# Online supplement to: Identification of the Linear Factor Model

Benjamin Williams [*]

June 5, 2018

## Contents

**B Details for examples in Section 2** **2**

B.1 $m^*$-dependence of idiosyncratic errors . . . . . 2

B.2 Block diagonal $\Delta$ . . . . . 4

B.3 Simultaneous equations system . . . . . 4

**C A model of human capital development** **5**

**D An Instrumental Variables Formulation** **6**

**E Identification of the factor distribution** **7**

**F Details of simulation** **8**

[*]George Washington University, Monroe Hall 309, 2115 G Street NW, Washington, DC 20052, bd-williams@gwu.edu

# B Details for examples in Section 2

## B.1 $m^*$-dependence of idiosyncratic errors

The rank condition on the factor loading matrix in Theorem 2.2 is stronger than necessary. The following theorem states that the same conclusion follows under a weaker rank condition.

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

$$
\begin{aligned}
&A_{1,\ldots,k;1,\ldots,k} \\
&A_{j+m^*+1,\ldots,j+m^*+k;1,\ldots,k}, \quad 1 \leq j \leq m-2k-2m^* \\
&\begin{pmatrix} A_{1,\ldots,j-(m-2k-2m^*);1,\ldots,k} \\ A_{j+m^*+1,\ldots,m-k-m^*;1,\ldots,k} \end{pmatrix}, \quad m-2k-2m^* < j \leq m-2k-2m^*+k \\
&A_{m-k+1,\ldots,m;1,\ldots,k} \\
&A_{j-m^*-k,\ldots,j-m^*-1;1,\ldots,k}, \quad 2k+2m^*+1 \leq j \leq m \\
&\begin{pmatrix} A_{k+m^*+1,\ldots,j-m^*-1;1,\ldots,k} \\ A_{(m-2k-2m^*)+j,\ldots,m;1,\ldots,k} \end{pmatrix}, \quad 2k+2m^*+1-k \leq j < 2k+2m^*+1
\end{aligned}
$$

This weaker rank condition allows, for example, for several measurements to only load on one particular factor. However, it does not allow, for example, a factor loading structure where only measurements with $j$ within $2m^*$ of each other load on a particular factor. When $j$ denotes time, this suggests that the model is poorly identified if the strength of one or more factors diminishes greatly over time. Identification is possible in this case though it would require a sufficiently large number (relative to $m^*$) of time periods that load on each dimension of $F_i$ and, as a result, would require $m > 2(k+m^*)+1$.

*Proof.* The proof consists of demonstrating identification of each element of the matrix $\Delta$ from equations of the form (5). Because $\Delta$ is symmetric I restrict attention to $\delta_{j_1 j_2}$ for $j_1 \leq j_2$. If $j_2 - j_1 \geq m^*+1$ then $\delta_{j_1 j_2} = 0$ by assumption.

First, consider any $j_1 \leq j_2$ such that $j_1 \leq m-2k-2m^*$ and $j_2 \leq m-k-m^*$. Let $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,j_1+m^*+1,\ldots,j_1+m^*+k}$. Then $\delta_{j_1 j_2}$ is the only nonzero element of the corresponding submatrix of $\Delta$ because $j_2 + (m^*+1) \leq m-k+1$ and $j_1 + m^* + k + (m^*+1) \leq m-k+1$. The corresponding equation can be solved for this parameter because $A_{j_1+m^*+1,\ldots,j_1+m^*+k;1,\ldots,k}$ and $A_{m-k+1,\ldots,m;1,\ldots,k}$ are full rank.

Second, consider $j_1 \leq j_2$ such that $k+m^*+1 \leq j_1$ and $2k+2m^*+1 \leq j_2$. Let $\Sigma^- = \Sigma_{j_2-m^*-k,\ldots,j_2-m^*-1,j_1;1,\ldots,k,j_2}$. Then $\delta_{j_1 j_2}$ is the only nonzero element of the corresponding submatrix of $\Delta$ because $k+(m^*+1) \leq j_1$ and $k+(m^*+1) \leq j_2 - m^* - k$. The corresponding equation can be solved for this parameter because $A_{1,\ldots,k;1,\ldots,k}$ and $A_{j_2-m^*-k,\ldots,j_2-m^*-1;1,\ldots,k}$ are full rank.

Third, consider $j_1 \leq j_2$ such that $j_1 \leq m-k-m^*$ and $k+m^*+1 \leq j_2$. Let $\Sigma^- = \Sigma_{1,\ldots,k,j_1;j_2,m-k+1,\ldots,m}$, noting that $j_2 \leq j_1 + m^*$ implies that $j_2 \leq m-k+1$. Then $\delta_{j_1 j_2}$ is the only nonzero element of the corresponding submatrix of $\Delta$ because $j_1 + (m^*+1) \leq m-k+1$ and $k+(m^*+1) \leq j_2$. The corresponding equation can be solved for this parameter because $A_{1,\ldots,k;1,\ldots,k}$ and $A_{m-k+1,\ldots,m;1,\ldots,k}$ are full rank.

If $m \geq 3m^* + 3k + 1$ then this completes the proof. Otherwise, two cases remain – (a) $m-2k-2m^* < j_1 < k+m^*+1$ and $j_2 < k+m^*+1$ and (b) $m-k-m^* < j_1 < 2k+2m^*+1$ and $j_2 < 2k+2m^*+1$.

For (a), proceed iteratively, starting with $j_1 = m-2k-2m^*+1$. Let $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1,j_1+m^*+1,\ldots,m-k-m^*}$, or $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1}$ if $k=1$.[1] The former is a $(k+1)\times(k+1)$ matrix because $m-k-m^*-(j_1+m^*+1)+1 = k-1$. Also, $\delta_{j_1 j_2}$, and $\delta_{1 j_1}$ are the only nonzero elements of the corresponding submatrix of $\Delta$ because $j_2+(m^*+1) \leq m-k+1$ and $m-k-m^*+(m^*+1) = m-k+1$. But $\delta_{1j_1}$ was identified in the first step above. The corresponding determinantal equation can be solved because $A_{1,j_1+m^*+1,\ldots,m-k-m^*;1,\ldots,k}$, $A_{1,\ldots,k;1,\ldots,k}$, and $A_{m-k+1,\ldots,m;1,\ldots,k}$ are full rank.

Next, for $j_1 = m-2k-2m^*+2$, let $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1,2,j_1+m^*+1,\ldots,m-k-m^*}$, or $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1,\ldots,k}$ if $k \leq 2$. The former is a $(k+1)\times(k+1)$ matrix because $m-k-m^*-(j_1+m^*+1)+1 = k-2$. Also, $\delta_{j_1 j_2}$, $\delta_{1j_1}$ and $\delta_{2j_1}$ are the only nonzero elements of the corresponding submatrix of $\Delta$ because $j_2+(m^*+1) \leq m-k+1$ and $m-k-m^*+(m^*+1) = m-k+1$. But $\delta_{1j_1}$ was identified in the first step above and $\delta_{2j_1}$ was either identified in that step too (if $m \geq 2k+2m^*+2$) or it was identified in the argument in the paragraph immediately preceeding this one (if $m < 2k+2m^*+2$). The corresponding determinantal equation can be solved because $A_{1,2,j_1+m^*+1,\ldots,m-k-m^*;1,\ldots,k}$, $A_{1,\ldots,k;1,\ldots,k}$, and $A_{m-k+1,\ldots,m;1,\ldots,k}$ are full rank.

We can continue for $j_1 = m-2k-2m^*+i$ by taking $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1,\ldots,i,j_1+m^*+1,\ldots,m-k-m^*}$, or $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;j_2,1,\ldots,k}$ if $k \leq i$. The former is a $(k+1)\times(k+1)$ matrix because $m-k-m^*-(j_1+m^*+1)+1 = k-i$. Also, $\delta_{j_1 j_2}$, and $\delta_{1j_1},\ldots,\delta_{ij_1}$ are the only nonzero elements of the corresponding submatrix of $\Delta$ because $j_2+(m^*+1) \leq m-k+1$ and $m-k-m^*+(m^*+1) = m-k+1$. But it can be shown that $\delta_{1j_1},\ldots,\delta_{ij_1}$ have been identified in a previous step. The corresponding determinantal equation can be solved because $A_{1,\ldots,i,j_1+m^*+1,\ldots,m-k-m^*;1,\ldots,k}$, $A_{1,\ldots,k;1,\ldots,k}$, and $A_{m-k+1,\ldots,m;1,\ldots,k}$ are full rank.

For (b), I proceed iteratively as with (a). Let $j_1 = 2k+2m^*+1-i$, starting with $i=1$ with

[1] Recall that we only need to consider the case where $j_1 \leq j_2 < j_1+m^*+1$ so $j_2, 1, j_1+m^*+1, \ldots, m-k-m^*$ are distinct integers.

$\Sigma^- = \Sigma_{j_1,k+m^*+1,\ldots,j_2-m^*-1,m-i+1,\ldots,m;1,\ldots,k,j_2}$, or $\Sigma^- = \Sigma_{j_1,m-k+1,\ldots,m;1,\ldots,k,j_2}$ if $k \leq i$. $\square$

## B.2 Block diagonal $\Delta$

The structure of $A$ and $\Delta$ in Theorem 2.3 satisfies the *groupwise row deletion property*. However, when some $A_g$ are not full rank, more groups (so $G > 3$) will be needed in order for the groupwise row deletion property to be satisfied. Suppose, for example, that $k = 2$ and $rank(A_1) = 1$. Then $G = 4$ is sufficient if $rank(A_g) = 2$ for $g = 2, 3, 4$. However, suppose $rank(A_2) = 1$ as well. Then the groupwise row deletion property is satisfied for $G = 4$ if and only if $rank((A_1', A_2')') = rank(A_3) = rank(A_4) = 2$. If groups 1 and 2 both load only on the first factor, for example, then the property can only be satisfied if $G \geq 5$. More generally, if the factor loading matrix $A$ has a block structure, it cannot mirror the block structure of $\Delta$.

*Proof of Theorem 2.4.* Consider any indices $1 \leq j_1 \leq j_2 \leq m$. If $\delta_{j_1 j_2} \neq 0$ then both corresponding measurements are in the same group, $g_0$. By the groupwise row deletion property there are two distinct sets of $k$ rows in $A$ that each form nonsingular $k \times k$ submatrices of $A$, $A_1 = A_{\ell_1,\ldots,\ell_k;1,\ldots,k}$ and $A_2 = A_{\ell_{k+1},\ldots,\ell_{2k};1,\ldots,k}$. These rows correspond to groups $g_1, \ldots, g_{2k}$ where $g_j \neq g_0$ for any $j$ and $g_j \neq g_{j'}$ for any $1 \leq j \leq k$ and $k+1 \leq j' \leq 2k$. Therefore, the only nonzero element of $\Delta^- = \Delta_{j_1,\ell_1,\ldots,\ell_k;j_2,\ell_{k+1},\ldots,\ell_{2k}}$ is $\delta_{j_1 j_2}$. Let $\Sigma^- = \Sigma_{j_1,\ell_1,\ldots,\ell_k;j_2,\ell_{k+1},\ldots,\ell_{2k}}$. Then $det(\Sigma^- - \Delta^-) = 0$ and this equation can be solved for $\delta_{j_1 j_2}$ because $A_1 \Phi A_2'$ is nonsingular. $\square$

## B.3 Simultaneous equations system

Section 2.2.3 briefly describes an identification strategy for the model

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

when $H$ is block diagonal. The following example demonstrates identification in a case where $H$ is not block diagonal. Suppose that $k = 1$, $m = 5$, and

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_{34} & 0 \\ 0 & 0 & \rho_{43} & 0 & 0 \\ 0 & \rho_{52} & \rho_{53} & 0 & 0 \end{pmatrix}$$

Further, suppose $Var(\tilde{u}_i)$ is diagonal. Then some tedious algebra shows that $\Delta$ has 6 zero elements $\delta_{jk}$ with $j \leq k$. They are $\delta_{1k}$ for $k = 2, \ldots, 5$, $\delta_{23}$ and $\delta_{24}$. The remaining elements of $\Delta$ are nonzero. It can then be shown that this model is identified if $a_1, a_2$ and either $a_3$ or $a_4$ are nonzero. The proof entails finding a system of $9$ determinantal equations that can be solved for the $9$ nonzero elements of $\Delta$ under this condition on $A$. Rank conditions on $A$ can also be stated in terms of $\tilde{A}$. In particular, it is satisfied in this case if $\tilde{a}_1 \neq 0$, $\tilde{a}_2 \neq 0$, and either $\tilde{a}_3/\tilde{a}_4 \neq \rho_{34}$ or $\tilde{a}_3/\tilde{a}_4 \neq \rho_{43}^{-1}$.

# C A model of human capital development

Consider a model where the $m$ measurements consist of $m_t$ measurements in period $t$ for $t = 1, \ldots, T$, $F_i = (F_{i1}', \ldots, F_{iT}')'$ where $dim(F_{it}) = k_t$, and measurements in period $t$ only load on factors $F_{it}$. Such a model can be written as

$$M_{ijt} = \psi_{jt}' F_{it} + u_{ijt}$$

Theorem 2.5 can be used to study identification in this class of models. I now apply this theorem for two cases of this model.

First, suppose $Cov(u_{ijt}, u_{ij't'}) = 0$ for all $j, j'$ and $t, t'$. In the stacked representation, this means that $\Delta = Var(u_i)$ is diagonal. Then suppose $m_t \geq k_t + 1$ and let $A_t = (\psi_{1t}, \ldots, \psi_{m_t t})'$ for each $t$. For each $j, t$ let $A_1^{(j,t)}$ denote a $k_t \times k_t$ submatrix of $A_t$ and let $A_2^{(j,t)}$ denote $k_t$ rows from the matrix $A = diag(A_1, \ldots, A_T)$ excluding any rows corresponding to $A_1^{(j,t)}$. If $A_1^{(j,t)} \Phi_t A_2^{(j,t)\prime}$ is full rank for each $t$, where $\Phi_t$ represents the $k_t$ rows of $\Phi$ corresponding to $F_{it}$, then the conditions of Theorem 2.5 are satisfied and $A\Phi A'$ is identified.

Second, suppose that for $j = 1, \ldots, k_t$, $Cov(u_{ijt}, u_{ij't}) = 0$ for all $1 \leq j \leq k_t$, $1 \leq j' \leq m_t$ but $Cov(u_{ijt}, u_{ij't})$ is unrestricted for $k_t < j, j' \leq m_t$. In addition, let $Cov(u_{ijt}, u_{ij't'}) = 0$ for all $j, j'$ when $t \neq t'$. This model is relevant, for example, if measurements $k_t + 1, \ldots, m_t$ were all administered on the same day while measurements $1, \ldots, k_t$ were administered on different days. In that case, this covariance structure represents a “day of test” effect. Similarly, if there are $k_t$ measurements that do not depend on classroom instruction then this allows for classroom or teacher effects to cause dependence across $u_{ijt}$ for the remaining measurements. Let $A_t^{(1)} = (\psi_{1t}, \ldots, \psi_{k_t t})'$. The conditions of Theorem 2.5 can be verified in this model if for each $t$ there is a $k_t \times k$ matrix $A_2^{(t)}$ consisting of $k_t$ rows from $A$ that do not correspond to period $t$ such that $A_t^{(1)} \Phi_t A_2^{(t)\prime}$ is full rank.

Next, this model provides an example of how restrictions on the relationship among different elements of $\Phi$ can also be helpful in resolving the scale. Suppose, for example that $T = 2$. Let $\Phi_{11}, \Phi_{21}, \Phi_{22}$ denote the corresponding blocks of $\Phi$ in partitioned form. If $F_{i2} = TF_{i1}$ for a $k_2 \times k_1$ matrix $T$, then $\Phi_{21} = T\Phi_{11}$. Along with normalizations of some of the elements of the matrix $T$, this type of restriction can be combined with other scale normalizations to avoid the need for scale normalizations on the variances or factor loadings corresponding to $F_{i2}$. Indeed, suppose the first $k$ rows of $A$ are $A_{1,...,k} = diag(A_1, A_2)$ where $A_1$ and $A_2$ are $k_1 \times k_1$ and $k_2 \times k_2$ lower triangular matrices so that $AL$ is identified and

$$A_{1,...,k}L = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} L_{11} & 0 \\ TL_{11} & L_{22} \end{pmatrix}$$

To simplify the argument, suppose that $A_1$ or $\Phi_{11}$ is known through normalizations. Then $A_2T$ is identified. If, in addition, $A_2$ is diagonal, $k_2 \leq k_1$, and $diag(T)$ is known then $A_2$ and $T$ are identified. Indeed, in that case all of $A$ and $\Phi$ are identified. A version of this result has been used in a model of human capital formation (Agostinelli and Wiswall, 2016, 2017).

# D An Instrumental Variables Formulation

Some of the results derived in Sections 2.1 and 2.2 can also be derived from an instrumental variables model. If the factor loading matrix $A$ satisfies the row deletion property then, for any $j$, the system of equations (2) can be written as

$$\begin{aligned} M_{ij} &= a_j F_i + u_{ij} \\ M_i^{(1)} &= A_1 F_i + ui^{(1)} \\ M_i^{(2)} &= A_2 F_i + u_i^{(2)} \\ M_i^{(3)} &= A_3 F_i + u_i^{(3)} \end{aligned} \tag{D.1}$$

where $A_1$ and $A_2$ are the two distinct sets of $k$ rows from $A$ guaranteed by row deletion, and $A_3$ consists of the remaining rows. Since $A_1$ is nonsingular, $F_i$ can be solved for in the second equation and plugged into the first, producing the reduced form equation, $M_{ij} = a_j A_1^{-1} M_i^{(1)} + u_{ij} - A_1^{-1} u^{(1)}$. The reduced form parameter $a_j A_1^{-1}$ is identified if $u_i^{(2)}$ is uncorrelated with $u_{ij}$ and $u_i^{(1)}$ since $Cov(M_i^{(1)}, M_i^{(2)}) = A_1 \Phi A_2''$ is nonsingular. In that case, $a_j A_1^{-1} = Cov(M_{ij}, M_i^{(2)}) \left( Cov(M_i^{(1)}, M_i^{(2)}) \right)^{-1}$. This is the formula for a just-identified instrumental variables regression of $M_{ij}$ on $M_i^{(1)}$ using $M_i^{(2)}$ as instruments. The additional measurements, $M_i^{(3)}$, can serve as additional instruments but are not needed for identification.

This argument provides some intuition for the row deletion property and the other rank conditions used in the results in Sections 2.1 and 2.2. These conditions can be interpreted as a requirement that there is enough variation in the measurements that one set of measurements can be used to proxy the factors while there are enough remaining measurements to serve as instruments. It becomes clear then when the order condition, $m \geq 2k + 1$, is not sufficient. For example, if $k = 2$

and $m = 5$, identification fails if all but one $M_{ij}$ depend only on the first dimension of $F_i$ because in that case the measurements do not contain enough signal on the second factor to serve as both proxy and instrument.

The link between the factor analysis model and IV regression is discussed in Hagglund (1982), who suggests two stage least squares as a computationally simpler estimator compared to the maximum likelihood estimator of the factor model. Instrumental variables estimation of factor models has also been considered in Madansky (1964) and Pudney (1982). This link is useful because it demonstrates identification under even weaker conditions on $\Delta$. Indeed note that Theorem 2.1, for example, requires that $u_{ij}$ is also uncorrelated with $u_i^{(1)}$. Furthermore, the above IV argument does not even require $F_i$ to be uncorrelated with $u_i^{(1)}$. On the other hand, it is not always obvious how to cobble together identification results on $a_j A_1^{-1}$ and $Var(u_{ij} - A_1^{-1} u_i^{(1)})$ for different $j$ to obtain identification of $A$, $\Phi$, $\Delta$ or other parameters of interest.

Consider, for example, the $m = 4$, $k = 1$ model previously studied where $\delta_{12}$ and $\delta_{34}$ are nonzero. I argued above that $\Delta$ is not identified in this model. Note, however that $a_{21}/a_{11}$ is identified from an IV regression with $M_{i2}$ as the dependent variable, $M_{i1}$ as the endogenous regressor and $M_{i3}$ or $M_{i4}$ as the instrument. Likewise, $a_{41}/a_{31}$ is identified from an IV regression. This could incorrectly be interpreted as showing that the model is identified "up to scale". However, while $a_{11} = 1$ is a scale normalization, also imposing that $a_{31} = 1$ implies the substantive restriction that $a_{31} = a_{11}$. While the problem is easy to see in this simple model, analogous problems that may arise in more complex models when applying an IV strategy may be less obvious.

On the other hand, the identification strategy used in the previous results in this paper is useful in some cases where an IV strategy is not sufficient. Consider a model with $k = 1$ and $m^*$-dependent errors for $m^* = 1$. I show above that this model is identified if $m = 5$ and $a_{11}, a_{31}$, and $a_{51}$ are nonzero. It is not possible, however, to identify $a_{21}/a_{j1}$ for any $j \neq 2$ from an IV regression in this case. For any $j$, there must be a third measurement, $M_{ij'}$, such that $u_{ij}$ is uncorrelated with $u_{i2}$ and $u_{ij'}$. If $u_{ij}$ is uncorrelated with $u_{i2}$ then $j > 3$. For $u_{ij}$ to also be uncorrelated with $u_{ij'}$ for $j > 3$ would require $j' \geq 6$. Thus identification arguments for factor models based on the IV formulation may lead to stronger conditions than necessary.

# E Identification of the factor distribution

I first state a theorem due to theorem due to Ben-Moshe (2018). Consider the model

$$\tilde{M}_i = \tilde{A}\tilde{F}_i + \tilde{u}_i. \tag{E.1}$$

Let $M_{i1}$ denote $m_1 < m$ components of $M_i$ and let $M_{i2}$ denote the remaining $m - m_1$ components. Let $\tilde{A}_1$ and $\tilde{A}_2$ denote the corresponding partition of the matrix $\tilde{A}$ and let $\tilde{u}_{i1}$ and $\tilde{u}_{i2}$ denote the corresponding partition of the vector $\tilde{u}_i$.

**Theorem E.1.** ***(Ben-Moshe, 2018)*** *Suppose that*

*(i)* $\tilde{F}_i$ *and* $\tilde{u}_i$ *are independent and* $E(\tilde{u}_{i1} \mid \tilde{u}_{i2}) = 0$*,*

*(ii)* $rank(\tilde{A}_1) = rank(\tilde{A}_2) = k$*, and*

*(iii)* $|E(\tilde{F}_i)| < \infty$ *and* $E(\exp(\iota \tilde{F}_i' v)) \neq 0$ *for all* $v \in \mathbb{R}^k$*, where* $\iota = \sqrt{-1}$*.*

*Then, if* $\tilde{A}$ *is a known matrix the distribution of* $\tilde{F}_i$ *is identified.*

Under the conditions of Theorem 2.1 there are natural partitions of $M_i$, $A$, and $u_i$. Indeed, both conditions (b) and (c) require two distinct sets of $k$ components of the vector $M_i$ where the corresponding rows of $A$ consist of two rank $k$ matrices and $Cov(u_{i1}, u_{i2}) = 0$. Consider any such partition and let $\tilde{M}_{i1}$ and $\tilde{M}_{i2}$ denote the two length $k$ vectors of components of $M_i$ and let $\tilde{M}_i = (\tilde{M}_{i1}', \tilde{M}_{i2}')'$. Let $A_1$ and $A_2$ denote the corresponding rows of $A$, let $\tilde{u}_{i1}$ and $\tilde{u}_{i2}$ denote the corresponding components of $u_i$ and let $\tilde{u}_i = (\tilde{u}_{i1}', \tilde{u}_{i2}')'$. Suppose that the assumption that $Cov(u_{i1}, u_{i2}) = 0$ and the maintained assumption that $Cov(F_i, \tilde{u}_i) = 0$ are strengthened to mean independence and independence, as stated in condition (i) of the above theorem. Further, suppose that $|E(F_i)| < \infty$ and $E(\exp(\iota F_i' v)) \neq 0$ for all $v \in \mathbb{R}^k$.

The above theorem still cannot be applied because, under the conditions of Theorem 2.1, $A$ is not identified. However, identification of $A \Phi A'$ implies identification of $A_2 A_1^{-1}$. Therefore, define $\tilde{F}_i = A_1 F_i$ and $\tilde{A}_1 = I_k$ and $\tilde{A}_2 = A_2 A_1^{-1}$. Then the conditions of the above theorem are satisfied and, hence, the distribution of $A_1 F_i$ is identified.

# F Details of simulation

In the simulations reported in Section 3., the data was generated by the model of equation (13) with $\varepsilon_{ij} \sim_{iid} N(0,1)$ for each $j$. In the first model, $d_X = 1$ and $F_i = (X_i, \theta_{i1}, \theta_{i2})' \sim_{iid} N(0, \Phi)$ where

$$\Phi = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix}$$

The coefficient values used were

$$\beta = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \alpha = \begin{pmatrix} 1 & 0 \\ 1 & \alpha_{22} \\ 0 & 1 \\ 0.5 & 1 \\ 1 & 0.5 \end{pmatrix}$$

If $\alpha_{22} = 0$ then $\beta$ is not identified. Two values of $\alpha_{22}$ were used – $0.2$ and $0.5$.

Let $Y_1 = (Y_{11}, \ldots, Y_{n1})'$ and similarly stack the other variables into $n \times 1$ vectors $Y_2, \ldots, Y_5, X$. The coefficient $\beta_3$ can be estimated via two stage least squares. Let

$$\hat{\psi}^{2sls} = (\tilde{X}'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{X})^{-1}\tilde{X}'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{Y}$$

where $\tilde{Y} = Y_3'$, $\tilde{X} = (\iota_n, X, Y_1, Y_2)$, and $\tilde{Z} = (\iota_n, X, Y_4, Y_5)$ and $\iota_n$ denotes a vector of $n$ ones.Then $\hat{\beta}_3^{2sls}$ is the second element in the coefficient vector $\hat{\psi}^{2sls}$.

This model can be estimated via maximum likelihood if sufficient normalizations are imposed. Let $\gamma$ denote the unrestricted parameters. Then let

$$\Sigma(\gamma) = A(\gamma)\Phi(\gamma)A(\gamma)' + \Delta(\gamma)$$

where

$$A(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{21} & \alpha_{22} \\ \beta_3 & 0 & 1 \\ \beta_4 & \alpha_{41} & \alpha_{42} \\ \beta_5 & \alpha_{51} & \alpha_{52} \end{pmatrix}, \qquad \Phi(\gamma) = \begin{pmatrix} \phi_x & \phi_{x\theta_1} & \phi_{x\theta_2} \\ \phi_{x\theta_1} & \phi_{\theta_1} & \phi_{\theta_1\theta_2} \\ \phi_{x\theta_1} & \phi_{\theta_1\theta_2} & \phi_{\theta_2} \end{pmatrix}$$

and $\Delta(\gamma) = diag(0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$. The log likelihood function is given by $\ell(\gamma) \propto -\ln(det(\Sigma(\gamma))) - trace(\Sigma(\gamma)^{-1}\hat{\Sigma})$ where $\hat{\Sigma}$ is the sample covariance matrix of $(X_i, Y_{i1}, \ldots, Y_{i5})'$. Then $\hat{\beta}_3^{mle}$ is the corresponding component of $\hat{\gamma} = \arg\max_\gamma \ell(\gamma)$.

In the second model simulated, $X_i = (X_{i1}, X_{i2}, X_{i3})'$ and $F_i = (X_i', \theta_{i1}, \theta_{i2})' \sim_{iid} N(0, \Phi)$

where

$$\Phi = \begin{pmatrix} 1 & r & r & 0 & 0.5 \\ r & 1 & r & 0 & 0.5 \\ r & r & 1 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0.5 & 0.5 & 0 & 1 \end{pmatrix}$$

and $r$ was specified in the first case as $0.8$ and in the second case as $0.999$. The specification for $\alpha$ was the same as in the first model and the coefficient vector for $X$ was

$$\beta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The maximum likelihood estimates were obtained by the same approach where now $\hat{\Sigma}$ is the sample covariance of $(X_{i1}, X_{i2}, X_{i3}, Y_{i1}, \ldots, Y_{i5})'$ and

$$A(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \alpha_{21} & \alpha_{22} \\ \beta_3 & 0 & 0 & 0 & 1 \\ 0 & \beta_4 & 0 & \alpha_{41} & \alpha_{42} \\ 0 & 0 & \beta_5 & \alpha_{51} & \alpha_{52} \end{pmatrix}$$

# References

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.

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

HAGGLUND, G. (1982): "Factor analysis by instrumental variables methods," *Psychometrika*, 47, 209–222.

MADANSKY, A. (1964): "Instrumental variables in factor analysis," *Psychometrika*, 29, 105–113.

PUDNEY, S. E. (1982): "Estimating Latent Variable Systems When Specification is Uncertain: Generalized Component Analysis and the Eliminant Method," *Journal of the American Statistical Association*, 77, 883–889.