Informational Content of Factor Structures in Simultaneous Binary Response Models*

Shakeeb Khan — Boston College

Arnaud Maurel — Duke University, NBER and IZA

Yichong Zhang — Singapore Management University

March 2022

Abstract

We study the informational content of factor structures in discrete triangular systems. Factor structures have been employed in a variety of settings in cross sectional and panel data models, and in this paper we formally quantify their identifying power in a bivariate system often employed in the treatment effects literature. Our main findings are that imposing a factor structure yields point identification of parameters of interest, such as the coefficient associated with the endogenous regressor in the outcome equation, under weaker assumptions than usually required in these models. In particular, we show that a "non-standard" exclusion restriction that requires an explanatory variable in the outcome equation to be excluded from the treatment equation is no longer necessary for identification, even in cases where all of the regressors from the outcome equation are discrete. We also establish identification of the coefficient of the endogenous regressor in models with more general factor structures, in situations where one has access to at least two continuous measurements of the common factor.

Keywords: Factor Structures, Discrete Choice, Causal Effects.

JEL: C14, C31, C35.

*First version: October 2019. We thank the Co-Editor, Simon Lee, an anonymous referee, Arthur Lewbel, Serena Ng, and seminar participants at Arizona State University, Emory, Michigan State, Shanghai University of Finance and Economics, University of Arizona, as well as conference participants at the 2015 SEA meetings for helpful comments. We also thank Zhangchi Ma and Qingsong Yao for excellent research assistance. Zhang acknowledges the financial support from Singapore Ministry of Education Tier 2 grant under grant MOE2018-T2-2-169 and the Lee Kong Chian fellowship.

1 Introduction

Factor models see widespread and increasing use in various areas of econometrics. This type of structure has been employed in a variety of settings in cross sectional, panel and time series models, and have proven to be a flexible way to model the behavior of and relationship between unobserved components of econometric models. The basic idea behind factor models is to assume that the dependence across the unobservables is generated by a low-dimensional set of mutually independent random factors. The applied and theoretical research employing factor structures in econometrics is extensive. In particular, these models are often used in the treatment effect literature as a way to identify the joint distribution of potential outcomes from the marginal distributions, and then recover the distribution of treatment effects from this joint distribution.[1] Factor models have been used in a number of different contexts in applied microeconomics. These include, among others, earnings dynamics (Abowd and Card, 1989; Bonhomme and Robin, 2010), estimation of returns to schooling and work experiences (Ashworth, Hotz, Maurel, and Ransom, 2021), as well as cognitive and non-cognitive skill production technology (Cunha, Heckman, and Schennach, 2010). Heckman and Vytlacil (2007a,b) provide various additional references. All of these papers, with the notable exception of Cunha et al. (2010), rely on linear factor models where the unobservables are assumed to be written as the sum of a linear combination of mutually independent factors and an idiosyncratic shock.

In this paper we bring together the literature on factor models with the literature on the identification and estimation of binary response models (Klein and Spady (1993); Lewbel (2000); Park and Phillips (2000); Blundell and Powell (2004)), in particular triangular binary choice models (Chesher (2005); Vytlacil and Yildiz (2007); Shaikh and Vytlacil (2011); Han and Vytlacil (2017)), by exploring the *informational content* of factor structures in this class of models.[2] Focusing on this class can be well motivated from both an empirical and theoretical perspective. From the former, many treatment effect models fit into this framework as treatment is typically a binary and endogenous variable in the system, whose effect on outcomes is often a parameter the econometrician wishes to conduct inference on. From a theoretical perspective, inference on this type of system can be complicated, if not impossible without strong parametric assumptions, which may not be reflected in the observed data. Imposing no restriction on the structure of endogeneity often fails to achieve identification of parameter, or at best only do so in sparse regions of the data, thus making inference impractical in practice. In this context, modeling the endogeneity between the selection and the outcome by a factor structure may be a useful "in-between" setting, which, at the very least, can be used to gauge the sensitivity of the parametric approach to their stringent assumptions.

[1] See also Abbring and Heckman (2007) for an extensive discussion of factor structures and prior studies using these models in the context of treatment effect estimation.

[2] See also recent work by Lewbel, Schennach, and Zhang (2020), who study the identification of a triangular linear model assuming that the disturbances are related through a factor model.

We start our analysis by imposing a particular factor structure to the two unobservables in our system of binary equations described in further detail in the next section, and explore the informational content of this assumption. We assume that the unobservables from the treatment equation (V) and the outcome equation (U) are related through the following factor model:

$$U = \gamma_0 V + \Pi \tag{1.1}$$

where Π is an unobserved random variable assumed to be distributed independently of V and γ_0 is a scalar parameter. This structure generalizes the canonical case where the unobservables (U, V) are jointly normally distributed, for which this relationship always holds. Our main finding is that there is indeed informational content of factor structures in the sense that, in contrast to prior literature - notably Vytlacil and Yildiz (2007) - one no longer requires an additional "non-standard" exclusion restriction, nor the strong support conditions on the covariates entering the outcome equation that are generally needed for identification in these models. Our identification results are constructive and translate directly into a rank based estimator of the coefficient associated with the binary endogenous variable, which we provide and study in a supplement to this paper.

While an appealing feature of the structure considered in Equation (1.1) is that it is a natural extension of the bivariate Probit specification that has often been considered in the literature, this model does impose significant restrictions on the nature of the dependence between the unobservables U and V. In the paper we extend this baseline specification by considering a linear factor structure of the form:

$$U = \gamma_0 W + \eta_1 \tag{1.2}$$
$$V = W + \eta_2 \tag{1.3}$$

where (W, η_1, η_2) are mutually independent unobserved random variables. We study the informational content of this extended factor structure in the context of triangular binary choice models and establish identification, assuming access to at least two continuous noisy measurements of the unobserved factor W. This setup has been used in a number of applications, in particular in labor economics. In these applications, the unobserved factor is typically interpreted as latent individual ability, about which several continuous noisy measurements are available from the data. This is the case of, for instance, Carneiro, Hansen, and Heckman (2003), Cunha et al. (2010), Heckman, Humphries, and Veramendi (2018) and Ashworth et al. (2021), who use components of the Armed Services Vocational Aptitude Battery test as measurements of cognitive ability.

The rest of the paper is organized as follows. In Section 2 we formally describe the triangular system with our factor structure, and discuss our main identification results for the parameters of interest in this model. Section 3 explores identification in more general factor structure models which involve multiple idiosyncratic errors, in a context where one has access to two continuous noisy measurements of the common unobserved factor. Section 4 concludes. We prove Theorems 2.1 and 3.1 in Sections A and B, respectively. In Section C, we establish the sharp identified set of α_0 when the support condition for point-identification is violated in the one-factor model and the necessary and sufficient condition for point-identification in the two-factor model with two continuous measurements of the common factor. The corresponding results are proved in Sections D and E. The Supplementary Material studies the asymptotic properties of a rank-based estimator for α_0 and explores its finite sample properties through some Monte Carlo simulation exercises.

Notation: throughout the paper we write $\mathbf{1}\{A\}$ to denote the usual indicator function that takes value 1 if event A happens, and 0 otherwise. We also denote by $d(U)$ and $d(U|V)$ the lengths of the support of random variable U, and the conditional support of U given V, respectively.

2 Triangular Binary Model with Factor Structure

2.1 Set-up and Main Identification Result

In this section we consider the identification of the following triangular binary model:

$$Y_1 = \mathbf{1}\{Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 Y_2 - U > 0\} \quad (2.4)$$

$$Y_2 = \mathbf{1}\{Z'\delta_0 - V > 0\} \quad (2.5)$$

where $Z \equiv (Z_1, Z_2)$ and (U, V) is a pair of random shocks. Z_2 and Z_3 provide the exclusion restrictions in the model, and the distribution of (Z_2, Z_3) is required to be nondegenerate conditional on $Z_1'\lambda_0 + Z_3'\beta_0$. We further assume that the error terms U and V are jointly independent of (Z_1, Z_2, Z_3). The endogeneity of Y_2 in (2.4) arises when U and V are not independent.

The above model, or minor variations of it, have often been considered in the recent literature. See for example, Vytlacil and Yildiz (2007), Abrevaya, Hausman, and Khan (2010), Klein, Shan, and Vella (2015), Vuong and Xu (2017), Khan and Nekipelov (2018) and references therein. A key parameter of interest[3] in our paper as is in much of the literature is α_0. In this paper we provide conditions under which the parameters of interest are point-identified. As such, our analysis complements alternative partial-identification approaches that have been proposed in the context of triangular binary models. See, in particular, Chiburis (2010), Shaikh and Vytlacil (2011), and Mourifié (2015).[4] As discussed in the aforementioned papers, the parameter α_0 is difficult, if

[3]As is always the case in models with binary outcomes, both the interpretation and the usefulness of regression coefficients warrant explanation. In the model considered here the coefficient on the treatment variable and the coefficients on exogenous variables in the binary outcome equation enable us to construct "equivalence classes" to answer important policy questions. For example consider the case where the dummy endogenous variable is job training, the exogenous regressor is years experience and the outcome variable is employment status. Knowing all coefficients would be informative on how many additional years of experience would be needed to compensate for a lack of training so the probability of being employed stays the same.

[4]In Section C in the supplement, we establish the sharp identified set of α_0 when the support condition for

not impossible to point identify and estimate without imposing parametric restrictions on the unobserved variables in the model, (U, V).

The difficulty of identifying α_0 in semi-parametric "distribution-free" models, and the sensitivity of its identification to misspecification in parametric models is what motivates the factor structure we add in this paper to the above model. Specifically, to allow for endogeneity in the form of possible non-zero correlation between U and V, we augment the model with the following equation:

$$U = \gamma_0 V + \Pi \tag{2.6}$$

where Π is an unobserved random variable, assumed to be distributed independently of (V, Z_1, Z_2, Z_3), and γ_0 is an additional unknown scalar parameter. Importantly, this type of factor structure always holds when the residuals of both equations are jointly normally distributed. Furthermore, this specification corresponds to the type of structure used in Independent Component Analysis (ICA), where V and Π are two mutually independent factors. This method has found many applications in various fields, including signal processing and image extraction; applications in economics include e.g., Hyvärinen and Oja (2000), Moneta, Hoyer, and Coad (2013) and Gourieroux, Monfort, and Renne (2017). While, in contrast to the ICA literature, the factors and the factor loadings are not the main objects of interest in our analysis, this dimension-reducing structure plays a key role in our identification results.

Our aim is to first explore identification of the parameters $(\alpha_0, \delta_0, \gamma_0, \beta_0, \lambda_0)$ under standard nonparametric regularity conditions on (V, Π). Note that the parameter δ_0 in the selection equation can be identified up to scale in various ways. See, for example, Klein and Spady (1993) and Han (1987), among others. We then impose the usual condition that one of δ_0's coordinates is equal to one to fix the scale. For simplicity, for the rest of the paper, we denote $X \equiv Z'\delta_0$ and assume X is observed. We further define $X_1 \equiv Z_1'\lambda_0 + Z_3'\beta_0$. However, we cannot identify λ_0 and β_0 beforehand. We propose instead to identify them along with α_0.

Our main identification result is based on the Assumptions **A1**-**A4** we state below:

A1 The first coefficient of λ_0 is normalized to one so that $\lambda_0 = (1, \lambda_{0,-1}^T)^T$. The parameter $\theta_0 \equiv (\alpha_0, \gamma_0, \lambda_{0,-1}, \beta_0)$ is an element of a compact subset of $\Re^{d_1+d_3+1}$, where d_1 and d_3 are the dimensions of Z_1 and Z_3, respectively.

A2 The vector of unobserved variables, (U, V, Π) is continuously distributed with support on a subset of $\Re^3$ and independently distributed of the vector (Z_1, Z_2, Z_3). Furthermore, we assume that the unobserved random variables Π, V are distributed independently of each other.

point-identification is violated. This result highlights that, except for the fact that the sign of α_0 is identified, we generally cannot say much about the value of $|\alpha_0|$. Related work by Shaikh and Vytlacil (2011) also provides partial identification results for a triangular binary model. That the bounds for α_0 are generally tighter in their analysis reflects the identifying power of the additional support restrictions that they impose.

A3 X is continuously distributed with absolute continuous density w.r.t. Lebesgue measure. Its density is bounded and bounded away from zero on any compact subset of its support.

A4 Let $Z_{1,-1}$ be all the coordinates of Z_1 except the first one, and $d = d_1 + d_3 + 1$. There exist $2d$ vectors $\{z_1^{(l)}, z_3^{(l)}, x^{(l)}\}_{l=1}^d$ and $\{\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{x}^{(l)}\}_{l=1}^d$ in the joint support of (Z_1, Z_3, X) such that

$$\alpha_0 + (z_{1,-1}^{(l)} - \tilde{z}_{1,-1}^{(l)})'\lambda_{0,-1} + (z_3^{(l)} - \tilde{z}_3^{(l)})'\beta_0 - \gamma_0(x^{(l)} - \tilde{x}^{(l)}) = \tilde{z}_{1,1}^{(l)} - z_{1,1}^{(l)},\ l = 1, \cdots, d$$

and $\text{rank}(\mathcal{M}) = d$, where

$$\mathcal{M} = \begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ x^{(1)} - \tilde{x}^{(1)} & \cdots & x^{(d)} - \tilde{x}^{(d)} \end{pmatrix}.$$

Before turning to our main identification result, a couple of remarks are in order.

Remark 2.1. *The first part of Assumption A1 is a standard scale normalization. Assumption A2 is also standard in this literature. The assumption that the instruments are independent of the unobservables can also be found in, among others, Abrevaya et al. (2010), Vytlacil and Yildiz (2007), Klein et al. (2015), and Khan and Nekipelov (2018). The assumption of independence between* Π *and* V *is also made in Bai and Ng (2002) and Carneiro et al. (2003).*

Remark 2.2. *Assumptions A3 and A4 impose some restrictions on the distributions of the covariates entering the selection and outcome equations, respectively. Specifically, Assumption A3 requires one component of the covariates* Z *entering the selection equation to be continuously distributed, which is often required in models with discrete outcomes. In contrast, Assumption A4 only requires some variation of* (Z_1, Z_3)*. In particular, the distribution of* (Z_1, Z_3) *cannot be degenerate but is allowed to be discrete. This assumption can be interpreted as a full rank condition, which ensures that the system of linear equations that delivers point identification has a unique solution.*

We now turn to our main identification result, Theorem 2.1, which concludes that under our stated conditions and our factor structure we can attain point identification of the vector of parameters θ_0.

Theorem 2.1. *Under Assumptions A1-A4,* θ_0 *is point identified.*

An important takeaway from this result, which we discuss further in Subsection 2.2 below, is that imposing the factor structure (2.6) yields point-identification under weaker support conditions when compared to the existing literature, and does not require a second exclusion restriction either. In particular, our model delivers point-identification of the parameters of interest even in situations where all of the regressors from the outcome equation are discrete. This indicates that, from the selection equation combined with the factor structure that we impose here, we can overturn the non-identification result of Bierens and Hartog (1988) which would apply to the outcome equation alone.

The proof of Theorem 2.1, which is reported in Section A in the Supplementary Appendix, relies on the fact that, for two observations (Z_1, Z_3, X) and $(\tilde{Z}_1, \tilde{Z}_3, \tilde{X})$,

$$\begin{aligned} & \partial_x P^{11}(Z_1, Z_3, X)/f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X})/f_V(\tilde{X}) = 0 \\ \iff & \alpha_0 + (Z_1 - \tilde{Z}_1)'\lambda_0 + (Z_3 - \tilde{Z}_3)'\beta_0 - \gamma_0(X - \tilde{X}) = 0, \end{aligned} \tag{2.7}$$

where $f_V(\cdot)$ is the pdf. of V, which is identified over the support of X, and $P^{ij}(z_1, z_3, x) \equiv Prob(Y_1 = i, Y_2 = j | Z_1 = z_1, Z_3 = z_3, X = x)$ ($\partial_x P^{ij}(z_1, z_3, x)$) denote the choice probability (partial derivative of the ij-choice probability with respect to the third argument), which are both identified from the data.

Remark 1. *This identification result can be extended to the case of a separable nonparametric factor model. Namely, consider the following relationship between unobserved components:*

$$U = g_0(V) + \tilde{\Pi} \tag{2.8}$$

where $\tilde{\Pi}$ is an unobserved random variable assumed to be distributed independently of V and all instruments. $g_0(\cdot)$ is an unknown function assumed to satisfy standard smoothness conditions. The parameter of interest is $(\alpha_0, \lambda_0, \beta_0)$, but now the unknown nuisance parameter in the factor equation is infinite dimensional. By replacing $\gamma_0 X$ by $g_0(X)$ in (2.7), we have

$$\begin{aligned} & \partial_x P^{11}(Z_1, Z_3, X)/f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X})/f_V(\tilde{X}) = 0 \\ \iff & \alpha_0 + (Z_1 - \tilde{Z}_1)'\lambda_0 + (Z_3 - \tilde{Z}_3)'\beta_0 - (g_0(X) - g_0(\tilde{X})) = 0. \end{aligned} \tag{2.9}$$

One can then establish identification after modifying the rank condition A4 by replacing $\gamma_0(x^{(l)} - \tilde{x}^{(l)})$ by $g_0(x^{(l)}) - g_0(\tilde{x}^{(l)})$.

Remark 2. *We assume rank invariance in (2.6). It is possible to relax such condition to rank similarity.*[5] *Specifically, we can consider the following model:*

$$\begin{aligned} Y_1 &= \mathbf{1}\{Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 Y_2 - U(Y_2) > 0\} \\ Y_2 &= \mathbf{1}\{Z'\delta_0 - V > 0\}, \end{aligned}$$

where

$$U(y_2) = \gamma_0 V + \Pi(y_2), \text{ for } y_2 = 0, 1.$$

[5] We thank the referee for pointing this out.

We further assume $(V, \Pi(1), \Pi(0))$ is continuously distributed with support on a subset of $\Re^3$ and independently distributed of the vector Z_1, Z_2, Z_3, V and $(\Pi(1), \Pi(0))$ are independent, $P(\Pi(1) \leq \pi) = P(\Pi(0) \leq \pi)$ for $\pi \in \Re$, and Assumptions **A1**, **A3**, *and* **A4** *hold. Then, we can identify θ_0 by a similar argument as the proof of Theorem 2.1.*

2.2 Connection with Prior Literature

We now discuss in detail how our setup and main identification result relates to the existing literature.

In a related work, Han and Vytlacil (2017) consider the identification of a generalized bivariate Probit model.[6] Our linear factor structure and the one-parameter copula model considered in Han and Vytlacil (2017) are not nested by each other. First, note that based on the factor structure, we can recover F_Π, the distribution of Π, as a function of (F_U, F_V, γ_0) by deconvolution. We can then write the copula of (U, V) as

$$F_{U,V}(F_U^{-1}(u), F_V^{-1}(v)) = \int_{-\infty}^{F_V^{-1}(v)} F_\Pi(F_U^{-1}(u) - \gamma_0 w; F_U, F_V, \gamma_0) f_V(w) dw = C(u, v; F_U, F_V, \gamma_0).$$

The copula depends not only on γ_0 but also on two infinite dimensional parameters (F_U, F_V). Thus, unlike Han and Vytlacil (2017), our factor structure cannot be characterized by a one-parameter copula. In addition, in order to achieve identification, Han and Vytlacil (2017) first nonparametrically identify the two marginals by assuming the existence of a full support regressor that is common to both equations.[7] In contrast, our approach does not rely on the existence of such a regressor. Under the factor structure assumed in our analysis, we bypass the nonparametric identification of the marginals as a whole and directly consider the identification of the structural parameters. It follows that our model cannot be nested by the one-parameter copula model considered by Han and Vytlacil (2017). On the other hand, there exist one-parameter copula models that cannot be decomposed into linear factor structures.[8] This implies that our model does not nest Han and Vytlacil (2017) either.

Our analysis also relates to Vytlacil and Yildiz (2007) and Vuong and Xu (2017), who consider the identification of α_0 in a triangular binary model. Our identification result, however, differs from theirs in important ways. Namely, denote $X = Z'\delta_0 = Z_1'\delta_{1,0} + Z_2'\delta_{2,0}$. Then, Assumption A4

[6] See also recent work by Han and Lee (2019) who study semiparametric estimation and inference in the framework considered by Han and Vytlacil (2017).

[7] Han and Vytlacil (2017) establish their identification of the coefficient on the endogeneous regressor (Theorems 4.2 and 5.1) under the assumption that the marginal distributions F_ε and F_ν are known. Then, they verify this condition by showing the identification of these two marginal distributions using large support common regressors.

[8] For instance, suppose that (U, V) has a Gaussian copula with correlation ρ, and that the marginal distributions of U and V are uniform $[0, 1]$. It then follows that, denoting by $\Phi(.)$ the standard normal cdf., $\left(\Phi^{-1}(U), \Phi^{-1}(V)\right)$ is bivariate normal with correlation ρ, which in turn yields the following non-linear relationship between U and V: $U = \Phi\left(\rho\Phi^{-1}(V) + W\right)$, where W is normally distributed and independent from V.

implies that we can find a pair of observations (z_1, z_2, z_3) and $(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ such that

$$z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \gamma_0(z_1'\delta_{1,0} + z_2'\delta_{2,0}) = \tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0 - \gamma_0(\tilde{z}_1'\delta_{1,0} + \tilde{z}_2'\delta_{2,0}). \quad (2.10)$$

In contrast, using our notation, Vytlacil and Yildiz (2007) require that one can find a pair of observations (z_1, z_2, z_3) and $(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ such that $z'\delta_0 = \tilde{z}'\delta_0$ and

$$z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 = \tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0. \quad (2.11)$$

Vuong and Xu (2017) do not assume the existence of Z_3. In our binary outcome setup, the functions $h(0, x, \tau)$ and $h(1, x, \tau)$ defined in Vuong and Xu (2017) are equal to $1\{x + F_{-U}^{-1}(\tau) \geq 0\}$ and $1\{x + \alpha + F_{-U}^{-1}(\tau) \geq 0\}$, respectively, where $x = z_1'\lambda_0$ and F_{-U} is the CDF of $-U$. Then, Vuong and Xu (2017, Assumption C'(ii)) requires that we can find z_1 and $\tilde{z}_1$ in the support of Z_1 so that for any τ_1, τ_2, if $1\{\tilde{z}_1'\lambda_0 + F_{-U}^{-1}(\tau_1) \geq 0\} = 1\{\tilde{z}_1\lambda_0 + F_{-U}^{-1}(\tau_2) \geq 0\}$, then $1\{z_1'\lambda_0 + \alpha_0 + F_{-U}^{-1}(\tau_1) \geq 0\} = 1\{z_1\lambda_0 + \alpha_0 + F_{-U}^{-1}(\tau_2) \geq 0\}$. Provided that the support of U nests the supports of $Z_1'\lambda_0$ and $Z_1'\lambda_0 + \alpha_0$, Vuong and Xu (2017, Assumption C'(ii)) is then equivalent to:[9]

$$z_1'\lambda_0 + \alpha_0 = \tilde{z}_1'\lambda_0. \quad (2.12)$$

Several remarks are in order. First, note that sufficient support conditions for the restrictions (2.10)–(2.12) are $d(Z_1'\lambda_0 + Z_3'\beta_0 - Z'\delta_0\gamma_0) \geq |\alpha_0|$, $d(Z_1'\lambda_0 + Z_3'\beta_0|Z'\delta_0) \geq |\alpha_0|$, and $d(Z_1'\lambda_0|Z'\delta_0) \geq |\alpha_0|$ with a positive probability, respectively, where $d(\cdot)$ denotes the "length" of its argument. These three support conditions are such that

$$d(Z_1'\lambda_0 + Z_3'\beta_0 - Z'\delta_0\gamma_0) \geq d(Z_1'\lambda_0 + Z_3'\beta_0|Z'\delta_0) \geq d(Z_1'\lambda_0|Z'\delta_0),$$

where the first and second inequalities are strict if Z_2 and Z_3 have at least one continuous component, respectively. Importantly, we show in Section C of the Supplement that for a version of the triangular binary model with univariate Z_2 and Z_3 and no common regressor Z_1, the support condition $d(Z_1'\lambda_0 + Z_3'\beta_0|Z'\delta_0) \geq |\alpha_0|$ is actually also necessary to the identification of the model without factor structure. This implies that by imposing our factor structure, one can identify values of α_0 in a region that cannot be identified in the model considered by Vytlacil and Yildiz (2007). Such region is characterized in Section C of the Supplement.

Second, it directly follows from these support conditions that, in the presence of a factor model and in contrast to both Vytlacil and Yildiz (2007) and Vuong and Xu (2017), variation in Z_2 helps in the identification of α_0. In that sense, the factor model allows to restore the intuition from standard IV approaches in linear models that variation in the instrument Z_2 is critical to the

[9] To see this, note that if, say, $z_1'\lambda_0 + \alpha_0 > \tilde{z}_1'\lambda_0$, then we can find τ_1, τ_2 such that $-z_1'\lambda_0 - \alpha_0 \leq F_{-U}^{-1}(\tau_1) < -\tilde{z}_1'\lambda_0$ and $F_{-U}^{-1}(\tau_2) < -z_1'\lambda - \alpha_0 < -\tilde{z}_1'\lambda_0$. This violates the above requirement, and thus, shows that Vuong and Xu (2017, Assumption C'(ii)) implies (2.12). On the other hand, if $z_1'\lambda_0 + \alpha_0 = \tilde{z}_1'\lambda_0$, then Vuong and Xu (2017, Assumption C'(ii)) holds trivially.

identification of the parameters of the outcome equation. Related to this, the support of Z_2 plays an important role in our identification analysis. In particular, if Z_2 is discrete, our identification strategy requires sufficient variation in the variables in the outcome equation, namely Z_1 and Z_3. In this case, our support requirement is equivalent to that assumed by Vytlacil and Yildiz (2007).

Third, another important aspect of Assumption A4 is that it does not impose any constraint on the variables from the outcome equation. Specifically, consider a case where the outcome equation does not contain a variable that is excluded from the selection equation (i.e., $\beta_0 = 0$), the regressor that is common to both equations, Z_1, is scalar and binary, and where $\lambda_0 = 1$. In this case, one can show that the identifying support conditions associated with Vytlacil and Yildiz (2007) (2.11) and Vuong and Xu (2017) (2.12) generally fail to hold, except for a finite set of values $\alpha_0 \in \{-1, 0, 1\}$. In contrast, our support restriction (2.10) holds under more general conditions: without any restriction on α_0 if one element of Z_2 is continuous with large support, and on a continuum of possible values for α_0 if one element of Z_2 is continuous with bounded support. In that sense, the factor structure replaces the need for a continuous component in (Z_1, Z_3) in the outcome equation.

Finally, at a high level, our identification strategy shares similarities with the Local Instrumental Variable (LIV) approach that has been proposed by Heckman and Vytlacil (2005) and further discussed by Carneiro and Lee (2009). In particular, our identifying restriction (2.7) can be alternatively derived from a local IV strategy applied to a potential outcomes model characterized by $Y_1(y_2) = \mathbf{1}\{Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 y_2 - U > 0\}$, with treatment given by $Y_2 = \mathbf{1}\{Z'\delta_0 - V > 0\}$. In contrast to the LIV literature though, we focus in our analysis on the structural parameter α_0 rather than on the marginal treatment effects. Our identification result shows that, by leveraging the identifying power of the factor structure, one can identify α_0 under weaker support restrictions than in the prior literature. In particular, our strategy makes it possible to use variation in $X = Z'\delta_0$ to identify α_0, even when all the components of Z_1 and Z_3 are discrete.[10]

3 Extended Factor Structure in the presence of Continuous Measurements

Up until now we have proposed identification and estimation results for a triangular system with a particular factor structure. A disadvantage of this structure is that it only includes one idiosyncratic shock (Π). We consider below an extension that addresses this limitation.

Namely, we consider the following model:

$$\begin{aligned} Y_1 &= \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\} \\ Y_2 &= \mathbf{1}\{X - V \geq 0\}, \end{aligned} \tag{3.1}$$

[10] An alternative approach to identifying this parameter can be found in Lewbel (2000). In his approach a second equation to model the endogenous variable is not needed, nor is the factor structure we impose. However, he imposes a strong support condition on a variable like Z_3 requiring that it exceeds the length of the unobservable U.

where $X_1 = Z_1'\lambda_0 + Z_3'\beta_0$, $X = Z'\delta_0$, $U = \gamma_0 W + \eta_1$, $V = W + \eta_2$, and (W, η_1, η_2) are mutually independent. In this setup, W can be interpreted as an unobserved confounder that satisfies the matching-on-unobservables condition $(Y_1(0), Y_1(1)) \perp\!\!\!\perp Y_2|W, X, X_1$ (Abbring and Heckman, 2007). Recall that, following the arguments in Section 2.1 above, we assume that X is observed. In addition, we assume two auxiliary continuous measurements

$$
\begin{aligned}
Y_3 &= \nu_0 W + \eta_3 \\
Y_4 &= \sigma_0 W + \eta_4,
\end{aligned}
\tag{3.2}
$$

where $(W, \eta_1, \eta_2, \eta_3, \eta_4)$ are mutually independent, and $\nu_0 \neq 0$.[11]

Our identification result is based on the following assumptions:

B0 The first coefficient of λ_0 is normalized to one so that $\lambda_0 = (1, \lambda_{0,-1}^T)^T$. The parameter $\theta_0 \equiv (\alpha_0, \gamma_0, \lambda_{0,-1}, \beta_0, \nu_0, \sigma_0)$ is an element of a compact subset of $\Re^{d_1+d_3+3}$, where d_1 and d_3 are the dimensions of Z_1 and Z_3, respectively. The vector of unobservables in the outcome and selection equations $(W, \eta_1, \eta_2, \eta_3)$ are independently distributed of the vector (Z_1, Z_2, Z_3). Both η_1 and η_2 are continuously distributed.

B1 $\gamma_0 \neq 0$. X is continuously distributed with absolute continuous density w.r.t. Lebesgue measure over the whole real line, conditionally on Z_1 and Z_3. The unconditional density of X is bounded and bounded away from zero on any compact subset of its support.

B2 W is not normally distributed or both η_3 and η_4 do not have a Gaussian component.

B3 $E(\eta_3) = E(\eta_4) = 0$, $E(|\eta_3|) < \infty$, and $E(|\eta_4|) < \infty$.

B4 $E(\exp(i\zeta\eta_2))$, $E(\exp(i\zeta\eta_3))$, and $E(\exp(i\zeta\eta_4))$ do not vanish for any $\zeta \in \Re$, where $i = \sqrt{-1}$.

B5 $E(\exp(i\zeta W)) \neq 0$ for all ζ in a dense subset of $\Re$.

B6 The distributions of W, η_2, and η_3 admit uniformly bounded densities $f_W(\cdot)$, $f_{\eta_2}(\cdot)$, and $f_{\eta_3}(\cdot)$ with respect to the Lebesgue measure that are supported on an interval (which may be infinite), respectively.

B7 Let $Z_{1,-1}$ be all the coordinates of Z_1 except the first one, and $d = d_1 + d_3 + 1$. There exist $2d$ vectors $\{z_1^{(l)}, z_3^{(l)}\}_{l=1}^d$ and $\{\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}\}_{l=1}^d$ in the joint support of (Z_1, Z_3) and $\{w^{(l)}\}_{l=1}^d, \{\tilde{w}^{(l)}\}_{l=1}^d$ such that

$$\alpha_0 + (z_{1,-1}^{(l)} - \tilde{z}_{1,-1}^{(l)})'\lambda_{0,-1} + (z_3^{(l)} - \tilde{z}_3^{(l)})'\beta_0 - \gamma_0(w^{(l)} - \tilde{w}^{(l)}) = \tilde{z}_{1,1}^{(l)} - z_{1,1}^{(l)}, \ l = 1, \cdots, d$$

[11] In practice, the continuous measurements might also depend on some observable characteristics. Our analysis goes through in this case after residualizing Y_3 and Y_4.

and $\text{rank}(\mathcal{M}) = d$, where

$$\mathcal{M} = \begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ w^{(1)} - \tilde{w}^{(1)} & \cdots & w^{(d)} - \tilde{w}^{(d)} \end{pmatrix}.$$

We now discuss these assumptions, before turning to the identification result. First, Assumption B0 is similar to Assumptions A1 and A2. We only need one of the idiosyncratic errors in the continuous measurements to be independent of the covariates because the other one is used to identify the distribution of the common factor W only. Second, as we assume in Assumption B1 that $\gamma_0 \neq 0$ and X has full support, the support condition

$$d(Z_1'\lambda_0 + Z_3'\beta_0 - \gamma_0 X) \geq |\alpha_0|.$$

holds automatically. The full support condition of X is necessary to identify the density of V, which is further used to identify the distribution of η_2. Assumption B1 reinforces this condition by supposing that X has full support conditional on Z_1 and Z_3, which is needed to identify the parameters from the outcome equation in a second step. Since $X = Z'\delta_0$ with $Z = (Z_1, Z_2)$, this is in turn equivalent to Z_2 having full support conditional on Z_1 and Z_3. Third, Assumptions B2–B6 imply Assumptions 1 to 4 in Hu and Schennach (2013). In practice we add the condition that the characteristic function of η_2 does not vanish, which is used for the deconvolution arguments in the proof of Theorem 3.1. We refer the reader to Hu and Schennach (2013) for more discussions of these assumptions.[12]

Theorem 3.1. *If* (3.1)–(3.2) *and Assumptions* **B0**–**B7** *hold, then* θ_0 *are identified.*

The proof of Theorem 3.1 can be found in Section B of the Supplement. Several remarks are in order. First, while we allow for a more general factor structure on the unobservables U and V, we also depart from our baseline specification by supposing that we have access to two continuous noisy measurements of the common factor W. This is a standard requirement in the nonparametric measurement error literature (Hu and Schennach, 2008). Besides, assuming access to a set of (selection-free) noisy measurements of the unobserved factors is also very standard in the evaluation literature. See, among many others, Carneiro et al. (2003), Heckman and Navarro (2007), Heckman and Vytlacil (2007a), and Cunha et al. (2010).

For instance, in applications in labor economics, the unobserved factor W often captures individual ability. This would apply, for example, to the evaluation of the effect of employment while in college (Y_2) on college graduation (Y_1). In this example, natural candidates for Z_2 are local labor market variables, including average wages and unemployment rate, while candidates for Z_1

[12]Note that Hu and Schennach (2013, Assumptions 5 and 6) hold automatically in our model with $\nu_0 \neq 0$.

include, among others, eligibility to financial aid programs providing tuition subsidy to students who maintain a minimum level of academic achievement.[13] In this context, cognitive skill measurements, such as the ASVAB test components that are available in the NLSY79 and NLSY97 surveys, are natural and often used candidates for the continuous measurements (Y_3, Y_4) (Ashworth et al., 2021).

Second, as is clear from the proof of Theorem 3.1, the key purpose of the continuous measurements is to identify the distribution of the common factor W. While we assume in this section that the measurement equations are linear, it is possible to identify θ_0 with a more general nonlinear system of continuous measurements, provided that the researcher has access to at least three such measurements. One can then combine Theorem 2 in Cunha et al. (2010) (Section 3.3, pp. 894-895), that yields identification of the distribution of W, with the proof of Theorem 3.1 in order to show identification of θ_0 for the case of nonlinear auxiliary measurements. Assuming access to a set of at least three measurements also makes it possible to relax the non-normality requirement imposed in Assumption B2.

Third, under the previous set of assumptions, the average treatment effect (ATE) is also identified. Key to this identification result is the full support condition on X given Z_1 and Z_3 (Assumption B1). Note that the conditional ATE given $X_1 = x_1$ is equal to $F_U(x_1 + \alpha_0) - F_U(x_1)$. In addition,

$$P(Y_1 = 1, Y_2 = 1|X_1 = x_1, X = x) = F_{U,V}(x_1 + \alpha_0, x).$$

One can let $x \to \infty$ so that

$$\lim_{x\to\infty} P(Y_1 = 1, Y_2 = 1|X_1 = x_1, X = x) = F_U(x_1 + \alpha_0).$$

Similarly,

$$\lim_{x\to-\infty} P(Y_1 = 1, Y_2 = 0|X_1 = x_1, X = x) = F_U(x_1).$$

This identifies the conditional and unconditional ATE.

Fourth, similar to the earlier discussions in Remark 2.2 and Section 2.2, Assumption B7 may still hold even when Z_3 is an empty set and Z_1 is discrete, since W is assumed to have full support. In such a case, identification primarily relies on the factor structure and the variation of the covariates in the selection equation, rather than that in the outcome equation. In this respect, this identification result is similar in spirit to Theorem 2.1 and different from the existing identification results in the literature for triangular binary models, e.g., Vytlacil and Yildiz (2007) and Vuong and Xu (2017). More generally, in Section C.2 in the supplement we establish that the factor model provides identification restrictions that are not otherwise available.[14]

[13] See Scott-Clayton (2011) for an evaluation of a program of this kind (PROMISE scholarship in West Virginia), and for a discussion of similar merit-based scholarship programs in place in other states.

[14] Specifically, we consider a version of the model (3.1), where we do not impose the factor structure and allow for an arbitrary (unknown to econometricians) dependence structure across the unobservables of the model. In this case,

Finally, we can relax the rank invariance condition to rank similarity by replacing $Y_1 = 1\{X_1 + \alpha_0 Y_2 - U \geq 0\}$ by $Y_1 = 1\{X_1 + \alpha_0 Y_2 - U(Y_2) \geq 0\}$. We then require $U(y_2) = \gamma_0 W + \eta_1(y_2)$ for $y_2 = 0, 1$. If Assumptions **B0–B7** hold with η_1 replaced by $(\eta_1(1), \eta_1(0))$ and $P(\eta_1(1) \leq e) = P(\eta_1(0) \leq e)$ for $e \in \Re$, then we can still identify θ_0 by a similar argument as the proof of Theorem 3.1.

4 Conclusion

In this paper, we explore the identifying power of linear factor structures in the context of simultaneous binary response models. We impose two alternative types of factor structures on the unobservables of the model. The first setup is a natural distribution-free extension of the bivariate Probit model, while the second model corresponds to a standard linear factor model with one common factor and two equation-specific idiosyncratic shocks. We establish that both factor models have identifying power in that they make it possible to relax some of the exclusion and support conditions typically required for identification in this class of models (Vytlacil and Yildiz, 2007). Overall, our analysis adds to our understanding of the identifying power of factor models, beyond their well known usefulness to recover the joint distribution of potential outcomes from the marginal distributions.

The work here opens areas for future research. The factor structure we assume could prove useful in more general nonlinear models. For instance, non-triangular discrete systems have shown to be an effective way to model entry games in the empirical industrial organization literature-see, for example, Tamer (2003). However, as shown in Khan and Nekipelov (2018), identification of structural parameters in these models can be even more challenging than for the triangular model considered in this paper, and furthermore, as shown recently in Khan and Nekipelov (2021), conducting valid uniform interest in all these models is very difficult. It would be useful to determine if factor structures on the unobservables could alleviate this problem. We leave this open question to future work.

we show non-identification of α_0 as long as $|\alpha_0| > b - a$, where $[a, b]$ denotes the conditional support of X_1 given X and, consistent with our Assumption B1, X has full support on the real line. However, by imposing the factor structure (and other conditions implied by B0–B7), Theorem 3.1 shows that α_0 is identified for this model even when $|\alpha_0| > b - a$.

A Proof of Theorem 2.1

Proof: Note that

$$P^{11}(z_1, z_3, x) = \int_{-\infty}^{x} F_\Pi(z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \gamma_0 v) f_V(v) dv$$
$$P^{10}(\tilde{z}_1, \tilde{z}_3, \tilde{x}) = \int_{\tilde{x}}^{+\infty} F_\Pi(\tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0 - \gamma_0 v) f_V(v) dv.$$

Taking derivatives w.r.t. the third argument of the LHS function, we obtain

$$\partial_x P^{11}(z_1, z_3, x)/f_V(x) = F_\Pi(z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \gamma_0 x)$$
$$-\partial_x P^{10}(\tilde{z}_1, \tilde{z}_3, \tilde{x})/f_V(\tilde{x}) = F_\Pi(\tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0 - \gamma_0 \tilde{x}).$$

By Assumption **A4**, we know that there exists pairs such that

$$Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 - \gamma_0 X = \tilde{Z}_1'\lambda_0 + \tilde{Z}_3'\beta_0 - \gamma_0 \tilde{X}.$$

Because $F_\Pi(\cdot)$ is monotone increasing, we have

$$\partial_x P^{11}(Z_1, Z_3, X)/f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X})/f_V(\tilde{X}) = 0$$
$$\Longleftrightarrow \alpha_0 + (Z_1 - \tilde{Z}_1)'\lambda_0 + (Z_3 - \tilde{Z}_3)'\beta_0 - \gamma_0(X - \tilde{X}) = 0$$

Note the LHS of the above display is identified from data. Denote $Z_{1,1}$ as the first element of Z_1, whose coefficient is set to one. The rest of Z_1 is denoted as $Z_{1,-1}$, whose coefficient is denoted as $\lambda_{0,-1}$. Then, we have

$$\alpha_0 + (Z_{1,-1} - \tilde{Z}_{1,-1})'\lambda_{0,-1} + (Z_3 - \tilde{Z}_3)'\beta_0 - \gamma_0(X - \tilde{X}) = \tilde{Z}_{1,1} - Z_{1,1}.$$

Then, by Assumption **A4**, we can find $(z_1^{(l)}, z_3^{(l)}, x^{(l)})_{l=1}^d$ and $(\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{x}^{(l)})_{l=1}^d$ such that

$$\text{rank}\begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ x^{(1)} - \tilde{x}^{(1)} & \cdots & x^{(d)} - \tilde{x}^{(d)} \end{pmatrix} = d.$$

Then, we can identify $(\alpha_0, \lambda_0, \beta_0, \gamma_0)$ by solving the linear system that

$$\begin{aligned}
\alpha_0 + (z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)})'\lambda_{0,-1} + (z_3^{(1)} - \tilde{z}_3^{(1)})'\beta_0 - \gamma_0(x^{(1)} - \tilde{x}^{(1)}) =& \tilde{z}_{1,1}^{(1)} - z_{1,1}^{(1)}, \\
&\vdots \\
\alpha_0 + (z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)})'\lambda_{0,-1} + (z_3^{(d)} - \tilde{z}_3^{(d)})'\beta_0 - \gamma_0(x^{(d)} - \tilde{x}^{(d)}) =& \tilde{z}_{1,1}^{(d)} - z_{1,1}^{(d)}.
\end{aligned}$$

This concludes the proof.

B Proof of Theorem 3.1

For notation simplicity, we write $\tilde{W} = \nu_0 W$, $\tilde{\sigma}_0 = \sigma_0/\nu_0$, $\tilde{\nu}_0 = 1/\nu_0$, and

$$\begin{aligned}
Y_2 &= \mathbf{1}\{X \geq \tilde{\nu}_0 \tilde{W} + \eta_2\} \\
Y_3 &= \tilde{W} + \eta_3 \\
Y_4 &= \tilde{\sigma}_0 \tilde{W} + \eta_4.
\end{aligned}$$

Because Assumptions B2–B6 hold, by applying Hu and Schennach (2013, Theorem 1) to Y_3 and Y_4, we can identify the densities for $\nu_0 W = \tilde{W}$, η_3, and η_4 as well as $\sigma_0/\nu_0 = \tilde{\sigma}_0$.

Then, we have

$$\begin{aligned}
\partial_{y_3} P(Y_2 = 1, Y_3 \leq y_3 | X = x) =& \partial_{y_3} \int F_{\eta_2}(x - \tilde{\nu}_0 w) F_{\eta_3}(y_3 - w) f_{\tilde{W}}(w) dw \\
=& \int F_{\eta_2}(x - \tilde{\nu}_0 w) f_{\eta_3}(y_3 - w) f_{\tilde{W}}(w) dw.
\end{aligned}$$

Applying Fourier transform w.r.t. y_3 on both sides, we have

$$\mathcal{F}(\partial_{y_3} P(Y_2 = 1, Y_3 \leq \cdot | X = x))(t) = \mathcal{F}(F_{\eta_2}(x - \tilde{\nu}_0 \cdot) f_{\tilde{W}}(\cdot))(t) \mathcal{F}(f_{\eta_3}(\cdot))(t),$$

where for a generic function $g(w)$,

$$\mathcal{F}(g(\cdot))(t) = \frac{1}{\sqrt{2\pi}} \int \exp(-2\pi i t w) g(w) dw.$$

Therefore,

$$\frac{\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2=1, Y_3 \leq \cdot | X=x))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)}\right)(w)}{f_{\tilde{W}}(w)} = F_{\eta_2}(x - \tilde{\nu}_0 w), \tag{B.3}$$

where for a generic function $g(w)$,

$$\mathcal{F}^{-1}(g(\cdot))(t) = \frac{1}{\sqrt{2\pi}} \int \exp(2\pi i t w) g(w) dw.$$

Note the LHS of (B.3) can be identified from data. We choose two pairs (x, w) and (x', w') such that $w \neq w'$ and

$$\frac{\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2=1, Y_3 \leq \cdot | X=x))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)}\right)(w)}{f_{\tilde{W}}(w)} = \frac{\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2=1, Y_3 \leq \cdot | X=x'))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)}\right)(w')}{f_{\tilde{W}}(w')}.$$

Then, given the monotonicity of F_{η_2}, we have

$$x - \tilde{\nu}_0 w = x' - \tilde{\nu}_0 w',$$

or

$$\tilde{\nu}_0 = (x - x')/(w - w'),$$

which is identified. Given the identification of $\tilde{\nu}_0$ and the distribution of $\tilde{W}$, we can identify the distribution of $W = \tilde{\nu}_0 \tilde{W}$. Recall $F_{\eta_1}(\cdot)$ and $f_{\eta_2}(\cdot)$ are the CDF and PDF of η_1 and η_2, respectively. Then, we have

$$P(Y_2 = 1 | X = x) = P(W + \eta_2 \leq x).$$

Because X has full support, we can identify the distribution of $W + \eta_2$. Then, it follows from standard deconvolution argument and the fact that the distribution of W is identified that we can identify the distribution of η_2. In addition, note that

$$\begin{aligned} P^{11}(z_1, z_3, x) =& P(Y_1 = 1, Y_2 = 1 | Z_1 = z_1, Z_3 = z_3, X = x) \\ =& \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \gamma_0 w) F_{\eta_2}(x - w) f_W(w) dw \end{aligned}$$

and

$$\begin{aligned} P^{10}(z_1, z_3, x) =& P(Y_1 = 1, Y_2 = 0 | Z_1 = z_1, Z_3 = z_3, X = x) \\ =& \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 - \gamma_0 w)(1 - F_{\eta_2}(x - w)) f_W(w) dw. \end{aligned}$$

Taking derivatives of $P^{11}(z_1, z_3, x)$ and $P^{10}(z_1, z_3, x)$ w.r.t. x, we have

$$\partial_x P^{11}(z_1, z_3, x) = \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - w) f_{\eta_2}(x - w) f_W(w) dw \tag{B.4}$$

and

$$-\partial_x P^{10}(z_1, z_3, x) = \int F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 - \gamma_0 w) f_{\eta_2}(x - w) f_W(w) dw. \tag{B.5}$$

Applying Fourier transform on both sides of (B.4) and (B.5), we have

$$\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot)) = \mathcal{F}(F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \cdot) f_W(\cdot))\mathcal{F}(f_{\eta_2}(\cdot)) \tag{B.6}$$

and

$$\mathcal{F}(-\partial_x P^{10}(z_1, z_3, \cdot)) = \mathcal{F}(F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 - \cdot) f_W(\cdot))\mathcal{F}(f_{\eta_2}(\cdot)).$$

Then, by (B.6), we can identify $F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \cdot)$ by

$$F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \gamma_0 \cdot) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))}\right)(\cdot)/f_W(\cdot).$$

Similarly, we can identify

$$F_{\eta_1}(z_1'\lambda_0 + z_3'\beta_0 - \gamma_0 \cdot) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(-\partial_x P^{10}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))}\right)(\cdot)/f_W(\cdot).$$

Because $F_{\eta_1}(\cdot)$ is monotone increasing, we have

$$\begin{aligned}
&\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))}\right)(w)/f_W(w) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(-\partial_x P^{10}(\tilde{z}_1, \tilde{z}_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))}\right)(\tilde{w})/f_W(\tilde{w}) \\
\iff &\alpha_0 + (z_1 - \tilde{z}_1)'\lambda_0 + (z_3 - \tilde{z}_3)'\beta_0 - \gamma_0(w - \tilde{w}) = 0
\end{aligned}$$

Then, by Assumption B7, we can find $(z_1^{(l)}, z_3^{(l)}, w^{(l)})_{l=1}^d$ and $(\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{w}^{(l)})_{l=1}^d$ such that

$$\text{rank}\begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ w^{(1)} - \tilde{w}^{(1)} & \cdots & w^{(d)} - \tilde{w}^{(d)} \end{pmatrix} = d.$$

Then, we can identify $(\alpha_0, \lambda_0, \beta_0, \gamma_0)$ by solving the linear system that

$$\begin{aligned}
\alpha_0 + (z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)})'\lambda_{0,-1} + (z_3^{(1)} - \tilde{z}_3^{(1)})'\beta_0 - \gamma_0(w^{(1)} - \tilde{w}^{(1)}) =& \tilde{z}_{1,1}^{(1)} - z_{1,1}^{(1)}, \\
&\vdots \\
\alpha_0 + (z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)})'\lambda_{0,-1} + (z_3^{(d)} - \tilde{z}_3^{(d)})'\beta_0 - \gamma_0(w^{(d)} - \tilde{w}^{(d)}) =& \tilde{z}_{1,1}^{(d)} - z_{1,1}^{(d)}.
\end{aligned}$$

This concludes the proof.

C Identification with and without Factor Structure

C.1 Identification Without Auxiliary Measurements

In this section, we discuss the information content of factor structure. For illustration purpose, we focus on the "condensed" model:

$$\begin{aligned}
Y_1 &= \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\} \\
Y_2 &= \mathbf{1}\{X - V \geq 0\}.
\end{aligned} \tag{C.7}$$

Assumption 1.
1. $(X_1, X) \perp (U, V)$.
2. (X_1, X) *are continuously distributed with absolute continuous joint density w.r.t. Lebesgue measure. The conditional support of* X_1 *given* X *is* $[a, b]$.
3. V *is continuously distributed over* $\Re$ *and its density w.r.t. Lebesgue measure exist.*

Theorem C.1. *If Assumption 1 holds, then* $|\alpha_0| \leq b - a$ *is necessary and sufficient for* α_0 *to be identified.*

We note that under Assumption 1, $|\alpha_0| \leq b - a$ is equivalent to the fact that we can find x_1 and $\tilde{x}_1$ in the support of X_1 such that $\alpha_0 = x_1 - \tilde{x}_1$.

Next, we assume, in addition to Assumption 1, the factor structure, i.e., (2.6) in Section 2. Our rank estimator can be written as an M-estimator

$$\hat{\theta} = \arg\max_{\theta} Q_n(\theta) \equiv \sum_{i \neq j} \hat{g}_{i,j}(\theta)$$

in which

$$\begin{aligned}
\hat{g}_{i,j}(\theta) \quad = \quad & [\mathbf{1}\{\partial_2 \hat{P}^{11}(X_{1,i}, X_i)/\hat{f}_V(X_i) + \partial_2 \hat{P}^{10}(X_{1,j}, X_j)/\hat{f}_V(X_j) \geq 0\}\mathbf{1}\{\Phi(X_{1,i}, X_i, X_{1,j}, X_j; \theta) \geq 0\} \\
+ \quad & \mathbf{1}\{\partial_2 \hat{P}^{11}(X_{1,i}, X_i)/\hat{f}_V(X_i) + \partial_2 \hat{P}^{10}(X_{1,j}, X_j)/\hat{f}_V(X_j) < 0\}\mathbf{1}\{\Phi(X_{1,i}, X_i, X_{1,j}, X_j; \theta) < 0\}],
\end{aligned}$$

with

$$\Phi(x_1, x, \tilde{x}_1, \tilde{x}; \theta) = x_1 + \alpha - \gamma x - (\tilde{x}_1 - \gamma \tilde{x}).$$

We will study the asymptotic properties of this estimator in Section **??**.

The information content explored by the M-estimator can be summarized as follows:

$$\mathcal{A}_2(\theta) = \{(X_1, \tilde{X}_1, X, \tilde{X}), \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta_0) \geq 0 > \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta)$$
$$\text{or } \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta_0) < 0 \leq \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta)\}.$$

Then we cannot distinguish, from the true parameter θ_0, all impostors in

$$\overline{\mathcal{A}}_2 = \{\theta : P(\mathcal{A}_2(\theta)) = 0\}.$$

In the condensed model, if $\text{Supp}(X_1, X) = [a, b] \times [c, d]$, then θ_0 is identified if $|\alpha_0| < b - a + |\gamma_0|(d - c)$. Recall Theorem C.1, without imposing factor structure, the necessary and sufficient condition for achieving identification is $|\alpha_0| \leq b - a$. Therefore, the blue area in the Figure below is the additional parts of parameter space that are identified with factor structure but not otherwise.

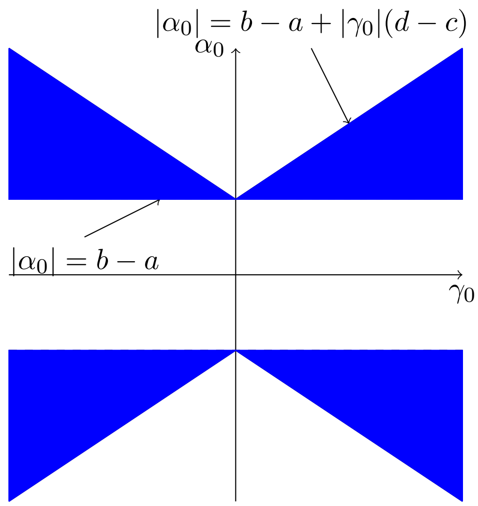

Figure 1: Identifying Power of Factor Structure

Theorem C.2. *Assumption 1 holds. When $|\alpha_0| > b - a$, the sharp identified set for α_0 is*

$$\mathcal{A}^* = \{\alpha : \alpha > b - a \text{ if } \alpha_0 > 0 \text{ and } \alpha < a - b \text{ if } \alpha_0 < 0\}.$$

Theorem C.2 highlights that, in the case without the factor structure and α_0 does not satisfy the parameter restriction, except for the fact that the sign of α_0 is identified, we actually cannot say much about the value of $|\alpha_0|$. When we assume the factor structure, the parameter is still not identified if $|\alpha_0| > b-a+|\gamma_0|(d-c)$. In addition, suppose $\alpha_0 > 0$. In this case, if we do not impose factor structure, by Theorem C.2, the sharp identified set is $\{\alpha : \alpha > b-a\}$ while with the factor structure, the identified set (not necessarily sharp) is $\alpha > b-a+|\gamma|(d-c)$. This implies, when identification fails in both cases, the blue area is also the extra identifying power on the identified set given by the factor structure.

C.2 Identification with two auxiliary measurements

Next, we expand our condensed model to include two continuous measurements. We show in this case, without the factor structure, α_0 is not identified. This is in contrast with the identification result established in Theorem 3.1.

Suppose in addition to (C.7), we also observe two continuous measurements of W denoted as Y_3 and Y_4. One example of such Y_3 and Y_4 are described in (3.2).

Assumption 2.
1. $(X_1, X) \perp (U, V, Y_3, Y_4)$.
2. (X_1, X) *are continuously distributed with absolute continuous joint density w.r.t. Lebesgue measure. The conditional support of* X_1 *given* X *is* $[a, b]$.
3. V *is continuously distributed over* $\Re$ *and its density w.r.t. Lebesgue measure exist.*

Theorem C.3. *If Assumption 2 holds, then* $|\alpha_0| \leq b-a$ *is necessary and sufficient for* α_0 *to be identified.*

The proof of Theorem C.3 is similar to that of Theorem C.1, and thus, is omitted. In the proof of Theorem C.1, we show that when $|\alpha_0| > b-a$, we can find an impostor $\alpha \neq \alpha_0$ and $\tilde{U}$ such that for any $x_1 \in [a, b]$ and any $v \in \text{Supp}(V)$, we have

$$
\begin{aligned}
&P(\tilde{U} \leq x_1 + \alpha | V = v) = P(U \leq x_1 + \alpha_0 | V = v) \\
&P(\tilde{U} \leq x_1 | V = v) = P(U \leq x_1 | V = v).
\end{aligned}
$$

This implies the conditional CDF of (Y_1, Y_2) given (X_1, X) under the DGPs (U, V, α_0) and $(\tilde{U}, V, \alpha)$ are the same, and thus, α_0 is observationally equivalent to the impostor α. Similarly, with the two continuous measurements, we can use the exact same construction of $\tilde{U}$ and α to show that, for any $x_1 \in [a, b]$ and $(v, y_3, y_4) \in \text{Supp}(V, Y_3, Y_4)$, we have

$$
\begin{aligned}
&P(\tilde{U} \leq x_1 + \alpha | V = v, Y_3 = y_3, Y_4 = y_4) = P(U \leq x_1 + \alpha_0 | V = v, Y_3 = y_3, Y_4 = y_4) \\
&P(\tilde{U} \leq x_1 | V = v, Y_3 = y_3, Y_4 = y_4) = P(U \leq x_1 | V = v, Y_3 = y_3, Y_4 = y_4).
\end{aligned}
$$

This implies the conditional CDF of (Y_1, Y_2, Y_3, Y_4) given (X_1, X) under the DGPs $(U, V, Y_3, Y_4, \alpha_0)$

and $(\tilde{U}, V, Y_3, Y_4, \alpha)$ are the same too. Such non-identification result holds even when X has full support.

D Proof of Theorem C.1

Denote $P^{ij}(x_1, x) = Prob(Y_1 = i, Y_2 = j | X_1 = x_1, X = x)$. Then

$$\begin{aligned} P^{11}(x_1, x) &= \int_{-\infty}^{x} F_U(x_1 + \alpha_0 | V = v) f(v) dv \\ P^{10}(\tilde{x}_1, x) &= \int_{x}^{+\infty} F_U(\tilde{x}_1 | V = v) f(v) dv. \end{aligned} \tag{D.8}$$

Taking derivatives w.r.t. the second argument of the the LHS function, we have

$$\begin{aligned} &\partial_2 P^{11}(x_1, x) = F_U(x_1 + \alpha_0 | V = x) f(x) \\ &\partial_2 P^{10}(\tilde{x}_1, x) = -F_U(\tilde{x}_1 | V = x) f(x). \end{aligned}$$

If $|\alpha_0| \leq b - a$, then there exists a pair $(x_1, \tilde{x}_1)$ such that $x_1 + \alpha_0 = \tilde{x}_1$. This pair can be identified by checking the equation below:

$$\partial_2 P^{11}(x_1, x)/f(x) + \partial_2 P^{10}(\tilde{x}_1, x)/f(x) = 0.$$

This concludes the sufficient part.

When $\alpha_0 < a - b$, for any $\alpha < \alpha_0$, we can define

$$\begin{aligned} &\tilde{U} = U + \alpha - \alpha_0 && \text{if} && U \leq b + \alpha_0 \\ &\tilde{U} = U && \text{if} && U > b + \alpha_0 \end{aligned}$$

Then for any $x_1 \in [a, b]$,

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(\tilde{U} \leq x_1 + \alpha, U \leq b + \alpha_0 | V = v) + P(\tilde{U} \leq x_1 + \alpha, U > b + \alpha_0 | V = v) \\ &= P(U \leq x_1 + \alpha_0 | V = v) \\ P(\tilde{U} \leq x_1 | V = v) &= P(\tilde{U} \leq x_1, U \leq b + \alpha_0 | V = v) + P(\tilde{U} \leq x_1, U > b + \alpha_0 | V = v) \\ &= P(U \leq b + \alpha_0, U \leq x_1 + \alpha_0 - \alpha | V = v) + P(b + \alpha_0 < U \leq x_1 | V = v) \\ &= P(U \leq b + \alpha_0 | V = v) + P(b + \alpha_0 < U \leq x_1 | V = v) \\ &= P(U \leq x_1 | V = v), \end{aligned}$$

where the third equality holds because, since $\alpha_0 < a - b$ and $\alpha < \alpha_0$, $b + \alpha_0 \leq x_1 + \alpha_0 - \alpha$ for

$x_1 \in [a,b]$. Let $G_{U,V}$ and $G_{\tilde{U},V}$ be the joint distribution of (U,V) and $(\tilde{U},V)$ respectively. Then the above calculation with (D.8) imply that $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ produce the identical pair $(P^{11}(x_1,x), P^{10}(x_1,x))$. In addition, the distribution of V is unchanged so that $P(Y_2 = 1|X = x)$ is identified from data. Therefore, $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ are observationally equivalent.

Similarly, when $\alpha_0 > b - a$, for any $\alpha > \alpha_0$, we can define

$$\begin{aligned} &\tilde{U} = U + \alpha - \alpha_0 && \text{if} && U > a + \alpha_0 \\ &\tilde{U} = U && \text{if} && U \leq a + \alpha_0 \end{aligned}$$

Then for any $x_1 \in [a,b]$,

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(\tilde{U} \leq x_1 + \alpha, U \leq a + \alpha_0 | V = v) + P(\tilde{U} \leq x_1 + \alpha, U > a + \alpha_0 | V = v) \\ &= P(U \leq a + \alpha_0 | V = v) + P(a + \alpha_0 < U \leq x_1 + \alpha_0 | V = v) \\ &= P(U \leq x_1 + \alpha_0 | V = v). \\ P(\tilde{U} \leq x_1 | V = v) &= P(\tilde{U} \leq x_1, U \leq a + \alpha_0 | V = v) + P(\tilde{U} \leq x_1, U > a + \alpha_0 | V = v) \\ &= P(U \leq x_1 | V = v), \end{aligned}$$

where we use the facts that $x_1 \leq a + \alpha_0$ and $x_1 - a < \alpha$ for $x_1 \in [a,b]$. So again, $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ are observationally equivalent.

E Proof of Theorem C.2

The sign of α_0 is identified by the data. In the following, we focus on deriving the results when $\alpha_0 > b - a$. By the proof of Theorem C.1, we have already shown that all $\alpha > \alpha_0$ is in the identified set. Now we consider $\frac{b-a+\alpha_0}{2} \leq \alpha < \alpha_0$.

$$\begin{aligned} &\tilde{U} = U + \alpha - \alpha_0 && \text{if} && U > a + \alpha \\ &\tilde{U} = U && \text{if} && U \leq a + \alpha \end{aligned}$$

Then for any $x_1 \in [a, b]$,

$$\begin{aligned}
P(\tilde{U} \le x_1 + \alpha | V = v) &= P(\tilde{U} \le x_1 + \alpha, U \le a + \alpha | V = v) + P(\tilde{U} \le x_1 + \alpha, U > a + \alpha | V = v) \\
&= P(U \le a + \alpha | V = v) + P(a + \alpha < U \le x_1 + \alpha_0 | V = v) \\
&= P(U \le x_1 + \alpha_0 | V = v). \\
P(\tilde{U} \le x_1 | V = v) &= P(\tilde{U} \le x_1, U \le a + \alpha | V = v) + P(\tilde{U} \le x_1, U > a + \alpha | V = v) \\
&= P(U \le x_1 | V = v) + P(U \le x_1 + \alpha_0 - \alpha, U > a + \alpha | V = v). \\
&= P(U \le x_1 | V = v).
\end{aligned}$$

Here note that the last equality is because $x_1 + \alpha_0 - \alpha \le b + \alpha_0 - \alpha \le a + \alpha$ if $\alpha \ge \frac{b-a+\alpha_0}{2}$. Denote $\alpha^{(1)} = \frac{b-a+\alpha_0}{2}$. Then we have shown that there exists $U^{(1)}(\alpha)$ which only depends on α such that for any $x_1 \in [a, b]$, any v and any $\alpha_0 > \alpha \ge \alpha^{(1)}$

$$\begin{aligned}
&P(U^{(1)}(\alpha) \le x_1 + \alpha | V = v) = P(U \le x_1 + \alpha_0 | V = v) \\
&P(U^{(1)}(\alpha) \le x_1 | V = v) = P(U \le x_1 | V = v).
\end{aligned}$$

In particular, there exists $U^{(1)}(\alpha^{(1)})$ such that

$$\begin{aligned}
&P(U^{(1)}(\alpha^{(1)}) \le x_1 + \alpha^{(1)} | V = v) = P(U \le x_1 + \alpha_0 | V = v) \\
&P(U^{(1)}(\alpha^{(1)}) \le x_1 | V = v) = P(U \le x_1 | V = v).
\end{aligned}$$

Now repeating the above construction but replacing U with $U^{(1)}$ and α_0 with $\alpha^{(1)}$, we have for any $\alpha^{(1)} > \alpha \ge \alpha^{(2)} \equiv \frac{b-a+\alpha^{(1)}}{2}$, there exists $U^{(2)}(\alpha)$ such that for any $x_1 \in [a, b]$, any v and any $\alpha^{(1)} > \alpha \ge \alpha^{(2)}$,

$$\begin{aligned}
&P(U^{(2)}(\alpha) \le x_1 + \alpha^{(2)} | V = v) = P(U^{(1)}(\alpha^{(1)}) \le x_1 + \alpha^{(1)} | V = v) = P(U \le x_1 + \alpha_0 | V = v) \\
&P(U^{(2)}(\alpha) \le x_1 | V = v) = P(U^{(1)}(\alpha^{(1)}) \le x_1 | V = v) = P(U \le x_1 | V = v).
\end{aligned}$$

This concludes that any α such that $\alpha_0 > \alpha \ge \alpha^{(2)}$ is in the identified set. In general, by repeating the procedure k times, we have that any α such that

$$\alpha_0 > \alpha \ge \alpha^{(k)} = (1 - \frac{1}{2^k})(b - a) + \frac{\alpha_0}{2^k}$$

is in the identified set. For any $\alpha > b - a$, there exists some finite k such that $\alpha > (1 - \frac{1}{2^k})(b - a) + \frac{\alpha_0}{2^k}$. This concludes the result that $\alpha > b - a$ is in the identified set.

Finally, since if $\alpha > b - a$, $\partial_2 P^{11}(x_1, x) + \partial_2 P^{10}(\tilde{x}_1, x) > 0$ for all pairs of (x_1, x) and $(\tilde{x}_1, x)$ while, if $\alpha \le b - a$, at least there exists one pair (x_1, x) and $(\tilde{x}_1, x)$ such that $\partial_2 P^{11}(x_1, x) + \partial_2 P^{10}(\tilde{x}_1, x) \le 0$. This implies $\alpha \le b - a$ is not in the identified set. Therefore, the sharp identified set when $\alpha_0 > b - a$ is $(b - a, \infty)$.

When $\alpha_0 < a - b$, a symmetric argument implies that the identified set is $(-\infty, a - b)$.

References

Abbring, J. and J. Heckman (2007). Econometrics evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics, Vol. 6B*. North Holland.

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

Abrevaya, J., J. Hausman, and S. Khan (2010). Testing for causal effects in a generalized regression model with endogenous regressors. *Econometrica 78*(6), 2043–2061.

Ashworth, J., V. J. Hotz, A. Maurel, and T. Ransom (2021). Changes across cohorts in wage returns to schooling and early work experiences. *Journal of Labor Economics 39*(4), 931–964.

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

Bierens, H. and J. Hartog (1988). Non-linear regression with discrete explanatory variables, with an application to the earnings function. *Journal of Econometrics 38*(3), 269–299.

Blundell, R. and J. Powell (2004). Endogeneity in binary response models. *Review of Economic Studies 73*.

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

Carneiro, P., K. Hansen, and 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.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica 73*(5), 1525–1550.

Chiburis, R. (2010). Semiparametric bounds on treatment effects. *Journal of Econometrics 159*(2), 267–275.

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

Gourieroux, C., A. Monfort, and J.-P. Renne (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics 196*(1), 111–126.

Han, A. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*(2–3), 303–316.

Han, S. and S. Lee (2019). Estimation in a generalization of a bivariate probit models with dummy endogenous regressors. *Journal of Applied Econometrics 34*(6), 994–1015.

Han, S. and E. J. Vytlacil (2017). Identification in a generalization of bivariate probit models with endogenous regressors. *Journal of Econometrics 199*(1), 63–73.

Heckman, J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Heckman, J. J., J. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*, S197–S246.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics 6*, 4779–4874.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics 6*, 4875–5143.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Hu, Y. and S. M. Schennach (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association 108*(501), 177–186.

Hyvärinen, A. and E. Oja (2000). *Independent component analysis: algorithms and applications*, Volume 13. Elsevier.

Khan, S. and D. Nekipelov (2018). Information structure and statistical information in discrete response models. *Quantitative Economics 9*(2), 995–1017.

Khan, S. and D. Nekipelov (2021). On uniform inference in nonlinear models with endogeneity. *Journal of Econometrics, forthcoming.*

Klein, R., C. Shan, and F. Vella (2015). Estimation of marginal effects in semiparametric selection models with binary outcomes. *Journal of Econometrics 185*(1), 82–94.

Klein, R. and R. Spady (1993). An efficient semiparametric estimator for binary response model. *Econometrica 61*(2), 387–421.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A., S. M. Schennach, and L. Zhang (2020). Identification of a triangular two equation system without instruments. Working Paper.

Moneta, A., D. E. P. O. Hoyer, and A. Coad (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics 75*(5), 705–730.

Mourifié, I. (2015). Sharp bounds on treatment effects in a binary triangular system. *Journal of Econometrics 187*(1), 74–81.

Park, J. and P. Phillips (2000). Nonstationary binary choice. *Econometrica 68*(5), 1249–80.

Scott-Clayton, J. (2011). Discussion of "simple estimators for invertible index models" by ahn et al. *Journal of Human Resources 46*(3), 614–46.

Shaikh, A. M. and E. Vytlacil (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica 79*(3), 949–955.

Tamer, E. (2003). Incomplete bivariate discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–167.

Vuong, Q. and H. Xu (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics 8*(2), 589–610.

Vytlacil, E. J. and N. Yildiz (2007). Dummy endogenous variables in weakly separable models. *Econometrica 75*(3), 757–779.