Econometric Theory, **34**, 2018, 134–165.
doi:10.1017/S026646661700007X

IDENTIFICATION OF JOINT DISTRIBUTIONS IN DEPENDENT FACTOR MODELS

DAN BEN-MOSHE
The Hebrew University of Jerusalem

This paper studies linear factor models that have arbitrarily dependent factors. Assuming that the coefficients are known and that their matrix representation satisfies rank conditions, we identify the nonparametric joint distribution of the unobserved factors using first and then second-order partial derivatives of the log characteristic function of the observed variables. In conjunction with these identification strategies the mean and variance of the vector of factors are identified. The main result provides necessary and sufficient conditions for identification of the joint distribution of the factors. In an illustrative example, we show identification of an earnings dynamics model with a subset of arbitrarily dependent income shocks. Closed-form formulas lead to estimators that converge uniformly and despite being based on inverse Fourier transforms have tight confidence bands around their theoretical counterparts in Monte Carlo simulations.

1. INTRODUCTION

In this paper, we consider the linear factor model,

$$\boldsymbol{Y} = A\boldsymbol{F} + \boldsymbol{U}, \tag{1}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_P)'$ is a vector of observed variables, $\boldsymbol{F} = (F_1, \ldots, F_K)'$ is a vector of unobserved factors, $\boldsymbol{U} = (U_1, \ldots, U_P)'$ is a vector of unobserved errors and A is a $P \times K$ matrix of known coefficients. We identify the nonparametric joint distribution of an arbitrarily dependent $\boldsymbol{F}$.

The starting assumption of the paper is that the matrix of coefficients A is either known or both identified and can be root-N consistently estimated. There is a large literature that investigates identification and estimation of these coefficients assuming uncorrelated or mutually independent unobservables using, for example, principal component analysis (see, e.g., Anderson and Rubin, 1956; Jöreskog, 1970; Jolliffe, 2002) or independent component analysis (see, e.g.,

This paper is based on the first chapter of my PhD Thesis and is a revised version of my job market paper. I am grateful to Rosa Matzkin and Jinyong Hahn for their generous support, advice, and guidance. I benefited greatly from the feedback of two anonymous referees and the editor Liangjun Su, comments at seminars at Haifa University, Hebrew University of Jerusalem, Tel Aviv University, UCL and UCLA and discussions with Stefan Hoderlein, Max Kasy, Arthur Lewbel, Áureo de Paula, Daniel Wilhelm and Martin Weidner. The support of a grant from the Maurice Falk Institute for Economic Research is gratefully acknowledged. Address correspondence to Dan Ben-Moshe, Department of Economics, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91905, Israel; e-mail: danbm@huji.ac.il.

 © Cambridge University Press 2017

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Hyvärinen, Karhunen, and Oja, 2004; Bonhomme and Robin, 2009; Comon and Jutten, 2010). Two recent papers that identify and root-N consistently estimate A while allowing for statistically dependent factors are Ben-Moshe (2016) and Williams (2015). They motivate the importance of allowing for dependent factors with models of child development, e.g., Conti, Heckman, and Urzua (2010) and Carneiro, Hansen, and Heckman (2003), in which outcome variables can be test scores, labor market outcomes, self-reported measures of health and body mass index, and unobserved factors can be cognitive and noncognitive capabilities, health shocks, and genetic endowment. Conti and Heckman (2010) and Cunha, Heckman, and Schennach (2010) argue that these factors can be statistically dependent and multidimensional. For example, cognitive capabilities could include language and mathematical intelligence, which may be statistically dependent on each other and with noncognitive capabilities such as communication skills. In turn, policy and program evaluations that measure success and failure by effects on observed outcomes are really interested in the effects on the joint distribution of these latent skills and unobserved health.

The factor model structure from (1) appears in many other economic settings. For example, Aigner, Hsiao, Kapteyn, and Wansbeek (1984), Chamberlain and Griliches (1975) and Goldberger (1972) discuss structural and reduced form multi-equation models with latent variables and how they can be reformulated within a factor model framework. Gorman (1981) and Lewbel (1991) analyze a model of demand using factor models. Factor models are used to forecast macroeconomic variables by Forni and Reichlin (1998) and Stock and Watson (1999), to model asset returns by Fama and French (1993) and Ross (1976) and to model spatial relationships between observations by Anselin (2013).

Another way to view (1) is as a random coefficient model in which the vector of observed variables represents individual outcomes, the vector of unobserved factors represents random coefficients and the matrix of known coefficients represents observed covariates. Beran, Feuerverger, and Hall (1996) and Hoderlein, Klemelä, and Mammen (2010) identify nonparametric distributions in a cross-section with independent random coefficients and covariates. Hoderlein, Holzmann, and Meister (2014) and Masten (2014) analyze identification of random coefficients in a triangular and simultaneous system of equations, respectively. In a panel data framework, where $\boldsymbol{Y}$ represents outcomes of individuals over time and $\boldsymbol{F}$ can represent unobserved heterogeneity across individuals and time, Graham and Powell (2012) identify means while Arellano and Bonhomme (2012) identify joint distribution of the random coefficients conditional on covariates using second-order partial derivatives of the log characteristic function of $\boldsymbol{Y}$. There are two main differences between Arellano and Bonhomme (2012) and the present paper. First, this paper provides necessary and sufficient conditions for identification and generalizes the model to allow subsets of factors to be arbitrarily dependent, with the main insight being that second-order partial derivatives (and covariances) are symmetric and lead to a simplification in the identifying rank condition. Second, this paper identifies the joint distribution of $\boldsymbol{F}$ using first-order,

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

instead of second-order, partial derivatives of the log characteristic function of $\boldsymbol{Y}$ under the weaker assumption of mean independent errors, which will lead to faster convergence rates of the estimators.

We have two strategies to identify the joint distribution of the factors that draw on the literature relating to solutions of functional equations (see, e.g., Khatri and Rao, 1968; Székely and Rao, 2000). The first strategy is analogous to identification of factor means and involves taking first-order partial derivatives of the log characteristic function of $\boldsymbol{Y}$. The main assumptions are that the vector of factors and the vector of errors are independent, that errors are mean independent of each other, allowing for heteroskedasticity, for example, and that A satisfies some rank conditions. It extends standard identification strategies in the measurement error literature—pioneered by Kotlarski (1967), extended by Schennach (2004) to allow some variables to be mean independent instead of mutually independent, and applied by Horowitz and Markatou (1996) and others—from a one dimensional factor to a multi-dimensional vector of factors.

The second identification strategy generalizes Bonhomme and Robin (2010) from assuming factors are mutually independent to allowing factors to be arbitrarily dependent, and is analogous to identification of factor covariances. Identification assumes that factors and errors are independent and assumes a full-rank condition on the matrix representation of the relationship between second-order partial derivatives of the log characteristic function of observables and unobservables. This strategy provides necessary and sufficient conditions for identification of the nonparametric joint distribution of the factors. The advantages of the first strategy over the second one, when it identifies the model, are that errors are allowed to be mean independent instead of mutually independent and better finite and asymptotic properties as suggested by uniform convergence rates and simulations.

In earnings dynamics models, permanent and transitory income shocks can be modeled within a factor model structure (see, e.g., Meghir and Pistaferri, 2011, and many references therein). Bonhomme and Robin (2010) show identification of a model with mutually independent unobserved income shocks; and we show that the model is identified even when fixed effects and transitory income shocks are allowed to be arbitrarily dependent. Browning, Ejrnaes, and Alvarez (2010) and Ejrnæs and Browning (2014) contend that these models should have more unobserved heterogeneity and that income shocks are correlated. For example, transitory income shocks may represent bonuses, which could be dependent across time periods.

The identification proofs are constructive, and so estimators replace population quantities with sample analogs in closed-form formulas of the joint distribution of factors. The estimators use the inverse Fourier transformation with a kernel regularization to alleviate the ill-posed inverse problem. Using methods from the empirical processes literature (see, e.g., Pollard, 1984; Geer, 2000), we derive uniform convergence rates that depend on the tail-behavior of characteristic functions.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

In Monte Carlo simulations, we compare estimators in a measurement error model with repeated measurements and add to the evidence in this literature that estimators perform well despite slow asymptotic convergence rates; and that even with small sample sizes and vanishing characteristic functions the medians of the estimates lie virtually on top of the target densities (see also Delaigle, Hall, and Meister, 2008). We also show in simulations of an earnings dynamics model with four outcomes and six dependent unobservables per individual, that the estimates of the joint and marginal distributions of income shocks are still good, suggesting that the estimators perform well even in models with several dependent unobservables.

The paper is organized as follows. Section 2 presents the assumptions and the main identification results. Section 3 shows identification of $\boldsymbol{F} \in \mathbb{R}^2$ when $\boldsymbol{Y} \in \mathbb{R}^3$ and when $\boldsymbol{Y} \in \mathbb{R}^4$ and in an earnings dynamics model with a subset of dependent factors. Section 4 constructs estimators and derives uniform convergence rates. Section 5 presents Monte Carlo simulations in a measurement error model with repeated measurements and in the earnings dynamics model from Section 3.3. Section 6 concludes. Details of all proofs are in the appendix.

We use the following standard notation. The mean of the random vector $\boldsymbol{Y}' = (Y_1, \ldots, Y_P)$, where $\boldsymbol{Y}'$ is the transpose of $\boldsymbol{Y}$, is $\boldsymbol{\mu_Y} = E[\boldsymbol{Y}] = (E[Y_1], \ldots, E[Y_P])'$ and the variance is $\Sigma_{\boldsymbol{Y}} = \text{Var}(\boldsymbol{Y})$ with elements $\sigma_{p_1,p_2} = \text{Cov}(Y_{p_1}, Y_{p_2})$, for all $1 \leq p_1, p_2 \leq P$. The log characteristic function of $\boldsymbol{Y}$ is $\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = \ln(E[\exp(i\boldsymbol{Y}'\boldsymbol{s})])$, where $i = \sqrt{-1}$, $\boldsymbol{s} \in \mathbb{R}^P$ and $\ln(\cdot)$ is the natural log restricted to its principal branch (similarly $\varphi_{\boldsymbol{Y}_1,\boldsymbol{Y}_2}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \ln(E[\exp(i\boldsymbol{Y}_1'\boldsymbol{s}_1 + i\boldsymbol{Y}_2'\boldsymbol{s}_2)])$). The gradient vector of all first-order partial derivatives of $\varphi_{\boldsymbol{Y}}$ is $\nabla\varphi_{\boldsymbol{Y}} = (\frac{\partial\varphi_{\boldsymbol{Y}}}{\partial s_1}, \ldots, \frac{\partial\varphi_{\boldsymbol{Y}}}{\partial s_P})'$. The Hessian matrix of all second-order partial derivatives of $\varphi_{\boldsymbol{Y}}$ is

$$\nabla^2\varphi_{\boldsymbol{Y}} = \begin{pmatrix} \frac{\partial^2\varphi_{\boldsymbol{Y}}}{\partial s_1^2} & \cdots & \frac{\partial^2\varphi_{\boldsymbol{Y}}}{\partial s_1 \partial s_P} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\varphi_{\boldsymbol{Y}}}{\partial s_P \partial s_1} & \cdots & \frac{\partial^2\varphi_{\boldsymbol{Y}}}{\partial s_P^2} \end{pmatrix}.$$

Let $\text{vec}(\Sigma_{\boldsymbol{Y}}) = (\sigma_{11}, \ldots, \sigma_{P1}, \sigma_{12}, \ldots, \sigma_{P2}, \ldots, \sigma_{1,P}, \ldots, \sigma_{PP})'$ denote the vectorization of $\Sigma_{\boldsymbol{Y}}$ and $\text{vech}(\Sigma_{\boldsymbol{Y}}) = (\sigma_{11}, \ldots, \sigma_{P1}, \sigma_{22}, \ldots, \sigma_{P-1P-1}, \sigma_{PP-1}, \sigma_{PP})'$ the half-vectorization of $\Sigma_{\boldsymbol{Y}}$, and let $\text{diag}(\Sigma_{\boldsymbol{Y}}) = (\sigma_{11}, \sigma_{22} \ldots, \sigma_{P-1P-1}, \sigma_{PP})'$.

The standard basis $\boldsymbol{e}_p = (0, \ldots, 0, 1, 0, \ldots, 0)'$ is the $P \times 1$ vector with a 1 in the p^{th} coordinate and 0's elsewhere. The matrix of zeros is denoted by $\mathbf{0}$ with its dimensions coming from context. The $K \times K$ identity matrix is denoted by I_K. The left pseudoinverse of the $P \times K$ matrix A with $\text{Rank}(A) = K$ is denoted by the $K \times P$ matrix $H_A = (A'A)^{-1}A'$ and satisfies $H_A A = I_K$. The augmented matrix $(A_1 : \cdots : A_J)$ appends the columns of $A_1, \ldots, A_J$. The Kronecker product of the matrices A and B is denoted by $A \otimes B$. Let $\|A\| = \max_{p,k} A_{pk}$ be the max norm of the matrix A.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The indicator function $I\{\cdot\}$ takes value 1 if the statement between the parentheses is true and 0 otherwise. We use $a_N = O(b_N)$ or $a_N \lesssim b_N$ for big-O notation so that there is some constant $C > 0$ such that $|a_N| \leq C|b_N|$.

2. IDENTIFICATION

In this section, we identify the mean, variance and nonparametric joint distribution of $\boldsymbol{F}$ from (1).

2.1. Identification of $\mu_{\boldsymbol{F}}$ and $\nabla\varphi_{\boldsymbol{F}}$

The mean of (1) and the first-order partial derivatives of its log characteristic function are,

$$\mu_{\boldsymbol{Y}} = (A : I_P)\begin{pmatrix}\mu_{\boldsymbol{F}}\\ \mu_{\boldsymbol{U}}\end{pmatrix}, \tag{2}$$

$$\nabla\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = (A : I_P)\nabla\varphi_{\boldsymbol{F},\boldsymbol{U}}(A'\boldsymbol{s}, \boldsymbol{s}). \tag{3}$$

The following assumption states that factors and errors are independent and that some errors are mean independent of other errors.

Assumption 2.1. Let $\boldsymbol{F}$ and $\boldsymbol{U}$ be independent and $E[\boldsymbol{U}_1|\boldsymbol{U}_2] = 0$ where $\boldsymbol{U}_1' = (U_1, \ldots, U_{P_1})$ and $\boldsymbol{U}_2' = (U_{P_1+1}, \ldots, U_P)$.

This assumption allows $\boldsymbol{F}$ to be arbitrarily dependent, $\boldsymbol{U}_1$ to be arbitrarily dependent and $\boldsymbol{U}_2$ to be arbitrarily dependent. The mean independence assumption also allows for some dependence between $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$. It is similar to Schennach (2004) who weakens independence assumptions to mean independence assumptions in a measurement error model. The assumption can be modified so that different unobservables are independent and mean independent.

The next assumption states that A can be divided into two disjoint submatrices each with rank K.

Assumption 2.2. Let $A = (A_1' : A_2')'$, where A_1 is a $P_1 \times K$ matrix and A_2 is a $(P - P_1) \times K$ matrix, and let $\text{Rank}(A_1) = \text{Rank}(A_2) = K$.[1]

Assumptions 2.1 and 2.2 identify $\mu_{\boldsymbol{F}}$ and $\nabla\varphi_{\boldsymbol{F}}$ by inverting (2) and (3) by left multiplying by $(H_{A_1} : \boldsymbol{0})$ and eliminating terms with $\boldsymbol{U}$ by setting $\boldsymbol{s}' = (\boldsymbol{0}, \boldsymbol{v}'H_{A_2})$,

$$\begin{aligned}\mu_{\boldsymbol{F}} &= (H_{A_1} : \boldsymbol{0})(\mu_{\boldsymbol{Y}} - \mu_{\boldsymbol{U}}) = (H_{A_1} : \boldsymbol{0})\mu_{\boldsymbol{Y}},\\ \nabla\varphi_{\boldsymbol{F}}(\boldsymbol{v}) &= (H_{A_1} : \boldsymbol{0})(\nabla\varphi_{\boldsymbol{Y}}(\boldsymbol{0}, \boldsymbol{v}'H_{A_2}) - \nabla\varphi_{\boldsymbol{U}}(\boldsymbol{0}, \boldsymbol{v}'H_{A_2}))\\ &= (H_{A_1} : \boldsymbol{0})\nabla\varphi_{\boldsymbol{Y}}(\boldsymbol{0}, \boldsymbol{v}'H_{A_2}).\end{aligned}$$

Next assume the regularity conditions of finite means and nonvanishing characteristic functions.

Assumption 2.3. Let $\mu_{\boldsymbol{F}}$ be finite and $E\left[e^{i\boldsymbol{F}'\boldsymbol{v}}\right] \neq 0$ for all $\boldsymbol{v} \in \mathbb{R}^K$.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The assumption is used to ensure that $\boldsymbol{\mu}_{\boldsymbol{F}}$ and $\partial\varphi_{\boldsymbol{F}}(\boldsymbol{v})/\partial v_k = iE[F_k e^{i\boldsymbol{F}'\boldsymbol{v}}]/E[e^{i\boldsymbol{F}'\boldsymbol{v}}]$, in the above equations, are finite. The assumption that characteristic functions do not vanish is standard in the measurement error literature (e.g., Schennach, 2007; Schennach and Hu, 2013) but can be weakened. Bondesson (1974) proves identification when characteristic functions satisfy a "short gap" condition that they do not vanish on any interval of length $L > 0$. Evdokimov and White (2012) allow characteristic functions to have a countable number of zeros. Zinde-Walsh (2014) bounds an integral instead of bounding each point that is being integrated over. Conceptually these assumptions are similar and state that $E[e^{i\boldsymbol{F}'\boldsymbol{v}}]$ cannot vanish on sets of nonzero measure (unless $E[F_k e^{i\boldsymbol{F}'\boldsymbol{v}}] = 0$ on these sets). In the simulations section, the density of a uniform distribution is estimated quite well even though its characteristic function has a countable number of zeros.

The following theorem identifies the mean and log characteristic function of the factors. It can be viewed as extending Schennach (2004) from a one-dimensional factor model to a multidimensional factor model.

THEOREM 2.1. *Suppose that (1), perhaps after relabeling equations, can be written as,*

$$\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{pmatrix} = \begin{pmatrix}A_1\\ A_2\end{pmatrix}\boldsymbol{F} + \begin{pmatrix}\boldsymbol{U}_1\\ \boldsymbol{U}_2\end{pmatrix},$$

where $\boldsymbol{Y}_1' = (Y_1, \ldots, Y_{P_1})$, $\boldsymbol{Y}_2' = (Y_{P_1+1}, \ldots, Y_P)$, $\boldsymbol{U}_1' = (U_1, \ldots, U_{P_1})$, $\boldsymbol{U}_2' = (U_{P_1+1}, \ldots, U_P)$, A_1 *is a* $P_1 \times K$ *matrix and* A_2 *is a* $(P - P_1) \times K$ *matrix, and that Assumptions 2.1 and 2.3 hold. If Assumption 2.2 holds then* $\boldsymbol{\mu}_{\boldsymbol{F}} = H_{A_1}\boldsymbol{\mu}_{\boldsymbol{Y}_1}$ *and*

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = \int_0^1 \sum_{k=1}^{K}\sum_{p=1}^{P_1} v_k H_{A_1 kp} \left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0}, \boldsymbol{v}'H_{A_2})'} \mathrm{d}u. \tag{4}$$

The next section provides necessary and sufficient conditions for identification of the variance and log characteristic function of $\boldsymbol{F}$. Later we give examples where $\varphi_{\boldsymbol{F}}$ is identified using the theorem in the next section but not the theorem above. The cost of necessity is that the assumption of mean independent errors is strengthened to mutually independent errors and identification will now use $\nabla^2\varphi_{\boldsymbol{F}}$ instead of $\nabla\varphi_{\boldsymbol{F}}$, which results in less robust estimators with slower convergence rates.

2.2. Identification of $\Sigma_{\boldsymbol{F}}$ and $\nabla^2\varphi_{\boldsymbol{F}}$

The variance of (1) and the second-order partial derivatives of its log characteristic function are,

$$\Sigma_{\boldsymbol{Y}} = (A : I_P)\Sigma_{\boldsymbol{F},\boldsymbol{U}}(A : I_P)', \tag{5}$$

$$\nabla^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = (A : I_P)\nabla^2\varphi_{\boldsymbol{F},\boldsymbol{U}}(A'\boldsymbol{s}, \boldsymbol{s})(A : I_P)'. \tag{6}$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The following independence assumption is used to simplify the above equations.

Assumption 2.4. Let $\boldsymbol{F}, U_1, \ldots, U_P$ be mutually independent.

The assumption can be weakened from mutually independent $\boldsymbol{F}, U_1, \ldots, U_P$ to subindependent $A\boldsymbol{F}$ and $\boldsymbol{U}$ and mutually independent $U_1, \ldots, U_P$. Subindependence is weaker than independence, and states that the distribution of $A\boldsymbol{F}$ and $\boldsymbol{U}$ is equal to the convolution of their distributions (see, e.g., Hamedani, Volkmer, and Behboodian, 2011; Schennach, 2013).[2]

The assumption implies that $\varphi_{\boldsymbol{F},\boldsymbol{U}}(A'\boldsymbol{s}, \boldsymbol{s}) = \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \sum_{p=1}^{P} \varphi_{U_p}(\boldsymbol{e}_p'\boldsymbol{s})$ so that

$$\Sigma_{\boldsymbol{F},\boldsymbol{U}} = \begin{pmatrix} \Sigma_{\boldsymbol{F}} & 0 \\ 0 & \Sigma_{\boldsymbol{U}} \end{pmatrix}, \quad \nabla^2 \varphi_{\boldsymbol{F},\boldsymbol{U}}(A'\boldsymbol{s}, \boldsymbol{s}) = \begin{pmatrix} \nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) & 0 \\ 0 & \nabla^2 \varphi_{\boldsymbol{U}}(\boldsymbol{s}) \end{pmatrix},$$

where $\Sigma_{\boldsymbol{U}}$ and $\nabla^2 \varphi_{\boldsymbol{U}}(\boldsymbol{s})$ are diagonal matrices with elements $\mathrm{Var}(U_p)$ and $\varphi''_{U_p}(\boldsymbol{e}_p'\boldsymbol{s})$ respectively.

Vectorizing (5) and (6) we obtain,

$$\mathrm{vec}(\Sigma_{\boldsymbol{Y}}) = (A \odot A)\mathrm{vech}(\Sigma_{\boldsymbol{F}}) + (I_P \star I_P)\mathrm{diag}(\Sigma_{\boldsymbol{U}}), \tag{7}$$

$$\mathrm{vec}(\nabla^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})) = (A \odot A)\mathrm{vech}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) + (I_P \star I_P)\mathrm{diag}(\nabla^2 \varphi_{\boldsymbol{U}}(\boldsymbol{s})), \tag{8}$$

where

$$A \odot A = \left(A_1 \otimes A_1 : \cdots : A_{k_1} \otimes A_{k_1+k_2} + A_{k_1+k_2} \otimes A_{k_2} : \cdots : A_K \otimes A_K\right), \quad k_1 < k_2, \tag{9}$$

is a matrix product with dimension $P^2 \times \frac{K(K+1)}{2}$ and

$$I_P \star I_P = (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 : \cdots : \boldsymbol{e}_P \otimes \boldsymbol{e}_P)$$

is the Khatri-Rao product (Khatri and Rao, 1968) with dimension $P^2 \times P$. The main insight behind the construction of $A \odot A$ in (9) is that covariances and second-order partial derivatives are symmetric, i.e., $\mathrm{Cov}(F_{k_1}, F_{k_2}) = \mathrm{Cov}(F_{k_2}, F_{k_1})$ and $\partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{v})/\partial v_{k_1} \partial v_{k_2} = \partial^2 \varphi_{\boldsymbol{F}}(\boldsymbol{v})/\partial v_{k_2} \partial v_{k_1}$, and so the corresponding columns can be added together.

To identify $\mathrm{vech}(\Sigma_{\boldsymbol{F}})$ and $\mathrm{vech}(\nabla^2 \varphi_{\boldsymbol{F}})$, we impose rank conditions that will be used to invert (7) and (8).

Assumption 2.5. Let (i) $\mathrm{Rank}((A \odot A) : (I_P \star I_P)) = \frac{K(K+1)}{2} + P$ and (ii) $\mathrm{Rank}(A) = K$.

The matrix $(A \odot A) : (I_P \star I_P)$ contains $P(P+1)/2$ unique rows. The redundant rows corresponding to the redundant equations $\mathrm{Cov}(Y_{p_1}, Y_{p_2})$ and $\mathrm{Cov}(Y_{p_2}, Y_{p_1})$, or $\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})/\partial s_{p_1} \partial s_{p_2}$ and $\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})/\partial s_{p_2} \partial s_{p_1}$, can be removed in practice. Assumption 2.5(i) implies that $K \leq P - 1$.

Next we impose regularity conditions.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Assumption 2.6. Let $\{U_p\}_{p=1}^P$ be nondegenerate, zero mean random variables, $\Sigma_{\boldsymbol{F}}$ be positive definite and finite and $E[e^{i\boldsymbol{F}'\boldsymbol{v}}] \neq 0$ for all $\boldsymbol{v} \in \mathbb{R}^K$.

THEOREM 2.2. *Suppose that (1) and Assumptions 2.4 and 2.6 hold. If Assumption 2.5 holds then* $\begin{pmatrix} \text{vech}(\Sigma_{\boldsymbol{F}}) \\ \text{diag}(\Sigma_{\boldsymbol{U}}) \end{pmatrix} = H_{(A\odot A):(I_P \star I_P)} \text{vec}(\Sigma_{\boldsymbol{Y}})$ *and*

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = i\boldsymbol{v}'\boldsymbol{\mu}_{\boldsymbol{F}} + \int_0^1 \int_0^{u_2} \boldsymbol{v}'\nabla^2\varphi_{\boldsymbol{F}}(u_1\boldsymbol{v})\boldsymbol{v}\, \mathrm{d}u_1 \mathrm{d}u_2, \tag{10}$$

where $\boldsymbol{\mu}_{\boldsymbol{F}} = H_A\boldsymbol{\mu}_{\boldsymbol{Y}}$ *and* $\begin{pmatrix} \text{vech}(\nabla^2\varphi_{\boldsymbol{F}}(\boldsymbol{v})) \\ \text{diag}(\nabla^2\varphi_{\boldsymbol{U}}(H_A'\boldsymbol{v})) \end{pmatrix} = H_{(A\odot A):(I_P \star I_P)}$ $\text{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(H_A'\boldsymbol{v}))$. *If Assumption 2.5(i) does not hold then* $\Sigma_{\boldsymbol{F}}$ *and hence* $\varphi_{\boldsymbol{F}}$ *are not identified and if Assumption 2.5(ii) does not hold then* $\varphi_{\boldsymbol{F}}$ *is not identified.*

2.3. A Generalization to Subsets of Arbitrarily Dependent Factors

This section presents identification of a more general factor model with subsets of arbitrarily dependent factors,

$$\boldsymbol{Y} = \sum_{j=1}^J A_j\boldsymbol{F}_j, \tag{11}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_P)'$, A_j is a $P \times K_j$ matrix and $\boldsymbol{F}_j = (F_{j1}, \ldots, F_{jK_j})'$ is an arbitrarily dependent random vector, for $j = 1, \ldots, J$. The model allows some of the unobserved factors to be errors. For example, if the errors are allowed to be arbitrarily dependent then let $\boldsymbol{F}_J = \boldsymbol{U}$ or if the errors are mutually independent, like in (1) with Assumption 2.4, then let $F_{J-P+p} = U_p$, for $p = 1, \ldots, P$.

THEOREM 2.3. *Suppose that (11) holds. If (i)* $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_J$ *are mutually independent, (ii)* $\text{Rank}((A_1 \odot A_1) : \cdots : (A_J \odot A_J)) = \sum_{j=1}^J K_j(K_j+1)/2$, *(iii)* $\text{Rank}(A_j) = K_j$, *for* $j = 1, \ldots, J$ *and (iv)* $\Sigma_{\boldsymbol{F}_j}$ *is positive definite and finite and* $E[e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}] \neq 0$, *for* $j = 1, \ldots, J$, *then* $\Sigma_{\boldsymbol{F}_j}$ *is identified and* $\varphi_{\boldsymbol{F}_j}$ *is identified up to mean, for* $j = 1, \ldots, J$. *Further, if* $\text{Rank}(A) = \sum_{j=1}^J K_j$ *then* $\boldsymbol{\mu}_{\boldsymbol{F}} = H_A\boldsymbol{\mu}_{\boldsymbol{Y}}$ *is identified.*

In the appendix we show that,

$$\text{vec}(\Sigma_{\boldsymbol{Y}}) = ((A_1 \odot A_1) : \cdots : (A_J \odot A_J))\text{vech}(\Sigma_{\boldsymbol{F}}),$$
$$\text{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})) = ((A_1 \odot A_1) : \cdots : (A_J \odot A_J))\text{vech}(\nabla^2\varphi_{\boldsymbol{F}}(A'\boldsymbol{s})),$$

where $((A_1 \odot A_1) : \cdots : (A_J \odot A_J))$ is a $P^2 \times \sum_{j=1}^J K_j(K_j+1)/2$ matrix, $\text{vech}(\Sigma_{\boldsymbol{F}}) = (\text{vech}(\Sigma_{\boldsymbol{F}_1})', \ldots, \text{vech}(\Sigma_{\boldsymbol{F}_J})')'$ and $\text{vech}(\nabla^2\varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) = (\text{vech}(\nabla^2\varphi_{\boldsymbol{F}_1}(A_1'\boldsymbol{s}))', \ldots, \text{vech}(\nabla^2\varphi_{\boldsymbol{F}_J}(A_J'\boldsymbol{s}))')'$.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

For example, if $F_1, \ldots, F_J$ are mutually independent scalars then the above equations reduce to,

$$\begin{aligned}\operatorname{vec}(\Sigma_{\boldsymbol{Y}}) &= ((A_1 \otimes A_1) : \cdots : (A_J \otimes A_J))\operatorname{diag}(\Sigma_{\boldsymbol{F}}) = (A \star A)\operatorname{diag}(\Sigma_{\boldsymbol{F}}),\\ \operatorname{vec}(\nabla^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})) &= ((A_1 \otimes A_1) : \cdots : (A_J \otimes A_J))\operatorname{diag}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) = (A \star A)\operatorname{diag}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})),\end{aligned} \tag{12}$$

where $A \star A = (A_1 \otimes A_1, \ldots, A_J \otimes A_J)$ is the Khatri-Rao product with dimension $P^2 \times J$. (12) is the identifying equation used by Bonhomme and Robin (2010).

We conclude this section by noting that the matrix $(A_1 \odot A_1) : \cdots : (A_J \odot A_J)$ has $\frac{P(P+1)}{2}$ unique rows so if the assumption $\operatorname{Rank}((A_1 \odot A_1) : \cdots : (A_J \odot A_J)) = \sum_{j=1}^{J} \frac{K_j(K_j+1)}{2}$ holds then a necessary condition for identification is

$$\sum_{j=1}^{J} K_j(K_j+1) \leq P(P+1).$$

3. EXAMPLES

In this section we present illustrative examples of identification using Theorems 2.1 and 2.2.

3.1. Identification of $\boldsymbol{F} \in \mathbb{R}^2$ when $\boldsymbol{Y} \in \mathbb{R}^4$

Suppose that,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}, \tag{13}$$

where $\boldsymbol{F}' = (F_1, F_2)$ and $\boldsymbol{U}' = (U_1, U_2, U_3, U_4)$ are assumed to be independent. Assume that $E[U_1|U_3, U_4] = E[U_2|U_3, U_4] = 0$. Then Assumption 2.1 is satisfied.

We identify $\varphi_{\boldsymbol{F}}$ using Theorem 2.1. The first-order partial derivatives of the log characteristic function of (13) are,

$$\begin{pmatrix} \frac{\partial \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1} \\ \frac{\partial \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2} \\ \frac{\partial \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_3} \\ \frac{\partial \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_4} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{pmatrix} \begin{pmatrix} \frac{\partial \varphi_{\boldsymbol{F}}(\boldsymbol{v})}{\partial v_1}\Big|_{\boldsymbol{v}=A'\boldsymbol{s}} \\ \frac{\partial \varphi_{\boldsymbol{F}}(\boldsymbol{v})}{\partial v_2}\Big|_{\boldsymbol{v}=A'\boldsymbol{s}} \end{pmatrix} + \begin{pmatrix} \frac{\partial \varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_1} \\ \frac{\partial \varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_2} \\ \frac{\partial \varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_3} \\ \frac{\partial \varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_4} \end{pmatrix}.$$

Assume that $\operatorname{Rank}(A_1) = \operatorname{Rank}(A_2) = 2$ where $A_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $A_2 = \begin{pmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{pmatrix}$ so that Assumption 2.2 is satisfied. Then,

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

$$\nabla\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = (H_{A_1} : \mathbf{0}) \begin{pmatrix} \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_3}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_4}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \end{pmatrix} + (H_{A_1} : \mathbf{0}) \begin{pmatrix} \frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_1}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_2}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_3}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_4}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \end{pmatrix}$$

$$= H_{A_1} \begin{pmatrix} \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \\ \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\Big|_{\boldsymbol{s}=(\mathbf{0},\boldsymbol{v}'H_{A_2})'} \end{pmatrix}$$

and, by the fundamental theorem of calculus,

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = \int_0^1 \boldsymbol{v}'\nabla\varphi_{\boldsymbol{F}}(u\boldsymbol{v})\mathrm{d}u$$

are identified.

3.2. Identification of $F \in \mathbb{R}^2$ when $Y \in \mathbb{R}^3$

Suppose now that

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}, \tag{14}$$

where $\boldsymbol{F}' = (F_1, F_2)$, U_1, U_2 and U_3 are assumed to be mutually independent. Then Assumption 2.4 is satisfied.

We identify $\varphi_{\boldsymbol{F}}$ using Theorem 2.2. The second-order partial derivatives of the log characteristic function of (14) are,

$$\mathrm{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})) = (A \odot A) : (I_3 \star I_3) \begin{pmatrix} \mathrm{vech}(\nabla^2\varphi_{\boldsymbol{F}}(A's)) \\ \varphi''_{U_1}(s_1) \\ \varphi''_{U_2}(s_2) \\ \varphi''_{U_3}(s_3) \end{pmatrix},$$

where,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \qquad (A \odot A) : (I_3 \star I_3) = \begin{pmatrix} a_{11}^2 & 2a_{11}a_{12} & a_{12}^2 & 1 & 0 & 0 \\ a_{11}a_{21} & a_{11}a_{22}+a_{12}a_{21} & a_{12}a_{22} & 0 & 0 & 0 \\ a_{11}a_{31} & a_{11}a_{32}+a_{12}a_{31} & a_{12}a_{32} & 0 & 0 & 0 \\ a_{11}a_{21} & a_{11}a_{22}+a_{12}a_{21} & a_{12}a_{22} & 0 & 0 & 0 \\ a_{21}^2 & 2a_{21}a_{22} & a_{22}^2 & 0 & 1 & 0 \\ a_{21}a_{31} & a_{21}a_{32}+a_{22}a_{31} & a_{22}a_{32} & 0 & 0 & 0 \\ a_{11}a_{31} & a_{11}a_{32}+a_{12}a_{31} & a_{12}a_{32} & 0 & 0 & 0 \\ a_{21}a_{31} & a_{21}a_{32}+a_{22}a_{31} & a_{22}a_{32} & 0 & 0 & 0 \\ a_{31}^2 & 2a_{31}a_{32} & a_{32}^2 & 0 & 0 & 1 \end{pmatrix}.$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The system contains the redundant rows $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1\partial s_2}$ and $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2\partial s_1}$, $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1\partial s_3}$ and $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_3\partial s_1}$, and $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2\partial s_3}$ and $\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_3\partial s_2}$.

Assume that $\text{Rank}((A \odot A) : (I_3 \star I_3)) = 6$ and $\text{Rank}(A) = 2$ so that Assumption 2.5 holds. Then,

$$\begin{pmatrix} \text{vech}(\nabla^2\varphi_{\boldsymbol{F}})(\boldsymbol{v}) \\ \varphi''_{U_1}(\boldsymbol{e}_1' H_A' \boldsymbol{v}) \\ \varphi''_{U_2}(\boldsymbol{e}_2' H_A' \boldsymbol{v}) \\ \varphi''_{U_3}(\boldsymbol{e}_3' H_A' \boldsymbol{v}) \end{pmatrix} = H_{(A\odot A):(I_3\star I_3)} \text{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(H_A'\boldsymbol{v}))$$

and, by the fundamental theorem of calculus,

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = i\boldsymbol{v}' H_A \boldsymbol{\mu}_{\boldsymbol{Y}} + \int_0^1 \int_0^{u_2} \boldsymbol{v}' \nabla^2 \varphi_{\boldsymbol{F}}(u_1\boldsymbol{v})\boldsymbol{v}\, \mathrm{d}u_1 \mathrm{d}u_2$$

are identified.

3.3. Identification of an Earnings Dynamics Model

Consider the earnings dynamics model from Bonhomme and Robin (2010, p. 494),

$$w_t = f + y_t^P + y_t^T, \qquad t = 1,2,3,4, \tag{15}$$

$$y_t^P = y_{t-1}^P + \varepsilon_t, \qquad t \geq 2, \tag{16}$$

$$y_t^T = \eta_t, \qquad \eta_1 = \eta_4 = 0, \tag{17}$$

where w_t is the residual of a regression of individual log earnings on a set of strictly exogenous covariates, f is the unobserved fixed effect, y_t^P is the unobserved persistent component, y_t^T is the unobserved transitory shock, and ε_t and η_t are innovations. The system in matrix notation is,

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \boldsymbol{F}, \tag{18}$$

where $\boldsymbol{F} = (f + y_1^P, \eta_2, \eta_3, \varepsilon_2, \varepsilon_3, \varepsilon_4)'$. Bonhomme and Robin (2010) difference out $f + y_1^P$ and identify all the other unobservables under the assumption of mutual independence. We identify the joint distribution of $\boldsymbol{F}$ allowing $(f + y_1^P, \eta_2, \eta_3)$ to be arbitrarily dependent. This dependence structure allows individual fixed effects, which can reflect occupational choice, and transitory income, such as bonuses, to be dependent, but alternative dependence structures are possible.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The log characteristic function of (18) is,

$$\varphi_{\boldsymbol{w}}(\boldsymbol{s}) = \varphi_{f+y_1^P,\eta_2,\eta_3}(s_1+s_2+s_3+s_4, s_2, s_3) + \varphi_{\varepsilon_2}(s_2+s_3+s_4) + \varphi_{\varepsilon_3}(s_3+s_4) + \varphi_{\varepsilon_4}(s_4).$$

The second-order partial derivatives are

$$\operatorname{vech}(\nabla^2\varphi_{\boldsymbol{w}}(\boldsymbol{s})) = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&0&1&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0\\ 1&2&0&1&0&0&1&0&0&0\\ 1&1&1&0&1&0&1&0&0&0\\ 1&1&0&0&0&0&1&0&0&0\\ 1&0&2&0&0&1&1&1&0&0\\ 1&0&1&0&0&0&1&1&0&0\\ 1&0&0&0&0&0&1&1&1&0 \end{pmatrix} \begin{pmatrix} \operatorname{vech}(\nabla^2\varphi_{f+y_1^P,\eta_2,\eta_3}(s_1+s_2+s_3+s_4, s_2, s_3))\\ \varphi''_{\varepsilon_2}(s_2+s_3+s_4)\\ \varphi''_{\varepsilon_3}(s_3+s_4)\\ \varphi''_{\varepsilon_4}(s_4) \end{pmatrix}, \tag{19}$$

where the redundant equations were removed. Inverting the above system of equations,

$$\operatorname{vech}(\nabla^2\varphi_{f+y_1^P,\eta_2,\eta_3}(\boldsymbol{v})) = \begin{pmatrix} \frac{1}{2}&0&0&\frac{1}{2}&0&0&0&0&0&0\\ -\frac{1}{2}&1&0&-\frac{1}{2}&0&0&0&0&0&0\\ -\frac{1}{2}&0&1&-\frac{1}{2}&0&0&0&0&0&0\\ \frac{1}{2}&-1&0&\frac{1}{2}&1&0&-1&0&0&0\\ \frac{1}{2}&0&-1&\frac{1}{2}&0&1&-1&0&0&0\\ \frac{1}{2}&0&-1&\frac{1}{2}&0&0&0&1&-1&0 \end{pmatrix} \times \operatorname{vech}(\nabla^2\varphi_{\boldsymbol{w}}(v_1, v_2, v_3, -v_2-v_3)),$$

$$\begin{pmatrix} \varphi''_{\varepsilon_2}(v)\\ \varphi''_{\varepsilon_3}(v)\\ \varphi''_{\varepsilon_4}(v) \end{pmatrix} = \begin{pmatrix} 0&-1&0&0&0&0&1&0&0&0\\ 0&1&-1&0&0&0&-1&0&1&0\\ -\frac{1}{2}&0&1&-\frac{1}{2}&0&0&0&0&-1&1 \end{pmatrix} \operatorname{vech}(\nabla^2\varphi_{\boldsymbol{w}}(0,0,0,v)).$$

By the fundamental theorem of calculus,

$$\varphi_{f+y_1^P,\eta_2,\eta_3}(\boldsymbol{v}) = \int_0^1\int_0^{u_2}\sum_{k_1=1}^{3}\sum_{k_2=1}^{3} v_{k_1}v_{k_2}\left.\frac{\partial^2\varphi_{f,\eta_2,\eta_3}(\boldsymbol{w})}{\partial v_{k_1}\partial v_{k_2}}\right|_{\boldsymbol{w}=(u_1\boldsymbol{v})} \mathrm{d}u_1\mathrm{d}u_2, \tag{20}$$

$$\varphi_{\varepsilon_k}(v) = \int_0^1\int_0^{u_2} v^2\varphi''_{\varepsilon_k}(u_1 v)\mathrm{d}u_1\mathrm{d}u_2, \qquad k=2,3,4. \tag{21}$$

The joint distribution of the unobservables is overidentified because there are ten equations derived from the partial derivatives of $\varphi_{\boldsymbol{Y}}$ in (19) and only nine unknown partial derivatives $\operatorname{vech}(\nabla^2\varphi_{f+y_1^P,\eta_2,\eta_3}), \varphi''_{\varepsilon_2}, \varphi''_{\varepsilon_3}, \varphi''_{\varepsilon_4}$ and so in principle additional dependencies or unobservables can be included. The model is not identified using Theorem 2.1 because Assumption 2.2 cannot hold.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

4. ESTIMATION

To estimate the mean $\boldsymbol{\mu}_{\boldsymbol{F}} = H_A \boldsymbol{\mu}_{\boldsymbol{Y}}$ and variance $\begin{pmatrix} \text{vech}(\Sigma_{\boldsymbol{F}}) \\ \text{diag}(\Sigma_{\boldsymbol{U}}) \end{pmatrix} = H_{(A \odot A):(I_P \star I_P)} \text{vec}(\Sigma_{\boldsymbol{Y}})$ from Theorems 2.1 and 2.2 we can replace $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\Sigma_{\boldsymbol{Y}}$ by their sample analogs. The estimators are easily seen to be root-N asymptotically normal by applying the central limit theorem and delta method. Therefore, in this section we focus on estimating the joint density of $\boldsymbol{F}$, which is constructed using sample analogs of population quantities from the closed-form formulas in Theorems 2.1 and 2.2 and inverse Fourier transforms. We show that the uniform convergence rate of an estimator based on first-order partial derivatives of the empirical log characteristic function is similar to Li and Vuong (1998), who also use first-order partial derivatives, and faster than an estimator based on second-order partial derivatives, which has the same rate as Bonhomme and Robin (2010) who also use second-order partial derivatives.

We impose the following conditions.

Assumption 4.1. Let (i) $\{\boldsymbol{Y}_n\}_{n=1}^N$ be independent identically distributed (i.i.d.) random draws from $\boldsymbol{Y}$, (ii) $\max_{p \in \{1,\ldots,P\}} E[\exp(|Y_p|\beta)] < \infty$ for some $\beta > 0$, (iii) $\boldsymbol{\kappa}(\boldsymbol{u}) = \prod_{k=1}^K \kappa(u_k)$, where $\kappa(u_k)$ is a q^{th}-order kernel with $q \geq 2$ and support $u_k \in [-1, 1]$ and (iv) $\Upsilon : \mathbb{R}^+ \to [0, 1]$ be an integrable decreasing function, such that $\Upsilon(\|\boldsymbol{s}\|) \leq |\phi_{\boldsymbol{Y}}(\boldsymbol{s})|$ for $\|\boldsymbol{s}\| = \max s_p$ large enough.

Let

$$\widehat{\varphi_{\boldsymbol{Y}}(\boldsymbol{s})} = \ln(E_N\left[\exp(i\boldsymbol{Y}'\boldsymbol{s})\right]), \qquad \widehat{\frac{\partial \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}} = \frac{i E_N\left[Y_p \exp(i\boldsymbol{Y}'\boldsymbol{s})\right]}{E_N\left[\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]},$$

$$\widehat{\frac{\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_{p_1} \partial s_{p_2}}} = \frac{E_N\left[Y_{p_1} \exp(i\boldsymbol{Y}'\boldsymbol{s})\right] E_N\left[Y_{p_2} \exp(i\boldsymbol{Y}'\boldsymbol{s})\right]}{\left(E_N\left[\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]\right)^2} - \frac{E_N\left[Y_{p_1} Y_{p_2} \exp(i\boldsymbol{Y}'\boldsymbol{s})\right]}{E_N\left[\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]},$$

where $E_N[\cdot] = \frac{1}{N}\sum_{n=1}^N [\cdot]$ is the sample analog of its population counterpart and $\widehat{\varphi_{\boldsymbol{Y}}}$ is the empirical log characteristic function.

Using the above expressions, $\widehat{\varphi}_{\boldsymbol{F}}$ is estimated either by the sample analog of $\varphi_{\boldsymbol{F}}$ from (4) in Theorem 2.1 that uses first-order partial derivatives of $\varphi_{\boldsymbol{Y}}$ or by the sample analog of $\varphi_{\boldsymbol{F}}$ from (10) in Theorem 2.2 that uses second-order partial derivatives of $\varphi_{\boldsymbol{Y}}$. Densities are then estimated on the compact interval $[-S_N, S_N]^K$,

$$\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) = \frac{1}{(2\pi)^K} \int e^{-i\boldsymbol{v}'\boldsymbol{u}} \widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) \phi_{\boldsymbol{\kappa}}(\boldsymbol{v}/S_N) \mathrm{d}\boldsymbol{v}, \tag{22}$$

where $\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) = e^{\widehat{\varphi}_{\boldsymbol{F}}(\boldsymbol{v})}$ and $S_N = N^\delta$, for some $0 < \delta < \infty$, so that Lemma 4.1 below holds. The Fourier transform of the kernel $\boldsymbol{\kappa}$, $\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}) = \int e^{i\boldsymbol{v}'\boldsymbol{u}} \boldsymbol{\kappa}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$, leads to relatively slow convergence rates but solves any irregularity problems by smoothing the estimator.

The convergence rate of $\widehat{f}_{\boldsymbol{F}}$ will depend on the rate of decay of $\phi_{\boldsymbol{Y}}$. The function Υ, in Assumption 4.1(iv), bounds the rate of decay of $\phi_{\boldsymbol{Y}}$. Like Fan (1991),

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

the convergence rate of $\widehat{f_{\boldsymbol{F}}}$ is faster if the distribution of $\boldsymbol{Y}$ is ordinary smooth, i.e., Υ is a polynomial function and slower if the distribution of $\boldsymbol{Y}$ is supersmooth, i.e., Υ is an exponential function.

To prove that $\widehat{f_{\boldsymbol{F}}}$ is uniformly consistent, we first uniformly bound the difference between a partial derivative of $\phi_{\boldsymbol{Y}}(\boldsymbol{s}) = E[e^{i\boldsymbol{Y}'\boldsymbol{s}}]$ and its sample analog.

LEMMA 4.1. *Let Assumption 4.1 hold. For $\alpha_1, \ldots, \alpha_P$ nonnegative integers and large enough N,*

$$\sup_{\boldsymbol{s}\in[-S_N,S_N]^P} \left| E_N\left[\prod_{p=1}^{P} Y_p^{\alpha_p} \exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right] - E\left[\prod_{p=1}^{P} Y_p^{\alpha_p} \exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]\right| < \varepsilon_N \qquad a.s.,$$

where $\varepsilon_N = O\left(\frac{\ln N}{N^{\frac{1}{2}}}\right)$ and $S_N = N^{\delta}$ for some finite $0 < \delta$.

Lemma 4.1 uniformly bounds the estimation error by $O\left(\frac{\ln N}{N^{\frac{1}{2}}}\right)$ provided that the size of the compact set $[-S_N, S_N]^P$ does not grow faster than some finite power of N. The bound, which is the same as Bonhomme and Robin (2010), is worse than Li and Vuong (1998) who assume bounded support and obtain the convergence rate $O\left(\frac{\ln\ln N}{N}\right)^{\frac{1}{2}}$.

The next theorem uses Lemma 4.1 to bound the estimation error of $\phi_{\boldsymbol{F}}$ and $f_{\boldsymbol{F}}$.

THEOREM 4.1. *Suppose that (1) and Assumptions 2.2, 2.4, 2.6 and 4.1 hold. Further, suppose that the matrix A is either known or identified and has root-N consistent estimator. For large enough N and for S_N that increases slowly enough,*

$$\sup_{\boldsymbol{v}\in[-S_N,S_N]^K} \left|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) - \phi_{\boldsymbol{F}}(\boldsymbol{v})\right| = O\left(\varepsilon_N S_N \int_0^1 \left(\Upsilon(uS_N\|H_{A_2}\|)\right)^{-2} \mathrm{d}u\right) \qquad a.s.,$$

where $\phi_{\boldsymbol{F}} = e^{\varphi_{\boldsymbol{F}}}$ comes from (4) in Theorem 2.1 and $\widehat{\phi}_{\boldsymbol{F}}$ is its sample analog, and

$$\sup_{\boldsymbol{v}\in[-S_N,S_N]^K} \left|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) - \phi_{\boldsymbol{F}}(\boldsymbol{v})\right| = O\left(\varepsilon_N S_N^2 \int_0^1 \left(\Upsilon(uS_N\|H_A\|)\right)^{-3} \mathrm{d}u\right) \qquad a.s.,$$

where $\phi_{\boldsymbol{F}} = e^{\varphi_{\boldsymbol{F}}}$ comes from (10) in Theorem 2.2 and $\widehat{\phi}_{\boldsymbol{F}}$ is its sample analog. Then,

$$\sup_{\boldsymbol{u}} \left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})\right| = O\left(S_N^K \cdot \sup_{\boldsymbol{v}\in[-S_N,S_N]^K} \left|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) - \phi_{\boldsymbol{F}}(\boldsymbol{v})\right| + \int_{\boldsymbol{v}\notin[-S_N,S_N]^K} \cdots \int |\phi_{\boldsymbol{F}}(\boldsymbol{v})| \mathrm{d}\boldsymbol{v} \right.$$
$$\left. + \frac{1}{S_N^{qK}} \int_{\boldsymbol{v}\in[-S_N,S_N]^K} \cdots \int |\phi_{\boldsymbol{F}}(\boldsymbol{v})| \prod_{k=1}^{K} |v_k|^q \mathrm{d}\boldsymbol{v}\right) \qquad a.s.$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

The uniform bounds on the convergence rate of $\widehat{\phi}_{\boldsymbol{F}}$ suggest that estimators based on first-order partial derivatives of the $\widehat{\varphi_{\boldsymbol{Y}}}$ converge faster than estimators based on second-order partial derivatives. The bound using first-order partial derivatives is similar to Li and Vuong (1998) and the bound using second-order partial derivatives is the same as Bonhomme and Robin (2010).

The first term in the convergence rate of $\widehat{f}_{\boldsymbol{F}}$ depends on the dimension of $\boldsymbol{F}$, the bandwidth S_N and the decay of the tail of $\phi_{\boldsymbol{Y}}$. The second and third terms depend on the smoothing kernel, the bandwidth S_N and the decay of the tail of $\phi_{\boldsymbol{F}}$. The first term increases with S_N and thus for this term to shrink to zero, S_N cannot grow too quickly. The second and third terms, in opposition, shrink as S_N grows. This tradeoff between small and large bandwidth is typical in nonparametric estimators.

To be more concrete, the rate at which S_N can grow, and hence the size of δ, is restricted by the decay of the tails of $\phi_{\boldsymbol{Y}}$ and $\phi_{\boldsymbol{F}}$. For example, suppose that the distributions of $\boldsymbol{Y}$ and $\boldsymbol{F}$ are ordinary smooth so that $\phi_{\boldsymbol{Y}}(\boldsymbol{s}) = O(\prod s_p^{-\beta})$, $\beta > 0$, $s_p > 0$, and $\phi_{\boldsymbol{F}}(\boldsymbol{v}) = O(\prod v_k^{-\theta})$, $\theta > 0$, $v_k > 0$, where for simplicity the parameters are the same across coordinates (see, e.g., Comte and Lacour, 2013). Then using the estimator based on first-order partial derivatives,

$$\sup_{\boldsymbol{u}} \left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})\right| = O\left(\frac{\ln(N)}{N^{\frac{1}{2}-\delta(K+1+2\beta K)}} + N^{-K\delta(\theta-1)}\right) \quad \text{a.s.,} \qquad \textbf{(23)}$$

so that $0 < \delta < \frac{1}{2(K+1+2\beta K)}$ and $1 < \theta$, and using the estimator based on second-order partial derivatives,

$$\sup_{\boldsymbol{u}} \left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})\right| = O\left(\frac{\ln(N)}{N^{\frac{1}{2}-\delta(K+2+3\beta K)}} + N^{-K\delta(\theta-1)}\right) \quad \text{a.s.,} \qquad \textbf{(24)}$$

so that $0 < \delta < \frac{1}{2(K+2+3\beta K)}$ and $1 < \theta$. Hence, $S_N = N^{\delta}$ can grow faster when the estimator is based on first-order partial derivatives.

To compare the convergence rates above with the rest of the literature let $K = 1$, $\delta = \frac{1}{2(1+2\beta+\theta)}$ in (23), $\delta = \frac{1}{2(2+3\beta+\theta)}$ in (24), and let $\beta = \theta$ approach infinity. Then (23) is $\sup_{\boldsymbol{u}} \left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})\right| = O\left(\frac{\ln N}{N^{\frac{1}{6}}}\right)$ and (24) is $\sup_{\boldsymbol{u}} \left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})\right| = O\left(\frac{\ln N}{N^{\frac{1}{8}}}\right)$. The convergence rate based on first-order partial derivatives of $\widehat{\varphi_{\boldsymbol{Y}}}$ is similar to Li and Vuong (1998), who obtain the convergence rate $O\left(\frac{\ln \ln N}{N^{\frac{1}{6}}}\right)$. The convergence rate based on second-order partial derivatives is exactly the same as Bonhomme and Robin (2010).

Finally, we mention that the growth of S_N is much slower when either $\phi_{\boldsymbol{Y}}$ or $\phi_{\boldsymbol{F}}$ is supersmooth and in fact S_N is often restricted to grow at a logarithmic rate

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

for convergence (see, e.g., Fan, 1991; Li and Vuong, 1998; Delaigle et al., 2008; Bonhomme and Robin, 2010).

5. MONTE CARLO SIMULATIONS

This section presents Monte Carlo studies of the finite sample properties of estimators in a measurement error model with repeated measurements and then in the earnings dynamics model from Section 3.3. The goal of this section is to present evidence that estimators based on first or second-order partial derivatives of the empirical log characteristic function perform well in finite samples even when some unobservables are dependent.

5.1. Measurement Error Model With a Repeated Measurement

In the first set of simulations, N i.i.d. observations $\{Y_{1n}, Y_{2n}\}_{n=1}^{N}$ are generated from the repeated measurement error model,

$$Y_1 = F + U_1, \tag{25}$$

$$Y_2 = F + U_2, \tag{26}$$

where Y_1 and Y_2 are observed measurements of the unobserved variable F, and U_1 and U_2 are independent measurement errors. We draw $1,000$ simulations of sample size $N \in \{50, 100, 250\}$ from one of the distributional specifications in Table 1.

The distributional specifications are chosen to illustrate the effects on estimation from features of a density that could be encountered in practice. Specifications 1 and 2 have F with two common parametric densities. Specification 3 has F with a bimodal distribution, which has a density with two peaks. Specification 4 has F with a uniform distribution, which has a discontinuous density. The characteristic functions of the bimodal and uniform distributions vanish on sets of zero measure, with the characteristic function of the uniform distribution having a countable number of zeros.

We compare three estimators of f_F using one of the following estimators for φ_F,

TABLE 1. Distributional specifications of (F, U_1, U_2) in the repeated measurement error model.

Specification	F	U_1	U_2
1	N(0,1)	N(0,1)	N(0,1)
2	Gam(5,1)	N(0,1)	N(0,1)
3	Bimodal	N(0,1)	N(0,1)
4	Uniform(0,2)	TN(0,$\frac{1}{4}$)	TN(0,$\frac{1}{4}$)

Notes: Bimodal is $\frac{1}{2}N(-2,1) + \frac{1}{2}N(2,1)$ and TN(0,$\frac{1}{4}$) is a truncated N(0,$\frac{1}{4}$) with lower bound -1 and upper bound 1.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

$$\widehat{\varphi}_F(v) = \int_0^v \frac{iE_N[Y_2 \exp(iuY_1)]}{E_N[\exp(iuY_1)]}\mathrm{d}u, \qquad \text{(FDE1)},$$

$$\widehat{\varphi}_F(v) = \widehat{\varphi}_{Y_2}(v) - \widehat{\varphi}_{U_2}(v) \text{ where}$$

$$\widehat{\varphi}_{U_2}(v) = \int_0^v \frac{iE_N[(Y_2 - Y_1)\exp(iuY_2)]}{E_N[\exp(iuY_2)]}\mathrm{d}u, \qquad \text{(FDE2)},$$

$$\widehat{\varphi}_F(v) = \int_0^v \int_0^{u_2} \left(\frac{E_N[Y_1 e^{iu_1(Y_1+Y_2)/2}]E_N[Y_2 e^{iu_1(Y_1+Y_2)/2}]}{(E_N[e^{iu_1(Y_1+Y_2)/2}])^2} - \frac{iE_N[Y_1 Y_2 e^{iu_1(Y_1+Y_2)/2}]}{E_N[e^{iu_1(Y_1+Y_2)/2}]} \right) \mathrm{d}u_1 \mathrm{d}u_2, \qquad \text{(SDE)}.$$

The first two estimators are based on first-order partial derivatives of the empirical log characteristic function and constructed using (4) and the third estimator is based on second-order partial derivatives of the empirical log characteristic function and constructed using (10). FDE1 is used, for example, in Li and Vuong (1998), FDE2 has not been used to our knowledge, and SDE is used by Bonhomme and Robin (2010). Details of how the estimators are constructed from identification is in the appendix.

The density f_F is estimated using (22) with $\widehat{\phi}_F$ smoothed by the Fourier transform of a second-order kernel, $\phi_\kappa(v) = (1 - v^2)^3 I\{v \in [-1, 1]\}$. We choose bandwidths that produce density estimates that look reasonable. The choices are not optimal and a range of bandwidths produce broadly similar results.[3]

Figures 1 and 2 compare FDE1, FDE2 and SDE and show the effects of sample size on medians, and 2.5–97.5% confidence bands.

Despite slow asymptotic convergence rates, all the estimators perform well—and better than expected—in finite samples (see also Delaigle, 2008), and even with small sample sizes, the medians lie almost on top of the target densities. For all distributional specifications, the target densities are tightly bound between the 2.5 and 97.5% confidence bands, with the bands shrinking with sample size.

5.2. Earnings Dynamics Model

In the second set of simulations, we consider the earnings dynamics model from (15)–(17) in Section 3.3, allowing $(f + y_1^P, \eta_2, \eta_3)$ to be dependent.

Let $\{\varepsilon_{2n}, \varepsilon_{3n}, \varepsilon_{4n}\}_{n=1}^N$ be i.i.d. draws from $N(0, I_3)$ and let $\{f_n + y_{1n}^P, \eta_{2n}, \eta_{3n}\}_{n=1}^N$ be i.i.d. draws from

$$\begin{pmatrix} f + y_1^P \\ \eta_2 \\ \eta_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix} \right).$$

First we estimate the marginal densities of ε_3 and ε_4 using the sample analogs of (21) from data generated by 100 simulations of sample size 500. Next we estimate the joint density of (η_2, η_3) using the sample analogs of (20) from data generated by 100 simulations of sample size 5,000. Like the

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

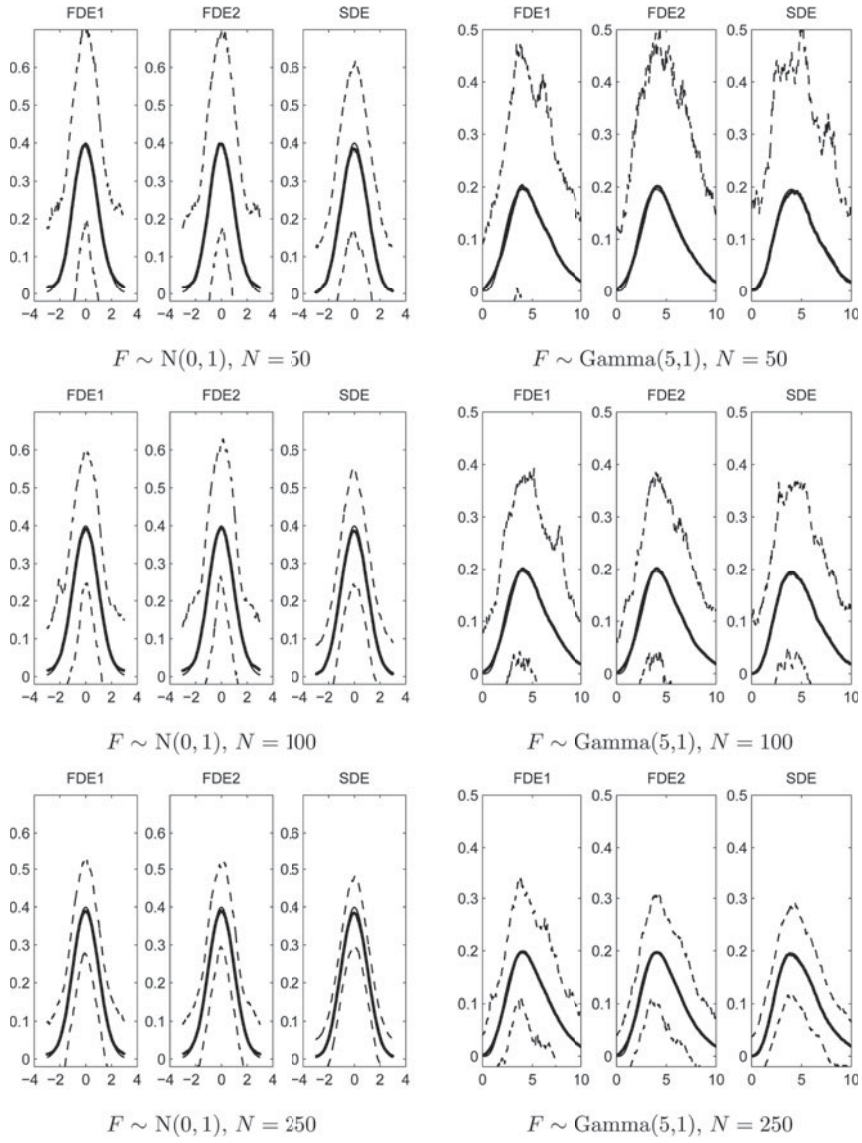


FIGURE 1. Estimates of f_F using FDE1, FDE2 and SDE in the repeated measurement error model.

Notes: Results are from 1,000 simulations. The thin solid lines are population quantities, the thick solid lines are medians and the dashed lines are the 2.5–97.5% confidence bands.

previous set of simulations, we use the second-order kernel with Fourier transform $\phi_\kappa(\boldsymbol{v}) = \prod_k (1 - v_k^2)^3 I\{v_k \in [-1, 1]\}$ and choose bandwidths that produce density estimates that look reasonable.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

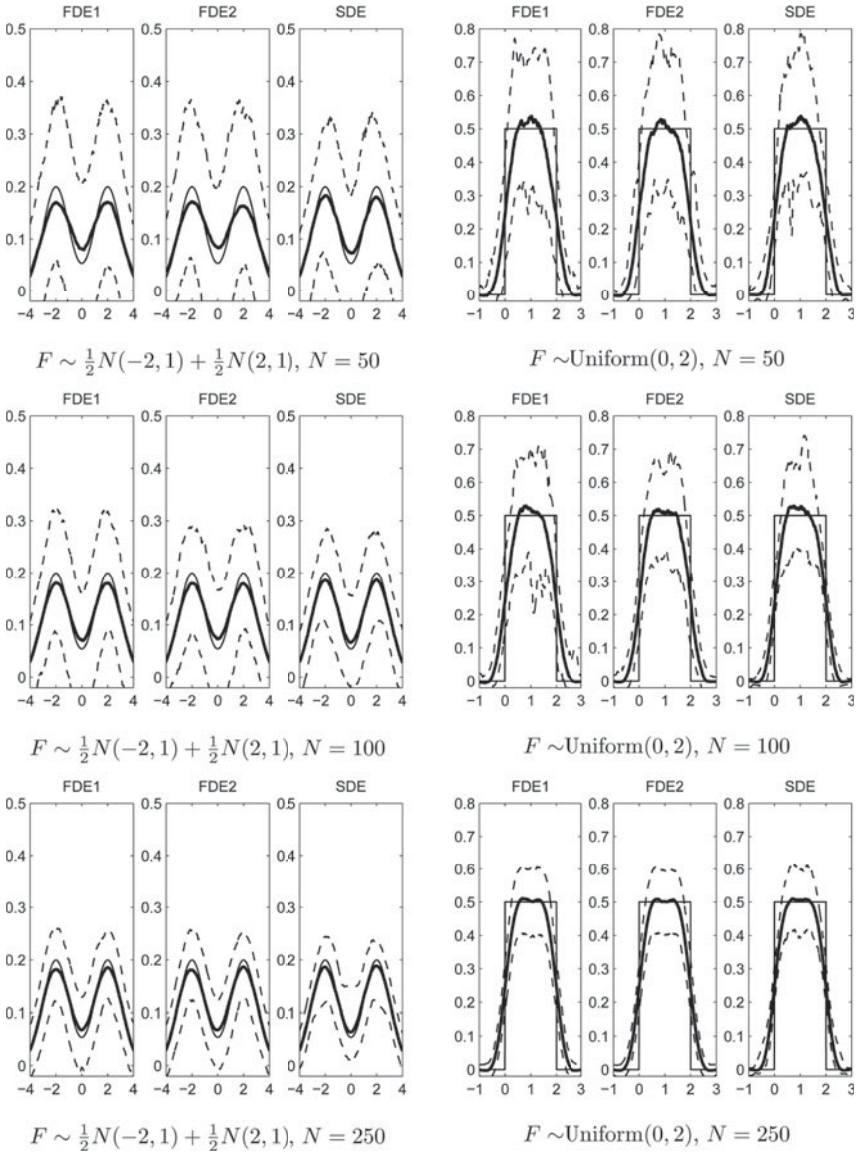


FIGURE 2. Estimates of f_F using FDE1, FDE2 and SDE in the repeated measurement error model.

Notes: Results are from 1,000 simulations. The thin solid lines are population quantities, the thick solid lines are medians and the dashed lines are the 2.5–97.5% confidence bands.

Figures 3(a) and 3(b) plot the medians and 10–90% confidence bands of the estimated marginal densities and the target standard normal densities. Figures 4(a) and 4(b) plot the median of the estimated joint density and the target jointly

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

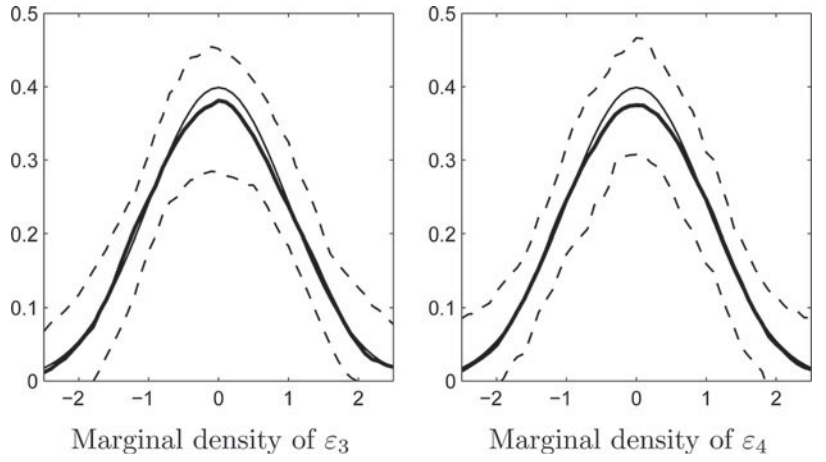


FIGURE 3. Estimates of f_{ε_3} and f_{ε_4} from the earnings dynamics model from (15)–(17).
Notes: Results are from 100 simulations of sample size 500. The thin solid lines are population quantities, the thick solid lines are medians and the dashed lines are the 10–90% confidence bands.

normal density. Due to the curse of dimensionality estimating the marginal densities well requires a smaller sample size than the joint density. Overall the estimates do a good job reproducing the target densities. This suggests that estimating densities from inverse Fourier transforms can be successfully applied to empirical problems, even when the number of unobservables is large and some of the unobservables are dependent.

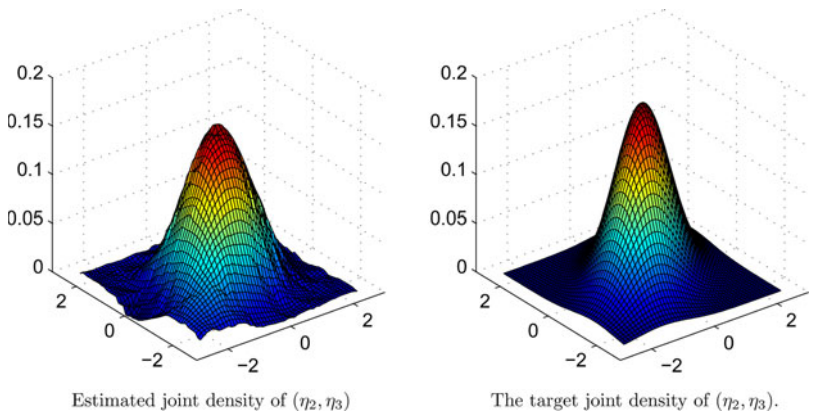


FIGURE 4. Estimates of f_{η_2,η_3} from the earnings dynamics model from (15)–(17).
Notes: Results are from 100 simulations of sample size 5,000.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

6. CONCLUSIONS

This paper identifies nonparametric joint distributions of unobserved dependent variables in a system of linear equations with noise. The identification strategies involve taking partial derivatives of the log characteristic function of observables and are analogous to identification of means and covariances. An illustrative example presents identification in an earnings dynamics model with dependent income shocks. The identification proofs provide closed-form formulas for the log characteristic function of the factors and estimators replace population quantities with sample analogs. The estimators converge uniformly and in finite sample simulations the medians of the estimates closely match the target densities.

The identification strategies can be extended to higher-order partial derivatives of $\varphi_{\boldsymbol{Y}}$, and analogously also to higher-order cumulants. Székely and Rao (2000) analyze this when unobservables are mutually independent. The drawback of such strategies is that log characteristic functions are identified up to higher-order moments. For example, identification using third-order partial derivatives determines the joint distribution of $\boldsymbol{F}$ up to variances. Using higher-order partial derivatives, however, could be used to identify and estimate other objects of interest like in independent component analysis.

The estimation results from Theorem 4.1 could possibly be used to test if $\boldsymbol{F}$ follows a particular distribution by using one of the test statistics $\sup_{\boldsymbol{u}} |\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v}) - \phi_{\boldsymbol{F}}(\boldsymbol{v})|$ or $\sup_{\boldsymbol{u}} |\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u}) - f_{\boldsymbol{F}}(\boldsymbol{u})|$, similar to a Kolmogorov-Smirnov test. Since we do not derive lower bounds on the convergence rates or asymptotic distributions, one would need to use a procedure like bootstrapping for proper inference.

In a companion paper, Ben-Moshe (2016) analyzes identification and estimation of coefficients in the linear factor model with arbitrarily dependent factors.

NOTES

1. Noting that $(H_{A_1} : \mathbf{0})A = (\mathbf{0} : H_{A_2})A = I_K$ and $(H_{A_1} : \mathbf{0}) \circ (\mathbf{0} : H_{A_1}) = \mathbf{0}$ (where $\circ$ is the Hadamard product), Assumption 2.2 can be somewhat weakened to the requirement that for each $k = 1, \ldots, K$ there exist $K \times P$ matrices F_A and G_A such that $F_A A = G_A A = I_K$ and $F_{Akp} G_{Akp} = 0$, for $p = 1, \ldots, P$. This is at the expense of transparency and additional notation.

2. $A\boldsymbol{F}$ and $\boldsymbol{U}$ are subindependent if $E[\exp(i\boldsymbol{s}'(A\boldsymbol{F} + \boldsymbol{U}))] = E[\exp(i\boldsymbol{s}'A\boldsymbol{F})]E[\exp(i\boldsymbol{s}'\boldsymbol{U})]$ for all $\boldsymbol{s}$, while $A\boldsymbol{F}$ and $\boldsymbol{U}$ are independent if $E[\exp(i\boldsymbol{s}_1' A\boldsymbol{F} + i\boldsymbol{s}_2'\boldsymbol{U})] = E[\exp(i\boldsymbol{s}_1' A\boldsymbol{F})]E[\exp(i\boldsymbol{s}_2'\boldsymbol{U})]$ for all $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. Thanks to Áureo de Paula for pointing this out to me.

3. Plug-in bandwidth selectors resulted in parts of f_F lying outside the 2.5–97.5% confidence bands (see Delaigle and Gijbels, 2004, for a discussion on bandwidth selection).

REFERENCES

Abadir, K. & J. Magnus (2005) *Matrix Algebra* (*Econometric Exercises*). Cambridge University Press, ISBN 9780521822893.

Aigner, D.J., C. Hsiao, A. Kapteyn, & T. Wansbeek (1984) Latent variable models in econometrics. *Handbook of Econometrics* 2, 1321–1393.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Anderson, T.W. & H. Rubin (1956) Statistical inference in factor analysis. In J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5, p. 1. Berkeley: University of California Press.

Anselin, L. (2013) *Spatial Econometrics: Methods and Models*, vol. 4. Springer Science & Business Media.

Arellano, M. & S. Bonhomme (2012) Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies* 79(3), 987–1020.

Ben-Moshe, D. (2016) Identification and Estimation of Coefficients in Dependent Factor Models. Working paper, The Hebrew University of Jerusalem.

Beran, R., A. Feuerverger, & P. Hall (1996) On nonparametric estimation of intercept and slope distributions in random coefficient regression. *The Annals of Statistics* 24(6), 2569–2592.

Bondesson, L. (1974) Characterizations of probability laws through constant regression. *Probability Theory and Related Fields* 30(2), 93–115.

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

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

Browning, M., M. Ejrnaes, & J. Alvarez (2010) Modelling income processes with lots of heterogeneity. *The Review of Economic Studies* 77(4), 1353–1381.

Carneiro, P., K. Hansen, & J.J. Heckman (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.

Chamberlain, G. & Z. Griliches (1975) Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers. *International Economic Review*, vol. 16, 422–449.

Comon, P. & C. Jutten (2010) *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic press.

Comte, F. & C. Lacour (2013) Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 49, pp. 569–609. Institut Henri Poincaré.

Conti, G. & J.J. Heckman (2010) Understanding the early origins of the education–health gradient a framework that can also be applied to analyze gene–environment interactions. *Perspectives on Psychological Science* 5(5), 585–605.

Conti, G., J. Heckman, & S. Urzua (2010) The education-health gradient. *The American Economic Review* 100(2), 234.

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

Delaigle, A. (2008) An alternative view of the deconvolution problem. *Statistica Sinica*, vol. 18, 1025–1045.

Delaigle, A. & I. Gijbels (2004) Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics & Data Analysis* 45(2), 249–267.

Delaigle, A., P. Hall, & A. Meister (2008) On deconvolution with repeated measurements. *The Annals of Statistics*, vol. 36, 665–685.

Ejrnæs, M. & M. Browning (2014) The persistent–transitory representation for earnings processes. *Quantitative Economics* 5(3), 555–581.

Evdokimov, K. & H. White (2012) Some extensions of a lemma of kotlarski. *Econometric Theory* 28(04), 925–932.

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

Fan, J. (1991) Global behavior of deconvolution kernel estimates. *Statistica Sinica* 1(2), 541–551.

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

Geer, S.A. (2000) *Empirical Processes in M-estimation*, vol. 6. Cambridge University Press.

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

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Gorman, W.M. (1981) Some engel curves. In A. Deaton (ed.), *Essays in the Theory and Measurement of Consumer Behaviour in Honor of Sir Richard Stone*, 7–29. New York: Cambridge University Press.

Graham, B.S. & J.L. Powell (2012) Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica*, vol. 80, 2105–2152.

Hamedani, G., H. Volkmer, & J. Behboodian (2011) A note on sub-independent random variables and a class of bivariate mixtures. *Studia Scientiarum Mathematicarum Hungarica* 49(1), 19–25.

Hoderlein, S., H. Holzmann, & A. Meister (2014) The Triangular Model with Random Coefficients. Working paper, Boston College.

Hoderlein, S., J. Klemelä, & E. Mammen (2010) Analyzing the random coefficient model nonparametrically. *Econometric Theory* 26(03), 804–837.

Horowitz, J.L. & M. Markatou (1996) Semiparametric estimation of regression models for panel data. *The Review of Economic Studies* 63(1), 145–168.

Hyvärinen, A., J. Karhunen, & E. Oja (2004) *Independent Component Analysis*, vol. 46. John Wiley & Sons.

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.

Khatri, C. & C.R. Rao (1968) Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 30, 167–180.

Kotlarski, I. (1967) On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics* 20(1), 69–76.

Lewbel, A. (1991) The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, vol. 59, 711–730.

Li, T. & Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* 65(2), 139–165.

Masten, M. (2014) Random Coefficients on Endogenous Variables in Simultaneous Equations Models. Tech. rep. Cemmap Working paper, Centre for Microdata Methods and Practice.

Meghir, C. & L. Pistaferri (2011) Earnings, consumption and life cycle choices. *Handbook of Labor Economics* 4, 773–854.

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer.

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

Schennach, S.M. (2004) Estimation of nonlinear models with measurement error. *Econometrica* 72(1), 33–75.

Schennach, S. (2007) Instrumental variables estimation of nonlinear errors-in-variables models. *Econometrica* 75, 201–239.

Schennach, S.M. (2013) Convolution Without Independence. Tech. rep. Cemmap Working paper, Centre for Microdata Methods and Practice.

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

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

Székely, G. & C. Rao (2000) Identifiability of distributions of independent random variables by linear combinations and moments. *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 62, 193–202.

Williams, B. (2015) Identification of the Linear Factor Model. Working paper, George Washington University.

Zinde-Walsh, V. (2014) Measurement error and deconvolution in spaces of generalized functions. *Econometric Theory* 30, 1207–1246.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

APPENDIX A: Proofs

The log characteristic function of (1) is,

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \sum_{p=1}^{P} \varphi_{\boldsymbol{U}}(\boldsymbol{s}), \tag{A.1}$$

where the functions are separated by Assumption 2.1 or 2.4.

A.1. Proof of Theorem 2.1

The mean of (1) and the first-order partial derivatives of (A.1) are,

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = A\boldsymbol{\mu}_{\boldsymbol{F}} + \boldsymbol{\mu}_{\boldsymbol{U}}, \tag{A.2}$$

$$\nabla\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = A\nabla\varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \nabla\varphi_{\boldsymbol{U}}(\boldsymbol{s}). \tag{A.3}$$

Note that (A.2) could be obtained by setting $\boldsymbol{s} = \boldsymbol{0}$ in (A.3). Now left multiply (A.2) and (A.3) by $(H_{A_1} : \boldsymbol{0})$ and use Assumptions 2.1 and 2.2,

$$\boldsymbol{\mu}_{\boldsymbol{F}} = (H_{A_1} : \boldsymbol{0})(\boldsymbol{\mu}_{\boldsymbol{Y}} - \boldsymbol{\mu}_{\boldsymbol{U}}) = (H_{A_1} : \boldsymbol{0})\boldsymbol{\mu}_{\boldsymbol{Y}} = H_{A_1}\boldsymbol{\mu}_{\boldsymbol{Y}_1},$$

$$\nabla\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = (H_{A_1} : \boldsymbol{0})(\nabla\varphi_{\boldsymbol{Y}_1,\boldsymbol{Y}_2}(\boldsymbol{0}, H'_{A_2}\boldsymbol{v}) - \nabla\varphi_{\boldsymbol{U}_1,\boldsymbol{U}_2}(\boldsymbol{0}, H'_{A_2}\boldsymbol{v})),$$

where $(H_{A_1} : \boldsymbol{0})\boldsymbol{\mu}_{\boldsymbol{U}} = H_{A_1}\boldsymbol{\mu}_{\boldsymbol{U}_1} = H_{A_1}E[E[\boldsymbol{U}_1|\boldsymbol{U}_2]] = \boldsymbol{0}$. The k^{th} row of $\nabla\varphi_{\boldsymbol{F}}(\boldsymbol{v})$ is,

$$\begin{aligned}
\frac{\partial\varphi_{\boldsymbol{F}}(\boldsymbol{v})}{\partial v_k} &= \sum_{p=1}^{P_1} H_{A_1 kp}\left(\left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'} - \left.\frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'}\right) \\
&= \sum_{p=1}^{P_1} H_{A_1 kp}\left(\left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'} - \frac{iE\left[U_p e^{i\boldsymbol{U}'_2H'_{A_2}\boldsymbol{v}}\right]}{E\left[e^{i\boldsymbol{U}'_2H'_{A_2}\boldsymbol{v}}\right]}\right) \\
&= \sum_{p=1}^{P_1} H_{A_1 kp}\left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'} - \sum_{p=1}^{P_1} H_{A_1 kp}\frac{iE\left[E[U_p|\boldsymbol{U}_2]e^{i\boldsymbol{U}'_2H'_{A_2}\boldsymbol{v}}\right]}{E[e^{i\boldsymbol{U}'_2H'_{A_2}\boldsymbol{v}}]} \\
&= \sum_{p=1}^{P_1} H_{A_1 kp}\left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'},
\end{aligned}$$

where the last equality follows by Assumption 2.1. Finally, using the formulas above, the fundamental theorem of calculus and the boundary condition $\varphi_{\boldsymbol{F}}(\boldsymbol{0}) = \ln E[e^0] = 0$,

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = \int_0^1 \boldsymbol{v}'\nabla\varphi_{\boldsymbol{F}}(u\boldsymbol{v})\mathrm{d}u = \int_0^1 \sum_{k=1}^{K}\sum_{p=1}^{P_1} v_k H_{A_1 kp}\left.\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_p}\right|_{\boldsymbol{s}=(\boldsymbol{0},\boldsymbol{v}'H_{A_2})'}\mathrm{d}u.$$

We conclude the proof by noting that

$$\left|\frac{\partial\varphi_{\boldsymbol{F}}(\boldsymbol{v})}{\partial v_k}\right| = \left|\frac{iE\left[F_k e^{i\boldsymbol{F}'\boldsymbol{v}}\right]}{E\left[e^{i\boldsymbol{F}'\boldsymbol{v}}\right]}\right| \leq \frac{\max_k E[|F_k|]}{|E\left[e^{i\boldsymbol{F}'\boldsymbol{v}}\right]|} < \infty,$$

where the last inequality follows by Assumption 2.3. Hence, $\varphi_{\boldsymbol{F}}$ is identified and the joint distribution of $\boldsymbol{F}$ is identified by the inverse Fourier transform.

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

A.2. Proof of Theorem 2.2

The variance of (1) and the second-order partial derivatives of (A.1) are,

$$\Sigma_{\boldsymbol{Y}} = A\Sigma_{\boldsymbol{F}}A' + \Sigma_{\boldsymbol{U}}, \tag{A.4}$$

$$\nabla^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = A\nabla^2\varphi_{\boldsymbol{F}}(A'\boldsymbol{s})A' + \nabla^2\varphi_{\boldsymbol{U}}(\boldsymbol{s}). \tag{A.5}$$

Note that (A.4) could be obtained by setting $\boldsymbol{s} = \boldsymbol{0}$ in (A.5). Now using Assumption 2.4 and that $\mathrm{Cov}(F_{k_1}, F_{k_2}) = \mathrm{Cov}(F_{k_2}, F_{k_1})$,

$$\begin{aligned}
\mathrm{Cov}(Y_{p_1}, Y_{p_2}) &= \mathrm{Cov}\left(\sum_{k=1}^{K} a_{p_1k}F_k + U_{p_1}, \sum_{k=1}^{K} a_{p_2k}F_k + U_{p_2}\right) \\
&= \sum_{k_1,k_2} a_{p_1k_1}a_{p_2k_2}\mathrm{Cov}(F_{k_1}, F_{k_2}) + \mathrm{Cov}(U_{p_1}, U_{p_2}) \\
&= \sum_{k=1}^{K} a_{p_1k}a_{p_2k}\mathrm{Var}(F_k) + \sum_{k_1<k_2} (a_{p_1k_1}a_{p_2k_2} + a_{p_1k_2}a_{p_2k_1})\mathrm{Cov}(F_{k_1}, F_{k_2}) + \mathrm{Cov}(U_{p_1}, U_{p_2}) \\
&= (a_{p_11}a_{p_21}, \ldots, a_{p_1k_1}a_{p_2k_2} + a_{p_2k_1}a_{p_1k_2}, \ldots, a_{p_1K}a_{p_2K}, 1)\begin{pmatrix} \mathrm{vech}(\Sigma_{\boldsymbol{F}}) \\ \mathrm{Cov}(U_{p_1}, U_{p_2}) \end{pmatrix},
\end{aligned} \tag{A.6}$$

where $\mathrm{Cov}(U_{p_1}, U_{p_2}) = 0$, for all $p_1 \neq p_2$, by Assumption 2.4. Similarly,

$$\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_{p_1}\partial s_{p_2}} = (a_{p_11}a_{p_21}, \ldots, a_{p_1k_1}a_{p_2k_2} + a_{p_2k_1}a_{p_1k_2}, \ldots, a_{p_1K}a_{p_2K}, 1)\begin{pmatrix} \mathrm{vech}(\nabla^2\varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) \\ \dfrac{\partial^2\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_{p_1}\partial s_{p_2}} \end{pmatrix},$$

where $\partial^2\varphi_{\boldsymbol{U}}(\boldsymbol{s})/\partial s_{p_1}\partial s_{p_2} = 0$, for all $p_1 \neq p_2$, by Assumption 2.4. Now vectorizing $\Sigma_{\boldsymbol{Y}}$,

$$\begin{aligned}
\mathrm{vec}(\Sigma_{\boldsymbol{Y}}) &= (A \otimes A)\mathrm{vec}(\Sigma_{\boldsymbol{F}}) + (I_P \otimes I_P)\mathrm{vec}(\Sigma_{\boldsymbol{U}}) = (A \odot A)\mathrm{vech}(\Sigma_{\boldsymbol{F}}) + (I_P \star I_P)\mathrm{diag}(\Sigma_{\boldsymbol{U}}) \\
&= ((A \odot A) : (I_P \star I_P))\begin{pmatrix} \mathrm{vech}(\Sigma_{\boldsymbol{F}}) \\ \mathrm{diag}(\Sigma_{\boldsymbol{U}}) \end{pmatrix},
\end{aligned}$$

where the first equality follows by the connection between vectorization of matrices and the Kronecker product (e.g., Abadir and Magnus, 2005) and the second equality by (A.2). Similarly,

$$\mathrm{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})) = ((A \odot A) : (I_P \star I_P))\begin{pmatrix} \mathrm{vech}(\nabla^2\varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) \\ \mathrm{diag}(\nabla^2\varphi_{\boldsymbol{U}}(\boldsymbol{s})) \end{pmatrix}.$$

To show sufficiency let Assumption 2.5 hold,

$$\begin{pmatrix} \mathrm{vech}(\Sigma_{\boldsymbol{F}}) \\ \mathrm{diag}(\Sigma_{\boldsymbol{U}}) \end{pmatrix} = H_{(A\odot A):(I_P\star I_P)}\mathrm{vec}(\Sigma_{\boldsymbol{Y}}),$$

$$\begin{pmatrix} \mathrm{vech}(\nabla^2\varphi_{\boldsymbol{F}}(\boldsymbol{v})) \\ \mathrm{diag}(\nabla^2\varphi_{\boldsymbol{U}}(H_A'\boldsymbol{v})) \end{pmatrix} = H_{(A\odot A):(I_P\star I_P)}\mathrm{vec}(\nabla^2\varphi_{\boldsymbol{Y}}(H_A'\boldsymbol{v})).$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Then using the above equation and the fundamental theorem of calculus,

$$\varphi_{\boldsymbol{F}}(\boldsymbol{v}) = i\boldsymbol{v}'\boldsymbol{\mu}_{\boldsymbol{F}} + \int_0^1 \int_0^{u_2} \boldsymbol{v}'\nabla^2\varphi_{\boldsymbol{F}}(u_1\boldsymbol{v})\boldsymbol{v}\,\mathrm{d}u_1\mathrm{d}u_2,$$

where $\boldsymbol{\mu}_{\boldsymbol{F}} = H_A\boldsymbol{\mu}_{\boldsymbol{Y}}$.

Finally,

$$\left|\frac{\partial^2\varphi_{\boldsymbol{F}}(\boldsymbol{v})}{\partial v_{k_1}\partial v_{k_2}}\right| = \left|\frac{E\left[F_{k_1}e^{i\boldsymbol{F}'\boldsymbol{v}}\right]E\left[F_{k_2}e^{i\boldsymbol{F}'\boldsymbol{v}}\right]}{\left(E\left[e^{i\boldsymbol{F}'\boldsymbol{v}}\right]\right)^2} - \frac{E\left[F_{k_1}F_{k_2}e^{i\boldsymbol{F}'\boldsymbol{v}}\right]}{E\left[e^{i\boldsymbol{F}'\boldsymbol{v}}\right]}\right| \leq \frac{\max_k E[F_k^2]}{\left(E\left[e^{iu\boldsymbol{F}'\boldsymbol{v}}\right]\right)^2} < \infty,$$

where the last inequality follows by Assumption 2.6.

Now to show necessity let Assumption 2.5 not hold and let $(\widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}})$ be observationally equivalent to $(\boldsymbol{F}, \boldsymbol{U})$. If Assumption 2.5(i) does not hold then there exists a nonzero $\boldsymbol{c} \in \mathbb{R}^{\frac{K(K+1)}{2}}$ and $\boldsymbol{b} \in \mathbb{R}^P$ such that

$$((A \odot A) : (I_P \star I_P))\begin{pmatrix}\boldsymbol{c}\\ \boldsymbol{b}\end{pmatrix} = (A \odot A)\boldsymbol{c} + (I_P \star I_P)\boldsymbol{b} = \boldsymbol{0}.$$

Let C be the $K \times K$ symmetric matrix with lower triangle equal to $\mathrm{vech}^{-1}(\boldsymbol{c})$ and let B be a $P \times P$ diagonal matrix with diagonal $\boldsymbol{b}$. Then,

$$ACA' + B = \boldsymbol{0}.$$

Let $\mathrm{Var}(\widetilde{U}_p) = \mathrm{Var}(U_p) - \delta b_p$, for $p = 1, \ldots, P$ so that $\Sigma_{\widetilde{\boldsymbol{U}}} = \Sigma_{\boldsymbol{U}} - \delta B$ and let $\Sigma_{\widetilde{\boldsymbol{F}}} = \Sigma_{\boldsymbol{F}} - \delta C$, where $\delta \in \mathbb{R}$ is small. By Assumption 2.6, $\mathrm{Var}(U_p) > 0$ and $\Sigma_{\boldsymbol{F}}$ is positive definite so δ can always be chosen arbitrarily small so that $\mathrm{Var}(\widetilde{U}_p) > 0$ and $\Sigma_{\widetilde{\boldsymbol{F}}}$ is positive definite. Then,

$$\begin{aligned}\Sigma_{\boldsymbol{Y}} &= A\Sigma_{\boldsymbol{F}}A' + \Sigma_{\boldsymbol{U}} = A\Sigma_{\boldsymbol{F}}A' + \Sigma_{\boldsymbol{U}} - \delta(ACA' + B)\\ &= A(\Sigma_{\boldsymbol{F}} - \delta C)A' + (\Sigma_{\boldsymbol{U}} - \delta B) = A\Sigma_{\widetilde{\boldsymbol{F}}}A' + \Sigma_{\widetilde{\boldsymbol{U}}}.\end{aligned}$$

So $\Sigma_{\widetilde{\boldsymbol{F}}}$ and $\Sigma_{\boldsymbol{F}}$ are observationally equivalent and we can construct observationally equivalent $(\widetilde{\boldsymbol{F}}, \widetilde{\boldsymbol{U}})$ and $(\boldsymbol{F}, \boldsymbol{U})$ by $\varphi_{\widetilde{\boldsymbol{F}}}(\boldsymbol{v}) = \varphi_{\boldsymbol{F}}(\boldsymbol{v}) - \delta\boldsymbol{v}'C\boldsymbol{v}$ and $\varphi_{\widetilde{U}_p}(v_p) = \varphi_{U_p}(v_p) - \delta b_p v_p$, for $p = 1, \ldots, P$, so that $\widetilde{\boldsymbol{F}}$ shifts the variance of $\boldsymbol{F}$ and $\widetilde{U}_p$ shifts the variance of U_p and thus the log characteristic functions are well-defined for small enough δ. Further,

$$\begin{aligned}\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) &= \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \sum_{p=1}^P \varphi_{U_p}(s_p) = \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \sum_{p=1}^P \varphi_{U_p}(s_p) - \delta\boldsymbol{s}'ACA'\boldsymbol{s} - \delta\boldsymbol{s}'B\boldsymbol{s}\\ &= \varphi_{\widetilde{\boldsymbol{F}}}(A'\boldsymbol{s}) + \sum_{p=1}^P \varphi_{\widetilde{U}_p}(s_p).\end{aligned}$$

Now if Assumption 2.5(ii) does not hold then there exists a nonzero $\boldsymbol{c} \in \mathbb{R}^K$ such that $A\boldsymbol{c} = \boldsymbol{0}$. Let $\widetilde{\boldsymbol{F}} = \boldsymbol{F} + \boldsymbol{c}F_1$ and $\widetilde{\boldsymbol{U}} = \boldsymbol{U}$. Then,

$$\begin{aligned}\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) &= \varphi_{\boldsymbol{F}}(A'\boldsymbol{s}) + \sum_{p=1}^P \varphi_{U_p}(s_p) = \ln E[\exp(i\boldsymbol{s}'(A\boldsymbol{F} + A\boldsymbol{c}F_1))] + \sum_{p=1}^P \varphi_{U_p}(s_p)\\ &= \ln E[\exp(i\boldsymbol{s}'A\widetilde{\boldsymbol{F}})] + \sum_{p=1}^P \varphi_{\widetilde{U}_p}(s_p) = \varphi_{\widetilde{\boldsymbol{F}}}(A'\boldsymbol{s}) + \sum_{p=1}^P \varphi_{\widetilde{U}_p}(s_p).\end{aligned}$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

A.3. Proof of Theorem 2.3

The log characteristic function of (11) is,

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = \sum_{j=1}^{q} \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}), \tag{A.7}$$

where the functions are separated by the assumption that $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_J$ are mutually independent. The variance of (11) and the second-order partial derivatives of (A.7) are,

$$\begin{aligned}
\Sigma_{\boldsymbol{Y}} &= \sum_{j=1}^{J} A_j \Sigma_{\boldsymbol{F}_j} A_j', \\
\nabla^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s}) &= \sum_{j=1}^{J} A_j \nabla^2 \varphi_{\boldsymbol{F}_j}(A_j'\boldsymbol{s}) A_j'.
\end{aligned}$$

Vectorizing these equations we obtain,

$$\begin{aligned}
\operatorname{vec}(\Sigma_{\boldsymbol{Y}}) &= ((A_1 \odot A_1) : \cdots : (A_J \odot A_J))\operatorname{vech}(\Sigma_{\boldsymbol{F}}), \\
\operatorname{vec}(\nabla^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})) &= ((A_1 \odot A_1) : \cdots : (A_J \odot A_J))\operatorname{vech}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})),
\end{aligned}$$

where $((A_1 \odot A_1) : \cdots : (A_J \odot A_J))$ is a $P^2 \times \sum_{j=1}^{J} \frac{K_j(K_j+1)}{2}$ matrix, $\operatorname{vech}(\Sigma_{\boldsymbol{F}}) = (\operatorname{vech}(\Sigma_{\boldsymbol{F}_1})', \ldots, \operatorname{vech}(\Sigma_{\boldsymbol{F}_J})')'$ and $\operatorname{vech}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) = (\operatorname{vech}(\nabla^2 \varphi_{\boldsymbol{F}_1}(A_1'\boldsymbol{s}))', \ldots, \operatorname{vech}(\nabla^2 \varphi_{\boldsymbol{F}_J}(A_J'\boldsymbol{s}))')'$. Now because $\operatorname{Rank}((A_1 \odot A_1) : \cdots : (A_J \odot A_J)) = \sum_{j=1}^{J} K_j(K_j+1)/2$ we can invert the above equations to obtain,

$$\begin{aligned}
\operatorname{vech}(\Sigma_{\boldsymbol{F}}) &= H_{(A_1 \odot A_1):\cdots:(A_J \odot A_J)} \operatorname{vec}(\Sigma_{\boldsymbol{Y}}), \\
\operatorname{vech}(\nabla^2 \varphi_{\boldsymbol{F}}(A'\boldsymbol{s})) &= H_{(A_1 \odot A_1):\cdots:(A_J \odot A_J)} \operatorname{vec}(\nabla^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})).
\end{aligned}$$

Then using the above equation, $\operatorname{Rank}(A_j) = K_j$ and the fundamental theorem of calculus,

$$\varphi_{\boldsymbol{F}_j}(\boldsymbol{v}_j) = i\boldsymbol{v}_j'\boldsymbol{\mu}_{\boldsymbol{F}_j} + \int_0^1 \int_0^{u_2} \boldsymbol{v}_j' \nabla^2 \varphi_{\boldsymbol{F}_j}(u_1 \boldsymbol{v}_j)\boldsymbol{v}_j \, \mathrm{d}u_1 \mathrm{d}u_2, \qquad j = 1, \ldots, J,$$

and if $\operatorname{Rank}(A) = \sum_{j=1}^{J} K_j$ then $\boldsymbol{\mu}_{\boldsymbol{F}} = H_A \boldsymbol{\mu}_{\boldsymbol{Y}}$.

Finally, because $\Sigma_{\boldsymbol{F}_j}$ is finite and $E\left[e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right] \neq 0$,

$$\begin{aligned}
\left| \frac{\partial^2 \varphi_{\boldsymbol{F}_j}(\boldsymbol{v}_j)}{\partial v_{k_1} \partial v_{k_2}} \right| &= \left| \frac{E\left[F_{jk_1} e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right] E\left[F_{jk_2} e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right]}{\left(E\left[e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right]\right)^2} - \frac{E\left[F_{jk_1} F_{jk_2} e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right]}{E\left[e^{i\boldsymbol{F}_j'\boldsymbol{v}_j}\right]} \right| \\
&\leq \frac{\max_k E\left[F_{jk}^2\right]}{\left(E\left[e^{iu\boldsymbol{F}_j'\boldsymbol{v}_j}\right]\right)^2} < \infty.
\end{aligned}$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

A.4. Proof of Lemma 4.1

Let $g_{\boldsymbol{s}}(\boldsymbol{Y}) = \prod_{p=1}^{P} Y_p^{\alpha_p} \exp(i\boldsymbol{Y}'\boldsymbol{s})$. To simplify notation, denote $\sigma_N^2 = E_N\left[\prod_{p=1}^{P} |Y_p|^{2\alpha_p}\right]$ and $\sigma^2 = E\left[\prod_{p=1}^{P} |Y_p|^{2\alpha_p}\right]$. Then,

$$
\begin{aligned}
&\Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P} |E_N[g_{\boldsymbol{s}}] - E[g_{\boldsymbol{s}}]| > \varepsilon\right) \\
&\quad = \Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P} |E_N[g_{\boldsymbol{s}}] - E[g_{\boldsymbol{s}}]| > \varepsilon \mid \sigma_N^2 \geq \gamma\right) \cdot \Pr\left(\sigma_N^2 \geq \gamma\right) \\
&\qquad + \Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P} |E_N[g_{\boldsymbol{s}}] - E[g_{\boldsymbol{s}}]| > \varepsilon \mid \sigma_N^2 < \gamma\right) \cdot \Pr\left(\sigma_N^2 < \gamma\right) \\
&\quad \leq \Pr\left(\sigma_N^2 \geq \gamma\right) + \Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P} |E_N[g_{\boldsymbol{s}}] - E[g_{\boldsymbol{s}}]| > \varepsilon \mid \sigma_N^2 < \gamma\right) \\
&\quad = B_1 + B_2.
\end{aligned}
$$

We now bound B_1 and B_2. Consider B_1,

$$
\begin{aligned}
\Pr\left(\sigma_N^2 \geq \gamma\right) &\leq \frac{E\left[e^{E_N\left[\prod_{p=1}^{P} |Y_p|^{2\alpha_p}\right]}\right]}{e^{\gamma}} = \frac{\prod_{n=1}^{N} E\left[e^{\prod_{p=1}^{P} |Y_p|^{2\alpha_p}/N}\right]}{e^{\gamma}} \\
&= \frac{\left(E\left[e^{\prod_{p=1}^{P} |Y_p|^{2\alpha_p}/N}\right]\right)^N}{e^{\gamma}} = \frac{\left(1 + \frac{E\left[\prod_{p=1}^{P} |Y_p|^{2\alpha_p}\right]}{N} + O\left(\frac{1}{N^2}\right)\right)^N}{e^{\gamma}} \leq \frac{e^{\sigma^2}}{e^{\gamma}},
\end{aligned}
$$

where the first inequality follows by Chernoff's bound and the last equality by Assumption 4.1(ii).

To bound B_2, we follow Pollard (1984) and Geer (2000). Define the L_1-covering number, $N_1(\varepsilon, \mathcal{Q}, \mathcal{G})$, as the smallest L for which there exist functions $g_1 \ldots, g_L$ such that $\min_l E_{\mathcal{Q}}\|g - g_l\| \leq \varepsilon$ for all $g \in \mathcal{G}$ where $\mathcal{Q}$ is a probability measure and $\mathcal{G}$ is a class of functions in $\mathcal{L}^1(\mathcal{Q})$. Now, we show that $N_1(\varepsilon, \mathcal{P}_{\mathcal{N}}, \mathcal{G}) \lesssim \left(\frac{S\sigma_N^2}{\varepsilon}\right)^P$ where $\mathcal{P}_N$ is the empirical probability measure and $\mathcal{G}$ is the class of functions defined as $\mathcal{G} = \{g_{\boldsymbol{s}}(\boldsymbol{Y}) : \boldsymbol{s} \in [-S, S]^P\}$ where $g_{\boldsymbol{s}}(\boldsymbol{Y}) = \prod_{p=1}^{P} Y_p^{\alpha_p} \exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)$. Discretize $[-S, S]^P$ into $L = \left(\frac{4SP\sigma_N^2}{\varepsilon}\right)^P$ points, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L$, by cutting $[-S, S]$ in each dimension into equidistant segments of length $\frac{\varepsilon}{2P\sigma_N^2}$. For any $\boldsymbol{s} \in [-S, S]^P$ there exists an l such that

$$
\begin{aligned}
&E_N\left|g_{\boldsymbol{s}}(\boldsymbol{Y}) - g_{\boldsymbol{s}_l}(\boldsymbol{Y})\right| \\
&\quad = E_N\left|\prod_{p=1}^{P} Y_p^{\alpha_p} \exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right) - \prod_{p=1}^{P} Y_p^{\alpha_p} \exp\left(i\boldsymbol{Y}'\boldsymbol{s}_l\right)\right| \\
&\quad = E_N\left|\prod_{p=1}^{P} Y_p^{\alpha_p} \cos\left(\boldsymbol{Y}'\boldsymbol{s}\right) + i\prod_{p=1}^{P} Y_p^{\alpha_p} \sin\left(\boldsymbol{Y}'\boldsymbol{s}\right) - \prod_{p=1}^{P} Y_p^{\alpha_p} \cos\left(\boldsymbol{Y}'\boldsymbol{s}_l\right) - i\prod_{p=1}^{P} Y_p^{\alpha_p} \sin\left(\boldsymbol{Y}'\boldsymbol{s}_l\right)\right|
\end{aligned}
$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

$$\leq E_N\left|\prod_{p=1}^{P} Y_p^{\alpha_p}\cos\left(\boldsymbol{Y}'\boldsymbol{s}\right)-\prod_{p=1}^{P} Y_p^{\alpha_p}\cos\left(\boldsymbol{Y}'\boldsymbol{s}_l\right)\right|+E_N\left|i\prod_{p=1}^{P} Y_p^{\alpha_p}\sin\left(\boldsymbol{Y}'\boldsymbol{s}\right)-i\prod_{p=1}^{P} Y_p^{\alpha_p}\sin\left(\boldsymbol{Y}'\boldsymbol{s}_l\right)\right|$$
$$\leq 2P\cdot\max_l\left\{\left|\boldsymbol{s}-\boldsymbol{s}_l\right|\right\}\cdot\sigma_N^2\leq\varepsilon.$$

It follows that the L_1-covering number satisfies $N_1(\varepsilon,P_N,\mathcal{G})\lesssim\left(\frac{S\sigma_N^2}{\varepsilon}\right)^P$.

For N large enough and $8\sigma^2/\varepsilon^2\leq N$, B_2 is now bounded using a symmetrization argument, Hoeffding's inequality, and the L_1-covering number,

$$\Pr\left(\sup_{\boldsymbol{s}\in[-S,S]^P}\left|E_N\left[g_{\boldsymbol{s}}\right]-E\left[g_{\boldsymbol{s}}\right]\right|>\varepsilon\mid\sigma_N^2<\gamma\right)\leq 8N_1\left(\varepsilon/8,P_N,\mathcal{G}\right)\exp\left(-\frac{N\varepsilon^2}{128\sigma_N^2}\right)$$
$$\lesssim\left(\frac{S\gamma}{\varepsilon}\right)^P\exp\left(-\frac{N\varepsilon^2}{128\gamma}\right).$$

These bounds for B_1 and B_2 imply that

$$\Pr\left(\sup_{\boldsymbol{s}\in[-S,S]^P}\left|E_N\left[g_{\boldsymbol{s}}\right]-E\left[g_{\boldsymbol{s}}\right]\right|>\varepsilon\right)\leq B_1+B_2\lesssim\frac{e^{\sigma^2}}{e^{\gamma}}+\left(\frac{S\gamma}{\varepsilon}\right)^P\exp\left(-\frac{N\varepsilon^2}{128\gamma}\right).$$

The last step is to apply the Borel–Cantelli Lemma. Index ε, S and γ by N and let

$$\gamma_N=\alpha\ln N,\qquad 1<\alpha,$$
$$S_N=N^{\delta},\qquad 0<\delta,$$
$$\varepsilon_N=C\frac{\ln N}{N^{\frac{1}{2}}},$$

where $128\alpha(P(\delta+1/2)+1)<C^2$. Then for all N large enough,

$$\Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P}\left|E_N\left[g_{\boldsymbol{s}}\right]-E\left[g_{\boldsymbol{s}}\right]\right|>\varepsilon_N\right)\lesssim\frac{e^{\sigma^2}}{e^{\gamma_N}}+\left(\frac{S_N\gamma_N}{\varepsilon_N}\right)^P\exp\left(-\frac{N\varepsilon_N^2}{128\gamma_N}\right)$$
$$\lesssim N^{-\alpha}+N^{P(\delta+1/2)-\frac{C^2}{128\alpha}}$$
$$<2N^{-1}.$$

Hence,

$$\sum_{N=1}^{\infty}\Pr\left(\sup_{\boldsymbol{s}\in[-S_N,S_N]^P}\left|E_N\left[g_{\boldsymbol{s}}\right]-E\left[g_{\boldsymbol{s}}\right]\right|>\varepsilon_N\right)\lesssim\sum_{N=1}^{\infty}\left(N^{-\alpha}+N^{P(\delta+1/2)-\frac{C^2}{128\alpha}}\right)<\infty.$$

The Borel–Cantelli lemma then implies that for all N large enough $\sup_{\boldsymbol{s}\in[-S_N,S_N]^P}\left|E_N\left[g_{\boldsymbol{s}}\right]-E\left[g_{\boldsymbol{s}}\right]\right|\leq\varepsilon_N$ a.s.

A.5. Proof of Theorem 4.1

Let ϕ and $\widehat{\phi}$ be based on (4) from Theorem 2.1. Denote $\phi_{\boldsymbol{Y}}(\boldsymbol{s})=iE\left[\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]$, $\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{s})=iE\left[Y_p\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]$, $\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{s})=iE_N\left[\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]$, and $\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{s})=iE_N\left[Y_p\exp\left(i\boldsymbol{Y}'\boldsymbol{s}\right)\right]$. For N large enough,

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

$$
\begin{aligned}
&\left|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})\right| \\
&\quad=\left|\sum_{k=1}^{K}\sum_{p=1}^{P_1} v_k H_{A_1 kp}\right|\cdot\left|\exp\left(\int_0^1 \frac{\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u\right)-\exp\left(\int_0^1 \frac{\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u\right)\right| \\
&\quad\lesssim S_N\cdot\left|\exp\left(\int_0^1 \frac{\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u\right)-\exp\left(\int_0^1 \frac{\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u\right)\right| \\
&\quad\lesssim S_N\cdot\left|\int_0^1 \frac{\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u-\int_0^1 \frac{\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\mathrm{d}u\right| \\
&\quad\lesssim S_N\cdot\int_0^1\left(\left|\frac{\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})-\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}{\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})}\right|\right. \\
&\qquad\qquad\left.+\left|\frac{\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})(\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})-\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v}))}{\left(\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})\right)^2}\right|\right)\mathrm{d}u \\
&\quad\le S_N\cdot\sup_{\boldsymbol{v}\in[-S_N,S_N]^K}\left|\widehat{\phi}_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})-\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})\right|\int_1^0\left|(\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v}))^{-1}\right|\mathrm{d}u \\
&\qquad+S_N\cdot\sup_{\boldsymbol{v}\in[-S_N,S_N]^K}\left|\widehat{\phi}_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})-\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})\right|\int_0^1\frac{|\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})|}{\left(\phi_{\boldsymbol{Y}}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})\right)^2}\mathrm{d}u \\
&\quad\lesssim\varepsilon_N S_N\int_0^1\left(\Upsilon(uS_N\|H_{A_2}\|)\right)^{-2}\mathrm{d}u,
\end{aligned}
$$

where the second $\lesssim$ follows by the Taylor expansion $e^x=e^{x_0}+O(|x-x_0|)$, the third $\lesssim$ follows by the Taylor expansion $\frac{x}{y}=\frac{x_0}{y_0}+O\left(|\frac{1}{y_0}(x-x_0)|+|\frac{x_0}{y_0^2}(y-y_0)|\right)$ and the last $\lesssim$ follows by Lemma 4.1, Assumption 4.1(iv) and $\left|\phi_{\boldsymbol{Y}}^{(p)}(\boldsymbol{0},uH'_{A_2}\boldsymbol{v})\right|=\left|iE\left[Y_p e^{iu\boldsymbol{Y}_2'H'_{A_2}\boldsymbol{v}}\right]\right|\le\max_{p\in\{1,\dots,P\}}E[|Y_p|]<\infty$.

Similarly, basing ϕ and $\widehat{\phi}$ on (10) from Theorem 2.2,

$$
\left|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})\right|\lesssim\varepsilon_N S_N^2\int_0^1\left(\Upsilon(uS_N\|H_A\|)\right)^{-3}\mathrm{d}u.
$$

By Assumption 4.1(iii), there exists m such that $\phi_{\boldsymbol{\kappa}}(\boldsymbol{v})=\prod_{k=1}^K(1+m(v_k)v_k^q)$ for all $\boldsymbol{v}\in[-1,1]^K$ and $\phi_{\boldsymbol{\kappa}}(\boldsymbol{v})=0$ otherwise and $m(v_k)$ is continuous for all $v_k\in[-1,1]$. Let $h_N=\frac{1}{S_N}$. For all $\boldsymbol{u}$ in the support of $f_{\boldsymbol{F}}$ and N large enough,

$$
\begin{aligned}
&\left|\widehat{f}_{\boldsymbol{F}}(\boldsymbol{u})-f_{\boldsymbol{F}}(\boldsymbol{u})\right| \\
&\quad=\left|\frac{1}{(2\pi)^K}\int e^{-i\boldsymbol{v}'\boldsymbol{u}}\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)\,\mathrm{d}\boldsymbol{v}-\frac{1}{(2\pi)^K}\int e^{-i\boldsymbol{v}'\boldsymbol{u}}\phi_{\boldsymbol{F}}(\boldsymbol{v})\mathrm{d}\boldsymbol{v}\right| \\
&\quad=\left|\frac{1}{(2\pi)^K}\int e^{-i\boldsymbol{v}'\boldsymbol{u}}\left(\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)-\phi_{\boldsymbol{F}}(\boldsymbol{v})\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)+\phi_{\boldsymbol{F}}(\boldsymbol{v})\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)-\phi_{\boldsymbol{F}}(\boldsymbol{v})\right)\mathrm{d}\boldsymbol{v}\right| \\
&\quad=\left|\frac{1}{(2\pi)^K}\int e^{-i\boldsymbol{v}'\boldsymbol{u}}\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)\left(\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})\right)\mathrm{d}\boldsymbol{v}+\frac{1}{(2\pi)^K}\int e^{-i\boldsymbol{v}'\boldsymbol{u}}\phi_{\boldsymbol{F}}(\boldsymbol{v})\left(\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)-1\right)\mathrm{d}\boldsymbol{v}\right|
\end{aligned}
$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

$$\leq \frac{1}{(2\pi)^K}\int |\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)|\,|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})| + \frac{1}{(2\pi)^K}\int |\phi_{\boldsymbol{F}}(\boldsymbol{v})|\,|\phi_{\boldsymbol{\kappa}}(\boldsymbol{v}h_N)-1|\,\mathrm{d}\boldsymbol{v}$$

$$\leq \int\limits_{\boldsymbol{v}\in[-S_N,S_N]^K}\cdots\int |\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})|\,\mathrm{d}\boldsymbol{v} + \int\limits_{\boldsymbol{v}\in[-S_N,S_N]^K}\cdots\int |\phi_{\boldsymbol{F}}(\boldsymbol{v})|\left|\prod_{k=1}^{K}(1+m(v_kh_N)(v_kh_N))^q)-1\right|\mathrm{d}\boldsymbol{v}$$
$$+ \int\limits_{\boldsymbol{v}\notin[-S_N,S_N]^K}\cdots\int |\phi_{\boldsymbol{F}}(\boldsymbol{v})|\,\mathrm{d}\boldsymbol{v}$$

$$\lesssim S_N^K \sup|\widehat{\phi}_{\boldsymbol{F}}(\boldsymbol{v})-\phi_{\boldsymbol{F}}(\boldsymbol{v})| + h_N^{qK}\int\limits_{\boldsymbol{v}\in[-S_N,S_N]^K}\cdots\int |\phi_{\boldsymbol{F}}(\boldsymbol{v})|\prod_{k=1}^{K}|v_k|^q\mathrm{d}\boldsymbol{v} + \int\limits_{\boldsymbol{v}\notin[-S_N,S_N]^K}\cdots\int |\phi_{\boldsymbol{F}}(\boldsymbol{v})|\mathrm{d}\boldsymbol{v},$$

where the $\lesssim$ follows because m is continuous and so bounded on $[-1,1]$.

A.6. Derivations for φ_F in the Measurement Error Model With Repeated Measurements from Section 5.1

The log characteristic function of (25) and (26) is,

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{s}) = \varphi_F(s_1+s_2)+\varphi_{\boldsymbol{U}}(\boldsymbol{s}). \tag{A.8}$$

The first-order partial derivatives are,

$$\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1} = \varphi_F'(s_1+s_2)+\frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_1}, \qquad \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2} = \varphi_F'(s_1+s_2)+\frac{\partial\varphi_{\boldsymbol{U}}(\boldsymbol{s})}{\partial s_2}.$$

Hence,

$$\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\bigg|_{\boldsymbol{s}=(u,0)'} = \varphi_F'(u)+\frac{iE[U_2\exp(iuU_1)]}{E[\exp(iuU_1)]} = \varphi_F'(u),$$

$$\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\bigg|_{\boldsymbol{s}=(0,u)'} - \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1}\bigg|_{\boldsymbol{s}=(0,u)'} = \frac{iE[U_2\exp(iuU_2)]}{E[\exp(iuU_2)]} - \frac{iE[U_1\exp(iuU_2)]}{E[\exp(iuU_2)]} = \varphi_{U_2}'(u).$$

Integrating out, we get FDE1 and FDE2 are, respectively, the sample analogs of

$$\varphi_F(v) = \int_0^v \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\bigg|_{\boldsymbol{s}=(u,0)'}du = \int_0^v \frac{iE[Y_2\exp(iuY_1)]}{E[\exp(iuY_1)]}\mathrm{d}u,$$

$$\varphi_F(v) = \varphi_{Y_2}(v)-\varphi_{U_2}(v), \text{ where}$$

$$\varphi_{U_2}(v) = \int_0^v\left(\frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_2}\bigg|_{\boldsymbol{s}=(0,u)'} - \frac{\partial\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1}\bigg|_{\boldsymbol{s}=(0,u)'}\right)du = \int_0^v \frac{iE[(Y_2-Y_1)\exp(iuY_2)]}{E[\exp(iuY_2)]}\mathrm{d}u.$$

Assuming U_1 and U_2 are independent, the second-order partial derivative of (A.8) with respect to s_1, s_2 is

$$\frac{\partial^2\varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1\partial s_2} = \varphi_F''(s_1+s_2).$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press

Integrating out, we get SDE is the sample analog of

$$\varphi_F(v) = iE[F] + \int_0^v \int_0^{u_2} \left. \frac{\partial^2 \varphi_{\boldsymbol{Y}}(\boldsymbol{s})}{\partial s_1 \partial s_2} \right|_{\boldsymbol{s}=(u_1/2, u_1/2)'} du_1 du_2$$

$$= iE[F] + \int_0^v \int_0^{u_2} \left(\frac{E[Y_1 e^{iu_1(Y_1+Y_2)/2}] E[Y_2 e^{iu_1(Y_1+Y_2)/2}]}{(E[e^{iu_1(Y_1+Y_2)/2}])^2} \right.$$

$$\left. - \frac{iE[Y_1 Y_2 e^{iu_1(Y_1+Y_2)/2}]}{E[e^{iu_1(Y_1+Y_2)/2}]} \right) \mathrm{d}u_1 \mathrm{d}u_2.$$

https://doi.org/10.1017/S026646661700007X Published online by Cambridge University Press