the parameters are identified in a neighborhood of the true parameter values. This is in contrast with global identification of the parameters.

[6] The Ledermann bound is equivalent to the inequality $m+k \leq (m-k)^2$ and ensures that the number of unrestricted parameters is no greater than $m(m+1)/2$.

conventions to refer to elements of an $s \times t$ matrix Q. Lowercase letters are used to refer to vectors or scalars – q_j refers to the j^{th} row and q_{jk} refers to the (j,k) element of the matrix. Capital letters are used to denote submatrices. While $Q_1, Q_2, ...$ may be used to refer to a list of submatrices of Q, $Q_{j_1,\ldots,j_p;k_1,\ldots,k_r}$ refers specifically to the $p \times r$ matrix formed from rows $j_1, \ldots, j_p$ and columns $k_1, \ldots, k_r$ of Q,

$$Q_{j_1,\ldots,j_p;k_1,\ldots,k_r} = \begin{pmatrix} q_{j_1k_1} & q_{j_1k_2} & \cdots & q_{j_1k_r} \\ \vdots & \vdots & \vdots & \vdots \\ q_{j_pk_1} & q_{j_pk_2} & \cdots & q_{j_pk_r} \end{pmatrix}$$

In addition, $Q_{j_1,\ldots,j_p}$ refers to the $p \times t$ matrix $Q_{j_1,\ldots,j_p;1,\ldots,t}$. Lastly, if matrices $Q_1, Q_2, \ldots$ have the same number of rows then $Q = (Q_1, Q_2, \ldots)$ refers to the matrix obtained by concatenating the columns.

2 Identification of the Factor Structure

In this section, I consider identification in the model

$$M_i = AF_i + u_i \tag{2}$$

where $M_i \in \mathbb{R}^m$ is a vector of measurements, $F_i \in \mathbb{R}^k$ is a vector of latent factors, and u_i is a vector of idiosyncratic errors. The elements of the $m \times k$ matrix A are referred to as factor loadings. The j^{th} measurement is said to load on the s^{th} factor if $a_{js} \neq 0$. The model of equation (1) fits in this framework by defining $M_i = (X_i', Y_i')'$ and $F_i = (X_i', \theta_i')'$ and imposing a priori restrictions on A and u_i; see Section 3.

I maintain the assumption that each of the common factors, $F_{is}, 1 \leq s \leq k$, is uncorrelated with each of the idiosyncratic errors, $u_{ij}, 1 \leq j \leq m$. This maintained assumption, along with equation (2), implies that

$$\Sigma = A\Phi A' + \Delta, \tag{3}$$

where $\Sigma = Var(M_i)$, $\Phi = Var(F_i)$ and $\Delta = Var(u_i)$. I study identification of A, Φ, and Δ based on this equation. It is not required for any of the results that M_i is observed directly, only that Σ can be consistently estimated.[7] In fact, as demonstrated in Theorem 2.1 below, identification

[7]Thus, the results apply in cases where only the categorical variables $Z_{ij} = \sum_{r=1}^{R} a_r \mathbf{1}(\tau_{j,r} M_{ij} \geq \tau_{j,r+1})$ are observed if the thresholds $\{\tau_{j,r}\}$ and the polychoric correlations can be consistently estimated in a first stage. See Ng (2015) for a recent discussion of estimation of factor models from categorical data.

is possible even when only some of the $m(m+1)/2$ unique components of the matrix Σ can be consistently estimated. See also Example 1.

2.1 Identification with nonzero idiosyncratic correlations

They key idea underlying identification based on equation (3) is that the rank of the matrix $A\Phi A'$ is equal to k, and, thus, any square submatrix of size $k+1$ is singular. Let Σ^- denote a square submatrix of Σ of size $k+1$ and let Δ^- denote the corresponding submatrix of Δ. According to equation (3),

$$\Sigma^- = A_1 \Phi A_2' + \Delta^- \tag{4}$$

where A_1 and A_2 are $(k+1) \times k$ submatrices of A. This representation holds regardless of the order in which the rows and columns are taken from Σ to form Σ^-.[8] Because $A_1 \Phi A_2'$ has rank at most k and is therefore rank-deficient,

$$det(\Sigma^- - \Delta^-) = 0 \tag{5}$$

The proof of Theorem 5.1 in AR uses this fact to prove identification under the assumptions that Δ is diagonal and that the model satisfies the row deletion property. The model satisfies the row deletion property if and only if (i) Φ is positive definite and (ii) when any row is removed from A, two nonoverlapping sets of k linearly independent rows can be formed from the remaining rows of A. Suppose that these conditions on Δ, Φ and A hold, and, for any j, let Σ^- and Δ^- denote the submatrices consisting of rows $j, \ell_1, \ldots, \ell_k$ and columns $j, \ell_{k+1}, \ldots, \ell_{2k}$ where $j, \ell_1, \ldots, \ell_{2k}$ are distinct integers. Then, if Δ is diagonal,

$$\Sigma^- - \Delta^- = \begin{pmatrix} \sigma_{jj} - \delta_{jj} & \Sigma_{j;\ell_{k+1},\ldots,\ell_{2k}} \\ \Sigma_{\ell_1,\ldots,\ell_k;j} & \Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} \end{pmatrix}$$

where σ_{jj} and δ_{jj} refer to the element in the j^{th} row and j^{th} column of Σ and Δ, respectively. The row deletion property implies that the indices $\ell_1, \ldots, \ell_{2k}$ can be chosen so that $rank(A_{\ell_1,\ldots,\ell_k}) = rank(A_{\ell_{k+1},\ldots,\ell_{2k}}) = k$. So $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = A_{\ell_1,\ldots,\ell_k} \Phi A'_{\ell_{k+1},\ldots,\ell_{2k}}$ is nonsingular and the standard

[8] Specifically, let $\Sigma^- = \Sigma_{j_1,\ldots,j_{k+1};j_{k+2},\ldots,j_{2k+2}}$ where $j_1, \ldots, j_{k+1}$ are distinct integers between 1 and m and $j_{k+2}, \ldots, j_{2k+2}$ is another set of distinct integers between 1 and m. Then $\Sigma^- = A_{j_1,\ldots,j_{k+1}} \Phi A'_{j_{k+2},\ldots,j_{2k+2}} + \Delta^-$ where $\Delta^- = \Delta_{j_1,\ldots,j_{k+1};j_{k+2},\ldots,j_{2k+2}}$.

Schur complement formula for the determinant of a partitioned matrix can be applied to obtain

$$det(\Sigma^- - \Delta^-) = det(\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}) \times \left((\sigma_{jj} - \delta_{jj}) - \Sigma_{j;\ell_{k+1},\ldots,\ell_{2k}}\Sigma^{-1}_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}\Sigma_{\ell_1,\ldots,\ell_k;j}\right) \tag{6}$$

which, together with equation (5), implies that

$$\delta_{jj} = \sigma_{jj} - \Sigma_{j;\ell_{k+1},\ldots,\ell_{2k}}\Sigma^{-1}_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}\Sigma_{\ell_1,\ldots,\ell_k;j}$$

Repeating this for each j shows that the matrix Δ, and hence the matrix $A\Phi A' = \Sigma - \Delta$, is identified.

This identification argument can be applied more generally in cases where Δ is not diagonal and/or all elements of Σ are not known. A system of nonlinear equations can be obtained by choosing different submatrices and stacking the corresponding equations given by (5). In many cases, as in the above argument, this system of equations can be solved explicitly. The most straightforward solution is when, as in the above argument, each determinantal equation involves only one unknown parameter. This case is summarized in the following theorem, a proof of which is supplied in the Appendix A.

Theorem 2.1. *Suppose that Φ is positive definite and that for each $1 \leq j_1 \leq j_2 \leq m$ one of the following conditions is satisfied.*

(a) $\sigma_{j_1j_2}$ *is known and* $\delta_{j_1j_2} = 0$.

(b) $\sigma_{j_1j_2}$ *is known and there exist distinct integers* $\ell_1, \ldots, \ell_{2k}$ *such that* $\Delta_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = 0$, $\Delta_{j_1;\ell_1,\ldots,\ell_k} = 0$, $\Delta_{\ell_{k+1},\ldots,\ell_{2k};j_2} = 0$*, and* $A_{\ell_1,\ldots,\ell_k}$ *and* $A_{\ell_{k+1},\ldots,\ell_{2k}}$ *are both full rank, and* $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}$, $\Sigma_{j_1;\ell_1,\ldots,\ell_k}$ *and* $\Sigma_{\ell_{k+1},\ldots,\ell_{2k};j_2}$ *are all known.*

(c) $\delta_{j_1j_2} = 0$ *and there exist distinct integers* $\ell_1, \ldots, \ell_{2k}$ *such that* $\Delta_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = 0$, $\Delta_{j_1;\ell_1,\ldots,\ell_k} = 0$, $\Delta_{\ell_{k+1},\ldots,\ell_{2k};j_2} = 0$, $A_{\ell_1,\ldots,\ell_k;1,\ldots,k}$ *and* $A_{\ell_{k+1},\ldots,\ell_{2k};1,\ldots,k}$ *are both full rank, and* $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}$, $\Sigma_{j_1;\ell_1,\ldots,\ell_k}$ *and* $\Sigma_{\ell_{k+1},\ldots,\ell_{2k};j_2}$ *are all known.*

Then Σ, Δ and $A\Phi A'$ are identified.

Remark 1: *Allowing for some elements of the covariance matrix Σ to be unknown is useful in two cases. First, it can be applied in cases where the full vector M_i is not observed in a single dataset but the missing elements are observed in an auxiliary dataset. See Piatek and Pinger (2016). Second, it can be applied in cases where some elements of M_i are counterfactual outcomes whose marginal distributions, but not joint distributions, may be identified (Carneiro et al., 2003). Both cases are demonstrated in Example 1 below.*

Remark 2: *Identification of the distribution of $A^* F_i$ for some $k \times k$ matrix A^* follows under the conditions of Theorem 2.1 if the zero correlation restrictions are strengthened to independence and if F_i has a finite mean and nonvanishing characteristic function by an application of Theorem 2.1 in Ben-Moshe (2018).*

Remark 3: *This result provides a way to derive identification of Δ and $A\Phi A'$ without imposing any normalizations. These normalizations can be considered in a second step to obtain identification of A and Φ or other features of the model (see Section 2.3).*

Remark 4: *The rank conditions in Theorem 2.1 are stated in terms of the underlying parameters. However, as is evident from the proof, it is sufficient for $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}$ to be nonsingular. Thus, this is a testable condition.*

Remark 5: *While the restrictions on Δ are imposed a priori, the identity of the indices $(\ell_1, \ldots, \ell_{2k})$ satisfying condition (b) or (c) does not need to be known. If the condition is satisfied for some $(\ell_1, \ldots, \ell_{2k})$ then one can search across all $(\ell_1, \ldots, \ell_{2k})$ for which (a) $\Delta_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = 0$, $\Delta_{j_1;\ell_1,\ldots,\ell_k} = 0$, $\Delta_{\ell_{k+1},\ldots,\ell_{2k};j_2} = 0$, and (b) $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}$, $\Sigma_{j_1;\ell_1,\ldots,\ell_k}$ and $\Sigma_{\ell_{k+1},\ldots,\ell_{2k};j_2}$ are known, and check if $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}$ is nonsingular.*

This result applies to a much broader class of models than the well-known AR result. It avoids the often tedious task of solving equation (3) for A, Φ, and Δ under particular normalizations. Though the notation is cumbersome, it can be fairly easy to implement. The result is limited in that it only considers identification from the determinantal equations of the form (5) that contain only one unknown parameter and it does not use the equations of this form that contain multiple unknown parameters. However, this situation still applies in many cases, and when it does not apply, the results can be naturally extended, though not in a way that lends itself to a general statement of identification. I now demonstrate the use of Theorem 2.1, and when the theorem does not apply, through three examples. I then study specific classes of examples.

Example 1 Consider a model where $m = 4$, $k = 1$ and σ_{12} is not known but the remaining elements of Σ are known. This situation could arise if two datasets are available where M_{i1} is observed in only the first dataset, M_{i2} is observed in only the second, and M_{i3} and M_{i4} are observed in both. It is also relevant in a sample selection model where M_{i1} and M_{i2} represent counterfactual outcomes.[9] Suppose Φ, which is a scalar in this case since $k = 1$, is strictly positive. Further, suppose all 4 factor loadings, $a_{j1}, j = 1, \ldots, 4$, are nonzero and that Δ is diagonal. After reordering

[9]Suppose $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ where D_i is a binary choice variable and Y_{1i} and Y_{0i} are potential outcomes. Suppose $D_i = \mathbf{1}(D_i^* \geq 0)$. Suppose W_i is a measurement or another outcome that is not affected by the choice D_i. Carneiro et al. (2003) show conditions under which the joint distribution of (Y_{1i}, D_i^*, W_i) and the joint distribution of (Y_{0i}, D_i^*, W_i) are identified. Thus, this model is applicable with $M_{i1} = Y_{1i}$, $M_{i2} = Y_{0i}$, $M_{i3} = W_i$, and $M_{i4} = D_i^*$.

the elements of M_i, equation (3) can be written as

$$\begin{pmatrix} \sigma_{11} & \sigma_{13} & \sigma_{14} & \mathit{\sigma_{12}} \\ \sigma_{13} & \sigma_{33} & \sigma_{34} & \sigma_{23} \\ \sigma_{14} & \sigma_{34} & \sigma_{44} & \sigma_{24} \\ \mathit{\sigma_{12}} & \sigma_{23} & \sigma_{24} & \sigma_{22} \end{pmatrix} = A\Phi A' + \begin{pmatrix} \delta_{11} & 0 & 0 & 0 \\ 0 & \delta_{33} & 0 & 0 \\ 0 & 0 & \delta_{44} & 0 \\ 0 & 0 & 0 & \delta_{22} \end{pmatrix}$$

where I use italics to indicate the elements of Σ that are unobserved. Note that the 3×3 upper left block of Σ is observed in one dataset and the 3×3 lower left block is observed in the other dataset.

The conditions of Theorem 2.1 can be easily verified. For any of the diagonal elements of Δ, δ_{jj}, apply condition (b) with ℓ_1 and ℓ_2 corresponding to two other measurements observed in the same dataset as M_j. This is the same argument used to prove the theorem of AR. The only other remaining pair for which condition (a) does not apply is $(j_1, j_2) = (1, 2)$. For this pair, condition (c) is satisfied with $(\ell_1, \ell_2) = (3, 4)$ because $\delta_{34} = \delta_{13} = \delta_{24} = 0$, a_{31} and a_{41} are nonzero, and σ_{34}, σ_{13}, and σ_{24} are known. Thus Σ, Δ, and $A\Phi A'$ are identified.

In particular, the covariance σ_{12} is identified. Note that this is identified from the equation $\sigma_{12} = \frac{\sigma_{13}\sigma_{24}}{\sigma_{34}}$. Using a factor model to identify dependence by reducing the dimension is not a new idea but the fact that this is possible without imposing any restriction on Φ or the factor loadings does not seem to have been previously recognized. It should also be noted that this argument is impossible if $\delta_{12} \neq 0$ though in that case $\sigma_{12} - \delta_{12}$ is still identified.

Example 2 In the previous example, a subset of the measurements formed an identified factor model. Specifically, (M_{i1}, M_{i3}, M_{i4}) satisfies all the conditions of Theorem 5.1 of AR, as does (M_{i2}, M_{i3}, M_{i4}). However, Theorem 2.1 can also be applied in cases where no subset of the measurements forms an identified factor model. For example, suppose $m = 6$, $k = 2$ and Σ is fully observed. Further, suppose that in addition to the diagonal elements of Δ, δ_{12}, δ_{34}, and δ_{56} are also nonzero. But the remaining elements of Δ are 0. For this case, equation (3) can be written as

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} & \sigma_{56} \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66} \end{pmatrix} = A\Phi A' + \begin{pmatrix} \delta_{11} & \delta_{12} & 0 & 0 & 0 & 0 \\ \delta_{12} & \delta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{33} & \delta_{34} & 0 & 0 \\ 0 & 0 & \delta_{34} & \delta_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{55} & \delta_{56} \\ 0 & 0 & 0 & 0 & \delta_{56} & \delta_{66} \end{pmatrix}$$

The conditions of Theorem 2.1 can then be verified for this example if $rank(A_{1,2}) = rank(A_{3,4}) = rank(A_{5,6}) = 2$. Consider first the index pair $(j_1, j_2) = (1, 1)$. Let $(\ell_1, \ldots, \ell_{2k}) = (3, 4, 5, 6)$.

Then condition (b) is satisfied because $\Delta_{3,4;5,6} = 0$, $\Delta_{1;3,4} = 0$ and $\Delta_{1;5,6} = 0$ and $A_{3,4}$ and $A_{5,6}$ are full rank.

For pairs $(1, 2)$ and $(2, 2)$, condition (b) in the theorem can also be applied with $(\ell_1, \ldots, \ell_{2k}) = (3, 4, 5, 6)$ since $\Delta_{2;3,4} = 0$ and $\Delta_{2;5,6} = 0$. Similarly, for pairs $(3, 3)$, $(3, 4)$ and $(4, 4)$, apply (b) with $(\ell_1, \ldots, \ell_{2k}) = (1, 2, 5, 6)$ and for the remaining pairs corresponding to nonzero elements of Δ apply (b) with $(\ell_1, \ldots, \ell_{2k}) = (1, 2, 3, 4)$. More general block structures like this are considered in Section 2.1.2.

Example 3 Lastly, before proceeding to consider some general classes of examples, I introduce an example where Theorem 2.1 does not apply but the model is nevertheless identified. Suppose that $k = 1$, $m > 3$, the off-diagonal elements of $\Delta_{1,2,3;1,2,3}$ are all 0, and that Φ, a_{11}, a_{21}, and a_{31} are all nonzero. For each $j > 3$, suppose that $\delta_{\ell j} = 0$ for some $\ell < j$. Lastly, suppose Σ is fully observed. One model that satisfies these conditions is

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix} = A\Phi A' + \begin{pmatrix} \delta_{11} & 0 & 0 & \delta_{14} \\ 0 & \delta_{22} & 0 & \delta_{24} \\ 0 & 0 & \delta_{33} & 0 \\ \delta_{14} & \delta_{24} & 0 & \delta_{44} \end{pmatrix}$$

Either condition (a) or condition (b) is satisfied for any (j_1, j_2) such that $1 \leq j_1 \leq j_2 \leq 3$ if a_{11}, a_{21}, and a_{31} are nonzero. Next consider the pair (j, j) for $j > 3$. If (ℓ_1, ℓ_2) is such that $\delta_{j,\ell_1} = \delta_{j,\ell_2} = 0$, as required by condition (b), then $\ell_1 = \ell_2$ because the model only restricts $\delta_{\ell,j}$ for a single value of ℓ. Since $\delta_{\ell,\ell} \neq 0$, it is not possible to satisfy condition (b).

However, identification is possible because, while $\delta_{\ell,\ell}$ is not 0, it is identified in a previous step. Indeed, let $\Sigma^- = \Sigma_{\ell,j;\ell,j}$ and let Δ^- denote the corresponding submatrix of Δ. Equation (5) holds and can be simplified to

$$(\sigma_{\ell\ell} - \delta_{\ell\ell})(\sigma_{jj} - \delta_{jj}) = \sigma_{j\ell}^2$$

Starting with $j = 4$, these equations can be solved iteratively if $a_{\ell 1} \neq 0$ since $\delta_{\ell\ell}$ is identified in a previous step and $\sigma_{\ell\ell} - \delta_{\ell\ell} = a_{\ell 1}^2 \Phi$.

Then, for any $j_1 < j_2$ such that $\delta_{j_1 j_2} \neq 0$, equation (5) for the submatrix with rows (ℓ, j_1) and columns (ℓ, j_2), for ℓ such that $\delta_{\ell,j_2} = 0$, can be written as

$$(\sigma_{\ell\ell} - \delta_{\ell\ell})(\sigma_{j_1 j_2} - \delta_{j_1 j_2}) = (\sigma_{\ell j_1} - \delta_{\ell j_1})\sigma_{\ell j_2}$$

Then $\delta_{\ell\ell}$ has been identified already so only $\delta_{j_1 j_2}$ and $\delta_{\ell j_1}$ are unknown. These equations can also be solved iteratively starting with a case where $\delta_{\ell j_1} = 0$.[10] This example again demonstrates the important link between the nonzero elements of Δ and the particular rank conditions required on the factor loading matrix, A. It is not the case that nonzero correlations in Δ come for free as long as m is sufficiently large and AR's row deletion property is satisfied.

2.1.1 m^*-dependence of idiosyncratic errors

Consider a model where $\delta_{j\ell} = 0$ if and only if $|j - \ell| > m^*$ for some nonnegative integer m^*. This type of dependence arises is satisfied by a moving average process but can also arise from a certain type of spatial error structure. The factor model of equation (2) based on the moment conditions (3) is still identified when the errors, u_i, satisfy m^*-dependence for $m^* > 0$, albeit under a stronger rank condition. I will first demonstrate this for $k = m^* = 1$.

Suppose, for example that $k = m^* = 1$ and $m = 5$. The submatrix $\Delta_{1,3;1,5}$ consists of only one nonzero element, δ_{11}. Therefore,

$$det(\Sigma_{1,3;1,5} - \Delta_{1,3;1,5}) = (\sigma_{11} - \delta_{11})\sigma_{35} - \sigma_{13}\sigma_{15}$$

Since $k = 1$, $det(\Sigma_{1,3;1,5} - \Delta_{1,3;1,5}) = 0$ which implies that $\delta_{11} = \sigma_{11} - \frac{\sigma_{13}\sigma_{15}}{\sigma_{35}}$ if $\sigma_{35} = a_3 a_5 \Phi \neq 0$. If a_1 is also nonzero then δ_{33} and δ_{55} are identified by an analogous argument.[11]

Next, δ_{12} is identified from the equation $det(\Sigma_{1,5;2,3} - \Delta_{1,5;2,3}) = (\sigma_{12} - \delta_{12})\sigma_{35} - \sigma_{13}\sigma_{25} = 0$, δ_{22} is identified from the equation $det(\Sigma_{1,2;2,5} - \Delta_{1,2;2,5}) = (\sigma_{12} - \delta_{12})\sigma_{25} - (\sigma_{22} - \delta_{22})\sigma_{15} = 0$, and δ_{23} from $det(\Sigma_{2,3;3,5} - \Delta_{2,3;3,5}) = (\sigma_{23} - \delta_{23})\sigma_{35} - (\sigma_{33} - \delta_{33})\sigma_{25} = 0$. Similarly, δ_{45} is identified from $det(\Sigma_{3,4;1,5} - \Delta_{3,4;1,5}) = 0$, δ_{44} is identified from $det(\Sigma_{1,4;4,5} - \Delta_{1,4;4,5}) = 0$, and δ_{34} from $det(\Sigma_{1,3;3,4} - \Delta_{1,3;3,4}) = 0$. Thus, Δ is identified if a_1, a_3, and a_5 are nonzero.

This identification argument can be extended as long as $m \geq 2(k + m^*) + 1$, as stated in the following theorem.

Theorem 2.2. *Suppose the errors u_i in the model of equation (2) satisfy m^*-dependence for $m^* > 0$ and $m \geq 2(k + m^*) + 1$. Then Δ and $A\Phi A'$ are identified based on the moment conditions (3) if Φ is positive definite and every $k \times k$ submatrix of A is full rank.*

This theorem is proved in the supplemental appendix under a weaker sufficient condition on A that still allows m to be as small as $2(k + m^*) + 1$. The weaker condition still requires loadings associated with particular each factor to be somewhat persistent over time, in order to separate

[10] Specifically, start with $j_2 = 4$. For any $j_1 < j_2 = 4$ it must be that $\delta_{\ell,j_1} = 0$ since $\ell, j_1 \leq 3$. Next, take $j_2 = 5$. Then either j_1 and ℓ are both less than 4, in which case $\delta_{\ell j_1} = 0$, or either ℓ or j_1 is equal to 4, in which case $\delta_{\ell j_1}$ was identified in the first step. This can be iterated until $j_2 = m$.

[11] If a_1, a_3, a_5 and Φ are nonzero then $\delta_{33} = \sigma_{33} - \frac{\sigma_{13}\sigma_{35}}{\sigma_{15}}$ and $\delta_{55} = \sigma_{55} - \frac{\sigma_{15}\sigma_{35}}{\sigma_{13}}$.

correlation due to the latent factors from correlation due to dependence among the idiosyncratic errors.

2.1.2 Block diagonal Δ

Another common type of dependence can be described by a cluster or group structure. Suppose there are groups of measurements, $g = 1, \ldots, G$ such that idiosyncratic errors are correlated within groups but not between groups. Suppose group g has m_g measurements and the measurements are ordered so that $j = 1, \ldots, m_1$ correspond to group 1, $j = m_1 + 1, \ldots, m_1 + m_2$ correspond to group 2, etc. Let A_g denote the $m_g \times k$ submatrix of A corresponding to group g and let Δ_g denote the submatrix of Δ corresponding to group g so that $\Delta = diag(\Delta_1, \ldots, \Delta_G)$.

Theorem 2.3. *Suppose the errors u_i in the model of equation (2) satisfy a group structure. Then Δ and $A\Phi A'$ are identified based on the moment conditions (3) if Φ is positive definite, $rank(A_g) = k$ for each g, and $G \geq 3$. The latter is a necessary condition for identification of Δ and $A\Phi A'$.*

The proof of Theorem 2.3, contained in the Appendix A, is similar to the proof of AR's Theorem 5.1. To accommodate cases where A_g is not full rank for some, or all, g, a more general result can be obtained from an extension of the row deletion property, the *groupwise row deletion property*. The loading matrix A satisfies the groupwise row deletion property if, when rows corresponding to any group g are removed there remain two sets of k linearly independent rows corresponding to two distinct sets of groups. That is, there are $k \times k$ submatrices A_1 and A_2 of A such that (a) each is nonsingular, (b) neither includes a row corresponding to group g, and (c) neither includes a row corresponding to the same group as a row in the other. Unlike the row deletion property, this is a property of A relative to the structure of Δ, and thus cannot be stated without specifying the group structure satisfied by Δ.

Theorem 2.4. *Suppose the errors u_i in the model of equation (2) satisfy a group structure. Then Δ and $A\Phi A'$ are identified based on the moment conditions (3) if Φ is positive definite and A satisfies the groupwise row deletion property.*

See the supplemental appendix for a proof of this theorem and further discussion.

2.1.3 Simultaneous equations system

Dependence between idiosyncratic errors also arises when equation (2) is the reduced form of a simultaneous equations system. Suppose that M_i satisfies the simultaneous equation system

$$M_i = HM_i + \tilde{A}F_i + \tilde{u}_i$$

Then, if $I_m - H$ is invertible, M_i satisfies the model of equation (2) with $A = (I - H)^{-1}\tilde{A}$ and $u_i = (I - H)^{-1}\tilde{u}_i$. Typically $\Delta = (I - H)^{-1}Var(\tilde{u}_i)(I - H)^{-1\prime}$ is not diagonal even if $Var(\tilde{u}_i)$ is. However, Δ may have enough nonzero elements to be identified.

Suppose, for example, that $Var(\tilde{u}_i)$ is diagonal and $H = diag(H_1, \ldots, H_S)$ where each H_s is $m_s \times m_s$. Then, if $I - H_s$ is invertible for each s, Δ is block diagonal as well since $(I - H)^{-1} = diag((I_{m_1} - H_1)^{-1}, \ldots, (I_{m_S} - H_S)^{-1})$. Since Δ is block diagonal, results in the previous section can be applied to derive identification conditions for the reduced form matrix Δ. See the supplemental appendix for an example of identification where H is not block diagonal. Once $A\Phi A'$ and Δ are identified, identification of $\tilde{A}$, H, and $Var(\tilde{u}_i)$ follows from standard arguments for simultaneous equations in combination with the arguments studied below in Section 2.3.

2.2 Dedicated measurements

In many empirical studies using factor models, overidentifying restrictions are imposed in the hope of making the components of F_i more interpretable (Conti et al., 2014; Cunha et al., 2010; Heckman et al., 2006). These studies assume a dedicated factor structure where some, or all, measurements load on only one factor. Bollen (1989), O'Brien (1994), Reilly (1995), Reilly and O'Brien (1996), and Conti et al. (2014) develop some rules for identification under such a structure. In this section, I consider identification under general overidentifying rank restrictions on the factor loading matrix, A, that include a dedicated factor structure as a special case. I seek minimal additional assumptions so that the model is identified, rather than imposing these restrictions in addition to standard identifying assumptions, such as those of AR. I also allow for some dependence among the idiosyncratic errors, as in Theorem 2.1

Before proceeding to the general result, I consider the following example. Suppose $k = 2$ and $m = 4$. Since $m < 2k + 1$, even if Δ is diagonal identification will generally fail. However, suppose that $a_{12} = a_{22} = a_{31} = a_{41} = 0$. That is, the first two measurements load only on F_{i1} and the second two measurements load only on F_{i2}. This is a dedicated measurement condition – the first two measurements are dedicated to the first factor and measurements 3 and 4 are dedicated to the second factor. Cunha et al. (2010), for example, study a dynamic factor model where 2 measurements are available each period for $T \geq 2$ periods and each is dedicated to a period-specific factor. In this simple example it is easy to see that if ϕ_{12}, and the four unrestricted factor loadings are all nonzero then Δ is identified.[12] This example demonstrates how overidentifying restrictions can change the nature of the identification problem. While fewer measurements are required than in the standard result for a 2 factor model, a new requirement is introduced – that $\phi_{12} \neq 0$.

[12]The usual argument proceeds by assuming that $a_{11} = a_{32} = 1$ and solving explicitly for the remaining parameters, starting with the equation $Cov(M_{i1}, M_{i3}) = \phi_{12}$.

In this example, the zero restrictions on the factor loadings imply that $rank(A_{1,2}) = rank(A_{3,4}) = 1 < k$. The approach introduced in this paper can be extended to consider reduced rank restrictions like this. Restricting the rank of submatrices of A has the advantage over imposing zeros that it does not impose a particular interpretation onto the factors. For example, if M_{i1} and M_{i2} are math tests and M_{i3} and M_{i4} are verbal tests then zero restrictions give the impression that F_{i1} represents math ability and F_{i2} represents verbal ability and that math ability does not influence performance on verbal tests and vice versa. A rank restriction, on the other hand, only implies that both math tests depend on the same linear combination of the two dimensions of ability.

Reduced rank restrictions imply that additional identifying equations of the form (5) hold. For some integer $1 \leq r < k$, let Σ^- denote a square submatrix of Σ of size $r + 1$ and let Δ^- denote the corresponding submatrix of Δ. According to equation (3),

$$\Sigma^- = A_1 \Phi A_2' + \Delta^- \tag{7}$$

where A_1 and A_2 are $(r+1)\times k$ submatrices of A. Suppose that $rank(A_1) = r$. Then $rank(A_1 \Phi A_2') \leq r$ so the $(r + 1) \times (r + 1)$ matrix $\Sigma^- - \Delta^- = A_1 \Phi A_2'$ is rank deficient and has zero determinant. Thus the reduced rank restriction on A_1 produces a new identifying equation, $det(\Sigma^- - \Delta^-) = 0$, that is not satisfied if $rank(A_1) > r$. For the two factor example above, let $\Sigma^- = \Sigma_{1,2;1,3}$. Since $A_{1,2}$ has rank 1 this equation can be solved for δ_{11} when Δ is diagonal since in that case it is the only nonzero element of Δ^-, as long as $\sigma_{23} = a_{21}a_{32}\phi_{12} \neq 0$.

We can use this idea to analyze a dedicated factor structure. In the dedicated factor structure, each row of A only has one nonzero entry. This is also known as a model with factor complexity one (Reilly, 1995). Also, suppose that Δ is diagonal. The dedicated factor structure means that the rows of A can be split into k different sets where each set of rows has rank 1. Then, suppose each set contains at least two rows. Let j, ℓ_1 correspond to two rows in the same set such that $A_1 = A_{j,\ell_1}$ has rank 1. Let ℓ_2 correspond to any row in a different set such that $a_{\ell_1} \Phi a_{\ell_2}' \neq 0$. Then if $\Sigma^- = \Sigma_{j,\ell_1;j,\ell_2}$ and Δ^- denotes the corresponding submatrix of Δ, δ_{jj} is identified from the equation

$$det(\Sigma^- - \Delta^-) = (\sigma_{jj} - \delta_{jj})\sigma_{\ell_1\ell_2} - \sigma_{j\ell_1}\sigma_{j\ell_2}$$

This equation holds because A_1 has rank 1 and it can be solved for δ_{jj} because $\sigma_{\ell_1\ell_2} = a_{\ell_1}\Phi A_{\ell_2}' \neq 0$.

The following theorem provides a general result for identification under reduced rank restrictions and non-diagonal Δ.

Theorem 2.5. *Suppose that Φ is positive definite and that for each $1 \leq j_1 \leq j_2 \leq m$ one of the following conditions is satisfied.*

(a) $\delta_{j_1 j_2} = 0$

(b) *there exist distinct integers* $\ell_1, \ldots, \ell_{2r}$ *such that* $rank\left(\begin{pmatrix} a_{j_1} \\ A_{\ell_1,\ldots,\ell_r} \end{pmatrix}\right) = r$*,* $\Delta_{\ell_1,\ldots,\ell_r;\ell_{r+1},\ldots,\ell_{2r}} = 0$*,* $\Delta_{j_1;\ell_1,\ldots,\ell_r} = 0$*,* $\Delta_{\ell_{r+1},\ldots,\ell_{2r};j_2} = 0$*, and*

$$rank(A_{\ell_1,\ldots,\ell_r}\Phi A_{\ell_{k+1},\ldots,\ell_{2r}}) = r.$$

(c) *there exist distinct integers* $\ell_1, \ldots, \ell_{2r}$ *such that* $rank\left(\begin{pmatrix} a_{j_2} \\ A_{\ell_1,\ldots,\ell_r} \end{pmatrix}\right) = r$*,* $\Delta_{\ell_1,\ldots,\ell_r;\ell_{r+1},\ldots,\ell_{2r}} = 0$*,* $\Delta_{j_1;\ell_1,\ldots,\ell_r} = 0$*,* $\Delta_{\ell_{r+1},\ldots,\ell_{2r};j_2} = 0$*, and*

$$rank(A_{\ell_1,\ldots,\ell_r}\Phi A_{\ell_{k+1},\ldots,\ell_{2r}}) = r.$$

Then Δ *and* $A\Phi A'$ *are identified, provided that all elements of* Σ *are known.*

Remark 6: *Reiersol (1950) and Wang (2016) also impose reduced rank restrictions. Specifically, both assume that each column of* A *has at least* $s \geq k$ *zeros. This implies that the submatrix formed by the rows with a zero in a particular column is rank deficient, with rank equal to* $k - 1$*. They impose these restrictions in addition to assuming the row deletion property of AR. The purpose of these assumptions is for identification of* A *and* Φ *once* Δ *and* $A\Phi A'$ *are identified from the row deletion property. This is discussed further in Section 2.3 below.*

Remark 7: *The zero restrictions in Reiersol (1950) and Wang (2016) are in unknown locations. This is useful because the location of the nonzero factor loadings can then be estimated along with the magnitude of these loadings. Conti et al. (2014), e.g., show that a factor model of complexity one is identified without knowing which column of each row of* A *is nonzero and use this to develop a strategy for estimating which factor each measurement loads on. By contrast, Theorem 2.5 requires that the submatrices of* A *satisfying reduced rank restrictions are known.*

Extending Theorem 2.5 to allow the location of the reduced rank restrictions to be unknown seems formidable because of the requirement that corresponding elements of Δ *must be* 0*. However, generally the location of such restrictions does not need to be known if there are testable implications of the assumption that specified submatrices have reduced rank. For example, consider the model with two factors and two dedicated measurements for each factor, where measurements* 1 *and* 2 *load on factor* 1 *and measurements* 3 *and* 4 *load on factor* 2 *and the unrestricted factor loadings are all nonzero. It is not necessary to assume a priori which measurements load on the same factor because* $\Sigma_{1,2;3,4}$ *is the only* 2×2 *submatrix of* Σ *which has rank* 1*. This indicates the proper grouping.*

It is easy to see that the dedicated factor structure with diagonal Δ satisfies the conditions of this theorem with $r = 1$. Also, Theorem 2.1 is a special case of this result with $r = k$ if Σ is known.[13] But it also allows for a range of models between these two extremes. In the supplemental appendix, I demonstrate this through a model from the literature on human capital formation.

2.3 Identifying economically meaningful parameters

While separating $A\Phi A'$ and Δ is often an important decomposition of variance into the portion due to F_i and the portion due to the idiosyncratic component, often this is not sufficient by itself to answer meaningful questions or test economic hypotheses. The issue is that the factor model suffers from a fundamental indeterminacy. For any nonsingular $k \times k$ matrix G, $(AG)(G^{-1}\Phi G^{-1\prime})(AG)' = A\Phi A'$. Therefore, if the goal is identification of Φ and A then additional restrictions, or normalizations, are required. In some cases these restrictions may arise from the underlying economic model naturally. When natural restrictions are not available, important objects of interest may still be identified. In this section, I briefly discuss these issues and provide an illustrative example.

Rotations Assume that $\Psi = A\Phi A'$ is identified. Suppose that $B'A$ is a lower triangular matrix for some known $m \times k$ matrix B. This includes the case where A itself is lower triangular by taking $B = (I_k, 0_{k\times m-k})'$. This is a standard assumption in factor analysis; alternatives include the restriction that $A'A$ is diagonal and the restriction that $A'\Delta^{-1}A$ is diagonal (see, e.g., Anderson and Rubin, 1956; Rao, 1955).

Suppose further that Φ is positive definite and that none of the elements on the diagonal of $B'A$ is zero. Then AL is identified where L is the unique Cholesky decomposition of Φ.[14]

Suppose in addition that either Φ is diagonal or $A_{1,\ldots,k}$, the first k rows of the matrix A, is diagonal. Under either of these conditions, A is identified up to scale.[15] These two restrictions can be combined in interesting ways as well. Suppose, for example, that $A_{1,\ldots,k}$ is block diagonal with T blocks, $A_1, \ldots, A_T$, where each block is $k_t \times k_t$. Then let $\Phi = (\Phi_{tt'})$ where $\Phi_{tt'}$ is $k_t \times k_{t'}$. Then A and Φ are identified up to scale if Φ_{tt} is diagonal for each t.

AR also describe an alternative scheme, due to Reiersol (1950), that imposes additional zero restrictions but does not require that the rows where these restrictions are satisfied is known. See

[13] Theorem 2.5 can, of course, also be extended to the case where some elements of Σ are unknown.

[14] See AR for a proof of this that involves solving algebraically for the elements of AL. Here is a shorter proof. *Proof.* If $B'A$ is lower triangular then so is $B'AL$ is lower triangular. Also, $B'AL(B'AL)' = B'A\Phi A'B$ so $B'AL$ is a Cholesky decomposition of $B'\Psi B$. Also, $B'A$ is nonsingular so $B'\Psi B$ is positive definite and therefore has a unique Cholesky decomposition. Thus, the matrix $B'AL$ is identified. Then, $B'\Psi = (B'AL)(AL)'$ and $B'AL$ is nonsingular so AL is identified as the transpose of $(B'AL)^{-1}B'\Psi$. □

[15] That is, $s_1A^1, \ldots, s_kA^k$ are identified, where $A^1, \ldots, A^k$ are the columns of A and $s_1, \ldots, s_k$ are unknown scalars.

Wang (2016) for a recent discussion of this result. Similarly, in a model of factor complexity one A and Φ are identified up to scale without assuming *a priori* which factor each measurement loads on, as shown by Conti et al. (2014).

Scale The scale problem is typically resolved through normalizations. If, for each $1 \leq s \leq k$, there is a j such that a_{js} is known then A and Φ are identified. The scale can be resolved through restrictions on Φ as well. For example, in either case, a normalization on a_{js} for some j can be replaced by a normalization on ϕ_{ss}. More interestingly, restrictions on the relationship among different elements of Φ can also be helpful in resolving the scale. See the supplemental appendix for an example drawn from Agostinelli and Wiswall (2016, 2017).

The lower triangular structure is often hard to justify when the individual factors take on specific economic or scientific meaning. The underlying economic model may suggest, for example, that each component of F_i should be present in all equations. When A naturally satisfies the lower triangular structure it may not be appealing to assume either that Φ is diagonal or that $A_{1,\ldots,k;1,\ldots,k}$ is diagonal. Lastly, even scale normalizations are not innocuous in that they can substantially change the interpretation of results (Agostinelli and Wiswall, 2016). Nevertheless, important interpretable features of the model are sometimes identified without imposing sufficient restrictions to identify the full model.

I now demonstrate this point through one particularly interesting result that has various applications. Suppose that

$$A = \begin{pmatrix} I_{k_1} & 0 \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix} \tag{8}$$

where A_{21} has $k - k_1 > 0$ rows. This is not even lower triangular if A_{22} is not lower triangular. Nevertheless, it can be shown that if A_{22} is invertible then both $A_{32}A_{22}^{-1}$ and $A_{31} - A_{32}A_{22}^{-1}A_{21}$ are uniquely determined from $A\Phi A'$. I summarize this as a theorem. A proof is provided in the Appendix A.

Theorem 2.6. *Suppose $A\Phi A'$ is identified and Φ is positive definite. If the factor loading matrix A can be written in the form of equation (8) where A_{22} is invertible then $A_{32}A_{22}^{-1}$ and $A_{31} - A_{32}A_{22}^{-1}A_{21}$ are identified.*

This result can be used in several ways. For example, if $A_{21} = 0$ then A_{31} is identified, regardless of any further restrictions on A_{22}. Or, if A_{22} is known then A_{32} is identified, regardless of any restrictions on A_{21}. Other restrictions can be used to identify A_{31} and/or A_{32} as well.

See Section 3. The result can also be used to test important economic hypotheses without any additional restrictions.

Consider, for example, a simplified version of the model studied in Cunha et al. (2005). Let M_{i1} denote an IQ score or another measure of cognitive ability. Let M_{ij} represents earnings in the $(j-1)$st period after schooling is completed ,for $2 \leq j \leq m-1$, and let M_{im} represent years of schooling.[16] Suppose F_{i1} represents cognitive ability and F_{i2} represents a separate factor that influences earnings. Thus $a_{12} = 0$ is indicated by the theory and $a_{11} = 1$ is an innocuous scale normalization but there are no additional restrictions. Then, for $2 \leq j, j' \leq m-1$, $a_{j'2}/a_{j2}$, which represents the relative influence of the "earnings factor" on earnings in different periods, is identified without requiring that F_{i1} and F_{i2} are uncorrelated or that F_{i1} is excluded from any equation.

The tests of the contents of inviduals' information set when making schooling decisions considered by Cunha et al. (2005) are also possible. First, if a_{21} and a_{22} are both nonzero it is possible to test the null hypothesis that F_{i1} and F_{i2} are both not in the information by testing the null that the identified parameters a_{m2}/a_{22} and $a_{m1} - a_{m2}(a_{21}/a_{22})$ are both 0. If this is rejected then either a_{m1} or a_{m2} must be nonzero, meaning that one of the two factors is in the information set. It is then possible to test the null hypothesis that only F_{i1} is in the identified set by test the null that $a_{m2}/a_{22} = 0$.

Finally, testing the null that only F_{i2} is in the identified set is more difficult. One could test the null that $\frac{a_{m2}}{a_{22}}(a_{m1} - a_{m2}(a_{21}/a_{22})) \leq 0$. If $a_{m1} = 0$ then this inequality is satisfied. However, this test only has power against alternatives where $\frac{a_{m1}}{a_{m2}} \geq \frac{a_{21}}{a_{22}}$. If the ratio a_{j1}/a_{j2} varies sufficiently with j then a more powerful procedure would be to test jointly that $\frac{a_{m2}}{a_{j2}}(a_{m1} - a_{m2}(a_{j1}/a_{j2})) \leq 0$ for all $2 \leq j \leq m-1$.

3 Identification of the model with observed regressors

In this section, I return to the general model of equation (1). Suppose that X_i represents a vector of d_X observed regressors and θ_i represents a vector of k_θ latent factors. I assume throughout that $Cov(X_{is}, \varepsilon_{ij}) = 0$ for all s, j, and $Cov(\theta_{i\ell}, \varepsilon_{ik}) = 0$ for all ℓ, k. I then consider identification based on the first and second moments of $M_i = (X_i', Y_{i1}, \ldots, Y_{iJ})'$.

If each X_{is} is uncorrelated with each $\theta_{i\ell}$ then identification is straightforward. In that case, β_j is identified from a regression of Y_{ij} on X_i since $Cov((\alpha_j \theta_{ij} + \varepsilon_{ij}), X_i) = 0$. Furthermore, this implies that the residuals, $Y_{ij} - \beta_j' X_i$, can be constructed and the results of Section 2 can be applied to the covariance matrix of the J residuals to obtain identification of the remaining parameters. In

[16] In Cunha et al. (2005), schooling is modeled using a threshold crossing model. The argument is still applicable if, for example, M_{i2} is a latent index such that schooling is given by $\mathbf{1}(M_{i2} \geq 0)$.

the remainder of this section I will be concerned with the case where some or all of the variables in X_i are correlated with the factors, θ_i.

Let $Y_i = (Y_{i1}, \ldots, Y_{iJ})'$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$. The model of equation (1) can be written in the form of equation (2) with $M_i = (X_i', Y_i')'$, $F_i = (X_i', \theta_i')'$, $u_i = (0, \varepsilon_i')'$ and

$$A = \begin{pmatrix} I_{d_X} & 0 \\ \beta & \alpha \end{pmatrix} \tag{9}$$

where α and β are constructed by stacking the vectors α_j and β_j as rows. Moreover, F_i is uncorrelated with u_i so that equation (3) is satisfied with

$$\Delta = \begin{pmatrix} 0 & 0 \\ 0 & \Delta_Y \end{pmatrix} \tag{10}$$

Lastly, the matrices Σ and Φ can also be written in partitioned form to conform with A and Δ.

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{XY}' & \Sigma_Y \end{pmatrix}, \qquad \Phi = \begin{pmatrix} \Phi_X & \Phi_{X\theta} \\ \Phi_{X\theta}' & \Phi_\theta \end{pmatrix}$$

Thus the system of equations given in (1) can be represented as a $k = d_x + k_\theta$ factor model with $m = d_X + J$.

In the remainder of this section, I study identification of the model of equation (1) based on the moment conditions (3) under the restrictions implied by (9) and (10). Following the analysis of Section 2, I first consider identification of Δ and $A\Phi A'$ from equations of the form (5). I then explore what additional restrictions are sufficient to identify A and Φ.

3.1 Identification of reduced form parameters

Applying Theorem 5.1 of AR, identification of Δ and $A\Phi A'$ follows if Φ is full rank, A satisfies the row deletion property, and Δ_Y is diagonal. Applying this directly, however, results in identification conditions that are more restrictive than necessary. Indeed this would suggest a requirement that $J \geq d_X + 2k_\theta + 1$.[17] Applying the AR results or the results in Section 2 directly fails to take advantage of the restrictions on Δ implied by the fact that X_i is observed.

An identified factor structure Because X_i is observed, Δ has *more* zero elements than in the examples considered in Section 2. Consider a submatrix of Σ of the form $\Sigma^- = Cov\left(\begin{pmatrix} X_i \\ Y_i^{(1)} \end{pmatrix}, \begin{pmatrix} X_i \\ Y_i^{(2)} \end{pmatrix}\right)$

[17]The row deletion property requires $m \geq 2k + 1$. Because $m = d_X + J$ and $k = d_X + k_\theta$, this inequality is equivalent to $J \geq d_X + 2k_\theta + 1$.

where $Y_i^{(1)}$ and $Y_i^{(2)}$ each consist of $k_\theta + 1$ elements of Y_i and let Δ^- represent the corresponding submatrix of Δ. Then $\Sigma^- - \Delta^-$ is a $(k+1)\times(k+1)$ matrix of rank k and hence $det(\Sigma^- - \Delta^-) = 0$. Moreover, since the only nonzero elements of Δ^- are in the submatrix $\Delta_Y^- = Cov(\varepsilon_i^{(1)}, \varepsilon_i^{(2)})$, if $\Sigma_X = Var(X_i)$ is full rank then

$$det(\Sigma^- - \Delta^-) = det(\Sigma_X) det\left(\Sigma_Y^- - \Delta_Y^-\right)$$

where $\Sigma_Y^- = Cov(Y_i^{(1)}, Y_i^{(2)}) - Cov(Y_i^{(1)}, X_i)\Sigma_X^{-1}Cov(X_i, Y_i^{(2)})$. Thus the only role that the presence of X_i in the model plays in terms of deriving sufficient conditions for identification of Δ and $A\Phi A'$ is that Σ_Y^- is a submatrix of the Schur complement of Σ with respect to Σ_X rather than simply a submatrix of Σ_Y. Increasing d_X, the number of observed covariates, does not affect any requirements on J.

To see this more directly, let E^* denote the linear projection operator.[18] Because $E^*(\varepsilon_i \mid X_i) = 0$,

$$Y_i - E^*(Y_i \mid X_i) = \alpha(\theta_i - E^*(\theta_i \mid X_i)) + \varepsilon_i$$

Thus, the observed covariates X_i have been removed, and, since ε_i is uncorrelated with $\theta_i - E^*(\theta_i \mid X_i)$, this leaves a linear factor model,

$$\Sigma_Y^S = \alpha\Phi_\theta^S\alpha' + \Delta_Y \tag{11}$$

where $\Sigma_Y^S = Var(Y_i - E^*(Y_i \mid X_i))$ is the Schur complement of Σ with respect to Σ_X and $\Phi_\theta^S = Var(\theta_i - E^*(\theta_i \mid X_i))$ is the Schur complement of Φ with respect to Φ_X. If $\Delta_Y = Var(\varepsilon_i)$ is identified in this factor model, then Δ and hence $A\Phi A' = \Sigma - \Delta$ are identified as well.

Exclusion restrictions In some cases the presence of X_i can reduce the restrictions required to identify Δ and $A\Phi A'$ if enough components of β are restricted to be 0. Such exclusion restrictions are common in practice. If some components of Y_i are "outcomes" while others are pure "measurements", many, if not all, of the regressors present in the outcome equations are excluded from the equations for the pure measurements. If j indexes time, so that the model is a panel data model, and $X_i = (X_{i1}', \ldots, X_{iJ}')'$ where $X_{ij} \in \mathbb{R}^q$ then typically $\beta = I_J \otimes \beta_0$ where β_0 is $q \times 1$. Even if the coefficients are time-varying and some lags or leads of X_{ij} are allowed, there will still be many exclusion restrictions.

Suppose that $d_X \geq k_\theta$. Fix $j = 1$ and suppose that at least k_θ elements of β_1 are 0. Further, let $Y_i^{(1)}$ denote a length k_θ subvector of $(Y_{i2}, \ldots, Y_{iJ})'$ with a corresponding coefficient matrix $\beta^{(1)}$

[18] For random vectors Z_1 and Z_2, $E^*(Z_2 \mid Z_1) = Cov(Z_2, Z_1)Var(Z_1)^{-1}Z_1$.

such that these same k_θ elements of each row of $\beta^{(1)}$ are also 0. Then let X_i^{exc} and X_i^{inc} denote the elements of X_i corresponding to the zero and nonzero elements of β_1, respectively. Thus the variables X_i^{exc} are excluded from the equation for Y_{i1} as well as the equation for $Y_i^{(1)}$. Then define $M_i^{(1)} = (X_i^{inc\prime}, Y_{ij}, Y_i^{(1)\prime})'$ and $M_i^{(2)} = (X_i^{inc\prime}, Y_{ij}, X_i^{exc,1})$ where $X_i^{exc,1}$ is a length k_θ subvector of X_i^{exc}. Let $M_i^{(1)} = A^{(1)}F + \varepsilon_i^{(1)}$ and $M_i^{(2)} = A^{(2)}F + \varepsilon_i^{(2)}$ where $A^{(1)}$ and $A^{(2)}$ are the appropriate submatrices of A, each with $k_\theta + dim(X_i^{inc}) + 1$ rows. Because of the restrictions on β_j and $\beta^{(1)}$, $rank(A^{(1)}) \leq k_\theta + dim(X_i^{inc})$. Therefore, if $\Sigma^- = Cov(M_i^{(1)}, M_i^{(2)})$ and Δ^- denotes the corresponding submatrix of Δ then $\Sigma^- - \Delta^- = A^{(1)}\Phi A^{(2)\prime}$ and $det(\Sigma^- - \Delta^-) = 0$. Let $M_i^{(11)}$ and $M_i^{(21)}$ denote the subvectors of $M_i^{(1)}$ and $M_i^{(2)}$ that exclude Y_{i1}. If the only nonzero element of Δ^- is $Var(\varepsilon_{i1})$ and if $Cov(M_i^{(11)}, M_i^{(21)})$ is full rank then the equation $det(\Sigma^- - \Delta^-) = 0$ uniquely identifies $Var(\varepsilon_{i1})$. The same argument can be repeated for $j > 1$ if there are sufficient exclusion restrictions.

Thus there is a tradeoff between these exclusion restrictions and restrictions on Δ_Y. If there are fewer exclusion restrictions, the above argument can be modified by redefining $M_i^{(21)}$ to include $Y_i^{(2)}$, a subvector of Y_i. Further, as I have already shown, if there are sufficient restrictions on Δ_Y then identification of $A\Phi A'$ and Δ_Y follows without any exclusion restrictions. On the other hand, with enough additional exclusion restrictions, identification of β is possible with no restrictions on Δ_Y. Starting with the same definition of $Y_i^{(1)}$, suppose that X_i^{inc} can be divided into $X_i^{inc,1}$ and $X_i^{inc,2}$ so that $X_i^{inc,1}$ is excluded from $Y_i^{(1)}$ and $X_i^{inc,2}$ is excluded from Y_{i1}.

$$
\begin{aligned}
Y_{i1} &= \beta_{01} - \alpha_1(\alpha^{(1)})^{-1}\beta_0^{(1)} + \beta_1 X_i - \alpha_1(\alpha^{(1)})^{-1}\beta^{(1)}X_i + \alpha_1(\alpha^{(1)})^{-1}Y_i^{(1)} + \varepsilon_{i1} - \alpha_1(\alpha^{(1)})^{-1}\varepsilon_i^{(1)} \\
&= \beta_1 X_i^{inc,1} + \psi_0 + \psi_1 X_i^{inc,2} + \psi_2 Y_i^{(1)} + \tilde{\varepsilon}_i
\end{aligned}
$$

The excluded variables X_i^{exc} can be used as instruments for $Y_i^{(1)}$. So the coefficients β_1, ψ_0, ψ_1, and ψ_2 are identified under an appropriate rank condition from the moments $E(\tilde{\varepsilon}_i X_i) = 0$, $E(\tilde{\varepsilon}_i) = 0$. This requires $k_\theta + dim(X_i^{inc,2})$ elements of X_i to be excluded from the equation for Y_{i1}. The argument is similar the identification argument in Ahn et al. (2013) in a panel data model with a factor structure. While β is identified by applying this argument for each j, generally Δ_Y, Φ, and α are not identified without further conditions.

3.2 Identification of the full model

Now consider identification of β, α, and Φ given $A\Phi A'$ and Δ. One approach is based on an application of Theorem 2.6. Let $A_1 = (\beta_{(1)}, \alpha_{(1)})$ denote k_θ rows of A and let $A_2 = (\beta_{(2)}, \alpha_{(2)})$

denote the remaining rows. Then

$$A = \begin{pmatrix} I_{d_X} & 0 \\ \beta_{(1)} & \alpha_{(1)} \\ \beta_{(2)} & \alpha_{(2)} \end{pmatrix}$$

By Theorem 2.6, if $A\Phi A'$ is identified and if $\alpha_{(1)}$ is full rank then $\alpha_{(2)}(\alpha_{(1)})^{-1}$ and $\beta_{(2)} - \alpha_{(2)}(\alpha_{(1)})^{-1}\beta_{(1)}$ are identified. A similar idea is used in Pudney (1981) and Heckman and Scheinkman (1987) using an instrumental variable approach (see the supplementary appendix). Heckman and Scheinkman (1987) showed, for example, that uniformity of the prices of different skills across sectors in a Gorman-Lancaster model for earnings could be tested using these reduced form parameters.

This shows how exclusion restrictions can be used to identify the coefficients on observed regressors. Suppose the equations can be reordered so that (a) the s^{th} regressor is excluded from the k_θ equations corresponding to $\beta_{(1)}$ and (b) $\alpha_{(1)}$ is full rank. The s^{th} column of $\beta_{(2)} - \alpha_{(2)}\alpha_{(1)}^{-1}\beta_{(1)}$ is equal to $\beta_{(2)}^{(s)} - \alpha_{(2)}\alpha_{(1)}^{-1}\beta_{(1)}^{(s)}$ where $\beta_{(1)}^{(s)}$ and $\beta_{(2)}^{(s)}$ denote the s^{th} columns of $\beta_{(1)}$ and $\beta_{(2)}$. Since $\beta_{(1)}^{(s)} = 0$ this implies that $\beta_{(2)}^{(s)}$ is identified. If this can be done for each $1 \leq s \leq d_X$ then the full coefficient matrix β is identified.

Alternatively, suppose the coefficients on the s^{th} regressor are restricted to be the same across $k_\theta + 1$ equations. Then the rows can be reordered so that $\beta_{(1)}^{(s)} = b_s \iota_{k_\theta}$ and $\beta_{(21)}^{(s)} = b_s$ for a scalar b_s and the remaining rows, $\beta_{(22)}$, are unrestricted, where the indices (21) and (22) correspond to a division of A_2 into a single row A_{21} and its remaining rows A_{22}. First, $\beta_{(21)}^{(s)} - \alpha_{(21)}\alpha_{(1)}^{-1}\beta_{(1)}^{(s)} = b_s(1 - \alpha_{(21)}\alpha_{(1)}^{-1}\iota_{k_\theta})$. Therefore, if $\alpha_{(1)}$ is full rank, $b_s(1 - \alpha_{(21)}\alpha_{(1)}^{-1}\iota_{k_\theta})$ and $\alpha_{(21)}\alpha_{(1)}^{-1}$ are both identified. If, in addition, $\alpha_{(21)}\alpha_{(1)}^{-1}\iota_{k_\theta} \neq 1$, b_s is identified. Moreover, $\beta_{(22)}^{(s)}$ is identified as well from $\beta_{(22)}^{(s)} - \alpha_{(22)}\alpha_{(1)}^{-1}\beta_{(1)}^{(s)} = \beta_{(22)}^{(s)} - \alpha_{(22)}\alpha_{(1)}^{-1} b_s \iota_{k_\theta}$ since the second term is identified. This argument does not apply to a standard fixed effects panel model since it requires the factor loadings to vary.

Beyond exclusion restrictions and homogeneity restrictions, identification of $\beta^{(s)}$ follows more generally from any restriction of the form $Q_{1s}\beta^{(s)} = Q_{0s}$ if the matrix

$$\begin{pmatrix} -\alpha_{(2)}\alpha_{(1)}^{-1} & I \\ \multicolumn{2}{c}{Q_{1s}} \end{pmatrix} \tag{12}$$

has full column rank. Evidently these restrictions can be combined to tailor the identification approach to the particular model.

When these applications of Theorem 2.6 do not fully identify the objects of interest, restrictions on the covariance between X_i and θ_i can also be used. The off-diagonal block of $A\Phi A'$ when written in partitioned form is $\beta\Phi_X + \alpha\Phi_{\theta X}$. Clearly if $\Phi_{\theta X} = 0$ then β is identified. But it can also be seen that if the s^{th} component of X_i is uncorrelated with θ_i then $\beta\Phi_X^s$, a weighted average of the coefficients, is identified. So if some but not all of the components of X_i are uncorrelated with θ_i then this result can be combined with the previous results to fully identify β. Or, similarly, if a weighted sum of the components of X_i is uncorrelated with θ_i then $\Phi_{\theta X} w = 0$ for some vector w so that $\beta \Phi_X w$ is identified. This is in the spirit of the Hausman and Taylor (1981) approach for panel data.[19]

Lastly, if $\Phi_{\theta X} \Phi_X^{(s)} = 0$ where $\Phi_X^{(s)}$ is the s^{th} column of Φ_X^{-1} then the s^{th} column of β is identified. This is a conditional mean independence assumption. It follows if $E^*(\theta_i \mid X_i) = E^*(\theta_i \mid X_{i,-s})$ where E^* represents the linear projection operator.

The analysis in Section 2 emphasizes the importance of the rank conditions for identification of Δ and $A\Phi A'$. Identification of β_j in equation (1) depends on these rank conditions but also depends on conditions related to whatever restrictions are used to identify β from $A\Phi A'$. When the latter restriction is that $Cov(X_i, \theta_i) = 0$ this rank condition is simply that $Var(X_i)$ is full rank. When exclusion restrictions are used, as in the interactive fixed effects model, the rank conditions require that the factor loading matrix corresponding to a particular set of equations be full rank. And more generally, the rank conditions take the form of equation (12). When the identification strategy is based on multiple different types of restrictions, a careful analysis of the rank conditions is particularly important.

3.3 Example

To demonstrate the value of these results I now provide a simple example. Consider a setup with $k = 2$ and $J = 5$ where the first 2 components of Y_i correspond to "measurements" while the remaining correspond to "outcomes". Correspondingly, assume that $\beta_1 = \beta_2 = 0$ but that there are no additional restrictions on β or α. Thus,

$$\begin{aligned}
Y_{i1} &= \alpha_{11}\theta_{i1} + \alpha_{12}\theta_{i2} + \varepsilon_{i1} \\
Y_{i2} &= \alpha_{21}\theta_{i1} + \alpha_{22}\theta_{i2} + \varepsilon_{i2} \\
Y_{i3} &= \beta_3' X_i + \alpha_{31}\theta_{i1} + \alpha_{32}\theta_{i2} + \varepsilon_{i3} \\
Y_{i4} &= \beta_4' X_i + \alpha_{41}\theta_{i1} + \alpha_{42}\theta_{i2} + \varepsilon_{i4} \\
Y_{i5} &= \beta_5' X_i + \alpha_{51}\theta_{i1} + \alpha_{52}\theta_{i2} + \varepsilon_{i5}
\end{aligned} \tag{13}$$

First, suppose the errors, ε_{ij}, are all mutually uncorrelated so that Δ_Y is diagonal, that α satisfies the row deletion property, and that Φ and Φ_X are both full rank. Then Δ_Y and $A\Phi A'$ are identified by equation (11). Next, let $\alpha_{(1)}$ and $\beta_{(1)}$ denote the first two rows of α and β and $\alpha_{(2)}$

[19]Hausman and Taylor (1981) show that, to identify the coefficient on a time-invariant regressor in a fixed effects panel model it is sufficient to assume that this regressor or a time series average of another regressor is uncorrelated with the fixed effect.

and $\beta_{(2)}$ the remaining 3 rows. As demonstrated above, if $\alpha_{(1)}$ is full rank then $\beta_{(2)} - \alpha_{(2)}\alpha_{(1)}^{-1}\beta_{(1)}$ is identified. Since $\beta_{(1)} = 0$, this implies that $\beta_{(2)}$ is identified. This argument fails, however, if $rank(\alpha_{(1)}) = 1$, as might be the case if the second factor is only captured by the behavioral outcomes and not by the measurements, and $\beta_{(2)}$ is not identified.

This can be seen through the IV formulation as well. If $\alpha_{(1)}$ is invertible then $\theta_i = \alpha_{(1)}^{-1}(Y_{i(1)} - \varepsilon_{i(1)})$, where $Y_{i(1)} = (Y_{i1}, Y_{i2})'$ and $\varepsilon_{i(1)} = (\varepsilon_{i1}, \varepsilon_{i2})'$. Plugging this into the equation for Y_{i3},

$$Y_{i3} = \beta_3' X_i + \alpha_3' \alpha_{(1)}^{-1} Y_{i(1)} + \varepsilon_{i3} - \alpha_3' \alpha_{(1)}^{-1} \varepsilon_{i(1)}$$

Identification follows by using (Y_{i4}, Y_{i5}) as an instrument. If, however, $\alpha_{(1)}$ is not full rank then either Y_{i4} or Y_{i5}, or both, must be used in proxying for θ_i. This introduces $\beta_4' X$ or $\beta_5' X$, or both, into the reduced form equation and it is not possible to separately identify β_3.

Estimation of β is also problematic if $\alpha_{(1)}$ is full rank, but just nearly so. If, for example, α_{12} and α_{22} are both close to 0, then the finite sample distribution of the estimator of $\beta_{(2)}$ will be distorted. Figure 1 demonstrates that the effect of weak identification on the distribution of both the two stage least squares and maximum likelihood estimators. The simulations show a substantial bias in the weakly identified model.[20]

One way to ensure that estimation of β is robust to this type of weak identification is to enforce additional exclusion restrictions if they are available. Suppose, for example, that $X_i = (X_{i1}', X_{i2}', X_{i3}')'$ where only X_{i1} has nonzero coefficients in the equation for Y_{i3}, only X_{i2} does for Y_{i4}, and only X_{i3} for Y_{i5}. Then β is identified, under the conditions already stated – that α satisfies the row deletion property and Φ and Φ_X are full rank. Panels (a) and (b) in Figure 2 demonstrate how this improves estimation of β even when X_{i1}, X_{i2}, and X_{i3} are fairly highly correlated. Panel (c) shows that the weak identification problem rears its head again, however, if they are too highly correlated. See the supplemental appendix for additional details regarding the simulations.

4 Conclusion

In this paper, I have provided a new framework for understanding identification in factor models. In doing so, this paper makes several important contributions. I separate the task of identification into two steps, thus demonstrating what is identified, and under what conditions, without imposing the sometimes unintuitive normalizations required in the second step. I have shown that there is an important tradeoff between exogeneity conditions (restrictions on Δ) and rank conditions (such as the row deletion property) in accomplishing the first step of the identification analysis.

[20] In the identified model, the 2SLS and MLE estimators both had a bias of 0.03. In the weakly identified model, the 2SLS and MLE estimators had bias of 0.67 and -0.3, respectively.

There is similarly a tradeoff in identification of the coefficient on observed covariates additionally between the restrictions and their related rank conditions. As demonstrated through numerous examples, this new approach can be used to show that many familiar models are identified under less restrictive assumptions than what is commonly imposed.

References

ABBRING, J. H. AND J. J. HECKMAN (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 72.

AGOSTINELLI, F. AND M. WISWALL (2016): "Estimating the Technology of Children's Skill Formation," WP #22442, National Bureau of Economic Research.

——— (2017): "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development," WP #22441, National Bureau of Economic Research.

AHN, S. C., Y. H. LEE, AND P. SCHMIDT (2001): "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics*, 101, 219–255.

——— (2013): "Panel data models with multiple time-varying individual effects," *Journal of Econometrics*, 174, 1–14.

AIGNER, D. J., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): "Latent variable models in econometrics," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, Elsevier, vol. 2 of *Handbook of Econometrics*, chap. 23, 1321–1393.

ANDERSON, T. AND H. RUBIN (1956): "Statistical Inference in Factor Analysis," in *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley: University of California Press, vol. V, 111–150.

BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

BEKKER, P. AND J. TEN BERGE (1997): "Generic Global Identification in Factor Analysis," *Linear Algebra and its Applications*, 264, 255–263.

BEKKER, P. A. (1989): "Identification in restricted factor models and the evaluation of rank conditions," *Journal of Econometrics*, 41, 5 – 16.

BEN-MOSHE, D. (2016): "Identification and Estimation of Coefficients in Dependent Factor Models," Working paper, The Hebrew University of Jerusalem.

——— (2018): "Identification of Joint Distributions in Dependent Factor Models," *Econometric Theory*, 34, 134–165.

BOLLEN, K. A. (1989): *Structural equations with latent variables*, Wiley Online Library.

BOLLEN, K. A. AND W. R. DAVIS (2009): "Two Rules of Identification for Structural Equation Models," *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 523 – 536.

BONHOMME, S. AND J.-M. ROBIN (2009): "Consistent noisy independent component analysis," *Journal of Econometrics*, 149, 12 – 25.

BRIGGS, N. E. AND R. C. MACCALLUM (2003): "Recovery of Weak Common Factors by Maximum Likelihood and Ordinary Least Squares Estimation," *Multivariate Behavioral Research*, 38, 25–56.

CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003): "2001 Lawrence R. Klein Lecture: Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice," *International Economic Review*, 44, 361–422.

CHAMBERLAIN, G. AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

CONTI, G., S. FRÜHWIRTH-SCHNATTER, J. J. HECKMAN, AND R. PIATEK (2014): "Bayesian exploratory factor analysis," *Journal of Econometrics*, 183, 31–57.

COX, G. (2017): "Weak Identification in a Class of Generically Identified Models with an Application to Factor Models," Working paper, Yale University.

CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57, 191–261.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78, 883–931.

DUNN, J. E. (1973): "A note on a sufficiency condition for uniqueness of a restricted factor matrix," *Psychometrika*, 38, 141–143.

FREYBERGER, J. (2017): "Non-parametric Panel Data Models with Interactive Fixed Effects," *The Review of Economic Studies*, forthcoming.

GOLDBERGER, A. S. (1972): "Structural equation methods in the social sciences," *Econometrica*, 40, 979–1001.

HAUSMAN, J. A. AND W. E. TAYLOR (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398.

HECKMAN, J. AND J. SCHEINKMAN (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies*, 54, 243–255.

HECKMAN, J., S. SCHENNACH, AND B. WILLIAMS (2011): "Matching on Proxy Variables," *Unpublished Manuscript*.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2016): "Dynamic treatment effects," *Journal of Econometrics*, 191, 276–292.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 24, 411–482.

HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating vector autoregressions with panel data," *Econometrica*, 1371–1395.

HU, Y. AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

KHAN, S., A. MAUREL, AND Y. ZHANG (2015): "Informational Content of Factor Structures in Simultaneous Discrete Response Models," working paper.

LAWLEY, D. AND A. MAXWELL (1971): *Factor analysis as a statistical method*, American Elsevier.

LEDERMANN, W. (1937): "On the rank of the reduced correlational matrix in multiple-factor analysis," *Psychometrika*, 2, 85–93.

MOON, H. AND M. WEIDNER (2010): "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Unpublished manuscript, University of Southern California*.

NG, S. (2015): "Constructing Common Factors from Continuous and Categorical Data," *Econometric Reviews*, 34, 1141–1171.

O'BRIEN, R. M. (1994): "Identification of Simple Measurement Models with Multiple Latent Variables and Correlated Errors," *Sociological Methodology*, 24, 137–170.

PIATEK, R. AND P. PINGER (2016): "Maintaining (Locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models," *Journal of Applied Econometrics*, 31, 734–755.

PUDNEY, S. E. (1981): "Instrumental variable estimation of a characteristics model of demand," *The Review of Economic Studies*, 48, 417–433.

RAO, C. (1955): "Estimation and tests of significance in factor analysis," *Psychometrika*, 20, 93–111.

REIERSOL, O. (1950): "On the identifiability of parameters in Thurstone's multiple factor analysis," *Psychometrika*, 15, 121–149.

REILLY, T. (1995): "A Necessary and Sufficient Condition for Identification of Confirmatory Factor Analysis Models of Factor Complexity One," *Sociological Methods and Research*, 23, 421 – 441.

REILLY, T. AND R. M. O'BRIEN (1996): "Identification of Confirmatory Factor Analysis Models of Arbitrary Complexity : The Side-by-Side Rule," *Sociological Methods and Research*, 24, 473 – 491.

SHAPIRO, A. (1985): "Identifiability of factor analysis: some results and open problems," *Linear Algebra and its Applications*, 70, 1 – 7.

SPEARMAN, C. (1904): "General intelligence, objectively determined and measured," *American Journal of Psychology*, 15, 201–293.

WANG, J. (2016): "Factor Analysis for High Dimensional Data," Ph.D. thesis, Stanford University.

WEGGE, L. L. (1996): "Local identifiability of the factor analysis and measurement error model parameter," *Journal of Econometrics*, 70, 351 – 382.

XIMÉNEZ, C. (2006): "A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis," *Structural Equation Modeling*, 13, 587–614.

Appendix

A Proofs

Proof of Theorem 2.1. For each j_1, j_2 with $1 \leq j_1 \leq j_2 \leq m$, I will (i) show that δ_{j_1,j_2} is identified by expressing it as a function of known elements of Σ and (ii) show that σ_{j_1,j_2} is identified by expressing it as a function of known elements of Σ. Then Δ and Σ are identified so $A\Phi A' = \Sigma - \Delta$ is as well.

First, if condition (a) holds then $\delta_{j_1,j_2} = 0$ and σ_{j_1,j_2} is identified because it is assumed to be known.

Now, suppose that condition (b) holds. First, σ_{j_1,j_2} is identified because it is assumed to be known. Then, let Σ^- and Δ^- denote the submatrices consisting of rows $j_1, \ell_1, \ldots, \ell_k$ and columns $j_2, \ell_{k+1}, \ldots, \ell_{2k}$. Then, since $\Delta_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = 0$, $\Delta_{j_1;\ell_1,\ldots,\ell_k} = 0$, $\Delta_{\ell_{k+1},\ldots,\ell_{2k};j_2} = 0$,

$$\Sigma^- - \Delta^- = \begin{pmatrix} \sigma_{jj} - \delta_{jj} & \Sigma_{j;\ell_{k+1},\ldots,\ell_{2k}} \\ \Sigma_{\ell_1,\ldots,\ell_k;j} & \Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} \end{pmatrix}$$

Since $A_{\ell_1,\ldots,\ell_k}$ and $A_{\ell_{k+1},\ldots,\ell_{2k}}$ are both full rank and Φ is nonsingular, it follows that $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = A_{\ell_1,\ldots,\ell_k}\Phi A'_{\ell_{k+1},\ldots,\ell_{2k}}$ is nonsingular and the Schur complement formula for the determinant of a partitioned matrix can be applied to obtain

$$\begin{aligned} 0 = det(\Sigma^- - \Delta^-) &= det(\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}}) \\ &\quad \times \left((\sigma_{j_1 j_2} - \delta_{j_1 j_2}) - \Sigma_{j_2;\ell_{k+1},\ldots,\ell_{2k}} \Sigma^{-1}_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} \Sigma_{\ell_1,\ldots,\ell_k;j_1} \right) \end{aligned} \tag{14}$$

which implies that

$$\delta_{j_1 j_2} = \sigma_{j_1 j_2} - \Sigma_{j_2;\ell_{k+1},\ldots,\ell_{2k}} \Sigma^{-1}_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} \Sigma_{\ell_1,\ldots,\ell_k;j_1}$$

The right hand side is known so $\delta_{j_1 j_2}$ and $c_{j_1 j_2} = \sigma_{j_1 j_2} - \delta_{j_1 j_2}$ are identified.

Lastly, suppose that condition (c) holds. Then $\delta_{j_1 j_2} = 0$ so it is sufficient to show that $\sigma_{j_1 j_2}$ is identified. Again, let Σ^- and Δ^- denote the submatrices consisting of rows $j_1, \ell_1, \ldots, \ell_k$ and columns $j_2, \ell_{k+1}, \ldots, \ell_{2k}$. Then $\Delta^- = 0$ and $\Sigma_{\ell_1,\ldots,\ell_k;\ell_{k+1},\ldots,\ell_{2k}} = A_{\ell_1,\ldots,\ell_k}\Phi A'_{\ell_{k+1},\ldots,\ell_{2k}}$ is nonsingular so the Schur complement formula for the determinant of a partitioned matrix can be applied to obtain

$$0 = det(\Sigma^- - \Delta^-) = det(\Sigma_{\ell_1,\dots,\ell_k;\ell_{k+1},\dots,\ell_{2k}}) \times \left((\sigma_{j_1j_2}) - \Sigma_{j_2;\ell_{k+1},\dots,\ell_{2k}}\Sigma^{-1}_{\ell_1,\dots,\ell_k;\ell_{k+1},\dots,\ell_{2k}}\Sigma_{\ell_1,\dots,\ell_k;j_1}\right) \tag{15}$$

which implies that

$$\sigma_{j_1j_2} = \Sigma_{j_2;\ell_{k+1},\dots,\ell_{2k}}\Sigma^{-1}_{\ell_1,\dots,\ell_k;\ell_{k+1},\dots,\ell_{2k}}\Sigma_{\ell_1,\dots,\ell_k;j_1}$$

The right hand side is known so $\sigma_{j_1j_2}$ is identified.

□

Proof of Theorem 2.3. Consider any indices $1 \leq j_1 \leq j_2 \leq m$. If $\delta_{j_1j_2} \neq 0$ then both corresponding measurements are in the same group, g. Since $G \geq 3$, there are two other groups g' and g'' distinct from g and from each other. Since $A_{g'}$ and $A_{g''}$ are both full rank they must each have a nonsingular $k \times k$ submatrix, denoted $A^-_{g'}$ and $A^-_{g''}$, respectively. Suppose $A^-_{g'}$ consists of rows $\ell_1, \dots, \ell_k$ from A and $A^-_{g''}$ consists of rows $\ell_{k+1}, \dots, \ell_{2k}$ from A. Finally, let Σ^- and Δ^- denote the $(k+1) \times (k+1)$ submatrices of Σ and Δ consisting of rows $j_1, \ell_1, \dots, \ell_k$ and columns $j_2, \ell_{k+1}, \dots, \ell_{2k}$. Then $det(\Sigma^- - \Delta^-) = 0$ and the only nonzero component of Δ^- is $\delta_{j_1j_2}$ since the others all correspond to intergroup, not intragroup, covariances. This equation can be solved for $\delta_{j_1j_2}$ since $A^-_{g'}\Phi A^{-\prime}_{g''}$ is nonsingular.

Now I will show that the model with $G = 2$ is not identified. Let $\tilde{\Phi} = \rho\Phi$ and let $\tilde{A}_1 = A_1$ and $\tilde{A}_2 = \rho^{-1}A_2$. This preserves the covariance structure between the two groups because $\tilde{A}_1\tilde{\Phi}\tilde{A}_2' = A_1(\rho\Phi)(\rho^{-1}A_2)' = A_1\Phi A_2'$. Furthermore, $\tilde{\Delta}_1 = A_1\Phi A_1' - \tilde{A}_1\tilde{\Phi}\tilde{A}_1' + \Delta_1 = (1-\rho)A_1'\Phi A_1 + \Delta_1$ is positive semidefinite and if Δ_2 is positive definite then ρ can be chosen close enough to 1 so that $\tilde{\Delta}_2 = (1-\rho^{-1})A_2\Phi A_2' + \Delta_2$ is also positive definite.

□

Proof of Theorem 2.6. In general, if A_1 denotes the first k rows of A and A_2 denotes the remaining rows of A then $A_2A_1^{-1}$ is identified if A_1 and Φ are both invertible because

$$A_2\Phi A_1'(A_1\Phi A_1')^{-1} = A_2A_1^{-1}$$

If A takes the form of equation (8) then

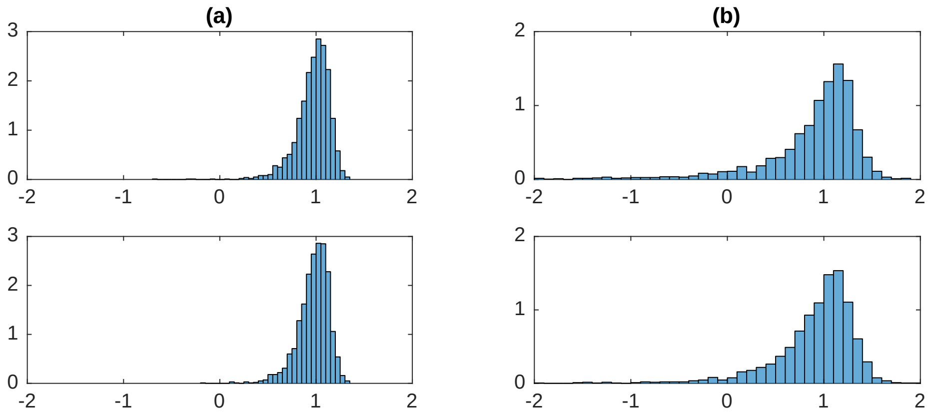
$$A_1^{-1} = \begin{pmatrix} I_{k_1} & 0 \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I_{k_1} & 0 \\ -A_{22}^{-1}A_{21} & A_{22}^{-1} \end{pmatrix}$$

and

$$\begin{aligned} A_2 A_1^{-1} &= \begin{pmatrix} A_{31} & A_{32} \end{pmatrix} \begin{pmatrix} I_{k_1} & 0 \\ -A_{22}^{-1} A_{21} & A_{22}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} A_{31} - A_{32} A_{22}^{-1} A_{21} & A_{32} A_{22}^{-1} \end{pmatrix} \end{aligned}$$

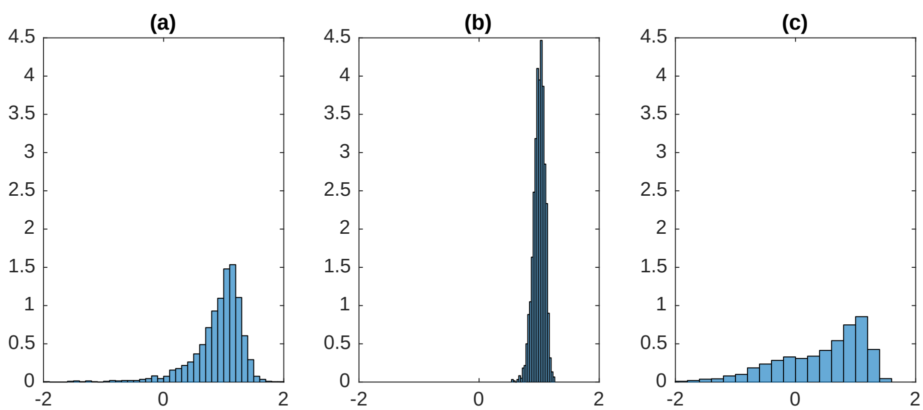
$\square$

Figure 1: Estimation based on first set of exclusion restrictions only



Notes: Two models were simulated 2000 times each. The results summarized in column (a) are from an identified model ($\alpha_{11} = 1, \alpha_{12} = 0, \alpha_{22} = 0.5$). The results summarized in column (b) are from a weakly identified model ($\alpha_{11} = 1, \alpha_{12} = 0, \alpha_{22} = 0.2$). In each column, the distribution of the two stage least squares estimator is in the top panel and the distribution of the maximum likelihood estimator is in the bottom panel. To improve presentation of the results, the histograms in column (b) are truncated; roughly 5% of the sample is dropped in the top panel and 1% in the bottom panel. All simulations were based on a sample size of $n = 1000$. The true value of the parameter was 1. See the appendix for more details.

Figure 2: Estimation based on both sets of exclusion restrictions



Notes: Three models were simulated 2000 times each. The results summarized in panel (a) are from a weakly identified model and are based only on the first set of exclusion restrictions. The results summarized in columns (b) and (c) use both sets of exclusion restrictions but are based on a model with the same factor loadings. In the model simulated for column (b), $corr(X_{ij}, X_{ij'}) = 0.8$. In the model simulated for column (c), $corr(X_{ij}, X_{ij'}) = 0.999$. To improve presentation of the results, the histograms in columns (a) and (c) are truncated; roughly 1% of the sample is dropped in panel (a) and .3% in panel (c). In each case, the distribution of the maximum likelihood estimates is plotted. All simulations were based on a sample size of $n = 1000$. The true value of the parameter was 1 and the means of the three distributions were 0.7023, 0.9993, and 0.4392, respectively. See the appendix for more details.