Identification and Estimation of Coefficients in Dependent Factor Models

Dan Ben-Moshe*

The Hebrew University of Jerusalem

March 2016

Abstract

We identify and estimate coefficients in a linear factor model in which factors are allowed to be arbitrarily dependent. Given a statistical dependence structure on the unobservables, rank conditions on the matrix representation and restrictions on coefficients, the unknown coefficients are identified under nonnormality assumptions. The identification strategy transforms the system of equations into a functional equation using log characteristic functions. By ruling out polynomial functions which correspond to normal distributions, we show that the unknown coefficients uniquely solve the functional equation. Identification is illustrated in the classical errors-in-variables model with arbitrarily dependent unobserved regressors and in a panel data moving average process in which subsets of the shocks are allowed to be arbitrarily dependent. We propose an extremum estimator based on second-order partial derivatives of the empirical log characteristic function that is root-n consistent and asymptotically normal. In Monte Carlo simulations the estimator produces similar results to a GMM estimator based on higher-order moments, and is more robust to different amounts of measurement error and distributional choices of unobservables.

JEL Classification: C14, C38

Keywords: *Linear factor models, multidimensional and statistically dependent unobservables, log characteristic functions, errors-in-variables, moving averages*

---

*Email: danbm@huji.ac.il. This paper is roughly based on the second chapter of my PhD thesis. I am grateful to Rosa Matzkin and Jinyong Hahn for their generous support, advice, and guidance. I benefited from comments at Hebrew University of Jerusalem, discussions with Stefan Hoderlein, Max Kasy, Saul Lach, Arthur Lewbel, Áureo de Paula, Daniel Wilhelm and Martin Weidner, and feedback from Xavier D'Haultfœuille. The support of a grant from the Maurice Falk Institute for Economic Research in Israel is gratefully acknowledged.

1 Introduction

In this paper, we consider identification and estimation in the linear factor model,

$$\boldsymbol{Y} = \sum_{j=1}^{q} A_j \boldsymbol{F}_j + \boldsymbol{U} = A\boldsymbol{F} + \boldsymbol{U}, \tag{1}$$

where $\boldsymbol{Y}$ is a vector of observed outcome variables, $A = (A_1 : \cdots : A_q)$ is a matrix of coefficients, $\boldsymbol{F}' = (\boldsymbol{F}_1', \ldots, \boldsymbol{F}_q')$ is a vector of unobserved factors, and $\boldsymbol{U}$ is a vector of unobserved errors.[1] The contribution of this paper is to allow each factor to have arbitrarily dependent components and the goal is to identify and estimate the unknown coefficients in the matrix A.

Principal component analysis (PCA) assumes that all the unobserved factor components are uncorrelated, and uses covariances to identify A up to scale normalizations and orthogonal transformations (e.g., Anderson and Rubin, 1956). This literature is often concerned with dimension reduction rather than identification. Estimation in this literature transforms the observed variables into linearly uncorrelated variables called principal components with each principal component having the largest possible variance (see, e.g., Jolliffe, 2002). Independent component analysis (ICA) assumes that the unobserved variables are mutually independent, and uses information from higher-order moments to identify A up to scale normalizations (e.g., Comon, 1994). Many estimation techniques in this literature can be viewed as maximizing a measure of nonnormality such as kurtosis or negentropy (see, e.g., Hyvärinen et al., 2004). The present paper generalizes these results from assuming mutually independent unobservables to allowing subsets of the unobserved variables to be arbitrarily dependent, and identifies A using characteristic functions. Estimation uses the empirical log characteristic function, which carries with it information about all moments and thus we believe should be more efficient than estimators that rely on the identifying power of skewness for example through third-order moments (e.g., Erickson et al., 2014; Lewbel, 1997), and can also be interpreted as maximizing a measure of nonnormality.

Many economic problems use factor models as part of the analysis. In microeconometrics, panel data models with measurement error and latent variables can be represented by factor models (see, e.g., Aigner et al., 1984), permanent and transitory income shocks can take factor model structure in the earnings dynamics literature (e.g., Bonhomme and Robin, 2010; Meghir and Pistaferri, 2004), unobserved skills can be modeled within a factor structure when analyzing human capital accumulation (e.g., Carneiro et al., 2003), and demand systems can be represented by factor models in consumer theory (e.g., Gorman, 1981; Lewbel, 1991). In macroeconometrics, Stock and Watson (1999) forecast inflation using factor models and Forni and Reichlin (1998) model cyclical variations in an economy by industry and country shocks within a factor model framework. In finance, Fama

[1] $(A_1 : \cdots : A_q)$ is the augmented matrix that appends the columns of $A_1, \ldots, A_q$.

and French's (1993) model of asset returns and Ross's (1976) arbitrage pricing theory describe asset returns in a factor structure.

The assumption of independent unobservables in structural factor models is often done for convenience rather than because the underlying unobserved variables are really believed to be unrelated to each other. In PCA, on the other hand, the principal components are uncorrelated by construction and, despite ex-post interpretations, in general do not represent any underlying independent structural objects. In the literature on human capital accumulation, for example, Attanasio (2015) argues that latent skills are complicated multidimensional objects that are related to one another in nontrivial ways. Cunha and Heckman (2008) and Cunha et al. (2010) use a factor model structure to model the relationship between observed measurements of skills and the underlying latent skills. In order to deal with the dependence of some of the unobservables they use repeated measurements, instrumental variables, exclusion restrictions and assumptions of uncorrelatedness, mean independence or independence. In our paper we allow for arbitrarily dependent nonparametrically distributed unobservables but as more unobservables are allowed to be dependent we will need to impose more restrictions on the matrix representation for identification. In the spirit of Carneiro et al. (2003), who impose that the upper triangle of the matrix representation is zero, suppose, for example, that there are three latent skills: cognitive, non-cognitive and motor skills. We do not want to impose distributional or independence restrictions on these unobserved skills. We also have some outcome variables such as test scores and physical characteristics that are linear combinations of these nonparametric arbitrarily dependent skills and measurement errors. Further, we may believe that some of these outcome variables, such as IQ test score, may exclude one of these skills and so we can decide to exclude motor skills, for example, from this equation.

We show identification under four assumptions. First, we assume a given statistical dependence structure on the unobservables, which potentially allows for all the components of $\boldsymbol{F}$ to be arbitrarily dependent. Second, we assume that the matrix representation satisfies certain rank conditions. Without this assumption the system can be reduced to have fewer unobservables and still satisfy the same rank conditions. Third, we impose restrictions on the coefficients. This can include scale normalizations and exclusion restrictions. Without this assumption A is identified up to nonsingular matrix multiplication. The assumption is the same one that is needed in the SVAR literature for identification (e.g., Sims, 1980). Lastly, we assume that the unobservables satisfy nonnormality assumptions. Nonnormality assumptions are common in factor models because the models, in general—and in particular the ones considered in this paper—are not identified from just the first two moments of the observables and require higher-order moments to provide identifying information.[2]

The identification strategy transforms Equation (1) into a functional equation by taking its

[2]With stronger rank conditions and more restrictions on the coefficients, Williams (2015) considers identification in factor models with dependent unobservables using second-order moments.

log characteristic function and simplifying it using the dependence structure. The rank conditions are then used to solve the functional equation. Finally, by ruling out polynomial functions by the nonnormality assumptions, and using the assumed matrix restrictions, we show that the unknown parameters uniquely solve the functional equation. The intuition for identification, as in most factor models, is that some of the unobservables are common to multiple equations while others are not and this, along with some independence, enables us to disentangle the effects of one variable from another.

There is a tradeoff between the number of restrictions on the matrix A and the amount of dependence on the unobservables. On the one extreme is the ICA model that has mutually independent unobservables and the rank conditions that no two columns are multiples of each other. In this case, assuming that none of the factors are normal identifies A up to scale normalizations. On the other extreme is the model with arbitrarily dependent $\boldsymbol{F}$ and a matrix A with full rank. In this case, assuming that no nonzero linear combinations of the components of $\boldsymbol{F}$ is normal identifies A up to a nonsingular matrix multiplication. These types of models with many restrictions on A, however, do appear in economics and are becoming increasingly popular. For example, in high dimensional models, a standard assumption is sparsity of the model representation (see, e.g., Fan et al., 2011, and references therein), which in factor models means that many factors are excluded from many equations (see, e.g., Zou et al., 2006). The complexity in our model, however, is not due to high dimension but rather the dependent unobservables. The exclusion restrictions allow the effects of the different variables to be isolated. As an extension, we consider the model in Equation (1) with weaker dependence and rank assumptions and show that the log characteristic function of the unobserved factors solves difference equations. If functional solutions to these equations are ruled out then A can be identified.

Probably the most well-known example of a model with many exclusion restrictions is the classical errors-in-variables model, where all the regressors are unobserved and measured with error. In this model each unobserved regressor is excluded from all but two equations; the main outcome equation and its observed measurement which is equal to the unobserved regressor plus measurement error. In most applications the unobserved regressors are statistically dependent; otherwise if a regressor is independent it may be possible to just include it as part of the error without worrying about dependence between the regressors and errors. Hence, in our setup we allow for all the unobserved regressors to be arbitrarily dependent and identify A as long as no linear combination of the unobserved regressors is normal.

For estimation, we observe many independent identically distributed vectors drawn from $\boldsymbol{Y}$ but assume that the dimensions of $\boldsymbol{Y}$ and $\boldsymbol{F}$ are fixed and known.[3] Although the identification procedure we use is not constructive (i.e., identification does not produce a mathematical expression

[3]There is a growing literature where the dimension of $\boldsymbol{Y}$ and number of draws from $\boldsymbol{Y}$ go to infinity (e.g., Bai and Ng, 2013) and / or the number of factors is unknown or large (e.g., Bai and Ng, 2002; Moon and Weidner, 2015).

for the unknown parameters), we will nevertheless be able to mimic the steps in identification to construct an extremum estimator based on second-order partial derivatives of the empirical log characteristic function, which contains information from second and higher order moments. We show that the extremum estimator is root-n consistent and asymptotically normal. Finite sample simulations suggest that the estimator could be a possible alternative to GMM estimators based on higher-order moments. The extremum and GMM estimators produce similar coefficient estimates with the extremum estimator more robust to different distributional choices of factors and different amounts of measurement error.

We apply our estimator to data on firm investment decisions. According to Tobin's q theory, these decisions should only depend on marginal returns to capital but many empirical studies suggest that financial constraints also play an important role in explaining investment decisions. Erickson and Whited (2000) provide evidence that a linear regression with classical measurement error in marginal returns to capital, corroborates Tobin's q theory. We show that the results are also sensitive to measurement error in financial constraints.

In the following section we present the main identification results, which include identification in the models with mutually independent and arbitrarily dependent factors. This is followed by two examples showing identification in the classical errors-in-variables model and a dynamic moving average process. We then introduce the estimator and show that it is root-n consistent and asymptotically normal. In finite sample simulations in the errors-in-variables model we compare our estimator to estimators that use higher-order moments and the ordinary least squares estimator. Finally, we apply the estimator to data on firm investment decisions and conclude.

2 Identification

Identification of the model in Equation (1) will rely on four assumptions. The first assumption describes the statistical dependence structure of the unobservables.

Assumption 2.1. *Let $\boldsymbol{F}_j \in \mathbb{R}^{p_j}$, for $j = 1, \ldots, q$, and $U_t \in \mathbb{R}$, for $t = 1, \ldots, T$. Let $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_q$, $U_1, \ldots, U_T$ be mutually independent.*

This assumption allows each vector $\boldsymbol{F}_j$ to have arbitrarily dependent components $F_{j1}, \ldots, F_{jp_j}$ but restricts the vectors to be mutually independent of each other.

The next assumption imposes rank conditions on the matrix representation of Equation (1).

Assumption 2.2. *Let A_j and $\widetilde{A}_j$ be $T \times p_j$ observationally equivalent matrices, for $j = 1, \ldots, q$.*

(i) $Rank(A_j : A_{j'}) = p_j + p_{j'}$, for all $j' \neq j$;

(ii) $Rank(A_j : \widetilde{A}_{j'}) = p_j + p_{j'}$, for all $j' \neq j$; and

(iii) $Rank(A_j : \boldsymbol{e}_t) = p_j + 1$ for $t = 1, \ldots, T$.[4]

[4]We adopt the standard notation $\boldsymbol{e}_t' = (0, \ldots, 0, 1, 0, \ldots, 0)$ for a column vector with a 1 in the t-th coordinate

The assumption serves two purposes. First, the assumption implies that $\text{Rank}(A_j) = p_j$ so that $A_j'A_j$ is invertible. Second, Assumptions 2.2(i) and (ii) maintain the distinction between factors and Assumption 2.2(iii) maintains the distinction between factors and errors and if Assumption 2.2(i) or Assumption 2.2(iii) fails then one of the columns can be replaces by a linear combination of the other columns. These assumptions are weaker than the standard assumption in factor models that A has full rank. Assumption 2.2(ii) is strong. However, in the case with arbitrarily dependent $\boldsymbol{F}$, i.e., $q = 1$, the assumption holds trivially (see Sections 2.2 and 3.1). In Section 2.1 we consider mutually independent $\boldsymbol{F}$, i.e., $p_j = 1$, in which case Assumption 2.2(ii) holds because at most one of the observationally equivalent columns can violate the assumption so if $\text{Rank}(A_j : \widetilde{A}_j) = 1$ then $\text{Rank}(A_j : \widetilde{A}_j') = 2$ (more generally if $\text{Rank}(A_j : \widetilde{A}_j) = p_j$ then Assumption 2.2(ii) will hold by Assumption 2.2(i)). In Section 3.2, Assumption 2.2(ii) will hold for all $A_j \in \mathcal{A}_j$ and $A_{j'} \in \mathcal{A}_{j'}$. In Section 4 we remove Assumption 2.2(ii), and weaken Assumptions 2.1 and 2.2(i), at the expense of imposing a high-level assumption for identification.

The following theorem provides necessary conditions that all observationally equivalent models must satisfy.

Theorem 2.1. *Let $\varphi_{\boldsymbol{F}_j}(\boldsymbol{u}) = \ln E[\exp(\mathrm{i}\boldsymbol{u}'\boldsymbol{F}_j)]$ be the log characteristic function of the random vector $\boldsymbol{F}_j \in \mathbb{R}^{p_j}$, where $\boldsymbol{u} \in \mathbb{R}^{p_j}$, $\mathrm{i} = \sqrt{-1}$ and $\ln(\cdot)$ is the natural log restricted to its principal branch. Let $P_k(\boldsymbol{u})$ be a polynomial of degree k in a neighborhood around the origin.*

Suppose that Equation (1) and Assumptions 2.1 and 2.2 hold. Let $(\widetilde{A}, \widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}})$ be observationally equivalent to $(A, \boldsymbol{F}, \boldsymbol{U})$. Then,

(i) $\varphi_{A_j\boldsymbol{F}_j}(\boldsymbol{s}) - \varphi_{\widetilde{A}_j\widetilde{\boldsymbol{F}}_j}(\boldsymbol{s}) = \sum_{j'\neq j}(\varphi_{A_{j'}\boldsymbol{F}_{j'}}(\boldsymbol{s}) - \varphi_{\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}}(\boldsymbol{s})) + \sum_{t=1}^{T}(\varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s})) = P_{2q-2T-2}(\boldsymbol{s})$;

(ii) $\varphi_{\widetilde{U}_t}(s_t) = \varphi_{U_t}(s_t) + P_{2q-2T-2}(s_t)$.

Let H and $\widetilde{H}$ be right pseudoinverses of A_j and $\widetilde{A}_j'$ respectively and let $\widetilde{\boldsymbol{\Pi}}$ be a basis for $\ker(\widetilde{A}_j')$ so that $A_j'H = I_{p_j}$, $\widetilde{A}_j'\widetilde{H} = I_{p_j}$ and $\widetilde{A}_j'\widetilde{\boldsymbol{\Pi}} = \boldsymbol{0}$. Then,

(iii) $\varphi_{\widetilde{\boldsymbol{F}}_j}(\boldsymbol{u}) = \varphi_{\widetilde{H}'A_j\boldsymbol{F}_j}(\boldsymbol{u}) + P_{2q-2T-2}(\boldsymbol{u})$;

(iv) $\varphi_{\boldsymbol{F}_j}(\boldsymbol{u}) = \varphi_{\boldsymbol{F}_j}(A_j'\widetilde{H}\widetilde{A}_j'H\boldsymbol{u}) + P_{2q-2T-2}(\boldsymbol{u})$; *and*

(v) $\widetilde{\boldsymbol{\Pi}}'A_j\boldsymbol{F}_j$ *is jointly normal.*

We provide some intuition for the proof, leaving the details to Section A.4 of the appendix. First, using the independence conditions from Assumption 2.1, the log characteristic function of Equation (1) is a functional equation that is the sum of the log characteristic functions of the factors. Then, using the rank conditions of Assumption 2.2, there is a linear combination of the outcome variables that does not depend on one of the factors and by taking a first difference the log characteristic function of this factor can be eliminated from the equation. Recursively choosing linear combinations of outcome variables and taking first differences results in a difference equation

and 0's elsewhere and with the dimension that is needed to make sense of the expression.

only in $\varphi_{\boldsymbol{F}_j}$. Finally, again using the rank conditions, the only solutions to this difference equation are polynomials.

Theorem 2.1 imposes functional restrictions on the relationship between the log characteristic functions of the underlying model and any observationally equivalent counterpart. The theorem can be used to construct an observationally equivalent model for some $\widetilde{A}_j \neq A_j$ by $\varphi_{\widetilde{U}_t}(s_t) = \varphi_{U_t}(s_t) + P_{2q-2T-2}(s_t)$ and $\varphi_{\widetilde{\boldsymbol{F}}_j}(\boldsymbol{u}) = \varphi_{\widetilde{H}' A_j \boldsymbol{F}_j}(\boldsymbol{u}) + P_{2q-2T-2}(\boldsymbol{u})$. The polynomial terms shift moments of the underlying random vectors, which can be seen by taking partial derivatives and setting the arguments equal to zero. As long as U_t and $\widetilde{H}' A_j \boldsymbol{F}_j$ have positive definite variances, i.e. U_t is not degenerate and no linear combination of $\widetilde{H}' A_j \boldsymbol{F}_j$ is degenerate, these polynomials can always be chosen so that the log characteristic functions are well-defined. Specifically, it is always possible to choose σ^2 and Σ and small enough δ so that $\varphi_{\widetilde{U}_t}(s_t) = \varphi_{U_t}(s_t) + \delta\sigma^2 s_t^2$ and $\varphi_{\widetilde{\boldsymbol{F}}_j}(\boldsymbol{u}) = \varphi_{\widetilde{H}' A_j \boldsymbol{F}_j}(\boldsymbol{u}) + \delta \boldsymbol{u}' \Sigma \boldsymbol{u}$ are well-defined log characteristic functions, where $\delta\sigma^2$ and $\delta \boldsymbol{u}' \Sigma \boldsymbol{u}$ are shifts in variances and covariances (and when $\delta < 0$, and Σ is semi-positive definite, these quadratic polynomials are independent normal terms). It then remains to check if the choices $\varphi_{\widetilde{U}_t}$, for $t = 1, \ldots, T$, and $\varphi_{\widetilde{\boldsymbol{F}}_j}$, for $j = 1, \ldots, q$, satisfy the first equality in condition (i).

Conditions (iv) and (v) place strong restrictions on the distribution of $\boldsymbol{F}_j$, when $A_j' \widetilde{H} \widetilde{A}_j' H \neq I_{p_j}$ or $A_j' \widetilde{\boldsymbol{\Pi}} \neq \boldsymbol{0}$, which are not satisfied by almost all of the well-known parametric distributions, with the notable exception of the normal distribution. Condition (iv) states that the difference in $\varphi_{\boldsymbol{F}_j}$ after a linear transformation of inputs is a polynomial. Condition (v) states that there are linear combinations of $\boldsymbol{F}_j$ that must be jointly normal.

The model is not identified even with the above functional restrictions that limit the set of observationally equivalent models. First, if the unobservables are jointly normal then all the conditions of Theorem 2.1 hold and in general there are infinitely many observationally equivalent models. Second, multiplying the factor $\boldsymbol{F}_j$ by a constant and dividing each element of A_j by the same constant results in an observationally equivalent model. In fact, it is easy to construct observationally equivalent models by $\widetilde{\boldsymbol{U}} = \boldsymbol{U}$ and any $(\widetilde{A}_j, \widetilde{\boldsymbol{F}}_j) = (A_j Q, Q^{-1} \boldsymbol{F}_j)$, where Q is a nonsingular matrix. This nonidentification by nonsingular matrix multiplication is the same in SVAR models.

The next assumption places restrictions on A_j.

Assumption 2.3. *Let $A_j \in \mathcal{A}_j$. There is no $\widetilde{A}_j \in \mathcal{A}_j$ such that A_j and $\widetilde{A}_j$ are right equivalent, where A_j and $\widetilde{A}_j$ are right equivalent if and only if there exists a $p_j \times p_j$ nonsingular matrix Q such that $\widetilde{A}_j = A_j Q$ (see, e.g., Robbin, 1995).*

The matrix A_j lies in the space $\mathcal{A}_j$, which could, for example, only include matrices with some specific elements equal to 1 (scale normalizations) or 0 (exclusion restrictions). While scale normalizations are necessary for identification and usually considered innocuous, exclusion restrictions of variables from some equations or equality of some coefficients are model specific and rely on economically relevant information about the problem being studied (see, e.g., Jöreskog, 1970). In

Section 3 we discuss the classical errors-in-variables model with many exclusion restrictions and a panel data moving average process which has exclusion restrictions and parameter constraints.

Assumption 2.3 states that if A_j is right multiplied by a nonsingular matrix then the result is no longer in $\mathcal{A}_j$. If the assumption does not hold then there is a nonsingular matrix $Q \neq I$ such that $\widetilde{A}_j = A_j Q \in \mathcal{A}_j$, $\widetilde{\boldsymbol{F}}_j = Q^{-1}\boldsymbol{F}_j$ and $\widetilde{A}_j \widetilde{\boldsymbol{F}}_j = A_j Q Q^{-1}\boldsymbol{F}_j = A_j \boldsymbol{F}_j$. Hence, $\widetilde{A}_j \neq A_j$ is observationally equivalent to A_j.

Without Assumption 2.3, A_j is identified up to right equivalent matrices. On the one extreme if $p_j = 1$, i.e., A_j is a column vector, and the unobservables are mutually independent then A_j is identified up to multiplication by a constant, i.e., a scale normalization (see Section 2.1). On the other extreme if $q = 1$, $T = p + 1$, i.e., A is a $(p+1) \times p$ matrix, and there is only one arbitrarily dependent factor $\boldsymbol{F} \in \mathbb{R}^p$ then A is identified up to multiplication by a p^2 nonsingular matrix or at least p^2 restrictions are needed for identification (see Section 2.2).

One of the implications of Assumption 2.3, which we use in the proof of the following lemma, is that A_j has no right equivalent matrix if and only if A'_j is identified by its null-space, i.e., if $\ker(\widetilde{A}'_j) = \ker(A'_j)$ then $\widetilde{A}'_j = A'_j$ (see Robbin, 2001).[5]

Lemma 2.1. *Suppose that Assumption 2.3 holds and Rank*$(A_j) = p_j$. *Then* $A_j \in \mathcal{A}_j$ *is the unique matrix that solves* $A'_j \boldsymbol{\Pi} = \boldsymbol{0}$ *where* $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1 : \ldots : \boldsymbol{\pi}_{T-p_j})$ *and* $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p_j}\}$ *is a basis of ker*(A'_j) *or equivalently* A'_j *uniquely solves* $A'_j \boldsymbol{\pi} = \boldsymbol{0}$ *for all* $\boldsymbol{\pi} \in$ *ker*(A'_j). *As a result at most* $p_j(T - p_j)$ *unknown parameters can be identified from this system of equations.*

For identification, we will show that $A'_j \widetilde{\boldsymbol{\pi}} = \boldsymbol{0}$ for all $\widetilde{\boldsymbol{\pi}} \in \ker(\widetilde{A}'_j)$, and by the first part of the lemma $\widetilde{A}'_j = A'_j$. For estimation of A_j, we will first estimate $\boldsymbol{\Pi} \in \text{Basis}(\ker(A'_j))$ by $\widehat{\boldsymbol{\Pi}}$ and show that it is root-n consistent and asymptotically normal, and then using the first part of the lemma we solve for A_j in terms of $\boldsymbol{\Pi}$ and use the delta method to show that the estimator $\widehat{A}_j$ that solves $\widehat{A}'_j \widehat{\boldsymbol{\Pi}} = \boldsymbol{0}$ is also root-n consistent and asymptotically normal.

The second part of the lemma bounds the number of parameters that can be identified and implies that the number of restrictions needed for identification of all the unknown parameters in A_j, which has Tp_j elements, must be at least $Tp_j - p_j(T - p_j) = p_j^2$. For example, if $p_j = 1$ then A_j is identified by a scale normalization. If $p_j = 2$ then A_j is identified by one scale normalization in each column and two exclusion restrictions.

The next assumption places distributional restrictions on the unobservables. The assumption will rule out well-defined log characteristic functions that satisfy the conditions in Theorem 2.1 so that $\widetilde{A}_j = A_j$.

Assumption 2.4.

(i) No nonzero linear combination of the components of $\boldsymbol{F}_j$ *is degenerate; and*

(ii) (a) No nonzero linear combination of the components of $\boldsymbol{F}_j$ *is normal; or*

[5] The kernel or null space of A'_j is $\ker(A'_j) = \{\boldsymbol{v} \in \mathbb{R}^T : A'_j \boldsymbol{v} = 0\}$.

(b) $\varphi_{A_{j'}\boldsymbol{F}_{j'}} - \varphi_{\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}}$*, for all* $j' \neq j$*, and* $\varphi_{U_t} - \varphi_{\widetilde{U}_t}$*, for* $t = 1, \ldots, T$*, do not differ by a polynomial of degree 2 or more in a neighborhood around the origin.*

Assumption 2.4(i) is a common assumption in linear models, e.g., no multicollinearity in linear regressions, and basically means that all variables carry unique information. Assumption 2.4(ii)(a) is a nonnormality assumption. Nonnormality assumptions are common in linear factor models (see, e.g., Hyvärinen et al., 2004; Rao, 1966; Stone, 2004) because if the factors are jointly normal then the model is in general not identified, and often there are infinitely many parameters that are observationally equivalent and can be constructed for any $\widetilde{A}_j \in \mathcal{A}_j$ using Theorem 2.1(iii). The idea behind the assumption is that the means and covariances of the observed variables are not enough for identification. Note, however, that for identification, we do not assume existence of any moments, and only that characteristic functions are continuous in a neighborhood of the origin, which is guaranteed by definition.

Assumption 2.4(ii)(a) could instead be stated as the log characteristic function of all nonzero linear combinations of $\boldsymbol{F}_j$ cannot be polynomials of degree 2 or more in a neighborhood around the origin because a log characteristic function that is a polynomial must correspond to a normal random variable (see Theorem A.1). However, Assumption 2.4(ii)(b) cannot be stated as $A_{j'}\boldsymbol{F}_{j'}$ and $\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}$ cannot differ by a normal term (see Section A.3 of the appendix for counterexamples).

The main identification result, which follows, uses Assumption 2.4(ii)(a) to identify A_j without identifying the rest of the model. Assumption 2.4(ii)(b), on the other hand, will be used to first identify $\sum_{j' \neq j} A_{j'}\boldsymbol{F}_{j'} + \boldsymbol{U}$ and then A_j, and allows nonzero linear combinations of the components of $\boldsymbol{F}_j$ to be normal.

Theorem 2.2. *Suppose that Equation* (1) *and Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then* A_j *is identified.*

We conclude this section by identifying the factor model first with mutually independent unobservables and then with arbitrarily dependent $\boldsymbol{F}$.

2.1 Mutually independent unobservables

Consider the factor model from Equation (1) with the ICA assumption that the factors and errors are mutually independent one-dimensional random variables (i.e., $p_1 = \ldots = p_q = 1$). It is well-known that if no component of $\boldsymbol{F}$ is normal then A is uniquely determined up to scale normalizations (see, e.g., Comon, 1994). We confirm this result using Theorem 2.2 with an assumption that is weaker than $\text{Rank}(A) = q$.

Corollary 2.1. *Suppose that*

$$\boldsymbol{Y} = \sum_{j=1}^{q} A_j F_j + \boldsymbol{U},$$

where A_j is a $T \times 1$ column vector and $F_1, \ldots, F_q, U_1, \ldots, U_T$ are mutually independent random variables. Assume that no column A_j is a multiple of another column $A_{j'}$ and that each column has at least two nonzero elements.

If (i) F_j is not degenerate and (ii)(a) F_j is not normal or (ii)(b) $\varphi_{F_{j'}} - \varphi_{\widetilde{F}_{j'}}$, for all $j' \neq j$, and $\varphi_{U_t} - \varphi_{\widetilde{U}_t}$, for $t = 1, \ldots, T$, do not differ by a polynomial of degree 2 or more in a neighborhood around the origin, then A_j is identified.

The assumption that columns of A are not multiples of other columns is to distinguish between factors and ensures that Assumptions 2.2(i) and 2.2(ii) hold. The assumption that A_j has at least two nonzero elements is to distinguish the factor from errors and ensures that Assumption 2.2(iii) holds. These assumptions, as in Rao (1966), are weaker than the more common assumption $\text{Rank}(A) = q$.

2.2 Arbitrarily dependent $\boldsymbol{F}$

Consider the factor model from Equation (1) where $\boldsymbol{F}' = (F_1, \ldots, F_p)$ is allowed to be arbitrarily dependent (i.e., $q = 1$).

Corollary 2.2. *Suppose that*

$$\boldsymbol{Y} = A\boldsymbol{F} + \boldsymbol{U},$$

where A is a $T \times p$ matrix and $\boldsymbol{F}$, $U_1, \ldots, U_T$ are mutually independent. Let $A \in \mathcal{A}$ and assume that A has no right equivalent matrix in $\mathcal{A}$. Assume Rank$(A, \boldsymbol{e}_t) = p + 1$, for $t = 1, \ldots, T$.

If (i) no nonzero linear combination of the components of $\boldsymbol{F}$ is degenerate and (ii)(a) no nonzero linear combination of the components of $\boldsymbol{F}$ is normal or (ii)(b) $\varphi_{U_t} - \varphi_{\widetilde{U}_t}$ does not differ by a polynomial of degree 2 or more in a neighborhood around the origin, for $t = 1, \ldots, T$, then A is identified.

By Lemma 2.1, the number of unknown parameters that can be identified is at most $p(T - p)$ and so we need at least p^2 restrictions on A. The classical errors-in-variables model that now follows is an example of a model with arbitrarily dependent $\boldsymbol{F}$.

3 Examples

In this section we provide two examples of identification. The first example is the classical errors-in-variables model in which all the unobserved regressors can be arbitrarily dependent. The matrix representation of the system contains many 0's with each unobserved regressor excluded from all but two equations. The second example identifies a panel data moving average process with subsets of dependent unobservables.

3.1 Errors-in-variables model

Consider the classical linear regression model with measurement error in all the regressors,

$$X_k = F_k + U_k \qquad k = 1, \dots, p, \tag{2}$$

$$Y = \sum_{k=1}^{p} \beta_k F_k + V, \tag{3}$$

which can be rewritten as,

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ \beta_1 & \dots & \dots & \beta_p \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_p \end{pmatrix} + \begin{pmatrix} U_1 \\ \vdots \\ U_p \\ V \end{pmatrix}. \tag{4}$$

The literature on measurement error in linear models is vast. Three excellent reviews are Cheng and Van Ness (1999), Fuller (2009) and Gillard (2010) (see also Carroll et al., 2006; Chen et al., 2011; Schennach, 2013, for the literature on measurement error in nonlinear models). In the single regressor errors-in-variables model, Reiersøl (1950) shows that a nonnormal unobserved regressor is sufficient for identification and Schennach and Hu (2013) show that if the regressor is normal then the model is identified if neither of the errors are divisible by normal distributions. In the multivariate errors-in-variables model, Kapteyn and Wansbeek (1983) show that if the errors are jointly normal then a sufficient condition for identification is that no linear combination of the regressors is normal and Bekker (1986) shows that the model is identified if and only if there does not exist a nonsingular matrix $(\boldsymbol{a} : B)$ such that $\boldsymbol{a}'\boldsymbol{F}$ is normal and independent of $B'\boldsymbol{F}$. Geary (1941) and later Cragg (1997), Dagenais and Dagenais (1997), Erickson and Whited (2012), Lewbel (1997) and Pal (1980) identify the coefficients from finite higher order moments using, for example, skewness or kurtosis. The intuition for requiring some nonnormality is that normal distributions are completely characterized by their means and covariances and Klepper and Leamer (1984) show that without higher-order moments the coefficients are at best partially identified.

Theorem 3.1. *Suppose that Equations* (2) *and* (3) *hold and assume that* $\boldsymbol{F}$, $U_1, \dots, U_p$ *and* V *are mutually independent with* $\boldsymbol{F}$ *arbitrarily dependent. Assume* $\beta_k \neq 0$, *for* $k = 1, \dots, p$.

If (i) no nonzero linear combination of the components of $\boldsymbol{F}$ *is degenerate and (ii)(a) no nonzero linear combination of the components of* $\boldsymbol{F}$ *is normal or (ii)(b)* $\varphi_V - \varphi_{\widetilde{V}}$ *and* $\varphi_{U_t} - \varphi_{\widetilde{U}_t}$, *for* $t = 1, \dots, p$, *do not differ by a polynomial of degree* 2 *or more in a neighborhood around the origin, then* $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ *is identified.*

Most of the entries of the matrix representation in Equation (4) are 0. We identify $p = \dim(\boldsymbol{\beta})$

parameters (out of the $(p+1)p$ elements of A), which by Lemma 2.1 is the most that can be identified for this dependence structure. As mentioned earlier sparsity is a common restriction in high dimensional models where the complexity of the model comes from the high dimensionality of the variables rather than statistically dependent unobservables as is the case in the errors-in-variables model.

3.2 Panel data moving average process

We now consider a moving average process in a short panel data framework. For example in earnings dynamics models, Blundell et al. (2008) and Hall and Mishkin (1982) analyze the effects of income shocks on consumption, and Abowd and Card (1989) and Meghir and Pistaferri (2004) are interested in the decomposition of earnings into permanent and transitory income components.

We show how some dynamic moving average processes are identified even when subsets of the shocks are allowed to be arbitrarily dependent. Consider the moving average process,

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_r \varepsilon_{t-r} \qquad t = 1, \ldots, T.$$

For illustrative purposes, we introduce a specific choice of T and dependence structure. Let $T = 2r$ and assume that $(\varepsilon_{-r+1}, \ldots, \varepsilon_0)$, $(\varepsilon_1, \ldots, \varepsilon_{r-1})$, $(\varepsilon_r, \ldots, \varepsilon_{2r-1})$ and ε_{2r} are mutually independent with each vector containing arbitrarily dependent components. Then,

$$\boldsymbol{Y} = A_1 \boldsymbol{F}_1 + A_2 \boldsymbol{F}_2 + A_3 \boldsymbol{F}_3 + \boldsymbol{e}_{2r} \varepsilon_{2r}, \tag{5}$$

where

$$A_1 = \begin{pmatrix} \theta_r & \ldots & \theta_1 \\ 0 & \ddots & \vdots \\ \vdots & \ddots & \theta_r \\ \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \theta_1 & \ddots & \mathbf{0} \\ \vdots & \ddots & 1 \\ \theta_{r-1} & \ldots & \theta_1 \\ \theta_r & \ddots & \vdots \\ \mathbf{0} & \ddots & \theta_{r-1} \\ \mathbf{0} & \mathbf{0} & \theta_r \\ 0 & \ldots & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} \mathbf{0} & \ldots & \mathbf{0} \\ 1 & \ddots & \vdots \\ \theta_1 & \ddots & 0 \\ \vdots & \ddots & 1 \\ \theta_r & \ldots & \theta_1 \end{pmatrix},$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_{2r} \end{pmatrix}, \quad \boldsymbol{F}_1 = \begin{pmatrix} \varepsilon_{-r+1} \\ \vdots \\ \varepsilon_0 \end{pmatrix}, \quad \boldsymbol{F}_2 = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{r-1} \end{pmatrix}, \quad \boldsymbol{F}_3 = \begin{pmatrix} \varepsilon_r \\ \vdots \\ \varepsilon_{2r-1} \end{pmatrix}.$$

Theorem 3.2. *Suppose that Equation* (5) *holds, and assume that* ε_{-r+1}, $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, $\boldsymbol{F}_3$ *and* ε_{2r} *are* *mutually independent with each vector containing arbitrarily dependent components. Assume* $\theta_k \neq 0$, *for* $k = 1, \ldots, r$, *and* $\theta_r - \theta_{r-1}\theta_1 - \theta_{r-2}(\theta_2 - \theta_1^2) - \ldots \neq 0$.

If (i) no nonzero linear combination of the components of $\boldsymbol{F}_3$ *is degenerate; and (ii)(a) no nonzero linear combination of the components of* $\boldsymbol{F}_3$ *is normal; or (ii)(b)*$\varphi_{A_{j'}\boldsymbol{F}_{j'}} - \varphi_{\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}}$, *for* $j' = 1, 2$, *and* $\varphi_{\varepsilon_{2r}} - \varphi_{\varepsilon_{2r}}$ *do not differ by a polynomial of degree* 2 *or more in a neighborhood around the origin, then* A_3 *is identified.*

The parameters were identified only using A_3 so that the model is overidentified. Overidentification can probably be used to weaken assumptions or, after the estimator is introduced, for testing or to improve efficiency.

4 Extension

In this section we consider the model in Equation (1) with weaker dependence and rank assumptions that allow a subset of the errors to be arbitrarily dependent and allow $\text{Rank}(A_j : A_{j'}) < p_j + p_{j'}$.

Assumption 4.1.

(i) $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_q, (U_1, \ldots, U_{\bar{t}}), U_{\bar{t}+1}, \ldots, U_T$ *are mutually independent;*

(ii) (a) Rank$(A_j : A_{j'}) \geq \max\{p_j, p_{j'}\} + 1$, *for* $j' \neq j$;

(b) Rank$(A_j : (I_{\bar{t}} : \mathbf{0})') \geq \max\{p_j, \bar{t}\} + 1$, *for* $j = 1, \ldots, q$;

(c) Rank$(A_j : \boldsymbol{e}_t) = p_j + 1$, *for* $j = 1, \ldots, q$ *and* $t = \bar{t} + 1, \ldots, T$; *and*

(iii) $A_j \in \mathcal{A}_j$ *has no right equivalent matrix in* $\mathcal{A}_j$, *for* $j = 1, \ldots, q$.

Assumption 4.1(i) allows $(U_1, \ldots, U_{\bar{t}})$ to be arbitrarily dependent. If one of the rank conditions does not hold then we can obtain a reduced observationally equivalent model with fewer unobservables and which still satisfies the same dependence assumptions and rank conditions above.

Theorem 4.1. *Suppose that Equation* (1) *and Assumption 4.1 hold. Then,*

$$\varphi_{\boldsymbol{F}_j}^{(2q)}\left(A'_{j,2q}\boldsymbol{s}\right) = P_0, \qquad j = 1, \ldots, q, \tag{6}$$

where $\varphi_{\boldsymbol{F}_j}^{(2q)}$, P_0 *and* $A_{j,2q}$ *are respectively equal to a* $2q^{th}$ *difference equation of* $\varphi_{\boldsymbol{F}_j}$, *a constant and a nonsingular matrix multiplied by* A_j, *and are defined formally in Section A.8 of the appendix.*

To prove Theorem 4.1, Equation (1) is transformed into a functional equation using its log characteristic function and the assumed dependence structure. The result is a linear combination of log characteristic functions of the factors and errors. Then, recursively applying first differences, which use the rank conditions, all the log characteristic functions are eliminated except for one, resulting in a difference equation in the log characteristic function of $\boldsymbol{F}_j$.

Unlike the theorems in the previous section, the weaker dependence structure and rank conditions of Assumption 4.1 do not provide an easy way to integrate out or solve Equation (6) to obtain an explicit expression for the functional form of $\varphi_{\boldsymbol{F}_j}$. In general, the functional solutions to the equation are problem specific and hard to solve, and often there are many, possibly infinite, solutions. However, identification of A_j is possible by a high-level assumption that rules out all families of functional solutions that correspond to a well-defined log characteristic functions.

When higher-order moments exist, the difference equation in Equation (6) can be replaced by an analogous partial differential equation. As an example, consider again the errors-in-variables model from Equations (2) and (3) allowing for $\boldsymbol{U}' = (U_1, \ldots, U_p)$ to be arbitrarily dependent and assuming finite second-order moments.

Corollary 4.1. *Suppose that Equations* (2) *and* (3) *hold and assume that* $\boldsymbol{F}$, $\boldsymbol{U}$ *and* V *are mutually independent with each vector containing arbitrarily dependent components. Assume* $\beta_k \neq 0$, *for* $k = 1, \ldots, p$, *and that all second-order moments are finite. Then,*

$$\sum_{k_1,k_2} \beta_{k_1}(\beta_{k_2} - \widetilde{\beta}_{k_2}) \left.\frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}}\right|_{\boldsymbol{u}=(I_p:\boldsymbol{0})\boldsymbol{v}+(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})s_{p+1}} = P_0,$$

where P_0 *is some constant.*

We can identify $\boldsymbol{\beta}$ by ruling out all log characteristic functions that solve the above differential equation. For example, if $p = 2$ then,

$$\beta_1(\beta_1 - \widetilde{\beta}_1) \left.\frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_1^2}\right|_{\boldsymbol{u}=(s_1,s_2)-(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})s_3} + (2\beta_1\beta_2 - \beta_1\widetilde{\beta}_2 - \widetilde{\beta}_1\beta_2) \left.\frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_1 \partial u_2}\right|_{\boldsymbol{u}=(s_1,s_2)-(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})s_3}$$
$$+ \beta_2(\beta_2 - \widetilde{\beta}_2) \left.\frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_2^2}\right|_{\boldsymbol{u}=(s_1,s_2)-(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})s_3} = P_0. \tag{7}$$

If $\beta_1/\beta_2 = \widetilde{\beta}_1/\widetilde{\beta}_2$ then the differential equation is parabolic and can be transformed into a heat equation. Otherwise the differential equation is hyperbolic and can be transformed into a wave equation. For identification we could, for example, rule out all solutions to the heat and wave equations or assume $\beta_2 = c\beta_1$ and rule out all solutions to the heat equation.

5 Estimation

Consider the model,

$$\boldsymbol{Y} = A\boldsymbol{F} + \boldsymbol{U}, \tag{8}$$

where A is a $T \times p$ matrix and $\boldsymbol{F}$ is arbitrarily dependent, i.e., $q = 1$ in Equation (1).

When the unobservables are mutually independent, the ICA literature estimates the coefficients of A using kurtosis (fourth-order moments), negentropy or maximum likelihood (see, e.g., Hyvärinen et al., 2004; Stone, 2004). In the economics literature Bonhomme and Robin (2009), for example, use second-to-fourth order moments for estimation. The reason we believe our estimator could be more efficient than estimators based on higher-order moments is because our estimator is based on second-order partial derivatives of the empirical log characteristic function which, by a Taylor expansion, has information from all higher-order moments.

We now mimic the steps in the identification proof to construct an estimator. First, partial derivatives of the empirical log characteristic function replace finite differences and eliminate the errors. Second, we choose arguments, or equivalently linear combinations of the observables, so that the log characteristic function of $\boldsymbol{F}$ is either a polynomial, ruled out by nonnormality, or the arguments are in the null-space of A'. The estimator of the null-space of A' takes the form of an extremum estimator based on second-order partial derivatives of the empirical log characteristic function. Finally, we use the estimation of the the null-space of A', that the null space identifies A', and Lemma 2.1 to construct an estimator for A'.

The estimator uses second-order partial derivatives of the empirical log characteristic function,

$$h_n(\boldsymbol{\pi}, v) = \left(\dots, \widehat{\mathrm{Cov}}(Y_{t_1}, Y_{t_2}) + \left. \frac{\partial^2 \widehat{\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}}{\partial s_{t_1} \partial s_{t_2}} \right|_{\boldsymbol{s}=\boldsymbol{\pi} v}, \dots \right), \qquad t_1 \neq t_2,$$

$$H_n(\boldsymbol{\Pi}, v)' = \left(h_n(\boldsymbol{\pi}_1, v), \dots, h_n(\boldsymbol{\pi}_{T-p}, v) \right),$$

where $\boldsymbol{\Pi} = \mathrm{vec}(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_{T-p})$ with $\boldsymbol{\pi}_j \in \mathbb{R}^T$ for $j = 1, \dots, T-p$ and

$$\widehat{\mathrm{Cov}}(Y_{t_1}, Y_{t_2}) = \frac{1}{n} \sum_{i=1}^n Y_{it_1} Y_{it_2} - \left(\frac{1}{n} \sum_{i=1}^n Y_{it_1} \right) \left(\frac{1}{n} \sum_{i=1}^n Y_{it_2} \right),$$

$$\left. \frac{\partial^2 \widehat{\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}}{\partial s_{t_1} \partial s_{t_2}} \right|_{\boldsymbol{s}=\boldsymbol{\pi} v} = \frac{\left(\frac{1}{n} \sum_{i=1}^n Y_{it_1} e^{\mathrm{i} v \sum_{t=1}^T \pi_t Y_{it}} \right) \left(\frac{1}{n} \sum_{i=1}^n Y_{it_2} e^{\mathrm{i} v \sum_{t=1}^T \pi_t Y_{it}} \right)}{\left(\frac{1}{n} \sum_{i=1}^n e^{\mathrm{i} v \sum_{t=1}^T \pi_t Y_{it}} \right)^2} - \frac{\frac{1}{n} \sum_{i=1}^n Y_{it_1} Y_{it_2} e^{\mathrm{i} v \sum_{t=1}^T \pi_t Y_{it}}}{\frac{1}{n} \sum_{i=1}^n e^{\mathrm{i} v \sum_{t=1}^T \pi_t Y_{it}}}.$$

In Section A.10 of the appendix we show that the population quantity of $H_n(\cdot)$ is equal to zero if and only if $\boldsymbol{\pi} \in \ker(A')$. Thus, we first estimate Basis($\ker(A')$) by the extremum estimator,

$$\widehat{\boldsymbol{\Pi}} = \underset{\substack{\boldsymbol{\Pi} \in \mathcal{P} \\ \{\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_{T-p}\} \text{ is a basis}}}{\operatorname{argmin}} \int_{\mathcal{V}} H_n(\boldsymbol{\Pi}, v)' W_n(v) H_n(\boldsymbol{\Pi}, v) dv, \tag{9}$$

where $W_n(v)$ is a random, positive definite matrix.

The objective function integrates over a minimum distance estimator which can be interpreted as a measure of nonnormality because a normal distribution has constant second-order partial derivatives and this measure will be equal to zero regardless of the argument choices. On the other hand, all other distributions have variation in their second-order partial derivatives, and the more variation the larger the measure. Variation in the second-order partial derivatives can be thought of as curvature around the optimum, which makes estimation easier.

In addition to the identification assumptions in the previous section, we now impose regularity conditions and assume random sampling.

Assumption 5.1. *We observe a sample* $\{\mathbf{Y}_i\}_{i=1}^n$ *of i.i.d. draws from* $\boldsymbol{Y}$*.*

Assumption 5.2.

(i) All $\mathbf{\Pi}_0 = vec(\boldsymbol{\pi}_{01}, \ldots, \boldsymbol{\pi}_{0T-p})$*, with* $\{\boldsymbol{\pi}_{01}, \ldots, \boldsymbol{\pi}_{0T-p}\} \in Basis(ker(A'))$*, belong to the interior of a compact set* $\mathcal{P}$ *and* $\mathcal{V} \subset \mathbb{R}$ *is a compact set containing* 0*;*

(ii) $\max_t E[\exp(|Y_t|\beta)] < \infty$ *for some* $\beta > 0$*, and* $E[e^{i\boldsymbol{\pi}'\boldsymbol{Y}v}] \neq 0$ *for all* $\boldsymbol{\pi} \in \mathcal{P}$ *and* $v \in \mathcal{V}$*;*

(iii) $\sup_{v\in\mathcal{V}} ||W_n(v) - W(v)|| \xrightarrow{p} 0$*; and*

(iv) $G(v) = \frac{\partial H_0(\mathbf{\Pi},v)}{\partial \mathbf{\Pi}}\Big|_{\mathbf{\Pi}=\mathbf{\Pi}_0}$ *has full rank.*

The following theorem states that $\widehat{\mathbf{\Pi}}$ is root-n consistent and asymptotically normal.

Theorem 5.1. *Suppose that Equation* (8)*, and Assumptions 2.1, 2.2, 2.3, 2.4(i), 2.4(ii)(a), and Assumptions 5.1 and 5.2 hold. Then,*

$$\sqrt{n}\left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_0\right) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

where $\Sigma = (\int_{\mathcal{V}} G(v)'W(v)G(v)dv)^{-1} \int_{\mathcal{V}} G(v)'W(v)\Omega(v)W(v)G(v)dv(\int_{\mathcal{V}} G(v)'W(v)G(v)dv)^{-1}$*. We show that* $\sqrt{n}H_n(\mathbf{\Pi}_0, v) \xrightarrow{d} N(\mathbf{0}, \Omega(v))$ *and define* $\Omega(v)$ *in Section A.10 of the appendix. The optimal weighting matrix is* $W(v) = \Omega(v)^{-1}$*.*

By Lemma 2.1, A uniquely solves $A'\mathbf{\Pi}_0 = \mathbf{0}$ so we can estimate A by $\widehat{A}'\widehat{\mathbf{\Pi}} = \mathbf{0}$, which because of linearity even has closed-form solutions for all the unknown parameters as functions of $\widehat{\mathbf{\Pi}}$. Finally, by the delta-method and Theorem 5.1, $\widehat{A}$ is root-n consistent and asymptotically normal. To be specific say the unknown parameters are $\boldsymbol{\theta}_0 = g(\mathbf{\Pi}_0)$ and we estimate $\mathbf{\Pi}_0$ by $\widehat{\mathbf{\Pi}}$ using Equation (9). Then we can estimate $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}} = g(\widehat{\mathbf{\Pi}})$ and by the delta method,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, D'\Sigma D)$$

where $D = \frac{\partial g(\mathbf{\Pi})}{\partial \mathbf{\Pi}}\Big|_{\mathbf{\Pi}=\mathbf{\Pi}_0}$ and Σ is the variance from Theorem 5.1.

6 Monte Carlo simulations

In this section we generate data $\{X_{1i}, X_{2i}, Y_i\}_{i=1}^n$ from the classical errors-in-variables model,

$$\begin{pmatrix} X_1 \\ X_2 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ V \end{pmatrix}, \tag{10}$$

where (F_1, F_2), U_1, U_2 and V are mutually independent and $(\beta_1, \beta_2) = (1, 1)$. We consider several distributional specifications for (F_1, F_2) and allow for various amounts of measurement error in U_1 and U_2, including no measurement errors. We also include a set of simulations with dependent (U_1, U_2).

As discussed earlier, under the assumption of nonnormality there are several estimators of (β_1, β_2) in the literature, including Dagenais and Dagenais (1997), Lewbel (1997) and Erickson et al. (2014) who use higher-order moments and Lewbel (2012) who uses heteroscedasticity. We compare (i) the extremum estimator based on the second-order partial derivatives (PD) of the empirical log characteristic function from the previous section,

$$(\widehat{\beta}_1, \widehat{\beta}_2) = \underset{(b_1,b_2)\in\mathcal{B}}{\operatorname{argmin}} \int_{\mathcal{V}} \left(\widehat{\operatorname{Cov}}(X_1, Y) + \left.\frac{\partial^2 \widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_1 \partial s_3}\right|_{\boldsymbol{s}=(b_1,b_2,-1)v} \right)^2 \tag{11}$$

$$+ \left(\widehat{\operatorname{Cov}}(X_2, Y) + \left.\frac{\partial^2 \widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_2 \partial s_3}\right|_{\boldsymbol{s}=(b_1,b_2,-1)v} \right)^2 + \frac{1}{2}\left(\widehat{\operatorname{Cov}}(X_1, X_2) + \left.\frac{\partial^2 \widehat{\varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}}{\partial s_1 \partial s_2}\right|_{\boldsymbol{s}=(b_1,b_2,-1)v} \right)^2 dv,$$

where $\mathcal{V} = [-0.5, 0.5]$ with (ii) a generalized method of moments estimator (GMM) that uses third, fourth and fifth-order cumulants and (iii) the ordinary least squares estimator (OLS) that ignores measurement error.[6]

For the PD estimator, we tried several different choices for the weights including exponentially and linearly increasing and decreasing weights and different choices for the relative weights on the terms. Only the relative weights had an impact on the results. We decided to put less weight on the last term because the equations for X_1 and X_2 do not contain the unknown parameters. We also experimented with the bandwidth (i.e., the choice of h in the interval $\mathcal{U} = [-h, h]$), which affected the estimates. However, like other estimators that require a bandwidth choice, there is a range of bandwidths that produce roughly the same estimates. For this particular data generating process, the estimates were similar even for much larger bandwidth choices. There is a tradeoff between small and large bandwidth. On the one hand, the characteristic function is estimated better closer to the origin. On the other hand identification may be coming from the tails of the characteristic functions. Finally, notice that $\left.\frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_j \partial s_{j'}}\right|_{\boldsymbol{s}=\boldsymbol{\pi}v} = \left.\frac{\partial^2 \varphi_{\boldsymbol{X},Y}(\boldsymbol{s})}{\partial s_j \partial s_{j'}}\right|_{\boldsymbol{s}=\boldsymbol{0}}$ where $\boldsymbol{\pi}' = (\beta_1, \beta_2, -1) \in \operatorname{Basis}(\ker(A'))$ so that we estimate β_1 by $\widehat{\beta}_1 = \widehat{\boldsymbol{\pi}}_1$ and β_2 by $\widehat{\beta}_2 = \widehat{\boldsymbol{\pi}}_2$.

Table 1 displays the biases, standard deviations (SD) and root mean squared errors (RMSE) of the PD, GMM and OLS estimators for β_1 and β_2 from 100 simulations of sample size $1,000$ with no measurement error ($U_1 = U_2 = 0$), with independent but small measurement errors (U_1 and U_2 drawn independently from $N(0, 1/4)$) and with independent and larger measurement errors (U_1

[6] The GMM estimator is the one derived and used by Erickson et al. (2014) and available for download at https://ideas.repec.org/c/boc/bocode/s457525.html or http://www.kn.owled.ge/#!downloads/c65q.

and U_2 drawn independently from $N(0,1)$). We separately also consider dependent measurement errors ((U_1, U_2) drawn from $N(\mathbf{0}, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$). The unobserved regressors F_1 and F_2 are drawn from a $Gamma(1,2)$ distribution. They are then adjusted to have variance 2 and covariance 1 by multiplying the generated variables by a Cholesky decomposition of the covariance matrix. The error term V is drawn from $N(0,1)$.

Table 1: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1,1)$ in Model (10) with F_1 and F_2 coming from $Gamma(1,2)$ distributions.

		β_1			β_2		
Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
PD	0	0.00	0.04	0.04	0.01	0.05	0.05
	1/4	-0.01	0.05	0.05	0.01	0.07	0.07
	1	-0.01	0.07	0.07	0.02	0.09	0.09
GMM	0	0.00	0.04	0.04	0.01	0.05	0.05
	1/4	0.00	0.07	0.07	-0.01	0.07	0.07
	1	0.00	0.09	0.09	-0.02	0.09	0.09
OLS	0	0.00	0.03	0.03	0.00	0.02	0.02
	1/4	-0.15	0.03	0.15	-0.14	0.02	0.14
	1	-0.25	0.03	0.25	-0.25	0.02	0.25

Notes: Results from 100 simulations of sample size $1,000$. U_1 and U_2 are drawn from independent normal distributions with variance σ_U^2.

The OLS estimates have the tightest confidence bands around β_1 and β_2 when there is no measurement error and are badly biased even with a small amount of measurement error with confidence bands very far from the true values of β_1 and β_2 and highest RMSEs. For example, when $\text{Var}(U_1) = 1/4$, about 11% of $\text{Var}(X_1)$, then the OLS estimates of β_1 are about 5 standard deviations away from β_1. The GMM and PD estimates have small RMSEs and are stable regardless of measurement error.

For robustness we have run the simulations using various choices of distributions with similar results: the OLS estimator performs the best when there is no measurement error and gets substantially worse even with small amounts of measurement error while the GMM and PD estimates have relatively small RMSEs and are stable for different amounts of measurement error. In Section A.11 of the appendix we present these results when F_1 and F_2 are drawn from t, χ^2, uniform and bimodal distributions. The GMM and PD estimators perform well even when the distributions are symmetric. The GMM estimates get somewhat worse as measurement error increases. The GMM estimator performs worse for the t distribution, which is probably due to its closeness to the normal distribution and fat tails. The PD estimator seems to be the most robust to measurement error and distributional choices.

Finally, we generate data with dependent measurement errors by drawing (U_1, U_2) from $N(\mathbf{0}, \Sigma)$ where $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. Table 2 shows results that are similar to Table 1 suggesting that the estimators are consistent. This in turn suggests that this particular set of distributional choices does not satisfy the differential equation in Equation (7) and that (β_1, β_2) is identified.

Table 2: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1, 1)$ in Model (10) with F_1 and F_2 coming from $Gamma(1, 2)$ distributions.

	β_1			β_2		
Estimator	bias	SD	RMSE	bias	SD	RMSE
PD	-0.01	0.08	0.08	0.01	0.10	0.10
GMM	0.00	0.09	0.09	-0.04	0.10	0.11
OLS	-0.34	0.03	0.34	-0.33	0.03	0.33

Notes: Results from 100 simulations of sample size 1,000. $(U_1, U_2) \sim N(\mathbf{0}, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

7 Application

In this section we apply our estimator to data on firm investment decisions, which according to Tobin's q theory (Brainard and Tobin, 1968; Tobin, 1969) should optimally only depend on the marginal value of capital replacement (marginal q). The theory, which is based on the arbitrage argument that firms will invest in capital if and only if it is profitable (Keynes, 1936) is supported by neoclassical theory of investment under convex adjustment costs (Hayashi, 1982).

Consider the linear model for firm investment decisions

$$Y = \beta q^* + U_Y, \tag{12}$$

where Y is the ratio of investment to capital, q^* is marginal q, and U_Y is an error.

Empirical studies that regress investment on observed measures of marginal q mostly fail to predict investment decisions (see, e.g., Blanchard et al., 1993). This has led to alternative theories of investment, endogenizing marginal q, nonlinear regressions, and at least the following two modifications to the linear model above: (a) including a variable for financial constraints in the regression (see, e.g., Fazzari et al., 2000),

$$Y = \beta_1 q^* + \beta_2 CF^* + U_Y, \tag{13}$$

where CF^* represents financial constraints and U_Y is an error and (b) accounting for measurement error in q^* (see, e.g., Erickson and Whited, 2000) by

$$q = q^* + U_q, \tag{14}$$

where q is the ratio of market to book value of equity plus liabilities, and U_q is measurement error. In addition to the above two modifications, we allow for measurement error in financial constraints,

$$CF = CF^* + U_{CF}, \tag{15}$$

where CF is cash flows and U_{CF} is measurement error.

Using data from Erickson et al. (2014) for the years 1992-1995,[7] we compare the PD, GMM and OLS estimators. Table 3 presents results for three models: Model 1 is based on Equations (12) and (14) and excludes a variable for financial constraints, Model 2 is based on Equations (13) and (14) and assumes that financial constraints are correctly measured, and Model 3 is based on Equations (13), (14) and (15) and allows both marginal q and financial constraints to be measured with error.

Ignoring measurement errors, the OLS coefficient estimates of the effects of marginal q and financial constraints on investment decisions are positive and significant at the 1% level in all the models.

In Model 1, that excludes financial constraints, the GMM and PD estimates are statistically similar, and agree that the effects of marginal q on investment decisions are positive and significant. These estimates are larger than the OLS estimates (the OLS estimator suffers from attenuation bias) and are somewhat different from year to year.

In Model 2, which assumes that financial constraints are correctly measured, the GMM and PD estimates of the coefficients on marginal q are all positive, significant and statistically similar, while the estimates of the coefficients on financial constraints are all insignificant. This, as Erickson et al. (2014) find, is evidence that investment decisions may not depend on financial constraints.

The results are ambiguous in Model 3, which allows measurement error in both marginal q and financial constraints. In this model the GMM estimate of the coefficient on marginal q is positive and significant in 1992 and the coefficient on financial constraints is negative and significant in 1992. The GMM estimates of all the other coefficients are insignificant. The PD estimates of the coefficients on marginal q are positive and significant in 1992 and 1993 and insignificant otherwise, while the PD estimates of the coefficients on financial constraints is positive and significant in 1994 and otherwise insignificant. This could suggest, for example, that the unobserved regressors and errors are dependent or that financial constraints are incorporated incorrectly into the model. One

[7] Erickson et al. (2014) collect the data from the 2012 Compustat Industrial Files. There are 2 850, 3 031, 3 279 and 3 425 observations in 1992, 1993, 1994 and 1995, respectively.

possible direction to pursue is a nonlinear model like Barnett and Sakellaris (1998) and Schennach and Hu (2013) that allows for measurement error in both marginal q and financial constraints.

Table 3: Estimates of firm investment decisions based on marginal q and financial constraints.

	β_1 (marginal q)			β_2 (financial constraints)		
Year	OLS	GMM	PD	OLS	GMM	PD
			Model 1: $Y = \beta_1 q^* + U_Y$, $q = q^* + U_q$			
1992	0.014**	0.015**	0.029**		—	
	(0.002)	(0.002)	(0.005)			
1993	0.016**	0.029**	0.032**		—	
	(0.001)	(0.008)	(0.006)			
1994	0.017**	0.036**	0.041**		—	
	(0.002)	(0.012)	(0.008)			
1995	0.017**	0.075*	0.072**		—	
	(0.003)	(0.037)	(0.017)			
			Model 2: $Y = \beta_1 q^* + \beta_2 CF^* + U_Y$, $q = q^* + U_q$			
1992	0.012**	0.017*	0.028**	0.077**	0.056	0.001
	(0.001)	(0.007)	(0.005)	(0.019)	(0.043)	(0.029)
1993	0.013**	0.039**	0.034**	0.064**	-0.047	-0.024
	(0.001)	(0.008)	(0.009)	(0.011)	(0.037)	(0.038)
1994	0.012**	0.040*	0.060**	0.111**	-0.017	-0.113
	(0.001)	(0.017)	(0.018)	(0.015)	(0.078)	(0.086)
1995	0.014**	0.055*	0.109**	0.106**	-0.023	-0.192
	(0.001)	(0.025)	(0.039)	(0.015)	(0.070)	(0.140)
			Model 3: $Y = \beta_1 q^* + \beta_2 CF^* + U_Y$, $q = q^* + U_q,\ CF = CF^* + U_{CF}$			
1992	0.012**	0.011**	0.030**	0.077**	-0.119*	-0.067
	(0.001)	(0.003)	(0.012)	(0.005)	(0.060)	(0.224)
1993	0.013**	0.008	0.046**	0.064**	0.029	-0.340
	(0.001)	(0.009)	(0.012)	(0.011)	(0.129)	(0.207)
1994	0.012**	0.003	0.000	0.111**	0.138	0.707**
	(0.002)	(0.007)	(0.013)	(0.015)	(0.142)	(0.203)
1995	0.014**	0.004	0.037	0.106**	0.061	0.435
	(0.002)	(0.005)	(0.039)	(0.015)	(0.090)	(0.410)

Notes: Standard errors are in parentheses under the estimates. Single and double asterisks indicate 5% and 1% significance, respectively.

8 Conclusion

We identify and estimate coefficients in the linear factor model $\boldsymbol{Y} = \sum_{j=1}^{q} A_j \boldsymbol{F}_j + \boldsymbol{U}$ allowing for each $\boldsymbol{F}_j$ to be arbitrarily dependent. For identification of A_j, we assume a statistical dependence structure, matrix rank conditions, that A_j has no right equivalent matrix and nonnormality assumptions on $\boldsymbol{F}_j$.

Our identification strategy transforms the system of linear equations into a functional equation using a log characteristic function and solves this equation. Despite identification not being constructive we can mimic its steps to produce an extremum estimator based on second-order partial derivatives of the empirical log characteristic function. The estimator is root-n consistent and asymptotically normal and has good finite sample properties in a Monte Carlo study where it is more robust to distributional choice and measurement error than a GMM estimator based on higher-order moments

The Monte Carlo simulations provide evidence that the extremum estimator could be viable alternative to some of the other available estimators but a more comprehensive comparison would be needed. Efficiency of the estimator can also be improved by a better, data-driven, choice of bandwidth and weighting matrix and perhaps combining second and higher-order partial derivatives of the empirical log characteristic function.

Estimation in the case $q \neq 1$ is more challenging. A possible route is to mimic identification using partial derivatives instead of finite differences and to use an extremum estimator based on higher-order partial derivatives of the empirical log characteristic function. The complication is that the estimator will need to be multi-stage like in identification and needs to jointly estimate $A_1, \ldots, A_q$.[8]

Another avenue for future work is inference. First, a necessary condition for no linear combination of unobservables to be normal is that no linear combination of the outcome variables is normal. There are many standard ways to test for normality but the results in this paper can be modified to provide an additional test based on second-order partial derivatives of the empirical log characteristic function. Second, the estimator might provide a way for testing certain restrictions on the matrix A like exclusion restrictions.

Finally, after the coefficients are estimated Ben-Moshe (2016) in a companion paper considers identification and estimation of the joint distribution of $\boldsymbol{F}$.

[8]In some concrete examples we were successful using such an estimator. Further, in a linear regression model with measurement error and repeated measurements we were even able to construct an extremum estimator based only on first-order partial derivatives of the empirical log characteristic function.

References

Abowd, J. M. and Card, D. (1989), 'On the covariance structure of earnings and hours changes', *Econometrica* **57**(2), 411–445.

Aigner, D. J., Hsiao, C., Kapteyn, A. and Wansbeek, T. (1984), 'Latent variable models in econometrics', *Handbook of econometrics* **2**, 1321–1393.

Anderson, T. W. and Rubin, H. (1956), Statistical inference in factor analysis, *in* 'Proceedings of the third Berkeley symposium on mathematical statistics and probability', Vol. 5, p. 1.

Attanasio, O. P. (2015), 'The determinants of human capital formation during the early years of life: Theory, measurement, and policies', *Journal of the European Economic Association* **13**(6), 949–997.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. and Ng, S. (2013), 'Principal components estimation and identification of static factors', *Journal of Econometrics* **176**(1), 18–29.

Barnett, S. A. and Sakellaris, P. (1998), 'Nonlinear response of firm investment to q: Testing a model of convex and non-convex adjustment costs', *Journal of Monetary Economics* **42**(2), 261–288.

Bekker, P. A. (1986), 'Comment on identification in the linear errors in variables model', *Econometrica: Journal of the Econometric Society* pp. 215–217.

Ben-Moshe, D. (2016), 'Identification of joint distributions in dependent factor models', *working paper* .

Blanchard, O., Rhee, C. and Summers, L. (1993), 'The stock market, profit, and investment', *The Quarterly Journal of Economics* **108**(1), 115–136.

Blundell, R., Pistaferri, L. and Preston, I. (2008), 'Consumption inequality and partial insurance', *The American Economic Review* pp. 1887–1921.

Bonhomme, S. and Robin, J.-M. (2009), 'Consistent noisy independent component analysis', *Journal of Econometrics* **149**(1), 12–25.

Bonhomme, S. and Robin, J.-M. (2010), 'Generalized non-parametric deconvolution with an application to earnings dynamics', *The Review of Economic Studies* **77**(2), 491–533.

Brainard, W. C. and Tobin, J. (1968), 'Pitfalls in financial model building', *The American Economic Review* pp. 99–122.

Carneiro, P., Hansen, K. T. and Heckman, J. J. (2003), '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**(2), 361–422.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006), *Measurement error in nonlinear models: a modern perspective*, CRC press.

Chen, X., Hong, H. and Nekipelov, D. (2011), 'Nonlinear models of measurement errors', *Journal of Economic Literature* **49**(4), 901–937.

Cheng, C.-L. and Van Ness, J. W. (1999), *Statistical regression with measurement error*, Vol. 6, Arnold London.

Comon, P. (1994), 'Independent component analysis, a new concept?', *Signal processing* **36**(3), 287–314.

Cragg, J. G. (1997), 'Using higher moments to estimate the simple errors-in-variables model', *Rand Journal of Economics* pp. S71–S91.

Cunha, F. and Heckman, J. J. (2008), 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation', *Journal of Human Resources* **43**(4), 738–782.

Cunha, F., Heckman, J. J. and Schennach, S. M. (2010), 'Estimating the technology of cognitive and noncognitive skill formation', *Econometrica* **78**(3), 883–931.

Dagenais, M. G. and Dagenais, D. L. (1997), 'Higher moment estimators for linear regression models with errors in the variables', *Journal of Econometrics* **76**, 193 – 221.

Erickson, T., Jiang, C. H. and Whited, T. M. (2014), 'Minimum distance estimation of the errors-in-variables model using linear cumulant equations', *Journal of Econometrics* **183**(2), 211–221.

Erickson, T. and Whited, T. M. (2000), 'Measurement error and the relationship between investment and q', *Journal of political economy* **108**(5), 1027–1057.

Erickson, T. and Whited, T. M. (2012), 'Treating measurement error in tobin's q', *Review of Financial Studies* **25**(4), 1286–1329.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of financial economics* **33**(1), 3–56.

Fan, J., Lv, J. and Qi, L. (2011), 'Sparse high-dimensional models in economics', *Annual Review of Economics* **3**(1), 291–317.

Fazzari, S. M., Hubbard, R. G. and Petersen, B. C. (2000), 'Investment-cash flow sensitivities are useful: A comment on kaplan and zingales', *Quarterly Journal of Economics* pp. 695–705.

Forni, M. and Reichlin, L. (1998), 'Let's get real: a factor analytical approach to disaggregated business cycle dynamics', *The Review of Economic Studies* **65**(3), 453–473.

Fuller, W. A. (2009), *Measurement error models*, Vol. 305, John Wiley & Sons.

Geary, R. C. (1941), Inherent relations between random variables, *in* 'Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences', JSTOR, pp. 63–76.

Ghurye, S. and Olkin, I. (1962), 'A characterization of the multivariate normal distribution', *The Annals of Mathematical Statistics* pp. 533–541.

Gillard, J. (2010), 'An overview of linear structural models in errors in variables regression', *REVSTAT–Statistical Journal* **8**(1), 57–80.

Gorman, W. M. (1981), 'Some engel curves', *Essays in the theory and measurement of consumer behaviour in honor of sir Richard Stone* .

Hall, R. E. and Mishkin, F. S. (1982), 'The sensitivity of consumption to transitory income: estimates from panel data on households', *Econometrica* **50**(2), 461–482.

Hayashi, F. (1982), 'Tobin's marginal q and average q: A neoclassical interpretation', *Econometrica: Journal of the Econometric Society* pp. 213–224.

Hyvärinen, A., Karhunen, J. and Oja, E. (2004), *Independent component analysis*, Vol. 46, John Wiley & Sons.

Inselberg, A. (1978), 'On determinants of toeplitz-hessenberg matrices arising in power series', *Journal of Mathematical Analysis and Applications* **63**(2), 347–353.

Jolliffe, I. (2002), *Principal component analysis*, Wiley Online Library.

Jöreskog, K. G. (1970), 'A general method for analysis of covariance structures', *Biometrika* **57**(2), 239–251.

Kapteyn, A. and Wansbeek, T. (1983), 'Identification in the linear errors in variables model', *Econometrica: Journal of the Econometric Society* pp. 1847–1849.

Keynes, J. M. (1936), 'The general theory of interest, employment and money'.

Khatri, C. and Rao, C. R. (1972), 'Functional equations and characterization of probability laws through linear functions of random variables', *Journal of Multivariate Analysis* **2**(2), 162–173.

Klepper, S. and Leamer, E. E. (1984), 'Consistent sets of estimates for regressions with errors in all variables', *Econometrica: Journal of the Econometric Society* pp. 163–183.

Lewbel, A. (1991), 'The rank of demand systems: theory and nonparametric estimation', *Econometrica: Journal of the Econometric Society* pp. 711–730.

Lewbel, A. (1997), 'Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r&d', *Econometrica* **65**(5), 1201–1214.

Lewbel, A. (2012), 'Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models', *Journal of Business & Economic Statistics* **30**(1), 67–80.

Marcinkiewicz, J. (1964), 'Collected papers, ed', *A. Zygmund, PWN, Warsaw* .

Meghir, C. and Pistaferri, L. (2004), 'Income variance dynamics and heterogeneity', *Econometrica* **72**(1), 1–32.

Moon, H. R. and Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Newey, W. K. and McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Pal, M. (1980), 'Consistent moment estimators of regression coefficients in the presence of errors in variables', *Journal of Econometrics* **14**(3), 349–364.

Rao, C. R. (1966), 'Characterisation of the distribution of random variables in linear structural relations', *Sankhyā: The Indian Journal of Statistics, Series A* pp. 251–260.

Reiersøl, O. (1950), 'Identifiability of a linear relation between variables which are subject to error', *Econometrica* pp. 375–389.

Robbin, J. W. (1995), *Matrix Algebra Using MINImal MATlab*, AK Peters.

Robbin, J. W. (2001), Equivalence of matrices. unpublished manuscript.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of economic theory* **13**(3), 341–360.

Schennach, S. and Hu, Y. (2013), 'Nonparametric identification and semiparametric estimation of classical measurement error models without side information', *Journal of the American Statistical Association* **108**, 177–186.

Schennach, S. M. (2013), Measurement error in nonlinear models: A review, *in* D. Acemoglu, M. Arellano and E. Dekel, eds, 'Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress', Econometric Society Monographs, Cambridge University Press.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica: Journal of the Econometric Society* pp. 1–48.

Stock, J. H. and Watson, M. W. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**(2), 293–335.

Stone, J. V. (2004), *Independent component analysis*, Wiley Online Library.

Tobin, J. (1969), 'A general equilibrium approach to monetary theory', *Journal of money, credit and banking* **1**(1), 15–29.

Williams, B. (2015), 'Identification of the linear factor model', *working paper* .

Zou, H., Hastie, T. and Tibshirani, R. (2006), 'Sparse principal component analysis', *Journal of Computational and Graphical Statistics* **15**(2), 265–286.

A Proofs

A.1 Preliminary theorems and lemma

The following theorem from Marcinkiewicz (1964) states that if the log characteristic function of a random variable is a polynomial then it must be normal. We will use this theorem in the proof of Theorem 2.2.

Theorem A.1. *(Marcinkiewicz, 1964). Let φ be a polynomial in a neighborhood of the origin. If φ is a log characteristic function then φ is of degree at most 2 and corresponds to a normal distribution.*

The following lemma on uniform convergence rates of moments, proved in Ben-Moshe (2016) and Bonhomme and Robin (2010), for example, is used to bound higher-order terms in the proof of Theorem 5.1.

Lemma A.1. *Let $\{W_i, \boldsymbol{Y}_i\}_{i=1}^n$ be independent identically distributed draws from $(W, \boldsymbol{Y})$. Assume that $E[\exp(|W|\beta)] < \infty$ for some $\beta > 0$. Let $\mathcal{U}$ and $\mathcal{P}$ be compact sets. Then for large enough n,*

$$\sup_{u\in\mathcal{U},\boldsymbol{\pi}\in\mathcal{P}} \left| \frac{1}{n}\sum_{i=1}^{n} W_i e^{\mathrm{i}u\boldsymbol{\pi}'\boldsymbol{Y}_i} - E\left[W e^{\mathrm{i}u\boldsymbol{\pi}'\boldsymbol{Y}}\right] \right| = O\left(\frac{\log n}{n^{\frac{1}{2}}}\right)$$

The following theorem from Khatri and Rao (1972), which generalizes the Skitovich-Darmois theorem and Ghurye and Olkin (1962), is used to prove identification in Theorem 2.2.

Theorem A.2. *(Khatri and Rao, 1972). Let φ_j be a continuous complex valued function of a real p_j-vector variable, for $j = 1, \ldots, q$. Let $P_k(\boldsymbol{s})$ be a polynomial of degree k in $\boldsymbol{s} \in \mathbb{R}^T$ in a neighborhood around the origin. Let A_j be a $T \times p_j$ matrix such that*

$$\sum_{j=1}^{q} \varphi_j(A_j'\boldsymbol{s}) = P_k(\boldsymbol{s}) \tag{16}$$

If Rank$(A_j) = p_j$, for $j = 1, \ldots, q$ and Rank$(A_j : A_{j'}) = p_j + p_{j'}$, for $j = 1, \ldots, r$ and $j' = r+1, \ldots, q$, then $\varphi_1(A_1'\boldsymbol{s}) + \ldots + \varphi_r(A_r'\boldsymbol{s})$ and $\varphi_{r+1}(A_{r+1}'\boldsymbol{s}) + \ldots + \varphi_q(A_q'\boldsymbol{s})$ are polynomials in $\boldsymbol{s}$ of degree at most $\max\{k, q-2\}$.

We provide some intuition for the proof, which repetitively applies the following two steps. First, by the rank conditions there is always an $\boldsymbol{s} = H_{11}\boldsymbol{u} + H_{21}\boldsymbol{v}$ where $H_1 = (H_{11} : \boldsymbol{\Pi}_1)$ is nonsingular and $A_1'H_{11} = I_{p_1}$ and $A_1'\boldsymbol{\Pi}_1 = \boldsymbol{0}$. Substituting this choice of $\boldsymbol{s}$ into Equation (16) leads to,

$$\varphi_1(\boldsymbol{u}) + \sum_{j=2}^{q} \varphi_j(A_j'(H_{11}\boldsymbol{u} + \boldsymbol{\Pi}_1\boldsymbol{v})) = P_k(H_{11}\boldsymbol{u} + \boldsymbol{\Pi}_1\boldsymbol{v}).$$

Now, φ_1 has fewer input variables than all the other functions. Second, φ_1 can be eliminated from the equation by taking first differences (or partial derivatives) with respect to $\boldsymbol{v}$. The resulting equation will satisfy the same rank conditions in the statement of the theorem except will have one less function. Repetitively using these two steps eliminates the functions $\varphi_1,\ldots,\varphi_r$ and results in an equation that is the sum of differences (or partial derivatives) of $\varphi_{r+1},\ldots,\varphi_q$ and equals a polynomial. Using the rank conditions, the difference equation can be recursively integrated out to give the result of the theorem.

A.2 Proof of Lemma 2.1

By Assumption 2.3, A_j has no right equivalent matrix if and only if A_j' has no left equivalent matrix if and only if A_j' is identified by its null-space, i.e., if $\ker(\widetilde{A}_j') = \ker(A_j')$ then $\widetilde{A}_j' = A_j'$ (see Robbin, 2001). Assumption 2.3 and $\dim(\ker(A_j')) = T - p_j$ (by the rank-nullity theorem) imply that A_j is the unique solution to $A_j'\mathbf{\Pi} = \mathbf{0}$ where $\mathbf{\Pi} = \left(\boldsymbol{\pi}_1 : \ldots : \boldsymbol{\pi}_{T-p_j}\right)$ and $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p_j}\}$ is any basis of $\ker(A')$ or equivalently A_j' uniquely solves $A_j'\boldsymbol{\pi} = 0$ for all $\boldsymbol{\pi} \in \ker(A_j')$.

Further, $A_j'\mathbf{\Pi} = \mathbf{0}$ provides $p_j(T-p_j)$ identifying equations; $T-p_j$ basis vectors each providing p_j identifying equations. Hence, for identification from this system the dimension of the unknown parameters in A_j can be at most $p_j(T-p_j)$.

A.3 Counterexamples: the difference in log characteristic functions is a polynomial but the underlying random variables are not normal

By Theorem A.1, if the log characteristic function of a random variable is a polynomial then the underlying random variable is normal. Nevertheless, if the difference in log characteristic functions is a polynomial then the underlying random variables need not be normal. The following examples are from Rao (1966) who attributes them to Dr. B. Ramachandran:

1. Consider the log characteristic functions $\varphi_{F_1}(s) = -|s| - s^4/4$ and $\varphi_{\widetilde{F}_1}(s) = -|s|$. Then $\varphi_{F_1}(s) - \varphi_{\widetilde{F}_1}(s) = -s^4/4$ is a quartic polynomial.
2. Consider the log characteristic functions,

$$\varphi_{F_1}(s) = \begin{cases} -2|t| - \frac{1}{2}t^2 & |t| \le \delta, \\ -\delta - \frac{1}{2}\delta^2 - |t| & |t| > \delta \end{cases} \qquad \varphi_{\widetilde{F}_1}(s) = \begin{cases} -2|t| & |t| \le \delta, \\ -\delta - |t| & |t| > \delta \end{cases}$$

   Then, $\varphi_{F_1}(s) - \varphi_{\widetilde{F}_1}(s) = \left\{ \begin{array}{ll} -\frac{1}{2}t^2 & |t| \le \delta, \\ -\frac{1}{2}\delta^2 & |t| > \delta \end{array} \right\}$ is a quadratic polynomial only in a neighborhood of the origin.

A.4 Proof of Theorems 2.1 and 2.2

Let $(\widetilde{A}, \widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}})$ be observationally equivalent to $(A, \boldsymbol{F}, \boldsymbol{U})$. The log characteristic function of Equation (1) is

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = \sum_{j=1}^{q} \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) + \sum_{t=1}^{T} \varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) = \sum_{j=1}^{q} \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_j'\boldsymbol{s}) + \sum_{t=1}^{T} \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s}),$$

where Assumption 2.1 was used to separate log characteristic functions. Hence,

$$\varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_j'\boldsymbol{s}) + \sum_{j'\neq j} \varphi_{\boldsymbol{F}_{j'}}(A_{j'}'\boldsymbol{s}) + \sum_{t=1}^{T} \varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \sum_{j'\neq j} \varphi_{\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{A}_{j'}'\boldsymbol{s}) - \sum_{t=1}^{T} \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s}) = 0. \quad (17)$$

By Assumption 2.2, $\text{Rank}(A_j) = p_j$, $\text{Rank}(A_j : A_{j'}) = \text{Rank}(A_j : \widetilde{A}_{j'}) = p_j + p_{j'}$, for $j' \neq j$, and $\text{Rank}(A_j, \boldsymbol{e}_t) = p_j + 1$ for all t. Applying Theorem A.2 to Equation (17),

$$\begin{aligned} \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_j'\boldsymbol{s}) - \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) &= \sum_{j'\neq j} (\varphi_{\boldsymbol{F}_{j'}}(A_{j'}'\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{A}_{j'}'\boldsymbol{s})) + \sum_{t=1}^{T} (\varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s})) \\ &= P_{2q+2T-2}(\boldsymbol{s}), \quad (18) \\ \varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s}) &= P_{2q+2T-2}(\boldsymbol{s}). \quad (19) \end{aligned}$$

Hence, conditions (i) and (ii) of Theorem 2.1 hold. Let $\boldsymbol{s} = \widetilde{H}\boldsymbol{u} + \widetilde{\boldsymbol{\Pi}}\boldsymbol{v}$, where $\boldsymbol{s} \in \mathbb{R}^T$, $\boldsymbol{u} \in \mathbb{R}^{p_j}$, $\boldsymbol{v} \in \mathbb{R}^{T-p_j}$, $\widetilde{H}$ is a $T \times p_j$ matrix, $\widetilde{\boldsymbol{\Pi}}$ is a $T \times (T - p_j)$ matrix with $(\widetilde{H} : \widetilde{\boldsymbol{\Pi}})$ nonsingular and $\widetilde{A}_j'\widetilde{H} = I_{p_j}$ and $\widetilde{A}_j'\widetilde{\boldsymbol{\Pi}} = \boldsymbol{0}$. The columns of $\widetilde{\boldsymbol{\Pi}}$ are a basis for $\ker(\widetilde{A}_j')$ and $\widetilde{H}$ is a right pseudoinverse of $\widetilde{A}_j'$. Plugging into Equation (18),

$$\begin{aligned} \varphi_{\widetilde{\boldsymbol{F}}_j}(\boldsymbol{u}) - \varphi_{\boldsymbol{F}_j}(A_j'\widetilde{H}\boldsymbol{u} + A_j'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) &= \sum_{j'\neq j} (\varphi_{\boldsymbol{F}_{j'}}(A_{j'}'\widetilde{H}\boldsymbol{u} + A_{j'}'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) - \varphi_{\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{A}_{j'}'\widetilde{H}\boldsymbol{u} + \widetilde{A}_{j'}'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v})) \\ &\quad + \sum_{t=1}^{T} (\varphi_{U_t}(\boldsymbol{e}_t'\widetilde{H}\boldsymbol{u} + \boldsymbol{e}_t'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\widetilde{H}\boldsymbol{u} + \boldsymbol{e}_t'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v})) \\ &= P_{2q+2T-2}(\boldsymbol{u}, \boldsymbol{v}). \quad (20) \end{aligned}$$

Let $\boldsymbol{v} = \boldsymbol{0}$,

$$\varphi_{\widetilde{\boldsymbol{F}}_j}(\boldsymbol{u}) - \varphi_{\boldsymbol{F}_j}(A_j'\widetilde{H}\boldsymbol{u}) = \sum_{j'\neq j} (\varphi_{\boldsymbol{F}_{j'}}(A_{j'}'\widetilde{H}\boldsymbol{u}) - \varphi_{\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{A}_{j'}'\widetilde{H}\boldsymbol{u})) + \sum_{t=1}^{T} (\varphi_{U_t}(\boldsymbol{e}_t'\widetilde{H}\boldsymbol{u}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\widetilde{H}\boldsymbol{u})) = P_{2q+2T-2}(\boldsymbol{u}).$$

Hence, condition (iii) of Theorem 2.1 holds. Now plug the above equation into Equation (18) and let $\boldsymbol{s} = H\boldsymbol{u}$, where $A_j'H = I_{p_j}$,

$$\varphi_{\boldsymbol{F}_j}(A_j'\widetilde{H}\widetilde{A}_j'H\boldsymbol{u}) - \varphi_{\boldsymbol{F}_j}(\boldsymbol{u}) = P_{2q+2T-2}(\boldsymbol{u}),$$

Hence, condition (iv) of Theorem 2.1 holds. Now let $\boldsymbol{u} = \boldsymbol{0}$ in Equation (20),

$$-\varphi_{\boldsymbol{F}_j}(A_j'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) = \sum_{j'\neq j}(\varphi_{\boldsymbol{F}_{j'}}(A_{j'}'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) - \varphi_{\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{A}_{j'}'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v})) + \sum_{t=1}^{T}(\varphi_{U_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\Pi}}\boldsymbol{v})) = P_{2q+2T-2}(\boldsymbol{v})$$

Hence, $\varphi_{\widetilde{\boldsymbol{\Pi}}'A_j\boldsymbol{F}_j}(\boldsymbol{v}) = P_{2q+2T-2}(\boldsymbol{v})$ is a polynomial and by Theorem A.1, the polynomial is quadratic and $\widetilde{\boldsymbol{\Pi}}'A_j\boldsymbol{F}_j$ is jointly normal. Hence, condition (v) of Theorem 2.1 holds.

Now let $\widetilde{\boldsymbol{\pi}} \in \ker(\widetilde{A}_j')$ and substitute $\boldsymbol{s} = \widetilde{\boldsymbol{\pi}}v$, where $v \in \mathbb{R}$, into Equations (18) and (19),

$$\begin{aligned}
\varphi_{\widetilde{\boldsymbol{\pi}}'A_j\boldsymbol{F}_j}(v) &= P_{2q+2T-2}(\widetilde{\boldsymbol{\pi}}v), & \\
\varphi_{\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{\boldsymbol{\pi}}v) - \varphi_{A_{j'}\boldsymbol{F}_{j'}}(\widetilde{\boldsymbol{\pi}}v) &= P_{2q+2T-2}(v), & j' \neq j, \\
\varphi_{U_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\pi}}v) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\pi}}v) &= P_{2q+2T-2}(v), & t = 1, \ldots, T.
\end{aligned}$$

By Theorem A.1, $\widetilde{\boldsymbol{\pi}}'A_j\boldsymbol{F}_j$ is normal. If Assumptions 2.4(i) and 2.4(ii)(a) hold then $\widetilde{\boldsymbol{\pi}}'A_j\boldsymbol{F}_j$ cannot be normal or degenerate. Hence, $A_j'\widetilde{\boldsymbol{\pi}} = \boldsymbol{0}$ for $\widetilde{\boldsymbol{\pi}} \in \ker(\widetilde{A}_j')$ and by Assumption 2.3 and Lemma 2.1, $\widetilde{A}_j = A_j$. Now, if instead Assumption 2.4(ii)(b) holds then,

$$\varphi_{\widetilde{\boldsymbol{\pi}}'A_j\boldsymbol{F}_j}(v) = \sum_{j'\neq j}(\varphi_{A_{j'}\boldsymbol{F}_{j'}}(\widetilde{\boldsymbol{\pi}}v) - \varphi_{\widetilde{A}_{j'}\widetilde{\boldsymbol{F}}_{j'}}(\widetilde{\boldsymbol{\pi}}v)) + \sum_{t=1}^{T}(\varphi_{U_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\pi}}v) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\widetilde{\boldsymbol{\pi}}v)) = P_1(v)$$

is degenerate. Hence, $A_j'\widetilde{\boldsymbol{\pi}} = \boldsymbol{0}$ for $\widetilde{\boldsymbol{\pi}} \in \ker(\widetilde{A}_j')$ and by Assumption 2.3 and Lemma 2.1, $\widetilde{A}_j = A_j$. This proves Theorem 2.2.

A.5 Proof of Corollary 2.1

No column A_j is a multiple of another column $A_{j'}$ so $\text{Rank}(A_j : A_{j'}) = 2$. Only one column of the observationally equivalent matrix $\widetilde{A}$, say $\widetilde{A}_j$, can satisfy $\text{Rank}(A_j : \widetilde{A}_j) = 1$. Each column A_j has at least two nonzero elements so $\text{Rank}(A_j : \boldsymbol{e}_t) = 2$ for all t. Hence, Assumption 2.2 holds.

Denote the element in the t^{th} row of A_j by a_{tj}. Assume, without loss of generality, that $a_{1j} \neq 0$. Let $\boldsymbol{\pi}' = (a_{tj}, 0, \ldots, 0, -a_{1j}, 0, \ldots, 0)$. Then, $A_j'\boldsymbol{\pi} = 0$ so $\boldsymbol{\pi} \in \ker(A_j')$. If A_j and $\widetilde{A}_j$ are right equivalent then $\boldsymbol{\pi} \in \ker(\widetilde{A}_j')$. Hence, $\widetilde{A}_j'\boldsymbol{\pi} = \widetilde{a}_{1j}a_{tj} - \widetilde{a}_{tj}a_{1j} = 0$ and $\widetilde{a}_{tj} = c_j a_{tj}$ where $c_j = \frac{\widetilde{a}_{1j}}{a_{1j}}$, for $t = 1, \ldots, T$. Hence, Assumption 2.3 holds up to the scale normalization c_j. The corollary follows by Theorem 2.2.

A.6 Proof of Theorem 3.1

Let

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ \beta_1 & \dots & \dots & \beta_p \end{pmatrix}$$

$\text{Rank}(A, \boldsymbol{e}_t) = p$, for $t = 1, \dots, p$, and $\text{Rank}(A, \boldsymbol{e}_{p+1}) = p$ because $\beta_t \neq 0$ so Assumption 2.2 holds.

Let $\boldsymbol{\pi}' = (-\beta_1, \dots, -\beta_p, 1)$. Then, $A'\boldsymbol{\pi} = \mathbf{0}$ so $\boldsymbol{\pi} \in \ker(A')$. If A and $\widetilde{A}$ are right equivalent then $\boldsymbol{\pi} \in \ker(\widetilde{A}')$. Hence, $\widetilde{A}'\boldsymbol{\pi} = (\widetilde{\beta}_1 - \beta_1, \dots, \widetilde{\beta}_p - \beta_p)' = \mathbf{0}$ and $\widetilde{\beta}_k = \beta_k$ for $k = 1, \dots, p$. Hence, Assumption 2.3 holds. The theorem follows by Corollary 2.2.

A.7 Proof of Theorem 3.2

$\text{Rank}(A_1 : A_3) = \text{Rank}(\widetilde{A}_1 : A_3) = 2r$, $\text{Rank}(A_2 : A_3) = \text{Rank}(\widetilde{A}_2 : A_3) = 2r - 1$ and $\text{Rank}(A_3 : \boldsymbol{e}_{2r}) = r + 1$ because $\theta_k \neq 0$, for $k = 1, \dots, r$, and $\theta_r - \theta_{r-1}\theta_1 - \theta_{r-2}(\theta_2 - \theta_1^2) - \dots \neq 0$ so Assumption 2.2 holds.[9]

Let $\boldsymbol{\pi}' = (\mathbf{0}, \theta_r - \theta_1\theta_{r-1} + \cdots, \dots, \theta_2 - \theta_1^2, \theta_1, -1)$. Then $A_3'\boldsymbol{\pi} = \mathbf{0}$ so $\boldsymbol{\pi} \in \ker(A_3)$. If A_3 and $\widetilde{A}_3$ are right equivalent then $\boldsymbol{\pi} \in \ker(\widetilde{A}_3')$. Hence,

$$\widetilde{A}_3'\boldsymbol{\pi} = \begin{pmatrix} \mathbf{0} & 1 & \widetilde{\theta}_1 & \dots & \widetilde{\theta}_r \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & 0 & 1 & \widetilde{\theta}_1 \end{pmatrix} \begin{pmatrix} \theta_r - \theta_1\theta_{r-1} + \dots \\ \vdots \\ \theta_2 - \theta_1^2 \\ \theta_1 \\ -1 \end{pmatrix} = \mathbf{0}.$$

From the last row, $\widetilde{\theta}_1 = \theta_1$. Then from the second last row, $\widetilde{\theta}_2 = \theta_2$ and recursively, $\widetilde{\theta}_k = \theta_k$ for $k = 1, \dots, r$. Hence, Assumption 2.3 holds. The theorem follows by Theorem 2.2.

[9] For $\text{Rank}(\widetilde{A}_2 : A_3) = 2r - 1$ we need a Hessenberg matrix to have full-rank (see Inselberg, 1978)

A.8 Proof of Theorem 4.1

Let $(\widetilde{A}, \widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}})$ be observationally equivalent to $(A, \boldsymbol{F}, \boldsymbol{U})$. The log characteristic function of Equation (1) is

$$\begin{aligned}\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) &= \sum_{j=1}^{q} \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) + \varphi_{U_1,\ldots,U_{\bar{t}}}((I_{\bar{t}} : \mathbf{0})\boldsymbol{s}) + \sum_{t=\bar{t}+1}^{T} \varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) \\ &= \sum_{j=1}^{q} \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_j'\boldsymbol{s}) + \varphi_{\widetilde{U}_1,\ldots,\widetilde{U}_{\bar{t}}}((I_{\bar{t}} : \mathbf{0})\boldsymbol{s}) + \sum_{t=\bar{t}+1}^{T} \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s}).\end{aligned}$$

where Assumption 4.1(i) was used to separate the log characteristic functions. Hence,

$$\sum_{j=1}^{q} \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) - \sum_{j=1}^{q} \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_j'\boldsymbol{s}) + \varphi_{\widetilde{\boldsymbol{F}}_{q+1}}(\widetilde{A}_{q+1}'\boldsymbol{s}) + \sum_{t=\bar{t}+1}^{T} (\varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s})) = 0, \tag{21}$$

where $\varphi_{\widetilde{\boldsymbol{F}}_{q+1}}(\widetilde{A}_{q+1}'\boldsymbol{s}) = \varphi_{U_1,\ldots,U_{\bar{t}}}(\widetilde{A}_{q+1}'\boldsymbol{s}) - \varphi_{\widetilde{U}_1,\ldots,\widetilde{U}_{\bar{t}}}(\widetilde{A}_{q+1}'\boldsymbol{s})$ and $\widetilde{A}_{q+1}' = (I_{\bar{t}} : \mathbf{0})$.

By Assumption 4.1(ii), $\text{Rank}(A_j) = p_j$, $\text{Rank}(A_j : \widetilde{A}_{j'}) \geq \max\{p_j, p_{j'}\} + 1$, $\text{Rank}(A_j : (I_{\bar{t}} : \mathbf{0})') \geq \max\{p_j, \bar{t}\} + 1$, $\text{Rank}(A_j : \boldsymbol{e}_t) = p_j + 1$, and, with at most one exception, $\text{Rank}(A_j : \widetilde{A}_{j'}) \geq \max\{p_j, \widetilde{p}_{j'}\} + 1$.

It is easy to see, applying Theorem A.2 for example, that $\varphi_{U_t}(\boldsymbol{e}_t'\boldsymbol{s}) - \varphi_{\widetilde{U}_t}(\boldsymbol{e}_t'\boldsymbol{s}) = P_{2q}(s_t)$ where $P_k(\boldsymbol{s})$ is a polynomial of degree k in $\boldsymbol{s}$.

We now follow Khatri and Rao (1972). Let $\boldsymbol{s} = H_{11}\boldsymbol{u} + \widetilde{\mathbf{\Pi}}_1\boldsymbol{v}$, where $\boldsymbol{s} \in \mathbb{R}^T$, $\boldsymbol{u} \in \mathbb{R}^{p_1}$, $\boldsymbol{v} \in \mathbb{R}^{T-p_1}$, H_{11} is a $T \times p_1$ matrix, $\widetilde{\mathbf{\Pi}}_1$ is a $T \times (T - p_1)$ matrix with $H_1 = (H_{11} : \widetilde{\mathbf{\Pi}}_1)$ nonsingular and

$$\widetilde{A}_1' H_{11} = I_{p_1} \text{ and } \widetilde{A}_1' \widetilde{\mathbf{\Pi}}_1 = \mathbf{0}.$$

Thus the columns of $\widetilde{\mathbf{\Pi}}_1$ are a basis for $\ker(\widetilde{A}_1')$. Note also that $\widetilde{\mathbf{\Pi}}_1$ is not empty because $T > p_1$ by Assumption 4.1(ii). Define

$$\begin{aligned}A_{j1} = H_1' A_j, \quad B_{j1} = \widetilde{\mathbf{\Pi}}_1' A_j, \quad \widetilde{A}_{j1} = H_1' \widetilde{A}_j, \quad \widetilde{B}_{j1} = \widetilde{\mathbf{\Pi}}_1' \widetilde{A}_j, \\ \varphi_{\boldsymbol{F}_j}^{(1)}(A_{j1}'\boldsymbol{s}) = \varphi_{\boldsymbol{F}_j}(A_{j1}'\boldsymbol{s} + B_{j1}'\boldsymbol{h}_1) - \varphi_{\boldsymbol{F}_j}(A_{j1}'\boldsymbol{s}), \\ \varphi_{\widetilde{\boldsymbol{F}}_j}^{(1)}(\widetilde{A}_{j1}'\boldsymbol{s}) = \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_{j1}'\boldsymbol{s} + \widetilde{B}_{j1}'\boldsymbol{h}_1) - \varphi_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}_{j1}'\boldsymbol{s}).\end{aligned}$$

Substituting the above into Equation (21),

$$\varphi_{\boldsymbol{F}_1}^{(1)}(A_{11}'\boldsymbol{s}) + \sum_{j=2}^{q} \varphi_{\boldsymbol{F}_j}^{(1)}(A_{j1}'\boldsymbol{s}) - \sum_{j=2}^{q+1} \varphi_{\widetilde{\boldsymbol{F}}_j}^{(1)}(\widetilde{A}_{j1}'\boldsymbol{s}) = P_{2q-1}(\boldsymbol{s}). \tag{22}$$

By Assumption 4.1(iii), $\widetilde{A}_1$ has no right equivalent matrix so if $B'_{11} = A'_1\widetilde{\mathbf{\Pi}}_1 = \mathbf{0}$ then $\widetilde{A}_1 = A_1$. Hence, one solution to Equation (21) is $\widetilde{A}_1 = A_1$. Now assume $\text{Rank}(B_{11}) \geq 1$. Equation (22) is of the same form as Equation (21) with one less function. Further, $\text{Rank}(A_{j1}) = \text{Rank}(A_j)$, $\text{Rank}(\widetilde{A}_{j1}) = \text{Rank}(\widetilde{A}_j)$, $\text{Rank}(A_{j1} : A_{j'1}) = \text{Rank}(A_j : A_{j'})$, $\text{Rank}(A_{j1} : \widetilde{A}_{j'1}) = \text{Rank}(A_j : \widetilde{A}_{j'})$, $\text{Rank}(B_{j1}) \geq 1$ and $\text{Rank}(\widetilde{B}_{j1}) \geq 1$ because ranks of matrices are retained after multiplication by the nonsingular matrix H_1. We now recursively eliminate $\varphi^{(\ell)}_{\widetilde{\boldsymbol{F}}_\ell}$ by substituting into the above equation,

$$\varphi^{(\ell)}_{\boldsymbol{F}_j}(A'_{j\ell}\boldsymbol{s}) = \varphi^{(\ell-1)}_{\boldsymbol{F}_j}(A'_{j\ell}\boldsymbol{s} + B'_{j\ell}\boldsymbol{h}_\ell) - \varphi^{(\ell-1)}_{\boldsymbol{F}_j}(A'_{j\ell}\boldsymbol{s}),$$
$$\varphi^{(\ell)}_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}'_{j\ell}\boldsymbol{s}) = \varphi^{(\ell-1)}_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}'_{j\ell}\boldsymbol{s} + \widetilde{B}'_{j\ell}\boldsymbol{h}_\ell) - \varphi^{(\ell-1)}_{\widetilde{\boldsymbol{F}}_j}(\widetilde{A}'_{j\ell}\boldsymbol{s}),$$

where $A_{j\ell} = H'_\ell A_{j\ell-1}$, $B_{j\ell} = \widetilde{\mathbf{\Pi}}'_\ell A_{j\ell-1}$, $\widetilde{A}_{j\ell} = H'_\ell \widetilde{A}_{j\ell-1}$, $\widetilde{B}_{j\ell} = \widetilde{\mathbf{\Pi}}'_\ell \widetilde{A}_{j\ell-1}$ and $H_\ell = (H_{\ell 1} : \widetilde{\mathbf{\Pi}}_\ell)$ is nonsingular and $\widetilde{A}'_{\ell\ell-1} H_{\ell 1} = I_{p_\ell}$ and $A'_{\ell\ell-1}\widetilde{\mathbf{\Pi}}_\ell = \mathbf{0}$. Further, as before, by the nonsingularity of H_ℓ all the rank conditions remain unchanged. After $q+1$ eliminations,

$$\varphi^{(q+1)}_{\boldsymbol{F}_1}(A'_{1q+1}\boldsymbol{s}) + \sum_{j=2}^{q} \varphi^{(q+1)}_{\boldsymbol{F}_j}(A'_{jq+1}\boldsymbol{s}) = P_{q-1}(\boldsymbol{s}).$$

As above, we recursively eliminate $\varphi^{(q+\ell)}_{\boldsymbol{F}_\ell}$, $\ell = 2, \ldots, q$, by substituting into the above equation,

$$\varphi^{(q+\ell)}_{\boldsymbol{F}_j}(A'_{jq+\ell}\boldsymbol{s}) = \varphi^{(q+\ell-1)}_{\boldsymbol{F}_j}(A'_{jq+\ell}\boldsymbol{s} + B'_{jq+\ell}\boldsymbol{h}_{q+\ell}) - \varphi^{(q+\ell-1)}_{\boldsymbol{F}_j}(A'_{jq+\ell}\boldsymbol{s}),$$

where $A_{jq+\ell} = H'_{q+\ell} A_{jq+\ell-1}$, $B_{jq+\ell} = \mathbf{\Pi}'_\ell A_{jq+\ell-1}$, and $H_{q+\ell} = (H_{q+\ell 1} : \mathbf{\Pi}_\ell)$ is nonsingular and $A'_{\ell q+\ell-1} H_{q+\ell 1} = I_{p_{\ell-1}}$ and $A'_{\ell q+\ell-1}\mathbf{\Pi}_\ell = \mathbf{0}$. After $q-1$ further eliminations,

$$\varphi^{(2q)}_{\boldsymbol{F}_1}(A'_{12q}\boldsymbol{s}) = P_0,$$

where this is a $(2q)^{th}$ difference equation and P_0 is some constant. Similarly we can get the difference equation $\varphi^{(2q)}_{\boldsymbol{F}_1}(A'_{12q}\boldsymbol{s}) = P_0$, for $j = 1, \ldots, p$ (and a difference equation for $\varphi_{U_1,\ldots,U_{\bar{t}}}(\cdot) - \varphi_{\widetilde{U}_1,\ldots,\widetilde{U}_{\bar{t}}}(\cdot)$).

A.9 Proof of Corollary 4.1

Let

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ \beta_1 & \dots & \dots & \beta_p \end{pmatrix}$$

Let $(\boldsymbol{F}, \boldsymbol{U}, V, A)$ and $(\widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}}, \widetilde{V}, \widetilde{A})$ be observationally equivalent. The log characteristic function of Equation (4) is

$$\varphi_{\boldsymbol{X}',Y}(\boldsymbol{s}) = \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \varphi_{\boldsymbol{U}}((I_p : \mathbf{0})\boldsymbol{s}) + \varphi_V(\boldsymbol{e}'_{p+1}\boldsymbol{s}) = \varphi_{\widetilde{\boldsymbol{F}}}(\widetilde{A}'\boldsymbol{s}) + \varphi_{\widetilde{\boldsymbol{U}}}((I_p : \mathbf{0})\boldsymbol{s}) + \varphi_{\widetilde{V}}(\boldsymbol{e}'_{p+1}\boldsymbol{s}),$$

where the independence assumptions are used to separate log characteristic functions. Hence,

$$\varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{F}}}(\widetilde{A}'\boldsymbol{s}) + \varphi_{\boldsymbol{U}}((I_p : \mathbf{0})\boldsymbol{s}) - \varphi_{\widetilde{\boldsymbol{U}}}((I_p : \mathbf{0})\boldsymbol{s}) + \varphi_V(\boldsymbol{e}'_{p+1}\boldsymbol{s}) - \varphi_{\widetilde{V}}(\boldsymbol{e}'_{p+1}\boldsymbol{s}) = 0. \tag{23}$$

Rank$(A, (I_p : \mathbf{0})') = p+1$ and Rank$(A, \boldsymbol{e}_{p+1}) = p+1$ so we can apply Theorem A.2, $\varphi_V(\boldsymbol{e}'_{p+1}\boldsymbol{s}) - \varphi_{\widetilde{V}}(\boldsymbol{e}'_{p+1}\boldsymbol{s}) = P_2(s_{p+1})$. The partial derivative of Equation (23) with respect to s_{p+1} is,

$$\sum_{k=1}^{p} \left(\beta_k \left. \frac{\partial \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{p+1}} \right|_{\boldsymbol{u}=(I_p:\mathbf{0})\boldsymbol{s}+\boldsymbol{\beta}s_{p+1}} - \widetilde{\beta}_k \left. \frac{\partial \varphi_{\widetilde{\boldsymbol{F}}}(\boldsymbol{u})}{\partial u_k} \right|_{\boldsymbol{u}=(I_p:\mathbf{0})\boldsymbol{s}+\widetilde{\boldsymbol{\beta}}s_{p+1}} \right) = P_1(s_{p+1}).$$

Substitute $\boldsymbol{s} = \boldsymbol{v} - \widetilde{\boldsymbol{\beta}}s_{p+1}$ into the above equation. The partial derivative with respect to s_{p+1} is,

$$\sum_{k_1,k_2} \beta_{k_1}(\beta_{k_2} - \widetilde{\beta}_{k_2}) \left. \frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}} \right|_{\boldsymbol{u}=(I_p:\mathbf{0})\boldsymbol{v}+(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})s_{p+1}} = P_0.$$

A.10 Proof of Theorem 5.1

Let

$$h_0(\boldsymbol{\pi}, v) = \left(\dots, \mathrm{Cov}(Y_{t_1}, Y_{t_2}) + \left. \frac{\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_{t_1} \partial s_{t_2}} \right|_{\boldsymbol{s}=\boldsymbol{\pi}v}, \dots \right),$$
$$H_0(\boldsymbol{\Pi}, v)' = \left(h_0(\boldsymbol{\pi}_1, v), \dots, h_0(\boldsymbol{\pi}_{T-p}, v) \right),$$

where $\mathbf{\Pi} = \text{vec}(\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p})$, $\boldsymbol{\pi}_j \in \mathbb{R}^T$ for $j = 1, \ldots, T-p$, and $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p}\}$ is a basis. Let

$$Q_n(\mathbf{\Pi}) = -\int_{\mathcal{V}} H_n(\mathbf{\Pi}, v)' W_n(v) H_n(\mathbf{\Pi}, v) dv,$$
$$Q_0(\mathbf{\Pi}) = -\int_{\mathcal{V}} H_0(\mathbf{\Pi}, v)' W(v) H_0(\mathbf{\Pi}, v) dv.$$

We now follow standard proofs of consistency and asymptotic normality of minimum distance estimators (e.g., Newey and McFadden, 1994).

Consistency. We show that Q_0 is maximized if and only if $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p}\} \in \text{Basis}(\ker(A'))$. Plugging Equation (8) into $h_0(\boldsymbol{\pi}, v)$,

$$\text{Cov}(Y_{t_1}, Y_{t_2}) + \frac{\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_{t_1} \partial s_{t_2}} = \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{t_1 k_1} a_{t_2 k_2} \left(\text{Cov}\left(F_{k_1}, F_{k_2}\right) + \left. \frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}} \right|_{\boldsymbol{u}=A'\boldsymbol{\pi} v} \right), \quad t_1 \neq t_2,$$

where a_{tk} represents the element in the t^{th} row and k^{th} column of the matrix A. Now if $\boldsymbol{\pi} \in \ker(A')$ then $\text{Cov}(F_{k_1}, F_{k_2}) + \left. \frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}} \right|_{\boldsymbol{u}=\mathbf{0}} = 0$ for all k_1 and k_2 so $h_0(\boldsymbol{\pi}, v) = \mathbf{0}$ and $H_0(\mathbf{\Pi}, v) = \mathbf{0}$. If $\boldsymbol{\pi} \notin \ker(A')$ but $\sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{t_1 k_1} a_{t_2 k_2} \left(\text{Cov}\left(F_{k_1}, F_{k_2}\right) + \left. \frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}} \right|_{\boldsymbol{u}=A'\boldsymbol{\pi} v} \right) = 0$ then by Assumption 2.2(ii), $\text{Cov}(F_{k_1}, F_{k_2}) + \left. \frac{\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{u})}{\partial u_{k_1} \partial u_{k_2}} \right|_{\boldsymbol{u}=A'\boldsymbol{\pi} v} = 0$ for all k_1, k_2 and v and so integrating out $\varphi_{\boldsymbol{\pi}' A \boldsymbol{F}}(u)$ is a quadratic. Hence, $\boldsymbol{\pi}' A \boldsymbol{F}$ is degenerate or normal, which violates Assumption 2.4(i) or 2.4(ii)(a). Hence, Q_0 achieves its maximum if and only if $\boldsymbol{\pi} \in \ker(A')$. Further, noting that $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p}\}$ is a basis by assumption and $\dim(\ker(A_j')) = T - p_j$ by the rank-nullity theorem and Assumption 2.2, we get Q_0 is maximized if and only if $\{\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{T-p}\} \in \text{Basis}(\ker(A'))$.

Next, let $\mathbf{\Pi} \in \mathcal{P}$ be a compact set. We show that $\sup_{\mathbf{\Pi} \in \mathcal{P}} |Q_n(\mathbf{\Pi}) - Q_0(\mathbf{\Pi})| \xrightarrow{p} 0$,

$$|Q_n(\mathbf{\Pi}) - Q_0(\mathbf{\Pi})| = \left| \int_{\mathcal{V}} \left(H_0(\boldsymbol{\pi}_j, v)' W(v) H_0(\boldsymbol{\pi}_j, v) - H_n(\boldsymbol{\pi}_j, v)' W_n(v) H_n(\boldsymbol{\pi}_j, v)\right) dv \right|$$
$$\leq \int_{\mathcal{V}} (|(H_0(\boldsymbol{\pi}_j, v) - H_n(\boldsymbol{\pi}_j, v))' W_n(v) (H_0(\boldsymbol{\pi}_j, v) - H_n(\boldsymbol{\pi}_j, v))|$$
$$+ |H_0(\boldsymbol{\pi}_j, v))' (W_n(v) + W_n'(v)) (H_n(\boldsymbol{\pi}_j, v) - H_0(\boldsymbol{\pi}_j, v))| + |H_0(\boldsymbol{\pi}_j, v)' (W_n(v) - W(v)) H_0(\boldsymbol{\pi}_j, v)|) dv$$
$$\leq \int_{\mathcal{V}} (||(H_0(\boldsymbol{\pi}_j, v) - H_n(\boldsymbol{\pi}_j, v)||^2 \; ||W_n(v)|| + 2||H_0(\boldsymbol{\pi}_j, v))|| \; ||W_n(v)|| \; ||H_n(\boldsymbol{\pi}_j, v) - H_0(\boldsymbol{\pi}_j, v))||$$
$$+ ||H_0(\boldsymbol{\pi}_j, v)||^2 \; ||W_n(v) - W(v)||) dv$$
$$\leq \sup_{v \in \mathcal{V}} ||(H_0(\boldsymbol{\pi}_j, v) - H_n(\boldsymbol{\pi}_j, v)||^2 \; \sup_{v \in \mathcal{V}} ||W_n(v)|| + \sup_{v \in \mathcal{V}} ||H_0(\boldsymbol{\pi}_j, v)||^2 \; \sup_{\boldsymbol{v} \in \mathcal{V}} ||W_n(v) - W(v)||$$
$$+ 2||H_0(\boldsymbol{\pi}_j, v))|| \; \sup_{v \in \mathcal{V}} ||W_n(v)|| \; \sup_{v \in \mathcal{V}} ||H_n(\boldsymbol{\pi}_j, v) - H_0(\boldsymbol{\pi}_j, v))||$$

where we repeatedly used the Cauchy-Schwartz and triangle inequalities. Hence, $\sup_{\mathbf{\Pi} \in \mathcal{P}} |Q_n(\mathbf{\Pi}) - Q_0(\mathbf{\Pi})| \to 0$ because $\mathcal{V}$ and $\mathcal{P}$ are compact, $W_n(v)$ is nonsingular and $\sup_{v \in \mathcal{V}} ||W_n(v) - W(v)|| \xrightarrow{p} 0$.

We now show consistency using (i) $Q_0(\mathbf{\Pi})$ is maximized if and only if $\mathbf{\Pi} = \mathbf{\Pi}_0$, where $\{\boldsymbol{\pi}_{01}, \ldots, \boldsymbol{\pi}_{0T-p}\} \in \text{Basis}(\ker(A'))$, (ii) $\mathcal{P}$ is compact, (iii) $Q_0(\mathbf{\Pi})$ is continuous and (iv) $\sup_{\mathbf{\Pi}\in\mathcal{P}} |Q_n(\mathbf{\Pi}) - Q_0(\mathbf{\Pi})| \to 0$.

For any $\varepsilon > 0$ we have,

$$Q_n(\widehat{\mathbf{\Pi}}) > Q_n(\mathbf{\Pi}_0) - \varepsilon/3, \qquad Q_0(\widehat{\mathbf{\Pi}}) > Q_n(\widehat{\mathbf{\Pi}}) - \varepsilon/3, \qquad Q_n(\mathbf{\Pi}_0) > Q_0(\mathbf{\Pi}_0) - \varepsilon/3$$

where the first inequality follows because $\widehat{\mathbf{\Pi}}$ maximizes Q_n and the second and third inequalities follow by (iv). Therefore,

$$Q_0(\widehat{\mathbf{\Pi}}) > Q_n(\widehat{\mathbf{\Pi}}) - \varepsilon/3 > Q_n(\mathbf{\Pi}_0) - 2\varepsilon/3 > Q_0(\mathbf{\Pi}_0) - \varepsilon$$

Now let $\mathcal{N}$ be an open subset containing all $\mathbf{\Pi}_0 \in \text{Basis}(\ker(A'))$. By (II), $\mathcal{P}\cap\mathcal{N}^C$ is compact so there exists $\mathbf{\Pi}^* \in \mathcal{P}\cap\mathcal{N}^C$ such that $Q_0(\mathbf{\Pi}^*) = \sup_{\mathbf{\Pi}\in\mathcal{P}\cap\mathcal{N}^C} Q_0(\mathbf{\Pi})$. Further by (I), $Q_0(\mathbf{\Pi}^*) < Q_0(\mathbf{\Pi}_0)$ for $\mathbf{\Pi}_0 \in \text{Basis}(\ker(A'))$. Now choose $\varepsilon = Q_0(\mathbf{\Pi}_0) - Q_0(\mathbf{\Pi}^*)$ so that $Q_0(\widehat{\mathbf{\Pi}}) > Q_0(\mathbf{\Pi}^*)$. Hence, $\widehat{\mathbf{\Pi}} \in \mathcal{N}$. This proves consistency i.e., that $\mathbf{\Pi} \xrightarrow{p} \mathbf{\Pi}_0$ for some $\mathbf{\Pi}_0$ with $\{\boldsymbol{\pi}_{01}, \ldots, \boldsymbol{\pi}_{0T-p}\} \in \text{Basis}(\ker(A'))$.

Asymptotic normality. The first order conditions of Equation (9) are,

$$0 = \int_{\mathcal{V}} \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widehat{\mathbf{\Pi}}} \right)' W_n(v) H_n(\widehat{\mathbf{\Pi}}, v) dv,$$

where could switch the derivative and integral by the dominated convergence theorem.

Now using the mean value theorem

$$H_n(\widehat{\mathbf{\Pi}}, v) = H_n(\mathbf{\Pi}_0, v) + \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widetilde{\mathbf{\Pi}}} \right) \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right),$$

where $\widetilde{\mathbf{\Pi}}$ is located between $\widehat{\mathbf{\Pi}}$ and $\mathbf{\Pi}_0$. We then have

$$0 = \int_{\mathcal{V}} \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widehat{\mathbf{\Pi}}} \right)' W_n(v) \left[H_n(\mathbf{\Pi}_0, v) + \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widetilde{\mathbf{\Pi}}} \right) \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right) \right] dv.$$

Rearranging,

$$\begin{aligned} \sqrt{n} \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right) = &- \left[\int_{\mathcal{V}} \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widehat{\mathbf{\Pi}}} \right)' W_n(v) \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widetilde{\mathbf{\Pi}}} \right) dv \right]^{-1} \\ &\times \int_{\mathcal{V}} \left(\left.\frac{\partial H_n(\mathbf{\Pi}, v)}{\partial \mathbf{\Pi}}\right|_{\mathbf{\Pi}=\widehat{\mathbf{\Pi}}} \right)' W_n(v) \left(\sqrt{n} H_n(\mathbf{\Pi}_0, v) \right) dv. \end{aligned} \tag{24}$$

Now we show that $\sqrt{n} H_n(\mathbf{\Pi}_0, v)$ is asymptotically normal. By a Taylor expansion of $H_n(\mathbf{\Pi}_0, v)$

around $H_0(\mathbf{\Pi}_0, v) = \mathbf{0}$,

$$H_n(\mathbf{\Pi}_0, v)' = (\dots, h_n(\boldsymbol{\pi}_{0j}, v), \dots) = \left(\dots, \widehat{\mathrm{Cov}}(Y_{t_1}, Y_{t_2}) + \left. \frac{\partial^2 \widehat{\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}}{\partial s_{t_1} \partial s_{t_2}} \right|_{\boldsymbol{s}=\boldsymbol{\pi}_{0j} v}, \dots \right)$$

$$= \left(\dots, \frac{1}{n}\sum_{i=1}^n Y_{it_1} Y_{it_2} - \left(\frac{1}{n}\sum_{i=1}^n Y_{it_1}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2}\right) + \frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i}\right)^2} \right.$$
$$\left. - \frac{\frac{1}{n}\sum_{i=1}^n Y_{it_1} Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i}}{\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i}}, \dots \right)$$

$$= \left(\dots, \frac{E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2} \left(\frac{1}{n}\sum_{i=1}^n Y_{it_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right) \right.$$
$$+ \frac{E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2} \left(\frac{1}{n}\sum_{i=1}^n Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)$$
$$+ \frac{1}{E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]} \left(\frac{1}{n}\sum_{i=1}^n Y_{it_1} Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_1} Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)$$
$$+ \frac{E\left[Y_{t_1} Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right] E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right] - 2E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right] E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^3} \left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)$$
$$+ O\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2 + O\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2$$
$$\left. + O\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2 + O\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1} Y_{it_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}_i} - E\left[Y_{t_1} Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j}'\boldsymbol{Y}}\right]\right)^2, \dots \right).$$

Hence, by the central limit theorem which applies because of Assumptions 5.1 and 5.2(ii), the manipulation above and Lemma A.1 which eliminates the higher-order terms,

$$\sqrt{n} H_n(\mathbf{\Pi}_0, v) \xrightarrow{d} N(\mathbf{0}, \Omega(v)) \tag{25}$$

where the terms in $\Omega(v)$ come from the expansion above and take the form,

$$\mathrm{Cov}\left(\frac{E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]\right)^2} \cdot Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}} + \frac{E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]\right)^2} \cdot Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}} + \frac{1}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]\right)^2} \cdot Y_{t_1} Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}} \right.$$
$$+ \frac{E\left[Y_{t_1} Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right] E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right] - 2E\left[Y_{t_1} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right] E\left[Y_{t_2} e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}}\right]\right)^3} \cdot e^{\mathrm{i}v\boldsymbol{\pi}_{0j_1}'\boldsymbol{Y}},$$

$$\frac{E\left[Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]\right)^2}\cdot Y_{t_3}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}+\frac{E\left[Y_{t_3}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]\right)^2}\cdot Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}+\frac{1}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]\right)^2}\cdot Y_{t_3}Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}$$
$$\left.+\frac{E\left[Y_{t_3}Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]-2E\left[Y_{t_3}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]E\left[Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]}{\left(E\left[e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right]\right)^3}\cdot e^{\mathrm{i}v\boldsymbol{\pi}_{0j_2}'\boldsymbol{Y}}\right). \tag{26}$$

Next, we show that $\sup_{v\in\mathcal{V},\boldsymbol{\Pi}\in\mathcal{P}}\left|\frac{\partial H_n(\boldsymbol{\Pi},v)}{\partial\boldsymbol{\Pi}}-\frac{\partial H_0(\boldsymbol{\Pi},v)}{\partial\boldsymbol{\Pi}}\right|\xrightarrow{p}\mathbf{0}$. By a Taylor expansion of

$$\frac{\partial h_n(\boldsymbol{\pi},v)}{\partial\pi_{t_3}}=\frac{\partial}{\partial\pi_{t_3}}\left(\left.\frac{\partial^2\widehat{\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}}{\partial s_{t_1}\partial s_{t_2}}\right|_{\boldsymbol{s}=\boldsymbol{\pi}v}\right)$$
$$=iv\frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_3}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)^2}$$
$$+iv\frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2}Y_{it_3}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)^2}$$
$$-2iv\frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_2}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_3}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)^3}$$
$$-iv\frac{\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_2}Y_{it_3}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}}{\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}}+iv\frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_2}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)\left(\frac{1}{n}\sum_{i=1}^n Y_{it_3}e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)}{\left(\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\sum_{t=1}^T\pi_tY_{it}}\right)^2}$$

around $\frac{\partial h_0(\boldsymbol{\pi},v)}{\partial\pi_{t_3}}$,

$$\frac{\partial h_n(\boldsymbol{\pi},v)}{\partial\pi_{t_3}}=\frac{\partial h_0(\boldsymbol{\pi},v)}{\partial\pi_{t_3}}+O\left(\left|\frac{1}{n}\sum_{i=1}^n e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}_i}-E\left[e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}}\right]\right|\right)+O\left(\sum_{t_1}\left|\frac{1}{n}\sum_{i=1}^n Y_{it_1}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}_i}-E\left[Y_{t_1}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}}\right]\right|\right)$$
$$+O\left(\sum_{t_1,t_2}\left|\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_2}e^{\mathrm{i}\boldsymbol{\pi}'\boldsymbol{Y}_i}-E\left[Y_{t_1}Y_{t_2}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}}\right]\right|\right)$$
$$+O\left(\sum_{t_1,t_2,t_3}\left|\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_2}Y_{it_3}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}_i}-E\left[Y_{t_1}Y_{t_2}Y_{t_3}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}}\right]\right|\right)$$
$$+O\left(\sum_{t_1,t_2,t_3,t_4}\left|\frac{1}{n}\sum_{i=1}^n Y_{it_1}Y_{it_2}Y_{it_3}Y_{it_4}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}_i}-E\left[Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}e^{\mathrm{i}v\boldsymbol{\pi}'\boldsymbol{Y}}\right]\right|\right)$$

and by Assumption 5.2(ii) and Lemma A.1,

$$\sup_{v\in\mathcal{V},\boldsymbol{\Pi}\in\mathcal{P}}\left|\frac{\partial H_n(\boldsymbol{\Pi},v)}{\partial\boldsymbol{\Pi}}-\frac{\partial H_0(\boldsymbol{\Pi},v)}{\partial\boldsymbol{\Pi}}\right|\xrightarrow{p}\mathbf{0}. \tag{27}$$

Further by Assumption 5.2(iii),

$$\sup_{v\in\mathcal{V}} |W_n(v) - W(v)| \xrightarrow{p} \mathbf{0}. \tag{28}$$

Finally, plugging Equations (25), (27) and (28) into (24) and using consistency and Slutsky's Theorem,

$$\sqrt{n}\left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left[\int_{\mathcal{V}} G(v)'W(v)G(v)dv\right]^{-1} \int_{\mathcal{V}} G(v)'W(v)\Omega(v)W(v)G(v)dv \left[\int_{\mathcal{V}} G(v)'W(v)G(v)dv\right]^{-1}\right)$$

where $G(v) = \frac{\partial H_0(\mathbf{\Pi},v)}{\partial \mathbf{\Pi}}\Big|_{\mathbf{\Pi}=\mathbf{\Pi}_0}$.

A.11 Additional simulation results

In this section we report results from additional simulation studies. The data is generated from the model in Equation (10) using the same data generating process as in Section 5 but with the unobserved regressors drawn from $\chi^2(10)$, $Unif(-1,1)$, $t(10)$ or bimodal distributions. The results are qualitatively similar: the RMSEs for the GMM and PD estimators are small and relatively stable for different amounts of measurement error, the PD estimator is the most robust to distributional choices and measurement error, and the OLS estimator has smallest RMSEs when there is no measurement error and the largest RMSEs even with small amounts of measurement error.

Table 4: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1,1)$ in Model (10) with F_1 and F_2 coming from $t(10)$ distributions.

			β_1			β_2	
Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
PD	0	0.02	0.08	0.08	0.01	0.10	0.10
	1/4	0.03	0.09	0.09	0.02	0.09	0.09
	1	0.02	0.06	0.06	0.00	0.08	0.08
GMM	0	0.03	0.15	0.15	-0.03	0.18	0.18
	1/4	-0.03	0.19	0.19	-0.10	0.25	0.26
	1	-0.10	0.21	0.22	-0.19	0.27	0.33
OLS	0	0.00	0.02	0.02	0.00	0.02	0.02
	1/4	-0.15	0.03	0.15	-0.14	0.03	0.14
	1	-0.25	0.03	0.26	-0.25	0.03	0.25

Notes: Results from 100 simulations of sample size 1,000. U_1 and U_2 are drawn from independent normal distributions with variance σ_U^2.

Table 5: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1, 1)$ in Model (10) with F_1 and F_2 coming from $Unif(-1, 1)$ distributions.

			β_1			β_2	
Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
PD	0	0.02	0.07	0.07	0.02	0.08	0.08
	1/4	0.02	0.06	0.07	0.02	0.06	0.06
	1	0.01	0.05	0.05	0.02	0.06	0.06
GMM	0	0.00	0.03	0.03	0.01	0.04	0.04
	1/4	-0.01	0.07	0.07	0.01	0.08	0.08
	1	-0.01	0.10	0.10	-0.05	0.13	0.13
OLS	0	0.00	0.02	0.02	0.00	0.03	0.03
	1/4	-0.14	0.03	0.14	-0.14	0.03	0.14
	1	-0.25	0.03	0.25	-0.25	0.03	0.25

Notes: Results from 100 simulations of sample size $1,000$. U_1 and U_2 are drawn from independent normal distributions with variance σ_U^2.

Table 6: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1, 1)$ in Model (10) with F_1 and F_2 coming from $\frac{1}{2}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(2, 1)$, i.e., bimodal distributions.

			β_1			β_2	
Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
PD	0	0.00	0.09	0.09	0.02	0.09	0.10
	1/4	0.02	0.07	0.07	0.01	0.06	0.06
	1	0.02	0.07	0.07	0.02	0.07	0.07
GMM	0	0.00	0.04	0.04	0.01	0.04	0.04
	1/4	-0.04	0.07	0.07	0.00	0.09	0.09
	1	-0.01	0.10	0.10	-0.04	0.13	0.14
OLS	0	0.00	0.03	0.03	0.00	0.03	0.03
	1/4	-0.14	0.03	0.15	-0.15	0.03	0.15
	1	-0.25	0.03	0.25	-0.25	0.03	0.25

Notes: Results from 100 simulations of sample size $1,000$. U_1 and U_2 are drawn from independent normal distributions with variance σ_U^2.

Table 7: Performances of the PD, GMM and OLS estimators for $(\beta_1, \beta_2) = (1, 1)$ in Model (10) with F_1 and F_2 coming from $\chi^2(10)$ distributions.

		β_1			β_2		
Estimator	σ_U^2	bias	SD	RMSE	bias	SD	RMSE
PD	0	0.00	0.06	0.06	-0.01	0.08	0.08
	1/4	0.01	0.09	0.09	0.00	0.13	0.13
	1	0.02	0.13	0.13	-0.01	0.18	0.18
GMM	0	0.00	0.06	0.06	-0.04	0.09	0.09
	1/4	0.01	0.10	0.10	-0.06	0.13	0.15
	1	0.00	0.14	0.14	-0.10	0.18	0.20
OLS	0	0.00	0.03	0.03	0.00	0.03	0.03
	1/4	-0.14	0.03	0.15	-0.14	0.03	0.14
	1	-0.25	0.03	0.25	-0.25	0.03	0.25

Notes: Results from 100 simulations of sample size $1,000$. U_1 and U_2 are drawn from independent normal distributions with variance σ_U^2.